# Projecting the impacts of climate change on wheat yields in the UK and Germany and assessing modeling uncertainties

Andrea Monica Del Rosario Ortiz

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

DEPARTMENT OF GEOGRAPHY

FACULTY OF SOCIAL SCIENCES

THE UNIVERSITY OF SHEFFIELD

June 2019

# Abstract

Climate change and variability are projected to negatively affect wheat production in Europe. The impacts of climate change are typically projected using global and regional climate models (GCMs and RCMs) and impact assessment tools such as crop models. However, this impact simulation chain can propagate uncertainty, as errors are introduced by GCMs, RCMs, and crop models. There are also many intermediate steps and decisions in the impact simulation process that are influenced by the different communities of practice that utilize climate and crop models. These differences in methods and approaches can also influence the range of future yield projections. Yield projections are thus considered inherently uncertain because of this cascade of uncertainty.

This interdisciplinary study projects the impacts of climate change on wheat yields in the UK and Germany, two key wheat-growing countries. Added value is found when using RCMs for downscaling temperature and precipitation simulations for the impact assessment. However, these GCM-RCM simulations are shown to have significant errors relative to observations, necessitating a bias correction (BC) step. Different BC methods are shown to be effective in improving simulations. Two BC calibration approaches, one that corrects RCM-only error and the other GCM-RCM error, are used to examine how different GCM-RCM combinations can affect projected changes in climate. Future climate projections are used in a multi-method crop modeling approach, and the uncertainty in the resulting yield projections is analyzed. Key findings are that wheat yields in the UK and Germany will be affected by changes in temperature and precipitation. However, these impacts are shown to be region-dependent and vary based on the crop modeling method, making the choice of crop modeling method a major contributor to uncertainty.

# Acknowledgments

There are many people and institutions that I am grateful for that have helped me produce this culmination of my education thus far. I would like to thank my family – Mama, Papa, Gian, Mara and Mikey – for believing in me no matter how near or far I am. This work is dedicated to my *Lola*, who would have probably not understood a thing in my thesis but would have been proud of me all the same.

My deepest thanks to Jeremy Grantham and the Grantham Centre for Sustainable Futures, and to my supervisors Julie Jones, Rob Freckleton, and Adam Scaife, for believing in me from the start, challenging me, and helping me through to the end. I have learned so much from all of you.

I would also like to thank Deborah Beck and the Geography Department administration team for their support. I am grateful for the friendship from colleagues and friends who became my family in Sheffield. To all those who have helped me attain the skills and learn the tools I needed for this thesis, thank you. Thank you to my friends and mentors for the encouragement to pursue and finish the PhD. Thank you to Colin Osborne and Andy Challinor for the expertise and support to make my work better.

Lastly, I would like to thank Bernardo – the Sam to my Frodo – because without his steadfast support, patience, and love I would not have been able throw this ring into Mordor, nor do a for loop for that matter. All this is for you and my (real) little Sam.

*Maraming salamat po sa Panginoon, at sa inyong lahat.*

# Contents


**1 Introduction** **1**

1.1 Background of the study . . . . . . . . . . . . . . . . . . . . . . 4

1.1.1 Global significance of wheat . . . . . . . . . . . . . . . . 4

1.1.2 Wheat production in Europe . . . . . . . . . . . . . . . . . 5

1.1.3 Sensitivity of wheat to weather and climate . . . . . . . 6

1.1.4 Risk of climate change and variability to wheat . . . . . 10

1.1.5 Typical approaches to investigate the impacts of climate change on wheat production . . . . . . . . . . . 11

1.1.6 Uncertainty as a major theme of the research . . . . . . 12

1.2 Diverging communities of practice . . . . . . . . . . . . . . . . 15

**2 Literature review and defining research gaps** **17**

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17

2.1.1 Chapter objectives . . . . . . . . . . . . . . . . . . . . . . 18

2.2 Uncertainty from climate models . . . . . . . . . . . . . . . . . 19

2.2.1 Introduction to climate models . . . . . . . . . . . . . . . 20

2.2.2 Sources of error in climate models . . . . . . . . . . . . 21

2.2.3 Evaluating climate model performance . . . . . . . . . . 26

2.2.4 Simulating regional climate . . . . . . . . . . . . . . . . 27

2.2.5 Bias correction to improve agreement of simulations with observations . . . . . . . . . . . . . . . . . . . . . . 31

2.2.6 Climate change projections and uncertainty . . . . . . . 32

2.3 Uncertainty from crop models . . . . . . . . . . . . . . . . . . . 33

2.3.1 Representing crop growth and responses to climate . . 33

2.3.2 Challenges in crop modeling approaches . . . . . . . . 37

2.4 Uncertainties in crop-climate model linkages . . . . . . . . . . 43

2.4.1 Relevance of downscaling in crop impact assessment . 43

2.4.2 Bias correction to improve climate model output . . . . 45

2.5 Identified research gaps and questions . . . . . . . . . . . . . . . 47
2.5.1 Bias correction, climate and crop projections . . . . . . 47
2.5.2 Multi-method comparisons . . . . . . . . . . . . . . . . . 48
2.5.3 Research questions and hypothesis . . . . . . . . . . . 51
2.5.4 Research aim . . . . . . . . . . . . . . . . . . . . . . . . 52
2.6 Structure of the thesis . . . . . . . . . . . . . . . . . . . . . . . . 53

**3 Evaluation and comparison of crop models and methods 55**
3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
3.1.1 Comparing crop modeling methods . . . . . . . . . . . 56
3.2 Chapter approach and objectives . . . . . . . . . . . . . . . . . 57
3.2.1 Considerations for the hindcast comparison . . . . . . . 57
3.2.2 Chapter research questions . . . . . . . . . . . . . . . . 60
3.3 Data and methods . . . . . . . . . . . . . . . . . . . . . . . . . 61
3.3.1 Overview of chapter experimental design . . . . . . . . 61
3.3.2 Sources of wheat production data . . . . . . . . . . . . 62
3.3.3 Sources of climate data . . . . . . . . . . . . . . . . . . 64
3.3.4 Evaluating the statistical crop-climate model . . . . . . . 65
3.3.5 Process-based model: CERES-Wheat . . . . . . . . . . 69
3.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 73
3.4.1 Yield trend analysis . . . . . . . . . . . . . . . . . . . . 74
3.4.2 Evaluating the statistical crop-climate model . . . . . . . 79
3.4.3 Testing the sensitivity of the PCM to environmental modifications . . . . . . . . . . . . . . . . . . . . . . . . 83
3.4.4 Crop model comparison: regional yield hindcasts . . . . 85
3.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
3.5.1 Using statistical crop modeling approaches . . . . . . . 87
3.5.2 Evaluation of the PCM approach . . . . . . . . . . . . . 89
3.5.3 Setting guidelines for crop modeling practices . . . . . . 95
3.6 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 98

**4 Evaluating the added value of downscaled GCM output 101**
4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 101
4.1.1 Simulations of the climate system with the CMIP ensemble of GCMs . . . . . . . . . . . . . . . . . . . . . 102
4.1.2 Added value of regional climate models . . . . . . . . . 104
4.2 Chapter approach and objectives . . . . . . . . . . . . . . . . . 105
4.2.1 Chapter research questions . . . . . . . . . . . . . . . . 106

4.3 Data and methods . . . . . . . . . . . . . . . . . . . . . . . . 106
4.3.1 Overview of chapter experimental design . . . . . . . . 106
4.3.2 Selection of GCMs, RCMs, GCM-RCM combinations . . 107
4.3.3 Statistical analyses and evaluation . . . . . . . . . . . 115
4.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117
4.4.1 Comparing climate model output to observations . . . . 117
4.4.2 Comparison of relative GCM and RCM performance: correlation . . . . . . . . . . . . . . . . . . . . . . . . 125
4.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . 129
4.5.1 Summary of findings . . . . . . . . . . . . . . . . . . . 129
4.5.2 Comparing of GCMs and GCM-RCM simulations . . . . 130
4.5.3 Connecting to the added value debate: advantages of regional climate and impact modeling . . . . . . . . . . 132
4.6 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . 135

**5 Evaluation of RCMs and the effect of bias correction on climate and yield simulations 137**

5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . 137
5.1.1 RCMs and bias correction in impact assessment . . . . 138
5.1.2 Comparing crop model approaches and past yield simulations . . . . . . . . . . . . . . . . . . . . . . . . 139
5.2 Chapter approach and objectives . . . . . . . . . . . . . . . 142
5.2.1 Chapter research questions . . . . . . . . . . . . . . . 142
5.3 Data and methods . . . . . . . . . . . . . . . . . . . . . . . 143
5.3.1 Overview of chapter experimental design . . . . . . . . 143
5.3.2 Observations of climate and yield . . . . . . . . . . . . 144
5.3.3 Simulations of temperature and precipitation from regional climate models . . . . . . . . . . . . . . . . . 145
5.3.4 Bias correction methods . . . . . . . . . . . . . . . . . 146
5.3.5 Crop modeling methods . . . . . . . . . . . . . . . . . 150
5.3.6 Methods of statistical evaluation . . . . . . . . . . . . 150
5.4 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . 153
5.4.1 Evaluating reanalysis-driven RCMs . . . . . . . . . . . 153
5.4.2 Results of the BC of RCM output . . . . . . . . . . . . 161
5.4.3 Comparing the effect of BC on past yield simulations generated by the SCCM . . . . . . . . . . . . . . . . . 176
5.4.4 Comparing the effect of BC on past yield simulations generated by the PCM . . . . . . . . . . . . . . . . . . 181
5.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . 184

5.5.1 Error in reanalysis-driven RCM simulations . . . . . . . 184
5.5.2 Improvement of crop yield simulations through BC . . . 187
5.5.3 Formulating a method for future yield projections . . . . 191
5.5.4 Revisiting the issues and limitations of BC . . . . . . . . 192
5.6 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . 197

**6 Projections of future temperature and precipitation 199**
6.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . 199
6.1.1 Bias correction and its contribution to uncertainty . . . . 200
6.1.2 Chapter approach and objectives . . . . . . . . . . . . . 201
6.2 Data and methods . . . . . . . . . . . . . . . . . . . . . . . 203
6.2.1 Climate models and future emission scenarios . . . . . 203
6.2.2 Bias correction method and approaches . . . . . . . . . 204
6.2.3 Future emission scenarios . . . . . . . . . . . . . . . . 206
6.2.4 Statistical analyses and evaluation . . . . . . . . . . . . 207
6.3 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 208
6.3.1 Projected temperature changes . . . . . . . . . . . . . . 208
6.3.2 Changes to total annual precipitation . . . . . . . . . . . 221
6.3.3 National-level changes to summer (JJA) climate indices 230
6.3.4 Regional changes to summer (JJA) climate indices . . . 236
6.4 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . 243
6.4.1 Summary of projected changes . . . . . . . . . . . . . . 244
6.4.2 Analyzing the effects of BC on projections . . . . . . . . 246
6.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . 257

**7 Multi-method comparison of the projected impacts of climate change on yield 259**
7.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . 259
7.1.1 Comparing the methods of crop models . . . . . . . . . 260
7.1.2 Chapter approach and objectives . . . . . . . . . . . . . 264
7.2 Data and methods . . . . . . . . . . . . . . . . . . . . . . . 265
7.2.1 Climate model output . . . . . . . . . . . . . . . . . . . 265
7.2.2 Review of selected crop models and their calibration . . 267
7.2.3 Yield comparison: statistical methods . . . . . . . . . . 269
7.2.4 Uncertainty decomposition . . . . . . . . . . . . . . . . 270
7.3 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 275
7.3.1 Crop yield projections: comparing the effect of crop model and BC methods . . . . . . . . . . . . . . . . . . 275

7.3.2 Regional comparison of crop yield projections . . . . . . 280
7.3.3 Summary of projected yield changes . . . . . . . . . . . 293
7.3.4 Uncertainty analysis . . . . . . . . . . . . . . . . . . . . . 296
7.4 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . 300
7.4.1 Outlook for future wheat yields . . . . . . . . . . . . . . 300
7.4.2 Comparing crop modeling methods: projected yield changes and trends . . . . . . . . . . . . . . . . . . . . 305
7.4.3 Revisiting evaluation results and their impacts on future yield projections . . . . . . . . . . . . . . . . . . . . . . 308
7.4.4 Advantages and disadvantages of different crop modeling approaches . . . . . . . . . . . . . . . . . . . 310
7.4.5 Other influences on uncertainty . . . . . . . . . . . . . . 313
7.4.6 Novel results and implications . . . . . . . . . . . . . . . 315
7.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . 316

**8 Conclusions 317**
8.1 Revisiting research gaps . . . . . . . . . . . . . . . . . . . . . 318
8.2 General discussion of results . . . . . . . . . . . . . . . . . . 319
8.3 Limitations of the study . . . . . . . . . . . . . . . . . . . . . 324
8.4 Recommendations and conclusion . . . . . . . . . . . . . . . 325
8.4.1 Concluding remarks . . . . . . . . . . . . . . . . . . . . . 327

**Appendix: Climate analysis and RCM evaluation 329**
A. Country, regional, and site climate analysis . . . . . . . 330
B. Seasonal analysis (Hot day index and JJA total precipitation) . . . . . . . . . . . . . . . . . . . . . . . . 338
C. Daily analysis of Tmax, Tmin and precipitation . . . . . 341
Bibliography . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 359

# List of Figures


1.1 Average wheat yields, Europe (1961-2013) . . . . . . . . . . . 5

1.2 Typical simulation process for the impacts of climate change on crops . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

1.3 Cascade of uncertainty (Wilby and Dessai, 2010) . . . . . . . . 13

1.4 Actual simulation process for the impacts of climate change on crops . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

3.1 Overview of Chapter 3 research design. . . . . . . . . . . . . . 62

3.2 Different models for UK and Germany wheat yield trends, 1961-2013 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 76

3.3 Regional yield trends, Germany, 1979-2014 . . . . . . . . . . . 77

3.4 Field-level wheat yield trends, Bad Lauchstädt, 1979-2014 . . . 78

3.5 Observed and SCCM simulated yields, UK and Germany, 1961-2013. . . . . . . . . . . . . . . . . . . . . . . . . . . . 80

3.6 Significant climate predictors for yield in German states. . . . . 81

3.7 Observed and SCCM simulated yields, four German states, 1980-2014. . . . . . . . . . . . . . . . . . . . . . . . . . . . 82

3.8 Simple climate sensitivity validation experiement with BL data and CERES-Wheat. . . . . . . . . . . . . . . . . . . . . . . . 83

3.9 Crop modeling method comparison: SCCM and PCM simulations of regional German wheat yields, 1981-2010. . . . 86

4.1 Overview of Chapter 4 research design. . . . . . . . . . . . . . 107

4.2 Mean annual bias of GCM and GCM-RCM simulations of temperature and precipitation, 1976-2005, UK. . . . . . . . . . 119

4.3 Mean annual bias of GCM and GCM-RCM simulations of temperature and precipitation, 1976-2005, Germany. . . . . . . 119

4.4 Mean annual bias of GCM and GCM-RCM simulations of temperature and precipitation, 1976-2005, DE2. . . . . . . . . 122

4.5 Mean annual bias of GCM and GCM-RCM simulations of temperature and precipitation, 1976-2005, DEA. . . . . . . . . 122

4.6 Mean annual bias of GCM and GCM-RCM simulations of temperature and precipitation, 1976-2005, DED. . . . . . . . . 123

4.7 Mean annual bias of GCM and GCM-RCM simulations of temperature and precipitation, 1976-2005, DEF. . . . . . . . . . . 123

4.8 Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for the UK. 126

4.9 Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for Germany. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 126

4.10 Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for DE2. . . 127

4.11 Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for DEA. . 127

4.12 Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for DED. . 128

4.13 Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for DEF. . . 128

5.1 Overview of Chapter 5 research design. . . . . . . . . . . . . . 144

5.2 Sample Taylor diagram . . . . . . . . . . . . . . . . . . . . . . . 152

5.3 Uncorrected RCM simulations, Tmax, Tmin and Precip, UK, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . . . . 155

5.4 Uncorrected RCM simulations, Tmax, Tmin and Precip, Germany, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . 156

5.5 Uncorrected RCM simulations, Tmax, Tmin and Precip, DE2, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . . . . 157

5.6 Uncorrected RCM simulations, Tmax, Tmin and Precip, DEA, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . . . . 158

5.7 Uncorrected RCM simulations, Tmax, Tmin and Precip, DED, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . . . . 159

5.8 Uncorrected RCM simulations, Tmax, Tmin and Precip, DEF, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . . . . 160

5.9 Bias-corrected RCM simulations, Tmax, Tmin and Precip, UK, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . . . . 164

5.10 Bias-corrected RCM simulations, Tmax, Tmin and Precip, Germany, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . 165

5.11 Bias-corrected RCM simulations, Tmax, Tmin and Precip, DE2, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . . 168

5.12 Bias-corrected RCM simulations, Tmax, Tmin and Precip, DEA, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . . 169

5.13 Bias-corrected RCM simulations, Tmax, Tmin and Precip, DED, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . . 170

5.14 Bias-corrected RCM simulations, Tmax, Tmin and Precip, DEF, 1981-2010 . . . . . . . . . . . . . . . . . . . . . . . . . 171

5.15 DE2 (South Germany) SCCM yield simulation with uncorrected and bias-corrected climate model output. . . . . . 177
5.16 DEA (West Germany) SCCM yield simulation with uncorrected and bias-corrected climate model output. . . . . . . . . . . . . 178
5.17 DED (East Germany) SCCM yield simulation with uncorrected and bias-corrected climate model output. . . . . . . . . . . . . 179
5.18 DEF (North Germany) SCCM yield simulation with uncorrected and bias-corrected climate model output . . . . . . 180
5.19 Comparison of yield observations, SCCM and PCM yield hindcasts . . . . . . . . . . . . . . . . . . . . . . . . . . . . 183

6.1 Overview of Chapter 6 research design. . . . . . . . . . . . . . 204
6.2 Ensemble BC temperature projections, UK. . . . . . . . . . . . 209
6.3 Ensemble BC temperature projections, Germany. . . . . . . . . 210
6.4 Ensemble BC temperature projections, DE2. . . . . . . . . . . 217
6.5 Ensemble BC temperature projections, DEA. . . . . . . . . . . 218
6.6 Ensemble BC temperature projections, DED. . . . . . . . . . . 219
6.7 Ensemble BC temperature projections, DEF. . . . . . . . . . . . 220
6.8 Ensemble BC precipitation projections, UK. . . . . . . . . . . . 223
6.9 Ensemble BC precipitation projections, Germany. . . . . . . . . 224
6.10 Regional ensemble BC precipitation projections. . . . . . . . . 229
6.11 Ensemble BC climate index projections, UK. . . . . . . . . . . . 232
6.12 Ensemble BC climate index projections, UK. . . . . . . . . . . . 234
6.13 Regional ensemble BC hot day climate index projections. . . . 240
6.14 Regional ensemble BC JJA precipitation climate index projections. . . . . . . . . . . . . . . . . . . . . . . . . . . . 241
6.15 Total annual precipitation, past and projected, Germany, CCLM-MPI. . . . . . . . . . . . . . . . . . . . . . . . . . . . 251
6.16 Annual average Tmin, UK, RCA-IPSL . . . . . . . . . . . . . . 252
6.17 Total annual precipitation, past and projected, UK, RACMO-ECEARTH . . . . . . . . . . . . . . . . . . . . . . . . . . . 252
6.18 Annual average Tmax, UK, RACMO-ECEARTH . . . . . . . . . 253

7.1 Overview of Chapter 7 research design. . . . . . . . . . . . . . 266
7.2 Research model of the cascade of uncertainty in yield projections. . . . . . . . . . . . . . . . . . . . . . . . . . . . 270
7.3 Coefficient of variation for SCCM yield projections, UK and Germany. . . . . . . . . . . . . . . . . . . . . . . . . . . . 277
7.4 Country level SCCM projections, comparing BC effect. . . . . . 278
7.5 Coefficient of variation for SCCM and PCM yield projections, German regions. . . . . . . . . . . . . . . . . . . . . . . . . 281

7.6 DE2 yield projections (SCCM, PCM and BC effect). . . . . . . . 282
7.7 DEA yield projections (SCCM, PCM and BC effect). . . . . . . 285
7.8 DED yield projections (SCCM, PCM and BC effect). . . . . . . 288
7.9 DEF yield projections (SCCM, PCM and BC effect). . . . . . . 291
7.10 Summary of regional projected yield changes: crop modeling method and BC comparison for RCP8.5. A * indicates that the mean of the SCCM and PCM change is significantly different (based on a t-test). . . . . . . . . . . . . . . . . . . . . . . . 295
7.11 Summary of regional projected yield changes: method and BC comparison for RCP2.6. A * indicates that the mean of the SCCM and PCM change is significantly different (based on a t-test). . . . . . . . . . . . . . . . . . . . . . . . . . . . . 295
7.12 Ternary plots to represent uncertainty partitions: GCM, BC and crop model. . . . . . . . . . . . . . . . . . . . . . . . . 297
7.13 Uncertainty partitions per region, interval and scenario. . . . . 298
7.14 Average projected yield changes, UK SCCM. . . . . . . . . . . 301
7.15 Average projected yield changes, Germany SCCM). . . . . . . 304
7.16 Projected yield changes, German regions. . . . . . . . . . . . . 304

A1 UK climate averages, 1961-2013. . . . . . . . . . . . . . . . . . 333
A2 Germany climate averages, 1961-2013. . . . . . . . . . . . . . 333
A3 South Germany (DE2) climate averages, 1979-2014. . . . . . . 334
A4 West Germany (DEA) climate averages, 1979-2014. . . . . . . 334
A5 East Germany (DED) climate averages, 1979-2014. . . . . . . 335
A6 North Germany (DEF) climate averages, 1979-2014. . . . . . . 335
A7 Bad Lauchstädt climate averages, 1978-2014. . . . . . . . . . 336
A8 Climate indices in the UK and Germany, 1961-2013. . . . . . . 336
A9 Summer (JJA) climate indices in German regions, 1979-2014. 337
A10 Empirical cumulative and probability distribution function plots for UK Tmax, Tmin and Precip. . . . . . . . . . . . . . . . . . . 347
A11 Empirical cumulative and probability distribution function plots for UK Tmax, Tmin and Precip. . . . . . . . . . . . . . . . . . . 348
A12 Empirical cumulative and probability distribution function plots for UK Tmax, Tmin and Precip. . . . . . . . . . . . . . . . . . . 349
A13 Empirical cumulative and probability distribution function plots for UK Tmax, Tmin and Precip. . . . . . . . . . . . . . . . . . . 350
A14 Empirical cumulative and probability distribution function plots for UK Tmax, Tmin and Precip. . . . . . . . . . . . . . . . . . . 351
A15 Empirical cumulative and probability distribution function plots for UK Tmax, Tmin and Precip. . . . . . . . . . . . . . . . . . . 352

A16 Taylor diagrams, daily Tmax, Tmin and Precip, UK and Germany353
A17 Taylor diagrams, daily Tmax, Tmin and Precip, German regions 354

# List of Tables


3.1 Selected German regions, and NUTS code. . . . . . . . . . . . . 64

3.2 Models used to determine yield evolution trends. . . . . . . . . 68

3.3 CERES-Wheat minimum data requirements (Jones et al., 2003). 71

3.4 CERES-Wheat (DSSAT) experimental values (Jones et al., 2003) and regional German cultivar coefficients (Nain and Kersebaum, 2007). . . . . . . . . . . . . . . . . . . . . . . . . 71

3.5 Summary evaluation statistics for national yield trend analysis. 76

3.6 Summary evaluation statistics for regional yield trend analysis. 77

3.7 Summary evaluation statistics for field-level yield trend analysis. 78

3.8 Summary statistics for national SCCMs. . . . . . . . . . . . . . 80

3.9 Coefficient values for regional German SCCMs. . . . . . . . . . 81

3.10 Yield responses to environmental modifications in experimental validation. . . . . . . . . . . . . . . . . . . . . . . 84

3.11 Summary statistics, CERES-Wheat (PCM) simulations compared to yield observations. . . . . . . . . . . . . . . . . . 87

4.1 GCM and RCM selection and combinations. . . . . . . . . . . . . 112

4.2 $RMSE$ between historical GCM Tmax, Tmin and Precip to observations. . . . . . . . . . . . . . . . . . . . . . . . . . . . 120

4.3 $RMSE$ between historical GCM-RCM simulations Tmax, Tmin and Precip to observations. . . . . . . . . . . . . . . . . . 120

4.4 $RMSE$ between historical GCM Tmax, Tmin and Precip to observations, German regions, 1976-2005. . . . . . . . . . . . 124

4.5 $RMSE$ between historical GCM-RCM Tmax, Tmin and Precip to observations, German regions, 1976-2005. . . . . . . . . . . 124

5.1 Selected regional climate models (RCMs) from EURO-CORDEX for evaluation simulations. . . . . . . . . . . . . 145

5.2 Statistical comparison of annual averages of maximum temperature from RCM evaluation simulations and observations, UK and Germany, 1981-2010. . . . . . . . . . . . . 163

5.3 Statistical comparison of annual averages of minimum temperature from RCM evaluation simulations and observations, UK and Germany, 1981-2010. . . . . . . . . . . . 163
5.4 Statistical comparison of total annual precipitation from RCM evaluation simulations and observations, UK and Germany, 1981-2010. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 166
5.5 Country-level KS statistics . . . . . . . . . . . . . . . . . . . . . 166
5.6 Statistical comparison of annual averages of maximum temperature from RCM evaluation simulations and observations, German regions, 1981-2010. . . . . . . . . . . . 172
5.7 Statistical comparison of annual averages of minimum temperature from RCM evaluation simulations and observations, German regions, 1981-2010. . . . . . . . . . . . 173
5.8 Statistical comparison of total annual precipitation from RCM evaluation simulations and observations, German regions, 1981-2010. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 174
5.9 German regional KS statistics . . . . . . . . . . . . . . . . . . . 175
5.10 Statistical evaluation between yield simulations, a yield hindcast generated with E-OBS, and yield observations, DE2 (South Germany). . . . . . . . . . . . . . . . . . . . . . . . . 177
5.11 Statistical evaluation between yield simulations, a yield hindcast generated with E-OBS, and yield observations, DEA (West Germany). . . . . . . . . . . . . . . . . . . . . . . . . . 178
5.12 Statistical evaluation between yield simulations, a yield hindcast generated with E-OBS, and yield observations, DED (East Germany). . . . . . . . . . . . . . . . . . . . . . . . . . 179
5.13 Statistical evaluation between yield simulations, a yield hindcast generated with E-OBS, and yield observations, DEF (North Germany). . . . . . . . . . . . . . . . . . . . . . . . . . 180

6.1 Paired GCM-RCM combinations for future climate projections with available RCP scenarios. . . . . . . . . . . . . . . . . . . 204
6.2 Summary of ensemble projected temperature changes (UK). . 209
6.3 Summary of ensemble projected temperature changes (Germany). . . . . . . . . . . . . . . . . . . . . . . . . . . . . 210
6.4 GCM-RCM annual projected Tmax changes for the UK, in °C and in percentage. . . . . . . . . . . . . . . . . . . . . . . . 214
6.5 GCM-RCM annual projected minimum temperature changes for the UK, in °C and in percentage. . . . . . . . . . . . . . . . 214
6.6 GCM-RCM annual projected Tmax changes for Germany, in °C and in percentage. . . . . . . . . . . . . . . . . . . . . . . 215
6.7 GCM-RCM annual projected minimum temperature changes for Germany, in °C and in percentage. . . . . . . . . . . . . . . 215
6.8 Summary of ensemble projected temperature changes (DE2). . 217

6.9 Summary of ensemble projected temperature changes (DEA). 218
6.10 Summary of ensemble projected temperature changes (DED). 219
6.11 Summary of ensemble projected temperature changes (DEF). 220
6.12 GCM-RCM annual projected precipitation changes for the UK, in mm and in percentage. . . . . . . . . . . . . . . . . . . . . . 223
6.13 GCM-RCM annual projected precipitation changes for Germany, in mm and in percentage. . . . . . . . . . . . . . . . . 224
6.14 GCM-RCM annual projected precipitation changes for German regions, in mm and in percentage. . . . . . . . . . . . 227
6.15 GCM-RCM projected changes to the number of hot days for the UK, in number of days and percentage. . . . . . . . . . . . 232
6.16 GCM-RCM projected JJA precipitation changes for the UK, in mm and in percentage. . . . . . . . . . . . . . . . . . . . . . . . 233
6.17 GCM-RCM projected changes to the number of hot days for Germany, in number of days and percentage. . . . . . . . . . . 234
6.18 GCM-RCM projected JJA precipitation changes for Germany, in mm and in percentage. . . . . . . . . . . . . . . . . . . . . . 235
6.19 GCM-RCM projected changes to the number of hot days in German regions, in days and in percentage. . . . . . . . . . . . 238
6.20 GCM-RCM projected changes to total JJA precipitation in German regions, im mm and and in percentage. . . . . . . . . 239
6.21 Summary of significant differences, with respect to uncorrected projections (RCP8.5 and 2.6). . . . . . . . . . . . . 242

7.1 UK SCCM crop yield projections. . . . . . . . . . . . . . . . . . 279
7.2 Germany SCCM crop yield projections. . . . . . . . . . . . . . 279
7.3 South Germany (DE2) SCCM crop yield projections. . . . . . . 283
7.4 South Germany (DE2) PCM crop yield projections . . . . . . . 283
7.5 West Germany (DEA) SCCM crop yield projections. . . . . . . 286
7.6 West Germany (DEA) PCM crop yield projections. . . . . . . . 286
7.7 East Germany (DED) SCCM crop yield projections. . . . . . . . 289
7.8 East Germany (DED) PCM crop yield projections. . . . . . . . 289
7.9 North Germany (DEF) SCCM crop yield projections. . . . . . . 292
7.10 North Germany (DEF) PCM crop yield projections. . . . . . . . 292
7.11 Significant yield differences, German regions . . . . . . . . . . 294
7.12 Fractional uncertainty, RCP8.5. . . . . . . . . . . . . . . . . . . . 299
7.13 Fractional uncertainty, RCP2.6. . . . . . . . . . . . . . . . . . . . 299

A1 Temperature trend analysis. . . . . . . . . . . . . . . . . . . . . . 331

A2 Statistical comparison between seasonal (summer, June-August or JJA) climate indices, UK and Germany, 1981-2010. (* is $p$ <0.05). . . . . . . . . . . . . . . . . . . . . . 338
A3 Statistical comparison between seasonal (summer, June-August or JJA) hot day index (days above 31°C), German regions, 1981-2010. (* is $p$ <0.05). . . . . . . . . . . 339
A4 Statistical comparison between seasonal (summer, June-August or JJA) total precipitation, German regions, 1981-2010. (* is $p$ <0.05). . . . . . . . . . . . . . . . . . . . . . 340
A5 Statistical comparison between daily values of RCM evaluation simulations and observations, UK and Germany, 1981-2010. (* is $p$ <0.05). . . . . . . . . . . . . . . . . . . . . . 341
A6 Kolmogorov-Smirnov (KS) test statistics on the distribution of daily maximum and minimum temperature, and precipitation from RCM evaluation simulations and observations, UK and Germany, 1981-2010. . . . . . . . . . . . . . . . . . . . . . . . 342
A7 Statistical comparison of daily values of maximum temperature from RCM evaluation simulations and observations, German regions, 1981-2010. (* is $p$ <0.05). . . . 343
A8 Statistical comparison of daily values of minimum temperature from RCM evaluation simulations and observations, German regions, 1981-2010. (* is $p$ <0.05). . . . . . . . . . . . . . . . . 344
A9 Statistical comparison of daily values of precipitation from RCM evaluation simulations and observations, German regions, 1981-2010. (* is $p$ <0.05). . . . . . . . . . . . . . . . . 345
A10 Kolmogorov-Smirnov (KS) test statistics on the distribution of daily maximum and minimum temperature, and precipitation from RCM evaluation simulations and observations, German regions, 1981-2010. . . . . . . . . . . . . . . . . . . . . . . . . 346
A11 SCCM yield simulations, DE2 (South Germany). . . . . . . . . 355
A12 SCCM yield simulations, DEA (West Germany). . . . . . . . . . 355
A13 SCCM yield simulations, DED (East Germany). . . . . . . . . . 356
A14 SCCM yield simulations, DEF (North Germany). . . . . . . . . 356
A15 PCM yield simulations, DE2 (South Germany). . . . . . . . . . 357
A16 PCM yield simulations, DEA (West Germany). . . . . . . . . . . 357
A17 PCM yield simulations, DED (East Germany). . . . . . . . . . . 358
A18 PCM yield simulations, DEF (North Germany). . . . . . . . . . 358

# Chapter 1

# Introduction

"The quest for food security can be the common thread that links the different challenges we face and helps build a sustainable future."

---

*José Graziano da Silva, United Nations Food and Agriculture Organization (FAO) Director-General*

Food security – generally defined as the availability and physical, social and economic access to food – is an underlying objective in fulfilling the global goals for sustainable development that are agreed upon by world nations. Although it is not an explicit Sustainable Development Goal, availability and access to clean, healthy food for all is tantamount to eliminating hunger and achieving better health, greater sustainability, and equality. However, attaining sufficient food for all is, and will continue to be, challenging. Current challenges to food production include natural resource constraints to growing food, its large environmental costs and demands, and the increasing consumption and dietary shifts of a growing global population.

Where and when food can be grown in the world is largely dictated by geography: namely, the type of climate and soil, human influence on the area, and cropping intensity of crop systems (Iizumi and Ramankutty, 2015, Beauregard and De Blois, 2014, Leff et al., 2004). Agricultural production, specifically crop production, demands a sufficient supply of solar radiation, appropriate temperatures, and adequate water necessary for plant growth, in addition to arable soil and land. The sensitivity of crops to climate means that extreme weather and changes to expected climate patterns – such as drought, heat waves, floods and tropical cyclones – can cause significant losses and damages to livelihood assets and crops (Porter et al., 2014). For example, climate-related disasters are among the main drivers of food insecurity, in the short- and long-term period after a climate hazard (Porter et al., 2014).

Human activity has greatly increased the concentration of greenhouse gases (GHGs) in the atmosphere. The rise in GHGs, together with other anthropogenic drivers, is extremely likely to have been the dominant cause of the observed warming since the mid-20th century (Summary for Policymakers, Intergovernmental Panel on Climate Change Fifth Assessment Report (IPCC AR5), 2013). An increase in the mean and/or variance of temperature results in potentially more of both hot and cold temperature extremes, whereas changes to the distribution of precipitation could result in an increase in mean precipitation (Cubasch et al., 2013). This could also increase heavy precipitation extremes and the duration of dry spells between precipitation events (Cubasch et al., 2013). While these extreme weather events may not be new to farmers and food producers, climate change may have impacts outside the realm of collective historical experience (Vermeulen et al., 2013).

Climate change and variability therefore have significant potential impacts on food security (Synthesis Report of the IPCC, 2014). Current climate variability already accounts for roughly a third of observed yield variability at the global scale for major crops such as maize, rice, wheat and soybean (Ray et al., 2015, Lobell and Field, 2007). In response to these projected changes, how can food producers prepare, adapt, or transform crop production systems in anticipation of climate change and its impacts? How and when will climate change affect food systems in a particular area, or a specific crop?

Numerous scientific studies have been conducted to find answers or provide evidence in response to these two questions, focusing on a variety of important agricultural crops, communities, and disciplinary emphasis (e.g. physical or economic impacts, plant crop growth and development, or sociopolitical responses). In answering these questions and developing the research niche, it is important to acknowledge the vast amount of knowledge on crop-climate relationships gained from existing studies, as well as remaining challenges in projecting the impacts of climate change.

This study is therefore motivated by the challenge posed by climate change on agricultural production, the need for knowledge on its projected impacts, and the opportunity to provide guidance to support climate change adaptation. In this thesis, climate and crop modeling are used to provide evidence of how important the impact assessment methods are in understanding how climate change will potentially affect wheat, which is one of the most important European food crops.

In addition, the focus of the work is to critically examine the current methods used in climate-crop research that are used to attain wheat yield projections. The aim of this chapter is firstly to discuss the importance of wheat production in Europe and the sensitivity of the wheat crop to different climatic variables, in order to highlight the significance of the study. In addition, the concepts of uncertainty and communities of practice are introduced, which are major themes for the research work because of how they can influence the outcomes of impact assessment studies.

## 1.1 Background of the study

### 1.1.1 Global significance of wheat

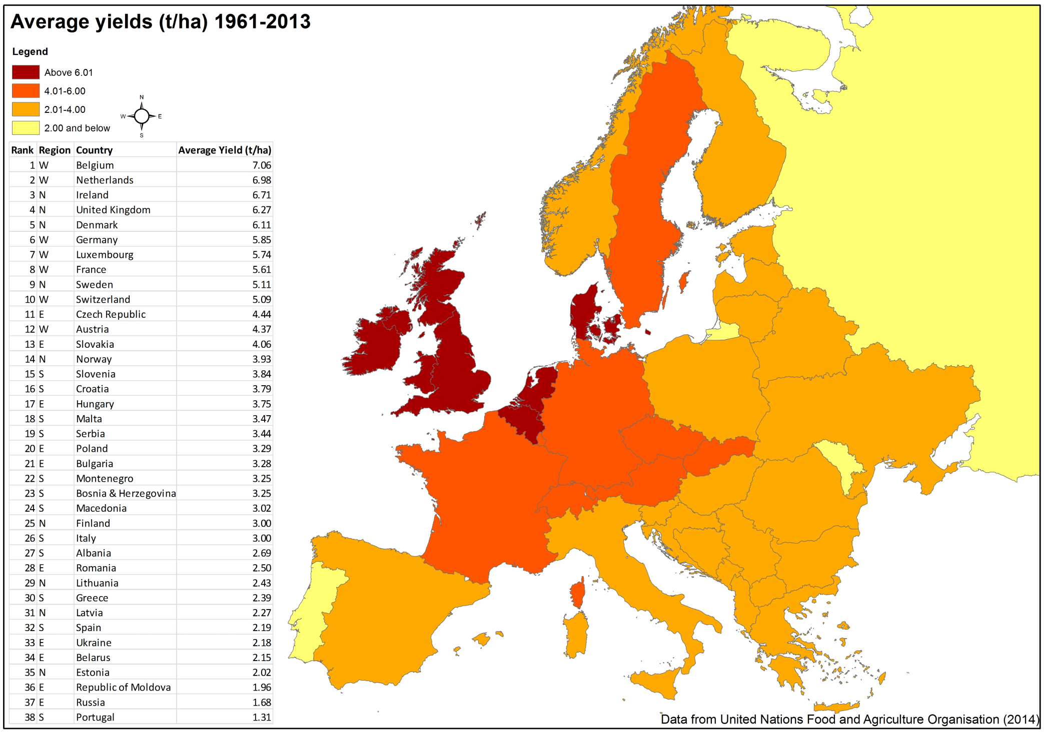
Wheat is one of the most important food and feed crops in the world. Approximately 21% of the world's food depends on wheat (*Triticum aestivum*) (Ortiz et al., 2008, Högy and Fangmeier, 2008). Wheat is a widely grown cereal crop, second only to rice in terms of production (Trnka et al., 2015). Humans directly consume more than 60% of wheat that is produced globally, thus wheat supplies approximately 20% of the energy and about 25% of the protein requirements of the world population (Högy and Fangmeier, 2008).

Where wheat is cultivated and grown successfully in the world is dependent on many factors, and one of these important influencing factors is climate. The relationship between climate and wheat is particularly significant to discuss because of concerns that climate change will adversely affect agricultural production, and thus food security. In addition to growing competition for land, water, and energy, the effects of climate change are seen as a further threat to food production (Godfray et al., 2010). Extreme weather events are already a significant challenge for grain producers, and they are predicted to increase in future climate scenarios (Barlow et al., 2015).

The global importance of wheat and its role in food security makes understanding the impacts of current and future climate change and variability a priority for food security research. In particular, wheat is an important crop in Europe, which is one of the largest producers of wheat in the world.

### 1.1.2 Wheat production in Europe

Europe is responsible for up to 25% of the global wheat area and 29% of global wheat production (Koehler et al., 2013, Trnka et al., 2014, 2015). Wheat is considered the main crop in France, the United Kingdom, and Germany (Gornott and Wechsung, 2016, Michel and Makowski, 2013, Semenov et al., 2012, Brisson et al., 2010). The favorable climate of Europe, in addition to intensive management practices, has contributed to regionally high yields in France, the UK and Germany in Western Europe, and also some of the highest yields globally (Wrigley et al., 2016). Based on data on wheat production and planted area from the United Nations Food and Agriculture Organization (FAO, 2014), wheat yields are shown to be highest in Western European countries (Fig. 1.1).




| Rank | Region | Country | Average Yield (t/ha) |
|---|---|---|---|
| 1 | W | Belgium | 7.06 |
| 2 | W | Netherlands | 6.98 |
| 3 | N | Ireland | 6.71 |
| 4 | N | United Kingdom | 6.27 |
| 5 | N | Denmark | 6.11 |
| 6 | W | Germany | 5.85 |
| 7 | W | Luxembourg | 5.74 |
| 8 | W | France | 5.61 |
| 9 | N | Sweden | 5.11 |
| 10 | W | Switzerland | 5.09 |
| 11 | E | Czech Republic | 4.44 |
| 12 | W | Austria | 4.37 |
| 13 | E | Slovakia | 4.06 |
| 14 | N | Norway | 3.93 |
| 15 | S | Slovenia | 3.84 |
| 16 | S | Croatia | 3.79 |
| 17 | E | Hungary | 3.75 |
| 18 | S | Malta | 3.47 |
| 19 | S | Serbia | 3.44 |
| 20 | E | Poland | 3.29 |
| 21 | E | Bulgaria | 3.28 |
| 22 | S | Montenegro | 3.25 |
| 23 | S | Bosnia & Herzegovina | 3.25 |
| 24 | S | Macedonia | 3.02 |
| 25 | N | Finland | 3.00 |
| 26 | S | Italy | 3.00 |
| 27 | S | Albania | 2.69 |
| 28 | E | Romania | 2.50 |
| 29 | N | Lithuania | 2.43 |
| 30 | S | Greece | 2.39 |
| 31 | N | Latvia | 2.27 |
| 32 | S | Spain | 2.19 |
| 33 | E | Ukraine | 2.18 |
| 34 | E | Belarus | 2.15 |
| 35 | N | Estonia | 2.02 |
| 36 | E | Republic of Moldova | 1.96 |
| 37 | E | Russia | 1.68 |
| 38 | S | Portugal | 1.31 |

*Figure 1.1: Average wheat yields, Europe (1961-2013).*

#### 1.1.2.1 Wheat production trends

Wheat yields in Europe have experienced a steady growth trend since the 1950s, mostly because of improvements to genetics and farm practices, such as the use of fertilizers, irrigation and mechanization (Michel and Makowski, 2013). However, it has been shown that wheat yields in Europe are experiencing stagnation or a 'yield plateau'. Analysis of wheat yield time series have shown that yields in Europe have been stagnated since the early to mid-1990s, particularly in countries which contribute significantly to the supply of wheat in Europe, such as France, Germany and England (Michel and Makowski, 2013, Brisson et al., 2010). Could this plateau already be caused by climate change? In order to respond to this question, firstly, there needs to be an understanding and review of the relationship between wheat yields and climate.

### 1.1.3 Sensitivity of wheat to weather and climate

Climate and weather directly affect the growth and development of food crops like wheat (Porter and Semenov, 2005), which is sensitive to extremes of temperature, excess and shortage of water, and other changes to optimal growing thresholds. Wheat is also sensitive to the occurrence of drought, late spring frosts and of severe winter frosts associated with inadequate snow cover (Trnka et al., 2015). The sensitivity of wheat to climate, particularly to temperature and precipitation, is seen as a primary reason for its vulnerability to climate change and variability. In this section, wheat physiology is briefly explained in terms of its sensitivity to temperature and precipitation.

#### 1.1.3.1 Yield responses to temperature

While wheat can be grown in a variety of climates around the world, it has an optimum temperature range between 17-23°C (Porter and Gawith, 1999).

Rising mean temperatures due to increased GHG emissions are anticipated to push wheat beyond optimal growing temperature ranges and subsequently reduce wheat yields. This is because warming temperatures accelerate the wheat crop towards maturity, thereby reducing the period of time that the crop has to accumulate grain mass; in addition, warming accelerates leaf aging and leaf death (Asseng et al., 2014, Hawkins et al., 2013a, Lobell et al., 2012, Asseng et al., 2011).

Wheat yield is determined by grain number and size, which are established around the flowering period (anthesis), a stage that is sensitive to high temperatures (Semenov and Shewry, 2011). Short periods of high temperatures around anthesis can substantially reduce the grain yield for heat-sensitive wheat cultivars due to heat-driven grain sterility and grain abortion (Barlow et al., 2015, Semenov et al., 2012), leading to poor yields. Wheat is also sensitive to extreme hot or cold temperatures, as it stops growing below 0°C and above 37°C (Porter and Gawith, 1999).

The large diversity of wheat varieties are also affected by climate in different ways. Generally, wheat varieties are qualitatively classified into two types: firstly, winter wheat, which has a low-temperature requirement called vernalization that is needed in order to commence flowering and thus have successful grain reproduction; secondly, spring wheat, which does not have this requirement (Li et al., 2013). Warming temperatures are important for parts of Europe that grow winter wheat, which are generally more high-yielding than spring wheat varieties (Thorup-Kristensen et al., 2009).

Compared to varieties of spring wheat, winter wheat has been shown to be more vulnerable to increasing temperatures during winter seasons because of its need for vernalization (Li et al., 2013). Vernalization requirements vary with the variety (cultivar) but are typically less than 8°C, with an optimum of 5-6°C for 2-4 weeks (Li et al., 2013). Extremely low temperatures, such as less than -15°C, can kill seedlings (Li et al., 2013, Porter and Gawith, 1999). Failed or insufficient vernalization in winter wheat can delay dormancy, which then delays the onset of the reproductive stage of winter wheat (Wang et al., 2015).

#### 1.1.3.2 Yield responses to precipitation

In addition to its sensitivity to warming and extreme temperatures, wheat is also sensitive to the absence or excess of precipitation, which are both negative influences on wheat yield. Shortage of water is a chief cause of variation in wheat yields in many parts of the world (Jamieson et al., 1998). The prolonged absence of water (drought) is the most significant environmental stressor to agriculture worldwide (Semenov and Shewry, 2011). Heavy or extreme precipitation can also have negative effects on wheat production, primarily due to the impacts of waterlogging, which can reduce yields by about 12-20% due to depleted oxygen in the ground (Li et al., 2016). This can result in insect infestations and plant diseases, causing crop losses and large economic costs (Li et al., 2016).

Rainfall also has indirect impacts on yield. Wet or cool weather can enhance disease occurrence and complicate crop management practices related to wheat harvest or sowing (Trnka et al., 2015). For example, when the ground is too wet, farmers have to decide between timely planting or harvest against the long-term compaction damage caused by driving on wet soil (Wolkowski and Lowery, 2008).

#### 1.1.3.3 Wheat responses to large-scale climate variability

Weather and climate also affect wheat grain quality. The preceding winter North Atlantic Oscillation (NAO) can have effects on the specific weight of grain – the higher the specific weight, the greater the weight of grain that can be loaded into a container (Kettlewell et al., 2003). The NAO is the large-scale alternation in air pressure between northern and southern regions of the North Atlantic Ocean (Kettlewell et al., 2003), and it is a leading pattern of weather and climate variability over the Northern Hemisphere (Hurrell and Deser, 2009).

Because of the influence of the NAO on winter surface climate, a strong association between the winter NAO index and specific weight ($r$=0.64) was found in a UK site (Kettlewell et al., 2003). When this mechanism was investigated further, however, it was reported that sunshine during grain growth and late summer precipitation during grain ripening, are the most important climatic factors determining specific weight of harvested UK wheat, meaning that NAO effects on the early life of the crop (i.e. during winter months) do not appear to have substantial effects on specific weight (Atkinson et al., 2005).

#### 1.1.3.4 Other factors and influences on yield

The complexity of real-world wheat systems means that there are many more factors apart from climate that affect wheat production. Factors such as genetics (that determine the wheat variety), soil, and management are important influences on yield. It has been reported that local factors such as farm and field management are reported to contribute more to yield variability than climate (Porter and Semenov, 2005). In addition, policy and agricultural practice changes in Europe, such as the extension of cultivated areas or the implementation of the Common Agricultural Policy (CAP) in the European Union (EU) are also thought to also have influenced yield stagnation (Brisson et al., 2010, Moore and Lobell, 2014). Therefore, while climate may play a role in European yield stagnation (Moore and Lobell, 2014), it is evident that the complexity and dynamics of agricultural production needs to be considered and analyzed when assessing the potential impacts of climate change.

### 1.1.4 Risk of climate change and variability to wheat

The sensitivity of wheat to climate is seen as a reason for the vulnerability of the crop to present and future climate variability and change. Some findings report that although there is heterogeneity across European wheat yield trends, climate has already been found to have negative impacts, an approximately 2.5% reduction in wheat yields since 1989 (Moore and Lobell, 2014). There is also evidence that long-term temperature and precipitation trends also account for approximately 10% of the stagnation in European wheat yields (Moore and Lobell, 2015).

Climate change is a threat to wheat production because it leads to changes in the frequency, intensity, spatial extent, duration, and timing of extreme weather and climate events (Special Report on Climate Extremes and the IPCC Fifth Assessment Working Group II Report, 2014, 2012). Extreme weather events such as heat waves, droughts, excessive cold, and heavy and prolonged precipitation can have significant impacts on agricultural production (van der Velde et al., 2012), making climate change and its potential impacts significant threats to agriculture. Increasing temperatures and drought incidence associated with global warming are posing serious threats to food security (Lobell et al., 2013).

Alongside climate change, the world's population and its consumption of food continue to increase. The global population is predicted to exceed nine billion by 2050 and there is increasing concern about the capability of agriculture to feed such a large population (Michel and Makowski, 2013). Climate change presents a considerable challenge in achieving the targeted 70% needed increase in world food production (Semenov et al., 2014). Thus, agricultural adaptation is needed in order to reduce the negative impacts of climate change on crop yields and to maintain food production (Tanaka et al., 2015). In order to offer evidence to support adaptation, this study focuses on climate change and its impacts on wheat yield as a measure of production.

### 1.1.5 Typical approaches to investigate the impacts of climate change on wheat production

The importance of wheat as a food staple and its sensitivity to climate have led to numerous studies in the field of crop-climate research to better understand and characterize the relationship between climate and crop yields. Many scientific studies have attempted to project the potential impacts of climate change and variability on different aspects of wheat, from development and phenology to yield quality and quantity (e.g. as reviewed by White et al., 2011). However, developing crop yield projections is not a straightforward process, although the typical impact assessment may seem simple. The crop yield projection simulation process involves the use of several types of models, including those from climate and crop science disciplines. The typical simulation process uses climate model output as input to crop impact models, which are used to project future changes to climate and crops (White et al., 2011), as shown in Fig. 1.2.

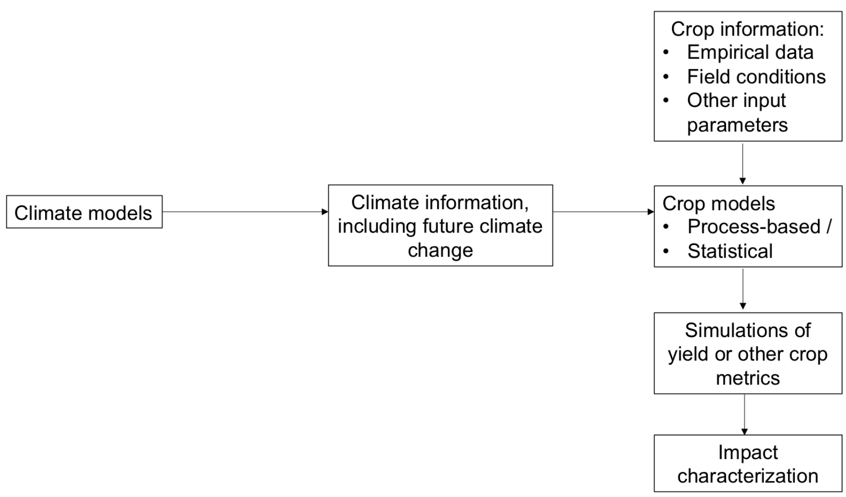



*Figure 1.2: Typical simulation process for the impacts of climate change on crops, where climate model output is used as input to a crop model to simulate the selected crop metric, such as yield.*

Crop models can be categorized into one of two approaches: process-based crop models – which represent crop development based on the mechanisms of plant interactions with weather and soil, as well as field management – or statistical models, which use empirical data on climate and yields to quantify relationships between them. In addition to the data and input needed by crop models to simulate the development and growth of crops like yield, they also need climate input, including scenarios for future climate change. Therefore, in order for crop models to be able to represent and project the impacts of climate change, agricultural and climate data are crucial (Ramirez-Villegas et al., 2013). Simulations of future climate are also dependent on models. These complex representations, or models, of the atmosphere and oceans are called general circulation models or global climate models (GCMs) (Maraun et al., 2015). Future outlooks of agricultural production and food security are therefore contingent on the skill of GCMs in reproducing seasonal rainfall and temperatures (Ramirez-Villegas et al., 2013).

While GCMs provide the capability to project climate change based on future emission scenarios of greenhouse gases (GHGs), and crop models likewise have been shown capable of simulating plant growth and development in many scientific analyses of the impacts of climate change, the projected impacts on crop yields are considered inherently uncertain (Asseng et al., 2013). But what is uncertainty and how does it affect yield projections?

### 1.1.6 Uncertainty as a major theme of the research

The definition of uncertainty in this work is adopted from the IPCC Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (IPCC SREX, 2012):

"[Uncertainty is] an expression of the degree to which a value

> or relationship is unknown. Uncertainty can result from lack of information or from disagreement about what is known or even knowable. Uncertainty may originate from many sources, such as quantifiable errors in the data, ambiguously defined concepts or terminology, or uncertain projections of human behavior."

Following this definition, it has been proposed that there are three general sources of uncertainty in approaches to understand the impacts of climate change on crops like wheat: (1) climate modeling, (2) crop modeling, and (3) the connections between them (Ruiz-Ramos and Mínguez, 2010). This is because crop models carry uncertainty from climate models and the methods used to link climate and crop models, leading to crop projections that have accumulated uncertainty. This uncertainty propagation through impact models is also known as the cascade of uncertainty (Wilby and Dessai, 2010, Fig. 1.3).

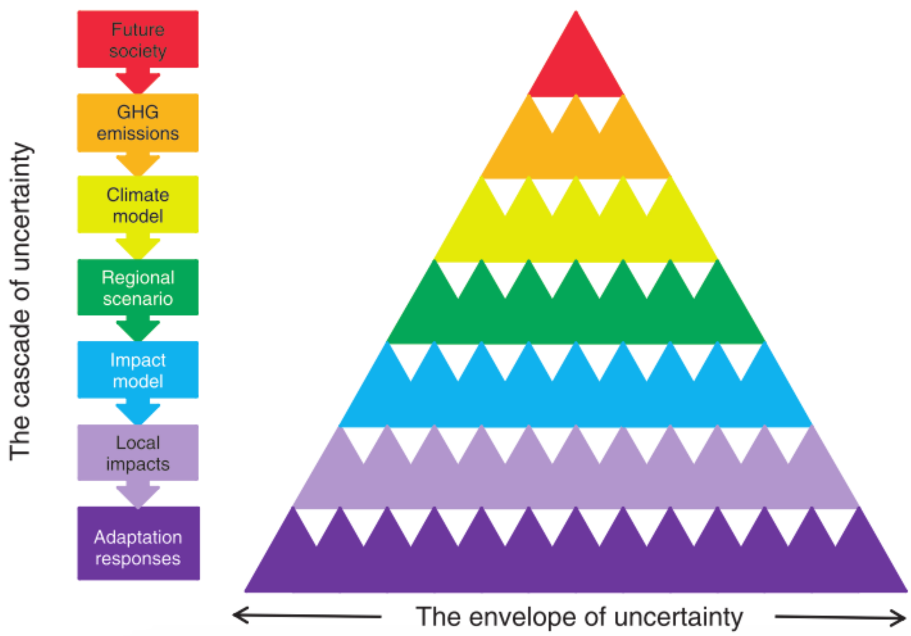



*Figure 1.3: A cascade of uncertainty proceeds from different future pathways, their associated greenhouse gas (GHG) concentrations, the resulting climate outcomes in global and regional models, and how these are translated into local impacts and ideal adaptation responses. The increasing number of triangles at each level symbolize expanding envelope of uncertainty. Figure from Wilby and Dessai (2010).*

The cascade of uncertainty begins with the choice between numerous plausible scenarios of future society and their potential GHG emissions, to climate models, to impact models and their different methods (Wilby and Dessai, 2010). Local impacts and possible adaptation options are at the bottom of the cascade, where uncertainty accumulates. Therefore, the typical or idealized process of impact assessment can propagate uncertainty. It is argued that the importance of wheat globally, and regionally in Europe, makes it crucial that methods that aim to understand future changes can accurately capture crop and climate relationships while identifying sources of error and uncertainty. This is important to be able to generate more confidence in them, as impact assessment studies may be used as evidence for the basis of adaptation.

Based on the work carried out in this research, the typical simulation approach to impact assessment (Fig. 1.2) is argued to have many intermediate steps and decisions that need to be made to proceed from one step to another, leading to an actual approach filled with numerous other processes. Each of these intermediate steps, such as downscaling, bias correction, utilizing different climate forcings (on the climate side) to crop model calibration and evaluation (on the crop modeling side) all contribute to the cascade of uncertainty in yield projections (Fig. 1.4).

Given the important role that uncertainty plays in decision-making (Vermeulen et al., 2013, Lemos and Rood, 2010, Wilby and Dessai, 2010), and the uncertainty cascade associated with impact assessment (Hawkins et al., 2013a, Wilby and Dessai, 2010, Challinor et al., 2010, Tsvetsinskaya and Mearns, 2003), there is a need to critically review current methods – and these intermediate steps in the typical assessment – for simulating the impacts of climate change with a focus, or lens, on uncertainty. Focusing on underlying uncertainties can also elucidate steps in the cascade where uncertainty can be better characterized, or where there are research gaps.

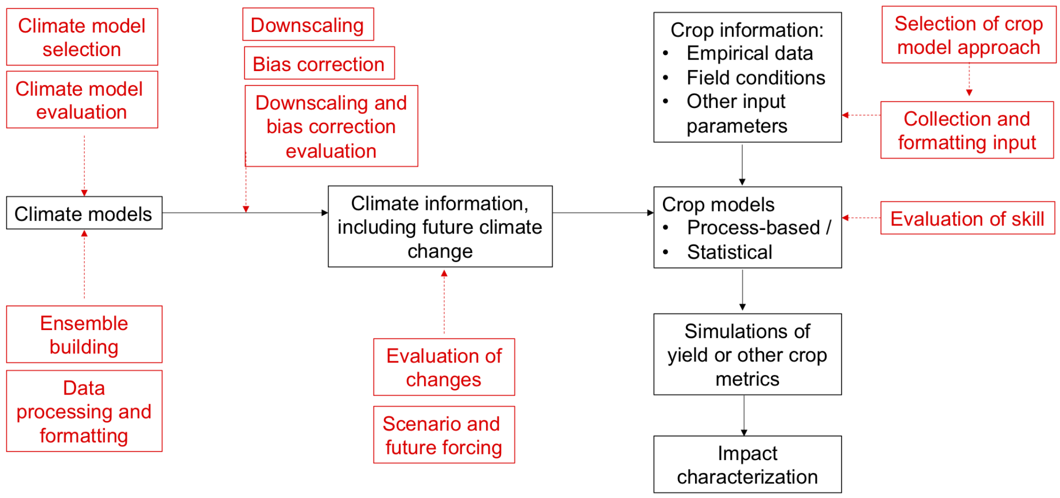



*Figure 1.4: Actual simulation process for the impacts of climate change on crops, where numerous intermediate steps to link climate model output to crop models are needed to generate the selected crop metric.*

## 1.2 Diverging communities of practice

Many of the intermediate steps in the 'actual' impact assessment cascade (Fig. 1.4) vary depending on the underlying reasons for the selection of the simulation approach, which also depends on the disciplinary orientation of the study. Because of the development trajectories of climate impact research, there are also different communities of practice that have developed around the use of climate and crop models. In this work, the definition of a community of practice is adopted as "a group [or groups] of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly" (Wenger, 1998).

Communities of practice in climate-crop modeling often use different or constrasting methods to simulate the impacts of climate change on sectors such as agriculture; for example, methods that are acceptable or common to one community may be a long-standing debate in the other. In addition, even within these disciplinary communities, there are smaller subgroups that approach impact assessment in different ways. For example, there are long-standing debates within the climate community on whether using techniques to derive higher-resolution climate model output can provide more useful information on future climate change in local and regional areas, which is discussed in Chapter 2.

In this project, additional differences between climate modeling communities of practice were found in how errors from GCMs are dealt with. For instance, while GCMs are powerful simulation tools, they are still simplifications of complex atmospheric and oceanic systems, so they are likely to contain errors relative to the real world. Some communities of practice within the climate modeling community believe that these errors should be corrected before their use in impact assessment; still others believe that doing so does not address underlying climate model error and in fact contributes to more uncertainty.

In addition, there are divisions between the communities of practice of crop modeling. As further explained in the next chapter, research around the relationship between crops and their environment is clearly divided between groups that use process-based models that can consider numerous factors, including genetics and management as influences on crop growth and development. Still other communities utilize relatively straightforward empirical/statistical models as a basis of impact assessments. Within these subgroups of crop modeling communities, there are also numerous frameworks, systems, and methods which may be vastly different from others even within the same disciplinary approach – all of which are here argued to also contribute to diverging outcomes in climate change impact assessment.

In the following chapter, these issues are expanded upon in a review of literature to better elucidate where research gaps are, and what can be undertaken in this research to address these gaps and fulfill reseach aims.

# Chapter 2

# Literature review and defining research gaps

## 2.1 Introduction

In the previous chapter, it was discussed how the combination of different methods and the numerous decisions made in the typical impact simulation process result in impact assessment projections, for example for crop yields, that are inherently uncertain due to the cascade of uncertainty. In addition, the diverging practices of research communities that utilize impact assessment methods may also contribute to a range of plausible future scenarios for climate change studies. These valuable differences in models, methods and communities in the field of impact assessment have resulted in a number of comparative analyses that focus on better understanding and characterizing uncertainty.

To further the knowledge gained from these studies, in this chapter, it is the objective to firstly outline what the differing methods and models are to provide a common research framework throughout the thesis, as well as to highlight important research gaps that can be investigated further.

### 2.1.1 Chapter objectives

In this chapter, the uncertainty of crop yield projections is discussed as stemming from three different sources, following the framing of Ruiz-Ramos and Mínguez (2010): from climate models, crop models, and the linkages between them. It is argued here that this framing indicates that apart from crop yield projections being 'inherently uncertain' (e.g. Asseng et al., 2013), the field of climate-crop impact assessment is also inherently multi- and inter-disciplinary. Impact studies combine the climate sciences with crop agronomy and agricultural sciences, and they also utilize many statistical tools to reach the end-goal of yield projections. This means that research in impact studies requires a common understanding of key concepts and tools.

To bring a common understanding to these important disciplinary concepts, in this review of literature, firstly, climate models are reviewed, in addition to a discussion of downscaling methods that are used to change global climate model simulations to a more regional scale. The method of bias correction, which is the use of statistical approaches to improve the output of climate models as a post-processing step, is also reviewed, in light of the ongoing discussion and criticism that it does not address underlying climate model error, nor that bias correction methods are appropriate for use in future climate projections.

The methods of crop modeling are also reviewed and compared. As previously indicated, crop modeling methods are roughly divided into process-based (or mechanistic) crop models and statistical approaches. Although both disciplinary approaches have been used extensively in crop-climate research, they are fundamentally different and efforts to directly compare their output are still relatively new in the research discipline. Lastly, the third source of uncertainty – the linkages between climate and crop models – is also discussed.

After this discussion, opportunities for sharpening the research focus are explored, with the intention to fulfill the research aims of offering evidence and recommendations to characterize sources of error in crop yield projections. The chapter ends with an overview of the research gaps, questions, and design that is a result of the review of related literature. Importantly, this chapter also defines the scope of what is investigated in the research study.

## 2.2 Uncertainty from climate models

Among the possible sources of crop projection uncertainty according to Ruiz-Ramos and Mínguez (2010), the first source is from climate models. GCMs generally show satisfactory performance for many large-scale features of climate (Flato et al., 2013), and through downscaling (through regional climate models, or RCMs) are also able to provide more spatial or temporal detail at a regional scale (e.g. for Europe, Kotlarski et al., 2014, Jacob et al., 2014). However, despite the immense simulation power and knowledge gained from GCMs, it is well known that climate models have numerous limitations that may lead to error and uncertainty in downstream impact projections.

For example, it has been shown that many biases in yield projections are related to errors in driving GCMs (Glotter et al., 2014). Climate models, while state-of-the-art tools, are limited in their capacity to realistically simulate all the components of the atmosphere and oceans, and thus need numerous parameterized processes (Flato et al., 2013). Uncertainties in the projections of climate change impacts on future crop yields derive from different sources in climate modeling, for example diverse GCM construction and parameterization, future emissions scenarios, and inherent or response uncertainty (Asseng et al., 2013).

In this section, the purposes and development of climate models are reviewed, giving focus to sources of error and uncertainty, developing regional climate simulations, and correcting errors in climate model output.

### 2.2.1 Introduction to climate models

According to the IPCC Fifth Assessment Report (Working Group I, Flato et al., 2013), climate models represent the most current understanding of the dynamics of the physical components of the climate system, particularly the atmosphere and oceans. Climate models are the primary tools that are used to investigate the responses of the climate system to various forcings, including future GHGs and aerosols. GCMs are used to make climate projections on seasonal to decadal time scales and for making projections of future climate for the coming century and beyond (Flato et al., 2013). Essentially, climate models are mathematical and physical expressions of the atmosphere and oceans. Numerical methods are then used to solve these discretized mathematical expressions, which are implemented on a grid (Flato et al., 2013).

Climate models are also essential in the detection and attribution of observed changes in climate. Observations unequivocally indicate that the earth has warmed (IPCC Summary for Policymakers, 2013). While a simple approach to detection and attribution of climate change would be to compare climate observations with model simulations driven with natural forcings to simulations driven with both natural and anthropogenic forcings, climate models must firstly be able to correctly simulate the response of the atmosphere and oceans (Bindoff et al., 2013). This is a strong assumption of the ability of GCMs. Therefore, while climate models may not get the exact magnitude of the response correctly, there is general consensus that models can simulate the shape – meaning large-scale patterns – of the response to external forcings (Bindoff et al., 2013).

Climate models are important tools in research because of their capabilities to represent complex atmospheric and oceanic processes, alongside their ability to be used for attribution experiments. Because of their importance in climate impact studies, therefore, it is crucial to evaluate the performance of these models, both individually and collectively (Flato et al., 2013). How well climate models perform is usually evaluated by comparing the model output to observations and analyzing the resulting difference. In this regard, the IPCC (2013) reports very high confidence that climate models are able to reproduce observed large-scale mean surface temperature patterns, and known large-scale climate features.

### 2.2.2 Sources of error in climate models

While climate model performance may be satisfactory, it is not without several shortcomings. A well-known example is that climate model simulations of precipitation perform less well compared to surface temperature (Flato et al., 2013). Errors in GCM simulations of precipitation can affect the simulated precipitation intensity. This can lead to a low number of dry days, which are compensated by too much drizzle (Piani et al., 2010). GCM errors can also result in biases in mean precipitation and poorly represented extreme events (Piani et al., 2010).

Some errors are a result of the coarse resolution (spatial and temporal grid characteristics) of large-scale climate models. The resolution of GCMs is usually of a grid cell resolution that is around 200 kilometers (Ekström et al., 2015), which is too coarse to resolve finer-scale features that affect crop growth and production. While recent developments have resulted in high-resolution GCMs with resolutions of approximately 50 km grid-point spacing, the large computational cost of running high-resolution models means that they have been performed at only a few research centers (Haarsma et al., 2016). Therefore, although the knowledge of atmospheric and ocean processes continues to advance alongside increases in the computational power needed to run GCMs, climate models still have limitations.

These limitations can be sources of error which propagate through to impact models. The sources of model error are summarized as the following: uncertainty in process representation, error propagation, sensitivity to resolution, uncertainty in observations, and other factors (IPCC, 2013).

**(1) Uncertainty in process representation**

There is still limited understanding of very complex processes of the atmosphere and oceans that need to be included in GCMs. In addition, it is a challenge to represent these complex processes mathematically and in a manner that preserves their physics (Flato et al., 2013). As a result, conceptual representatations, or parameterizations, are needed to include processes that occur at spatial and/or temporal scales that are not explicitly resolved (Flato et al., 2013). A wide range of processes must be parameterized, for example those associated with atmospheric convection, clouds, aerosols, ocean and sea ice dynamics, as well as radiation (Flato et al., 2013).

In particular, the representation of clouds, a key component in the atmospheric system, is considered problematic. This is due to several reasons, for example that the simulation of clouds with GCMs involves many nonlinear processes spanning a large range of spatial and temporal scales (Lauer and Hamilton, 2013). When modeling the response of global mean surface temperature to the doubling of atmospheric carbon dioxide ($CO_2$), clouds are among the leading causes of uncertainty in estimates produced by GCMs (Tan et al., 2016). Therefore, while the latest climate models now include more cloud and aerosol processes and their interactions (Flato et al., 2013), there remains low confidence in the representation of clouds and other complex processes in models.

**(2) Error propagation**

Biases can be propagated by climate models that use parameterized processes. Parameterization of physical processes in the atmosphere or oceans, for example those dealing with clouds, could lead to errors or uncertainties in simulations of radiation (Flato et al., 2013). Other examples, such as biases in the position of storm tracks, are partly due to sea-surface temperature (SST) biases in simulations, which are related to problems with the simulated location of warm waters such as the Gulf Stream and Kuroshio Current (Booth et al., 2017, Greeves et al., 2007, Keeley et al., 2012). For instance, a cold SST bias in the Pacific and a lack of El Niño Southern Oscillation (ENSO) variability lead to large changes in the Pacific storm track (Greeves et al., 2007). This mispositioning of storm tracks leads to errors that may propagate to climate simulations, leading to errors in the simulation of precipitation, for example (Wilby et al., 2009).

**(3) Sensitivity to resolution**

Some aspects of the climate system are found to be dependent on the scale – the horizontal or vertical resolution – of climate model simulations. While higher model resolution generally leads to mathematically more accurate models, it entails higher computational costs, and does not necessarily translate to more reliable simulations (Flato et al., 2013). However, higher-resolution models have been shown to improve the representation of the Gulf Stream, and Kuroshio Current (Haarsma et al., 2016, Ma et al., 2016, Kirtman et al., 2012), which as mentioned in the previous paragraph, are important in climate dynamics and have influence over other simulated climate variables. Higher resolution in the atmospheric component of some models has been shown to improve features such as storm tracks and extratropical cyclones, extreme precipitation, and tropical cyclone intensity and structure (Haarsma et al., 2016, Flato et al., 2013), making the resolution of GCMs an important influence on the presence and size of errors in simulations.

**(4) Uncertainty in observations**

In some cases, the observations used to evaluate climate model simulations are of insufficient length or quality. The normal and accepted length of climate data is 30 years, as defined by the World Meteorological Organization (WMO, Arguez and Vose, 2011). A shortage or lack of high-quality observational data to compare with model simulations, for example in topographically diverse or remote areas, can make the evaluation of model performance challenging (Flato et al., 2013). Recent advances include greater use of remote sensing (e.g. satellite imagery) technologies. Satellite data has provided major advances in understanding the climate system and its changes through added observational data, and satellite data are frequently used with climate models to simulate the dynamics of the climate system and to improve climate projections (Yang et al., 2013). However, this is also limited by spatial sampling over long periods of time, biases in sensors, and the need to also validate with other observations (Yang et al., 2013).

**(5) Other factors that contribute to uncertainty**

Aside from the uncertainty that can arise from numerous parameter values, uncertainty from climate models also arises because of different model formulations, internal variability, or boundary conditions (Flato et al., 2013). Model simulations are also affected by how they are forced, for example uncertainties in GHGs, aerosols emissions, or land use change, can all affect model results (Flato et al., 2013). Different statistical methods used in model evaluation can also lead to differences in the assessment of model quality (Flato et al., 2013).

In the case of future climate projections, scenario uncertainty also arises because future development and emission trajectories are also uncertain. Future scenarios of GHGs and their radiative forcings are used with climate models to generate estimates of climate change. These scenarios are currently implemented through the use of the representative concentration pathways (RCPs). Succeeding the emission scenarios described in the IPCC Special Report on Emission Scenarios (SRES), the RCPs are a set of four pathways developed for the climate modeling community as a basis for long-term and near-term modeling experiments (van Vuuren et al., 2011).

RCPs provide information on possible development trajectories for the climate change forcings (Moss et al., 2010, van Vuuren et al., 2011) and are currently used by the latest IPCC Fifth Assessment Report and recent impact studies. RCPs contain not only emission and $CO_2$ concentration trajectories, they also consider land-use, and can be used to explore alternative energy and technology futures (van Vuuren et al., 2011, Moss et al., 2010). Future climate simulations forced by the RCPs are then used to drive numerous different impact models, such as hydrological, economical, or crop models. For example, the RCPs are used with impact models from the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP) (Warszawski et al., 2014), which seeks to compare different impact models in various sectors of interest.

Based on all these differences and sources of potential uncertainty, it is clear that making use of climate models and their output requires, at least, a systematic and scientific method of comparing and evaluating climate models.

### 2.2.3 Evaluating climate model performance

Despite the limitations of climate models, climate models have continued to show significant improvement (IPCC, 2013). Some of these improvements come from the better understanding and subsequent incorporation of other important components of the earth system into climate models. For example, the current state-of-the-art climate models are called Earth System Models (ESMs), which expand on GCMs by including the representation of various biogeochemical cycles such as the carbon cycle or the sulfur cycle. Some features of the land surface are also included, such as vegetation, soil type and water bodies (Flato et al., 2013). The climatic effects of these land surface variables can be profound: for example, it has been suggested that changes in the state of the land surface, in particular soil moisture, may have played an important part in the severity and length of the 2003 European drought (García-Herrera et al., 2010, Fischer et al., 2007).

As a means to characterize these limitations which lead to uncertainties and error, ensemble approaches are frequently used. These ensemble approaches can be either Multi-model Ensembles (MMEs) or Perturbed Parameter (or Physics) Ensembles (PPEs). MMEs are created from existing model simulations from various climate models where a multi-model mean can be calculated. In contrast, PPEs are created to assess uncertainty based on a single model and benefit from the explicit control on parameters (Flato et al., 2013). By controlling different parameters for a single model in a PPE, statistical methods can determine which parameters are the main drivers of uncertainty across the ensemble.

A prominent example of an MME is the Coupled Model Intercomparison Project, now in Phase 6 (CMIP6). CMIPs are a central element of national and international assessments of climate change, for example the IPCC AR5 from 2013 (Eyring et al., 2016). However, even MMEs have limitations, for example, their evaluation can be confounded by the fact that some climate models share a common lineage and thus share common biases (Flato et al., 2013).

#### 2.2.3.1 Improving climate models

Improvements to climate models are also driven by increases in high-performance computing capabilities. Increased computing power has enabled the investigation of the impacts of increased resolution of climate models on simulated mean climate and its variability. The High Resolution Model Intercomparison Project (HighResMIP), part of CMIP6, uses climate models with increased horizontal resolution (Haarsma et al., 2016). Models in HighResMIP have shown significant improvements in the simulation of aspects of large-scale circulation phenomena such as the ENSO, tropical instability waves, the Gulf Stream, and their respective influences on the atmosphere (Haarsma et al., 2016). Other large-scale features such as the global water cycle, snow cover, the Atlantic inter-tropical convergence zone (ITCZ), the jet stream, storm tracks, and Euro-Atlantic blocking have also shown improvements with higher-resolution models (Haarsma et al., 2016).

### 2.2.4 Simulating regional climate

The capability of GCMs to represent and simulate climate implies that their output could potentially be directly used in impact models. However, as discussed, their resolution is too coarse, as a resolution of 100-250 km is too large to resolve features that are important at regional scales. For example, GCM precipitation output cannot be used to directly force impact models without some form of prior calibration of the uncorrected climate model output (Hawkins et al., 2013a, Piani et al., 2010). Because of the coarse scale of GCMs, methods to bridge the scale gap are needed (Ekström et al., 2015). These methods are known as downscaling methods.

Downscaling attempts to resolve the scale discrepancy between GCM grid cell resolution and the fine-scale resolution required for local and regional impact assessment (Maraun et al., 2010).

#### 2.2.4.1 Downscaling methodologies

Downscaling methodologies are generally divided into dynamical downscaling and statistical downscaling, which have significantly different approaches to providing more spatial or temporal detail to GCM simulations. Dynamical downscaling nests a regional climate model (RCM) into the GCM to represent the physical processes of the atmosphere with a higher grid box resolution (Maraun et al., 2010). RCM simulations are richer in spatial and temporal detail compared to GCMs (Flato et al., 2013), so they are often used to provide more detailed information for a particular geographical region. The analysis of several downscaled climate model experiments from the Coordinated Regional Downscaling Experiment over Europe (EURO-CORDEX, Jacob et al., 2014) confirmed the ability of RCMs to capture the features of European climate, including its variability in space and time (Kotlarski et al., 2014).

Alternatively, statistical downscaling involves deriving empirical relationships (as transfer functions) to link large-scale atmospheric variables and local/regional weather conditions (Kotlarski et al., 2014). Statistical downscaling is a popular alternative for use in impact studies because of its relative ease of use, and its performance is comparable to output from RCMs (Eden and Widmann, 2014). Statistical downscaling methods also constitute a range of techniques to provide regional or local detail.

However, while both these methods are valid approaches to bridge the scale gap, they are vastly different with regard to complexity. Thus, the single umbrella term of 'climate downscaling' can be somewhat simplistic, as different downscaling methods can produce information with dissimilar properties with regard to the climate change signal contained in the GCM output, and to what is required by the end-user (Ekström et al., 2015). Depending on the needs of the end-user, one particular method may be considered more suitable. In order to find common ground between dynamical and statistical downscaling, recent climate impact applications suggest that a combination of the two approaches is optimal (Kotlarski et al., 2014, Maraun et al., 2010). This combination of statistical and dynamical downscaling is called Model Output Statistics (MOS).

For a chosen climate model, MOS infers a correction function between a simulated climate variable and its corresponding observed value in the present-day climate, and applies this correction function to a future simulation with the same model (Wong et al., 2014). The comparison of statistical, dynamical downscaling and their combination is the focus of projects such as VALUE, which is a comprehensive effort to assess the credibility of regional climate change scenarios (Maraun et al., 2015).

#### 2.2.4.2 Increasing importance of regional climate simulations

The use of MOS is already common in numerical weather prediction, and it is gaining more prominence in downscaling climate change scenarios (e.g. Eden and Widmann, 2014). However, dynamical downscaling of GCM output is already considered a well-established and standard technique for the generation of regional climate change scenarios (Kotlarski et al., 2014). Apart from their role in the development of climate scenarios, RCMs have become important tools that help to advance the understanding of regional-scale climate processes (Kotlarski et al., 2014). RCMs are thus increasingly important tools in assessing the impacts of climate change. For example, progress in RCM research has made the use of RCM simulations for hydrological studies more attractive (Teutschbein and Seibert, 2012).

The progress of RCM intercomparison projects such as those available with EURO-CORDEX have made many RCM simulations available and more accessible for use in impact studies, for example for crop production. Because crop modeling studies rely on the accuracy of climate input data, they are sensitive to the downscaling method (Ramarohetra et al., 2015). The common experimental design of EURO-CORDEX can therefore facilitate comparison between other RCMs used to simulate European climate in impact studies for crop production.

However, systematic GCM errors have been shown to propagate into RCM output, leading to errors in simulations (Glotter et al., 2014). For example, if the GCM misplaces storm tracks, this leads to errors in the simulation of precipitation by an RCM (Wilby et al., 2009). Even RCM simulations do not often agree with local observations, and their output are not directly useful for assessing impacts at the catchment scale for hydrological impact studies (Teutschbein and Seibert, 2012). It is thus argued that this lack of agreement of simulations to observations – error or bias – is critical to the confidence in the impact assessment studies that utilize them. Based on the literature review, the quality of regional climate simulations depends not only on the validity of the physics and methods behind the RCM or downscaling technique, but also, and perhaps more critically, on the quality and realism of the boundary information from the GCM.

Error and bias from climate models can have effects on projections of impacts, as climate information is cascaded from one step to the next. The number of permutations of emission scenario, climate model, and downscaling method also proliferates uncertainty at each stage of the simulation process (Wilby and Dessai, 2010). Although downscaling does result in simulations with a higher spatial resolution (approximately 11-25 km for the EURO-CORDEX experiments), RCMs often inherit the biases from the GCMs (Maraun et al., 2010). Therefore, while the simplest way may be to directly utilize the uncorrected GCM or RCM output for driving impact models, such as crop models for agricultural production, there are biases between the simulations and reality which should be corrected (Hawkins et al., 2013a).

### 2.2.5 Bias correction to improve agreement of simulations with observations

Bias correction (BC), also sometimes referred to as statistical downscaling, post-processing, or calibration, addresses these errors in climate model simulations by improving their mean, distribution or shape to bring them closer to observations. BC is an attempt to make the GCM output more realistic (Hawkins et al., 2013a, Piani et al., 2010). Based on the literature review, how BC affects climate change projections is an important question in many scientific studies, particularly because BC is thought to modify the physical consistency and climate change signal (e.g. Maraun et al., 2017, Hempel et al., 2013, Piani et al., 2010). These issues with BC are an important point of discussion, and are revisited later in the Chapter, and across other chapters as a key focus of the research.

In a review of methods for hydrological impact studies, Teutschbein and Seibert (2012) describe some approaches to BC of climate model output: methods can range from simple scaling approaches like linear scaling, which makes RCM simulations agree with the monthly mean values of observations, to methods like quantile-quantile mapping that involve modifying the shape of the distribution of simulations. Extensions to linear methods can also correct the variance, while distribution or quantile mapping attempts to remove quantile-dependent biases (Maraun, 2013).

Stochastic weather generators (WGs) are another method of post-processing climate simulations. They figured prominently as a common method for modifying weather variables in a review of methodologies for simulating impacts of climate change on crop production (White et al., 2011). WGs such as the Long Ashton Research Station Weather Generator (LARS-WG) (Semenov and Barrow, 1997) – the most common WG used in crop modeling studies (White et al., 2011) – are capable of generating a synthetic weather time series with the statistical properties of observations, in order to generate long enough records of weather simulations, or fill in gaps in existing records (Mehan et al., 2017, Semenov and Barrow, 1997).

Although WGs were first developed for hydrological application, they have been widely used to investigate the influence of weather conditions on crop yields, and continue to figure in impact studies (e.g. Mehan et al., 2017). However, despite the utility of weather generators and suggested continued research and support for investigating stochastic WGs for BC (e.g. Maraun, 2016), they are not investigated in the research. Rather, in this thesis, linear, variance or quantile-quantile mapping BC methods that directly modify uncorrected projections are used.

### 2.2.6 Climate change projections and uncertainty

In summary, climate models are a useful way of simulating past, present and future climate, conducting attribution experiments, and contributing to the scientific understanding of our complex atmosphere and oceans. However, due to their limitations, and different development and structural designs, GCMs can have and propagate error, leading to uncertainty in climate change projections and further 'downstream' into crop projections. For example, it has been found that errors in yield projections are dominated by GCM systematic errors (Glotter et al., 2014), meaning that GCM error cascades into impact projections.

Additional uncertainty in climate change projections ranges from uncertainty in future emissions of GHGs, the range GCM responses to these specified emissions, combined with the natural, internal variability of the climate (Hawkins et al., 2013b, Hawkins and Sutton, 2009, 2011). In addition, there is uncertainty in the choice of calibration method in producing climate data for the impact model, including BC procedures (Hawkins et al., 2013b). Projections of future crop production accumulate uncertainties from GCMs and RCMs, which are passed onto their simulations of climate, and eventually to crop model simulations, in addition to being affected by the uncertainties of the crop models and methods themselves. In the following section, crop models are also focused on with an uncertainty lens.

## 2.3 Uncertainty from crop models

Uncertainty in crop yield simulations can come from the chosen crop impact models and methods that aim to realistically simulate plant growth and responses to climate. Uncertainty is introduced by the chosen crop model(s) approach and representation of crop growth and development. In this section, two fundamentally different approaches of crop modeling are compared and critically reviewed in order to break down and understand the uncertainty that results from using crop models to assess the impacts of climate change.

### 2.3.1 Representing crop growth and responses to climate

Because of their ability to model crop growth and development based on climate and other input variables, crop models are considered essential tools in the assessment of climate change impacts to local and global food production (Asseng et al., 2014). Most climate change impact studies use a similar methodological approach using crop models which may seem relatively 'straightforward' in principle (White et al., 2011, p.357). Firstly, GCMs are used to generate climate projections for time periods many years or decades into the future based on a selected emissions scenario. The output of climate models is used as input to a crop model that is able to translate the relationship between climate and crops into measurable impacts. Following the generation of future climate from climate models, one provides the crop model with the field conditions, crop information and future climate model simulations from GCMs. The crop model is then 'run' and outputs are then compared to observations or different crop simulations (White et al., 2011).

However, while the typical modeling approach may seem straightfoward, in reality there are numerous and nontrivial intermediate processes that are needed to link climate and crop models. Linking climate model output to crop models involves processes such as scaling the GCM output due to its typically coarse resolution to a scale that is used by crop models. In addition to this downscaling step to provide information at the appropriate spatio-temporal scale, corrections to errors in climate simulations are also often required in order to improve them relative to observations.

The chosen climate variables that need to be downscaled and corrected depend largely on the choice of crop modeling method and crop model, as there are numerous approaches to simulate crop growth and development and their responses to climate. In this review of literature, studies have been shown to use different crop modeling methods with varying complexity, input demands, and corresponding output related to crop processes.

In general, crop models are categorized as either process-based crop models (PCMs), or statistical crop-climate models (SCCMs), which are fundamentally different approaches to understanding crop yield (Lobell and Asseng, 2017, Liu et al., 2016). PCMs attempt to provide explanations of crop systems' behavior relative to changes in the environment (Angulo et al., 2013), in contrast to statistical approaches which link observed crop parameters to climate (Lobell and Burke, 2010). Because of their contrasting approaches, it is valuable to compare crop modeling methods and models, for example through their simulated yields, in order to understand how different crop modeling methods contribute to error and uncertainty.

Despite the fact that both approaches seek to quantify the impacts of climate change on agricultural productivity, there have been relatively few attempts to systematically compare findings from both approaches (Moore et al., 2017). While both PCMs and SCCMs are prevalent in the field of crop-climate research, there continues to be a clear divide between the communities that use them. For instance, the results from either approach are often published in different disciplinary journals (Lobell and Asseng, 2017) and it is only recently that there have been large-scale scientific efforts to methodically compare their differences (e.g Liu et al., 2016), among other method comparison studies (e.g. Soltani et al., 2016, Watson et al., 2015). Based on scientific research using crop models, the contrasting approaches and mechanisms of (1) PCMs and (2) SCCMs are as follows:

**(1) Process-based crop models (PCMs)**

Process-based crop models, also known as crop simulation models, dynamic growth models, or mechanistic crop models, are considered the state-of-the-art tools for simulating crop growth and development. They represent the most current understanding of crops, and they can integrate knowledge on physiology, agronomy, soil science and agrometeorology (Shi et al., 2013). PCMs attempt to explain not only the relationship between crop parameters and simulated variables, but also the mechanisms of the described processes that are relevant to plant growth and development (Palosuo et al., 2011). PCMs are able to consider dynamic interactions between genotype, environment, and management factors (Angulo et al., 2013). They use mathematical equations to describe physiological, physical and chemical processes to simulate crop growth and development over time (Shi et al., 2013). PCMs are considered useful tools in climate impact studies as they deal with multiple climate factors and how they interact with crop processes (Asseng et al., 2014).

There are numerous crop models that have been developed in research centers around the world that are available to the end-user for a variety of crops, climates, and simulation objectives, each with their own set of modeled processes, input parameters, and generated output. Therefore, the choice in crop model may lead to uncertainty, which is a major component of uncertainty in yield projections and is considered the most difficult source of uncertainty to quantify (Palosuo et al., 2011). In order to compare crop model structure, similar to MMEs used in climate studies, there are several studies that evaluate how different PCMs perform compared to one another for various crops, locations and crop variables (e.g. Angulo et al., 2014, Palosuo et al., 2011, Jamieson et al., 1998). A large global MME of crop models that is used for future climate change assessments is the Agricultural Model Intercomparison and Improvement Project (AgMIP), which focuses on using PCMs. AgMIP uses seven global gridded crop models for a coordinated set of simulations of global crop yields under climate change (Rosenzweig et al., 2013, 2014).

**(2) Statistical approaches to modeling crop-climate relationships**

Statistical approaches are also used to model climate and crop relationships. Statistical models are often thought to be capable of assessing climate change impacts on crop production rapidly and across large datasets (Rosenzweig et al., 2013). Statistical approaches are based on empirical data of crop yields and weather to develop models to relate these variables to each other (Lobell and Asseng, 2017, Lobell and Burke, 2010). Rather than plant processes being explicitly modeled, as with a

PCM, crop and climate data are used to calibrate regression equations that describe the relationship between crops (typically an aspect of crop production, such as crop yield), and climate. Statistical approaches and models have become increasingly common in recent years with the growing availability of data on both weather and crops (Lobell and Asseng, 2017).

Various types of SCCMs have been used for the analysis of yield time series. Linear regression has been used in many studies but other regression models, such as quadratic regression, bi-linear, tri-linear, and linear-plus-plateau models, have been used in a smaller number of papers (Michel and Makowski, 2013). SCCMs can be used in a variety of scales via three methods: through time series methods, which are based on time series data from a single point or area; panel methods, which are based on variations both in time and space; lastly, cross-section methods which are based on variations in space (Lobell and Burke, 2010).

SCCMs have been used at a variety of scales, locations, and crops, for example: maize in France and Africa (e.g. Hawkins et al., 2013a, Lobell and Burke, 2010, respectively), soybean, cotton and maize in the United States (e.g. Schlenker and Roberts, 2009), and at a global scale for several crops (e.g. Ray et al., 2015).

### 2.3.2 Challenges in crop modeling approaches

While both PCMs and SCCMs have been used in numerous scientific studies, their performances, and thus confidence in their yield simulations, are still constrained by several limitations. For instance, these methods have significant differences between their structure, represented processes, and calibration needs. In order to further contrast the two approaches, some of the main limitations of both PCMs and SCCMs – summarized as differences in calibration parameter demands, scale mismatch, upscaling parameters, aggregation error, and stationarity – are discussed next.

**(1) Calibration parameter differences**

In many of the studies reviewed for this chapter, PCMs have been considered the primary tools for simulating crop growth and development. However, a caveat of their powerful simulation capability is that they require extensive fine-scale input data in order to function. PCMs often require data on cultivar, management, and soil conditions that are unavailable in many parts of the world (Lobell and Burke, 2010). Even in the presence of such data, PCMs can be difficult to calibrate because of large numbers of parameters (Lobell and Burke, 2010). For example, the well-evaluated and used CERES-Wheat crop model which is part of the Decision Support Systems for Agrotechnology Transfer (DSSAT, Jones et al., 2003) requires fine-scale information on soil type, planting depth, row spacing, and several genetic parameters that relate to specific cultivars (crop varieties).

As an alternative to PCMs, SCCMs are considered to have advantages due to their limited reliance on field calibration data, and their transparent assessment of model uncertainties (Lobell and Burke, 2010). However, because of their relative simplicity, SCCMs have difficulty offering process-level understanding and testing of adaptation strategies, so extrapolation beyond the observations is considered risky (Rosenzweig et al., 2013). Some crop modeling studies call for 'appropriate complexity' for both SCCMs and PCMs (e.g. Hawkins et al., 2013a, Challinor et al., 2009). Models need to be complex enough to adequately represent cropping systems. However, an 'appropriate' level of complexity is called for, because the more processes are simulated (or the more statistical predictors used), the greater the number of potential interactions between them and the number of parameters that require calibration, thereby increasing the potential for error (Challinor et al., 2009).

The next three limitations of crop models are closely linked together, and have to do with the scale where PCMs and SCCMs are generally applied.

**(2) Scale mismatch**

PCMs require a large amount of fine-scale calibration data because they were originally developed to provide decision support at the field scale. Despite this original intended scale design, PCMs have become common tools in assessing agricultural impacts and adaptation to climate variability and change at larger scales beyond the field (Palosuo et al., 2011). This results in a scale mismatch between their design and implementation. A scale mismatch leads to a number of persistent challenges, such as the need to upscale input parameters on crop growth and development from the field to other larger scales (Angulo et al., 2013, Palosuo et al., 2011). Scaling up parameters and input data means that valuable information on the environmental conditions where crops are grown could be lost and smoothed in the process. However, most PCMs that are available 'off-the-shelf' for general impact assessment end-users are set at the field scale, although they have been applied at many scales larger than this (e.g. Challinor et al., 2017). A potential research gap is thus identified here: how does the application of PCMs at larger (regional) scales affect yield projections?

In contrast, statistical approaches can be used to directly link various scales of observations of crops and climate, and are often used at a regional or country scale where they are thought to perform well (e.g. Lobell and Burke, 2010). While SCCMs may be limited for finer-scale response compared to PCMs, the plausibility of field-scale simulations of climate change impacts on yield should also be questioned, since projections at field scales are extremely uncertain (Lobell and Burke, 2010).

**(3) Upscaling parameters**

Related to the scale difference, upscaling and deriving parameters to larger scales is considered a significant challenge with the PCM approach.

It has been hypothesized that many large-scale crop model applications that assess climate impacts and adaptation options for crops involve huge uncertainties related to the model parameters and structure (Palosuo et al., 2011). In addition, the reproducibility of crop model outputs is also an issue, as input parameters for PCMs are not always harmonized nor documented in a transparent manner (Balkovič et al., 2013).

To obtain realistic simulations, it is recommended that crop model parameters should be derived from field experiments where measurements were taken (Therond et al., 2011). However, there are a limited number of field experiments, due to their extensive and costly nature (Farina et al., 2011). This may lead to calibration of PCMs with commonly used field-based experimental datasets, which may propagate further bias (Balkovič et al., 2013). In their experiment, Therond et al. (2011) did additional calibration work for their simulations to match regional phenology (plant developmental stage) dates. While this improved the yield simulations, this was still not sufficient to reproduce regional observed yields (Therond et al., 2011), which highlights significant shortcomings in upscaling PCM parameters.

SCCMs are thought to have advantages over PCMs in this respect due to their limited reliance on field experiment data for calibration, and their transparent assessment of model uncertainties (Lobell and Burke, 2010).

**(4) Aggregation errors**

Also related to scale and upscaling, an additional source of error in using both PCMs and SCCMs at large scales is data aggregation. Most large-scale applications of PCMs have some way of considering the spatial variability of input data such as climate, soil characteristics and management practices, often through some form of data aggregation (Angulo et al., 2013). There is a large diversity and heterogeneity of the environmental conditions, such as soil and weather, that agricultural crops are grown and managed in. Aggregating and integrating spatially heterogeneous data inevitably results in spatial and temporal biases (Balkovič et al., 2013). This again connects to the potential for more research related to the application of field-based models at larger scales.

For statistical approaches, it is thought that there is a reduction of error when aggregating to larger scales because the relationship between weather and yields is more appropriately described by simple functions at coarse scales, rather than at fine scales (Lobell and Burke, 2010). This means that working and aggregating at larger scales with SCCMs "cancels out" many of the individual errors at individual fields (Lobell and Burke, 2010).

**(5) Stationarity and explanatory power**

One of the largest issues that remains with the use of statistical approaches is the issue of stationarity. This problem arises because of how SCCMs are trained and calibrated on a specific timeframe to describe crop-climate relationships. The validity of empirical-statistical methods under climate change is limited by the necessity of using data outside the range for which the models were fitted (Challinor et al., 2009). In contrast to PCMs which detail plant processes, SCCMs also have no explanatory power to enable understanding as to why certain changes have occurred (Challinor et al., 2009). Thus, statistical approaches should be used with caution when projecting impacts at long lead times (Osborne et al., 2013).

SCCMs are also considered less adaptable to different conditions over both time and space, such as changing $CO_2$ concentrations or growth being limited by water rather than radiation, and vice versa (Challinor et al., 2004). A significant difference between process-based and statistical studies is that the former tend to include the effects of $CO_2$ increases, whereas statistical models typically do not (Lobell and Asseng, 2017). The response of plants to $CO_2$ is of particular importance, as GHG concentrations continue to increase due to human activity, such as through fossil fuel emissions. Carbon dioxide fertilization has numerous effects on crop production: more $CO_2$ could result in increased rates of photosynthesis, decreased water use and various effects on crop leaf area index, biomass, radiation use efficiency and harvest index (Challinor and Wheeler, 2008).

Because of its potential to increase crop productivity for particular crops, it has been recommended that estimating the extent of the $CO_2$ effect is important because of its potential to stimulate plant growth, thus providing more food for an increasing global population (Vanuytrecht et al., 2012). It is clear that a major research need is for SCCMs to incorporate $CO_2$ effects. Additionally, both PCMs and SCCMs need to improve their treatment of ozone as well (Lobell and Asseng, 2017).

**(6) Other limitations and uncertainties**

There are other limitations associated with both crop modeling approaches, such as problems of co-linearity between predictor variables like temperature and precipitation (Hawkins et al., 2013a, Lobell and Burke, 2010). It is also well-known that PCMs are also limited in their capacity to simulate crop responses to extremes (Semenov et al., 2014), pests and diseases (Liu et al., 2016), all of which are important in current and future climate conditions. While SCCMs may be able to include indirect effects of climatic variability, such as those related to pests and diseases, it remains that statistical approaches do not directly consider processes inherent to crop growth (Liu et al., 2016).

Regardless of the crop modeling approach – whether process-based or based on statistical relationships – choosing or adjusting the climate model output to the right scale, and taking steps to make simulations or processes more realistic can help to manage and characterize uncertainty. The process of crop modeling involves numerous issues of data availability, quality and translating input data from climate models to the appropriate scale and detail needed by crop impact models (White et al., 2011), making the third step of linking climate and crop models another significant component of crop projections' uncertainty.

## 2.4 Uncertainties in linking climate model output and crop models

The final source of uncertainty in projections of the impacts of climate change on crop yields stems from the need to do numerous intermediate steps in order to link climate models with crop models. This linkage requires bridging the scale gap between climate models and crop models, and also making climate simulations more realistic prior to their use in crop models. In addition, the development and availability of diverse climate and crop models means that these intermediate steps are dependent on the research questions and choices of methods of the end-users.

In this section, the methods that are used to resolve the scale gap as well as address errors – downscaling and BC – are discussed. The use of multi-model ensembles is also discussed in the context of the differences between crop model approaches and the growing need for multi-method ensembles.

### 2.4.1 Relevance of downscaling in crop impact assessment

As discussed in Section 2.2.4, downscaling methods are used to produce finer-scale climate information from the coarse-resolution GCMs.

Downscaling methods are able to translate a coarse-scale GCM output to finer scaled information on climate change (Ekström et al., 2015), thus providing climate information at scales more relevant for plant processes. Because of the increasing availability of downscaled climate simulations, for example through EURO-CORDEX, the use of dynamically downscaled GCM output is considered a well-established and standard technique for the generation of regional climate change scenarios (Kotlarski et al., 2014).

The development and increasing utility of downscaled climate model output has not gone unnoticed by the impact assessment community – for example, because there is an assumption that rainfall projections represent a key bottleneck to reducing uncertainties in projections of climate change impacts on agriculture, there has been support of downscaling methods in crop modeling research (Lobell and Burke, 2008).

It is argued that this assumption, that finer-scale climate information is "better", is related to the poor ability of climate models to simulate precipitation, which is extremely important to crops. Many of the biggest shortfalls in crop production have been as a result of droughts caused by anomalously low precipitation (Lobell and Burke, 2008). Because crop modeling studies rely on the accuracy of climate input data, it is also argued that the choice of downscaling method, or RCM, is important. There are advantages and disadvantages associated with RCMs, which are more computationally expensive than GCms alone. Therefore, for this study, it is important to ask and demonstrate whether (1) RCMs are skillful in capturing important climate variables such as temperature and precipitation, and (2) have increased skill and utility over GCMs for the study sites and climate variables needed for crop projections.

### 2.4.2 Bias correction to improve climate model output

When linking climate and crop models in order to assess the potential impacts of climate change on agriculture at a regional or local level, plausible climate projections are required. To address this, in addition to downscaling, many crop modeling studies also use BC. This is because even small biases in climate input can have significant consequences, especially since crops have physical and/or biological thresholds that are critical for their successful growth and development (Ruiz-Ramos et al., 2016). For example, extreme temperatures can negatively affect wheat development: both heat stress and frost can decrease wheat yield (Barlow et al., 2015, Semenov and Shewry, 2011, Porter and Gawith, 1999). Therefore, there need to be minimal biases in the climate simulations used for crop models to avoid over- or underestimation of impacts.

Using BC can improve impact assessments, and may in fact be considered necessary to obtain reliable future changes and design robust ensembles (Macadam et al., 2016, Ruiz-Ramos et al., 2016). As a result, it is often recommended that calibration or correction strategies should be part of impact assessment, and that yield estimates made with both GCMs and RCMs require some form of correction to climate inputs (Ho et al., 2012, Glotter et al., 2014).

The use of BC in crop modeling applications has been shown to be useful in a crop modeling case performed in Iberia for maize for the near (2021-2050) and far future (2051-2100) (Ruiz-Ramos et al., 2016). The use of bias-corrected climate simulations in a crop model improved simulated crop phenology and yield, relative to crop projections simulated with uncorrected climate input driven by GCMs (Ruiz-Ramos et al., 2016). In another crop modeling study that also investigated maize, it was found that no climate model output could reproduce crop yields unless BC was first applied to climate simulations (Glotter et al., 2014).

#### 2.4.2.1 Criticism of bias correction

However, significant criticism of BC in the context of impact modeling also exists. For example, Glotter et al. (2014) also found that while the computationally expensive RCMs were able to correct some biases from GCMs in the simulations, most errors in yield were dominated by broad-scale systematic GCM errors, and after correction, RCM-driven and GCM-driven yields were indistinguishable. Because of this, the study was critical of the utility of downscaling. In addition, there have also been arguments that although current BC methods might improve the applicability of climate simulations for impact assessment, BC cannot improve low model credibility – in fact it may even hide a lack of credibility or reduce credibility (Maraun et al., 2017).

While the use of downscaling and BC have been shown to be useful in reducing error and characterizing uncertainty in yield projections, it remains a fact that BC cannot overcome major model errors. The naive application of correction methods might even result in ill-informed adaptation decisions (Maraun et al., 2017). This means that beyond the technical capability provided by downscaling, and that of BC to improve climate model simulations, a critical analysis of the skill, usefulness, and effects of downscaling and BC should be taken into consideration.

It is clear from a review of the scientific literature on climate change projections and impact assessment that there is a common and continuous demand that GCMs are improved and the root sources of error from GCMs are addressed. While improving GCMs cannot be feasibly covered by the work in this thesis, there are still remaining research gaps that deal with uncertainty in the typical modeling chain of GCM to RCM to crop model that can be addressed, which are discussed in the next section alongside key research questions.

## 2.5 Identified research gaps and questions

Based on the literature review, there are several opportunities to conduct more research in the field of climate change impact studies on crops, and these are discussed and summarized in this section. These research gaps are also important to be taken in the contexts of uncertainty and communities of practice (discussed in Chapter 1): many processes and methods that are assumed to be standard are often highly debated in and between communities of practice from different disciplines that interact in climate change impact assessments, for example using RCMs, PCMs over SCCMs, and the practice of BC.

### 2.5.1 Bias correction, climate and crop projections

Based on this review, further investigation is needed on the effects of BC on crop projections, specifically to identify how (1) the choice of BC method affects yield projections, particularly through different crop modeling methods, and (2) what biases are being addressed (GCM error, GCM-RCM error, or RCM error) by correction methods.

There are numerous ways to achieve BC, ranging from simple scaling to more complex distribution mapping (e.g. Teutschbein and Seibert, 2012), which are all capable of improving simulations. While several studies have investigated the effect of BC methods on hydrology (e.g. Teutschbein and Seibert, 2012), or for future crop yield projections (e.g. Macadam et al., 2016, Glotter et al., 2014, Koehler et al., 2013, Hawkins et al., 2013a) they are rarely investigated jointly with crop modeling method. This is significant because different crop modeling methods may utilize climate model simulations in non-linear methods. For example, errors in simulated precipitation were found to propagate through, and even enhanced by, non-linear processes that simulated stream flow (Hwang et al., 2014).

It is thus valuable to investigate how different RCMs and BC methods affect yield projections relative to actual yield observations through different crop modeling methods. Additionally, BC as a process is also discussed critically because of its potential to modify the climate change signal, among other effects.

For the second gap related to BC, in the typical impact assessment method that uses downscaled GCM output as input to the crop model, potential errors are introduced by the GCM and also by the RCM. It could be possible that a well-performing GCM is paired with a poorly-performing RCM for the climate variables needed for the simulation, and vice versa. However, when bias-corrected, future climate projections typically use historical GCM (or GCM-RCM) output as a calibration period for the correction. Historical simulations are simulated by a free-running GCM, which does not assimilate observations and thus does not match the temporal evolution of atmospheric states in the real world (Eden et al., 2014). Can BC potentially be a way to identify how biases from the choice of GCM-RCM selection can affect both climate and crop projections?

### 2.5.2 Multi-method comparisons

Because of how different the communities of practice around crop modeling are, another research gap is the comparison of PCMs and SCCMs in multi-method approaches. With regard to crop modeling uncertainty, crop model comparison projects such as AgMIP, which is within the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP) (Rosenzweig et al., 2013, Warszawski et al., 2014), is a current large-scale attempt to systematically explore differences between crop models, however it is limited to PCMs.

This is significant, because as discussed in Section 2.3, there are large methodological differences between PCMs and SCCMs. The differences in approach of PCMs and SCCMs means that there is 'method uncertainty' in their assessment of the impacts of climate change on crop production and yields. While studies between SCCMs and PCMs are increasing (e.g. Lobell and Asseng, 2017, Moore et al., 2017, Liu et al., 2016), it remains to be seen whether a particular crop modeling method or crop model emerges as an optimal method for future yield projections.

As a means to address this uncertainty, just as MMEs and multi-model means have proven to be useful in understanding and characterizing uncertainty, multi-method ensembles can enable the quantification of method uncertainty (Liu et al., 2016). Multi-method ensembles, coupled with multi-model ensembles of climate models, can allow for better comparison of yield projections across different growing conditions; based on these recommendations, a multi-method approach is adopted for the experimental design of the research.

#### 2.5.2.1 Characterizing and communicating uncertainty

Lastly, the communication and characterization of uncertainty is an area within crop-climate studies that is an area that constantly benefits from new and added knowledge, as well as opportunities for further investigation. Although using multi-model and multi-method ensembles may help to characterize uncertainty, there are still numerous in-between steps such as the different parameters and calibration steps for both the climate and crop models that may have impacts on yield projections. The argument that crop yield projections are inherently uncertain (e.g. Asseng et al., 2013) may be perceived as a bottleneck to decision-making for climate change adaptation. However, key papers (e.g. Vermeulen et al., 2013, Lemos and Rood, 2010) have refuted this 'uncertainty fallacy', as policymakers and decision-makers are quite accustomed to making large decisions under considerable uncertainty.

The literature review reveals that it has become important to address uncertainty in impact assessments. Dealing with uncertainty can be complex, but it is important to address as it helps the discourse on whether yield projections of climate change impacts can be useful, given their large uncertainties. This is particularly significant because uncertainty, and how it is dealt with, can critically affect the way climate projections move from useful to usable (Lemos and Rood, 2010).

The reduction of uncertainty – and proper communication of what it means – is thus of great concern within the climate impacts community. In this regard, there is an opportunity to use the research questions and results to address and communicate uncertainty in a transparent way. There are also opportunities to identify where sharpening of the research in the future is possible and where the study has limitations. By doing so, the research can contribute to the knowledge of uncertainty and the larger picture of making crop yield projections move from uncertain to useful.

#### 2.5.2.2 Summary of research gaps

In summary, there are several ways to focus the research, based on the identified research gaps and remaining challenges:

- Related to the climate modeling community of practice, it is important to go beyond the assumption that RCMs automatically provide more skill or information; therefore assessing whether RCMs are skillful in capturing important climate variables such as temperature and precipitation, in particular vis-à-vis to their driving GCMs;
- Investigating the effect of the BC method on climate and yield projections, particularly given that there are a number of ways to perform BC, and to investigate its effect on climate change projections;
- Differentiating the error contributed by the choice of the GCM and

RCM in a way to understand how this affects future projected climate changes;

- Further investigating how PCMs and SCCMs differ in their structure, limitations, and simulations of past and future yields, giving particular focus to issues of scale, calibration, and upscaled parameters;
- In general, characterizing the contribution of the 'intermediate' steps in the linkages between climate and crop models (See Figs. 1.2-1.4); and
- Using uncertainty as a way to understand the limitations of current methods to assess the impacts of climate change on agriculture, and to discuss this with the aim of providing useful information.

There are also opportunities to further investigate how solar radiation, an important factor in PCMs, can be bias corrected, as well as how $CO_2$ can be better accounted for in statistical crop modeling approaches. However, these are not covered by the work in the thesis in order to better focus on the following research questions outlined below.

### 2.5.3 Research questions and hypothesis

1. What climate variables are most important to wheat yields in the European study sites?
2. How well do climate models – the state-of-the-art tools of simulating the earth's atmospheric processes and projecting future changes – capture and represent climate variables that are relevant to crop growth and development? In particular, does the computationally expensive use of regional climate models add any skill or value to the coarser climate model output?
3. How do the two main methods of simulating crop yields, namely process-based models and statistical approaches, compare to each other in concept and in application, e.g. simulating observed yields

from the past? What are their differences in how they represent crop-climate relationships?

4. How do different methods of correcting biases in climate model output compare when applied to past and future climate simulations?
5. What are future climate projections for the chosen geographical areas, and how do different climate models, downscaling and bias correction methods affect these projected changes?
6. How do the simulation and modeling methods chosen affect the projections of wheat yields under different future climate change scenarios?
7. What are the sources of uncertainty in crop yield projections, and how can these sources be quantified?
8. Based on the results of the study, what areas need more focus on in future work?

These research questions and the hypothesis have informed the selection of data, methods, and the framing of results to be discussed in the thesis. Each chapter addresses one of these research questions, in addition to other relevant sub-questions that aid in answering a main research question. The hypothesis, based on the literature review, is that wheat yields in Europe will be negatively affected by climate change because of the sensitivity of wheat to climate; however, it is theorized that the severity of projected yield decreases depends largely on the methods, models, and climate change scenarios chosen.

### 2.5.4 Research aim

The research aim of this project is to produce projections of future wheat yields that consider the potential effects of climate change, while transparently communicating the limitations and uncertainties of the methods used to generate them. In this way, steps to better characterize uncertainty while understanding potential yield changes are offered, in order to inform agricultural adaptation and add to the body of knowledge in the interdisciplinary field of crop-climate studies.

## 2.6 Structure of the thesis

Because of the diversity of the interdisciplinary methods used in this study, each chapter contains introduction and methods sections specific to the chapter objectives and research questions. The thesis is made up of eight chapters, including (**Chapter 1**) which introduced the research project, key features of wheat physiology, and concepts such as uncertainty and communities of practice.

This chapter, **Chapter 2**, critically reviewed methods that aim to understand the impacts of climate change on yield, with a particular lens on the 'cascade of uncertainty' (e.g. as defined by Wilby and Dessai, 2010). Research gaps and questions that guide the research are identified, as well as the lines between what the research is able to feasibly investigate.

**Chapter 3** compares two different approaches to crop yield modeling, namely a statistical crop-climate model an a process-based crop model. Their respective ability to simulate yields is evaluated and compared through a simulation of past yield with observational climate as input.

**Chapter 4** investigates the performance of global and regional climate models in simulating temperature and precipitation relative to past climate observations, and whether downscaling provides any added value over the typically coarse-scaled global climate model output.

**Chapter 5** evaluates regional climate models, and compares how effective different bias correction methods are in improving climate simulations. These corrected climate model outputs are used in a comparison with past yield simulations from different crop modeling methods (from Chapter 3), in order to formulate a methodology for future climate simulations and yield projections.

**Chapter 6** is a comparison of how simulations of future climate change from different models and scenarios project changes in temperature and precipitation relative to one another. Based on the results of these future climate projections, the limitations of using bias correction to future projections are also discussed, considering the different sources of error that these methods are – and are not able – to address.

**Chapter 7** focuses on the development and comparison of future wheat yield projections, utilizing an uncertainty decomposition method to partition the sources of uncertainty into yield projections. By doing so, the aim of this chapter is to produce yield simulations while quantifying the contribution of uncertainty by each step in the impact simulation cascade.

**Chapter 8** seeks to bring together the results in order to recall their context and contribution to address to the main research questions, offer steps in moving forward, and discuss the limitations of the study.

An appendix is also available after conclusion chapter to provide further detail on simulations, plots, and other calculations that were not included, but referred to, in the main text.

# Chapter 3

# Evaluation and comparison of crop models and methods

## 3.1 Introduction

As discussed in Chapters 1 and 2, the sensitivity of wheat to climate is of great concern in a warming world. With present-day climate change and variability already having significant impacts on crop production, the climate and weather hazards brought by future climate change will indubitably pose greater risks to wheat production. It is therefore important to assess and quantify the relationship of crop growth and development to temperature and precipitation in order to determine the extent of influence that climate and climate change have on yields.

Given the two distinct approaches to modeling crop production, it is valuable to ask how, and how well, each approach represents and integrates the relationship between temperature, precipitation, and yield. The work of this chapter compares two different crop modeling approaches in reproducing wheat yields in the UK and Germany in order to answer research questions on crop-climate relationships.

The work of this chapter is particularly important because the skill of the crop models in reproducing regional yields is crucial for the work in future chapters when generating yield projections with both crop modeling approaches. Because of the significant differences between statistical and process-based crop models, each have their own advantages and limitations, as well as different communities of practice that may influence how and for what purpose a particular crop modeling method is applied. Their differences in skill, and the larger implications of the differences between the research communities that use them, are explored and critiqued in this chapter.

### 3.1.1 Comparing crop modeling methods

As discussed in Section 2.3.1, there are various possible approaches to determine the influence climate has on yields, and these have been used for a number of different crops and locations. In the case of statistical crop-climate models (SCCMs), typical approaches are using multiple regression, with different climate variables – usually indices of precipitation and temperature – as predictors (e.g. Ray et al., 2015, Martín et al., 2015, Moore and Lobell, 2014, Lobell and Burke, 2010, Hawkins et al., 2013a). The complexity of the SCCM may also be increased by including non-climatic predictors such as soil, and considering a non-linear trend (e.g. Michel and Makowski, 2013, Kristensen et al., 2011, Brisson et al., 2010).

In contrast, process-based crop models (PCMs) seek to represent the physical mechanisms of crop development. PCMs typically operate with a daily time step and dynamically calculate various crop and soil properties (Lobell and Asseng, 2017). Their general mechanisms are as follows: the timing of key events such as floral initiation, anthesis and physiological crop maturity are usually predicted by integrating a developmental rate, $R$, over time. This rate is determined by a potential rate of crop development, $R_{pot}$, which is modified by environmental factors such as temperature ($T$), photoperiod ($P$), vernalization ($V$), and other abiotic stresses ($Z$) (Boote et al., 2013, Equation 3.1). In the data and methods section, the specific processes represented by the chosen PCM for the study are discussed.

$$R = R_{pot} * f(T, P, V...Z) \tag{3.1}$$

*Equation 3.1: General mechanism of process-based crop models, where $R$ is the developmental rate integrated over time modified by factors (f). (Boote et al., 2013)*

## 3.2 Chapter approach and objectives

In this chapter, the main objective is to compare the simulations of yield from a SCCM and PCM driven with the same past climate observations. The aim of this comparison is to evaluate the crop models for their ability to capture past yields before they are used to project future wheat yields in subsequent chapters, considering the differences, limitations and associated uncertainties of either approach (See Section 2.3.2). This comparison is important because of the different ways climate-crop relationships are represented in each approach, and how their individual skill can influence the confidence and robustness of future yield projections.

### 3.2.1 Considerations for the hindcast comparison

While the chapter's main objective is to objectively compare and evaluate the skill of the SCCM and PCM, there are several considerations that make this comparison fairly complex. For instance, the scale of comparison: the SCCM is evaluated and used at the country and regional scale, while the PCM is used at the regional scale with initial validation at the site scale. While SCCMs have already been shown to perform well at the regional scale (e.g. Lobell and Burke, 2010), the use of PCMs at the regional scale is challenging because of the original field-level design of PCMs, which require climate inputs of high accuracy at high spatio-temporal resolution, typically more than the resolution of GCMs (Glotter et al., 2014, Ramirez-Villegas et al., 2013), in addition to the input-intensive design of PCMs (Lobell and Burke, 2010). Therefore, an important consideration in this chapter is how the PCM is used at larger scales to be comparable to the SCCM.

#### 3.2.1.1 Using field-level PCMs at greater scales

Field-level PCMs are often applied to make regional- or larger-scale simulations of yield (e.g. for the same countries in the study, the UK (Cho et al., 2012) and for Germany, (Nain and Kersebaum, 2007)). There are a number of 'typical' approaches performed in order to circumvent this scale discrepancy (See e.g. Ewert et al., 2011). One approach is that the PCM is evaluated or validated at the field level. After a satisfactory validation step, the PCM is then used directly the regional level, following the hypothesis that if yields 'reasonably match' reported regional yields, then this would provide validation for the use of crop growth models in predicting regional yields under a variety of climate, economic and management scenarios (Huffman et al., 2015).

While crop models are designed and used with careful evaluation, calibration, and knowledge on the mechanisms they represent, it may still be an outcome that evaluation for a particular crop or domain may result in a poor evaluation outcomes. For example, in their work, a validation exercise at the field level with the Rothamsted field-level experiment (for application at the UK regional level) did not show very high agreement between CERES-Wheat and the observed data because a generic cultivar without specific tuning was applied (Cho et al., 2012). In another case, the discrepancies between simulated and observed yields using CERES-Wheat following this field-level validation approach were deemed too high (e.g.

Langensiepen et al., 2008) and future yield projections were not generated.

Despite these poor evaluation results, however, PCMs are still often applied at scales beyond their original design (e.g. Cho et al., 2012, where the PCM was still applied at the regional scale to make yield projections into the future over the UK). An alternative approach to the field-validation involves direct calibration with regional-level parameters, although recognizing that this may also be a source of error due to aggregation of fine-scale field characteristics. Regional statistics are usually insufficient to derive parameters for crop models as they do not represent field-scale conditions for which the models have been originally developed (Balkovič et al., 2013, Therond et al., 2011), where regional-level data of finer calibration parameters for the PCM DSSAT is unavailable (e.g. spikelet length, date of emergence, date of heading).

#### 3.2.1.2 Critiqiuing practices at the regional scale

These typical simulation practices lead to questions of what exactly 'reasonably matched' in terms of validation means – given that good practice in crop modeling underpins accurate risk quantification (Challinor et al., 2017) – what are good practices for the validation step of field scale models used at spatial scales greater than those at which they were developed for? Furthermore, if the alternative regional calibration is used, are previous studies and the existence of CERES-Wheat regional wheat calibration parameters (e.g. (Nain and Kersebaum, 2007)), enough to justify the use of the PCM at larger scales? Acknowledging the limitations of the approach – and thinking about them critically – is important to the research questions outlined in the following section, and the chapter's design is structured to allow for this critical comparison not only between results, but between the underlying assumptions of the different methods.

However, it is reiterated here that the scale of interest of the study is the at the larger regional scale (country- to state-level) because it is a scale that climate projections are most available and reliable (Lobell and Burke, 2010). Encouraging results of the ability of downscaled climate models to simulate European climate with EURO-CORDEX (e.g. Kotlarski et al., 2014, Jacob et al., 2014) are also used to justify the choice of regional scales in the chapter and in the study.

### 3.2.2 Chapter research questions

The main research question addressed by this chapter is: How do the two main methods of modeling crop yields compare to each other in simulating observed yields from the past? In addition to the main research question, the following research questions are also investigated in this chapter, in light of the considerations of scale and regional-scale evaluation:

**(1)** How have wheat yields changed in the UK and Germany in recent decades?

**(2)** What is the statistical relationship between wheat yields, temperature, and precipitation for the UK and Germany?

**(3)** How well does a statistical crop-climate model, developed based on observations of climate, perform in reproducing wheat yields from the past?

**(4)** How well does a process-based model perform in reproducing yields from the past? Are the represented processes within a mechanistic plant growth model adequate in describing crop growth and yield?

**(5)** How do these recreated past yields ('yield hindcasts') from two different crop modeling methods compare to each other, and to observations?

**(6)** Lastly, and perhaps most critically, how do the 'standard' crop modeling procedures behind each method affect the credibility and robustness of crop simulations? Being cognizant of aggregation error and uncertainty – particularly the limitations of a field-based model being applied to a regional level – are there opportunities to improve skill, or refine these standardized methods?

## 3.3 Data and methods

### 3.3.1 Overview of chapter experimental design

The sensitivity of wheat to climate change and variability has led to numerous studies which attempt to project their possible impacts on wheat production. However, the typical approach to modeling and projecting crop yields is considered highly uncertain, and multiple steps of the impact assessment process contribute to this uncertainty. There are also opportunities within the chapter, and research overall, to critically assess how crop modeling is typically implemented. In response to this, the research of the chapter is designed to address uncertainty from the crop modeling method. The chapter research design is shown in Figure 3.1.

The chapter begins with firstly determining current trends in wheat production in the UK and Germany, two major wheat-growing countries, as well as an analysis of their recent climate records (in the Appendix). Following this, the SCCM is evaluated for its skill in reproducing past yield. A field-level PCM is tested with a simple sensitivity experiment using data from a long-term field experiment in Germany. Due to limitations of data availability at the regional scale for the UK, the regional analysis of yield and climate is limited to four regions in Germany that are chosen to represent north, west, east and south Germany climates.

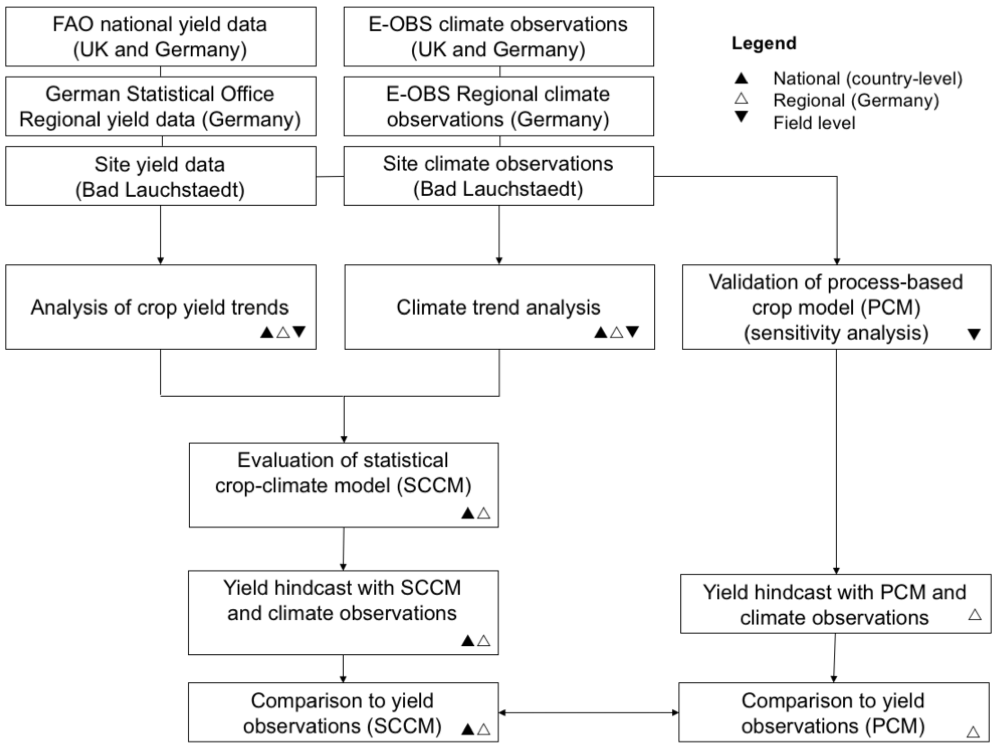



*Figure 3.1: Overview of Chapter 3 research design.*

## 3.3.2 Sources of wheat production data

### 3.3.2.1 National yield data

The UK and Germany are chosen as sites for data analysis because of their high wheat production and yields (See Chapter 1). The wheat yields of the UK and Germany are also potentially stagnated in terms of growth (Moore and Lobell, 2015, Brisson et al., 2010). National, regional and site-level wheat yield data from the UK and Germany are obtained from different secondary sources. National wheat production figures from 1961-2013 are obtained from the United Nations Food and Agriculture Organization (FAO, 2014), which also has data on cultivated area. These figures are used to calculate average yield in tons per hectare (t/ha). A limitation of the FAO data is that it does not distinguish between winter and spring wheat, which differ in their planting time. However, while this limitation has been noted, winter wheat is far more commonly planted rather than spring wheat (Thorup-Kristensen et al., 2009).

#### 3.3.2.2 Regional yield data

For regional data, wheat production and planted area data for the UK are available only for the period 2000-2015 at the time of the analysis. Following the convention of 30 years for the length of climate data for robust analysis, the length of yield data of only sixteen years is insufficient for regional yield analysis in the UK. However, longer-term regional data for Germany is available from winter wheat production records from *Land- und Forstwirtschaft und Fischerei* reports from the *Statistisches Bundesamt* (German Federal Statistics Office, Destatis, 2018). Because of this data limitation, the direct comparison of the output of the SCCM and PCM is only possible for Germany.

For consistency, the NUTS classification (Nomenclature of territorial units for statistics) is used to geographically identify regions in Germany. NUTS is a system used to geographically label the economic divisions and territory of the EU. Four German federal states are chosen to represent the northern, western, eastern, and southern wheat-growing regions: Bayern (South Germany, NUTS code DE2), Nordrhein-Westfalen (West Germany, DEA), Sachsen (East Germany, DED) and Schleswig-Holstein (North Germany, DEF). Yield data from these regions is available from 1979-2014 (36 years) (Table 3.1).

#### 3.3.2.3 Field-level data: Bad Lauchstädt

In addition, a long-term field experiment (LTFE) is chosen for the analysis to validate the PCM at a local scale, based on its long time period of observations and data availability. LTFE provide the possibility of

*Table 3.1: Selected German regions, and NUTS code.*

| Region | Code |
|---|---|
| Bayern, S Germany | DE2 |
| Nordrhein-Westfalen, W Germany | DEA |
| Sachsen, E Germany | DED |
| Schleswig-Holstein, N Germany | DEF |

investigating the effect of fertilization on yields, soil parameters and ecosystem functions (Merbach and Schulz, 2013), and are a rich source of data that can be used to test sustainability, study cropping dynamics, and their impacts on agriculture and the environment (Ortiz et al., 2008).

The static fertilization experiment in Bad Lauchstädt (BL, Merbach and Schulz, 2013) in Sachsen-Anhalt in Central Germany has been running since 1902, and data from 1978-2014 is used for some sensitivity analysis at the field scale. The BL experiment has eight strips where the crops are grown simultaneously every year in a rotation between sugar beets, spring barley, potatoes and winter wheat. The soil is haplic chernozem (loamy). Grain yield data from 1978 onwards is available. The plots were treated with no fertilizer (NoFert), different amounts of farm yard manure (FYM1 with 20t/ha and FYM2 with 30 t/ha) and combined in a factorial manner with mineral fertilizer (NPK).

### 3.3.3 Sources of climate data

#### 3.3.3.1 National and regional gridded climate data

In terms of climate data, daily values of maximum and minimum temperature (Tmax and Tmin), as well as precipitation are taken as observations from E-OBS, which is a gridded dataset of land-only daily high-resolution estimates of these climate variables in Europe (Haylock et al., 2008). E-OBS is used at 0.5° regular grid cell resolution. E-OBS was developed as part of the European Union ENSEMBLES project, an ensemble prediction system for climate change based on high resolution global climate models. E-OBS daily values are compiled from observations from various weather observation stations around Europe (Haylock et al., 2008). E-OBS daily values are calculated with a three-step process of interpolation, first with monthly precipitation and mean temperature because of an insufficient number and heterogenous distribution of European weather stations (Haylock et al., 2008).

E-OBS is used as climate observations in the study as it was developed for the European domain. Since climate variables for the UK and Germany at both country and regional levels are needed, E-OBS is deemed most appropriate for the study. Land-based grid cells over the UK and Germany are intersected with country and regional grid lat-lon boundaries based on the NUTS grid, and then aggregated to obtain country and regional climate observation averages. At the site level, the BL experiment has meteorological data for Tmax and Tmin, precipitation, humidity, and radiation for 1978-2014.

#### 3.3.3.2 Other climate data

Additional climate data apart from E-OBS is needed, considering the climate requirements of CERES-Wheat. For solar radiation, regional downward solar radiation estimates were taken from ERA-Interim (Dee et al., 2011) for the four German regions in the study.

### 3.3.4 Evaluating the statistical crop-climate model

#### 3.3.4.1 Accounting for yield changes

The SCCM used in this chapter is a generalized additive model patterned after crop-climate studies for for maize yield (Hawkins et al., 2013a and Lobell and Burke, 2010) which both use multiple linear regression to derive parameters for temperature and precipitation indices as predictors for yield. An adapted model based on their work is used in this chapter, where it has been modified by several changes: firstly, different time series models are fitted to yield data in order to determine the best-fit model of yield evolution over the recent past decades. Although yield changes have been typically described by a linear trend, a cubic regression spline is used in the work of Hawkins et al. (2013a) to represent the increase in expected yield due to improving technology. This avoids the assumption that the technology trend is linear with time (Hawkins et al., 2013a).

In other crop modeling studies, this trend has been found to be better described by quadratic and linear-plus-plateau (LPP) models, which tend to work better in cases of yield stagnation, which was the case in France from 1996 onwards (Brisson et al., 2010). Following the work of Brisson et al. (2010) and Michel and Makowski (2013), in addition to fitting a linear model, time series are fit with pairs of straight lines to test whether a 'stagnating' model fits the yield data at a national level.

#### 3.3.4.2 Using hot days and summer precipitation as predictors of yield

Another adaptation to the general SCCM is the choice of climate predictors. Based on the importance of heat stress on yields during anthesis, the proposed model has two main climate predictors for yield: firstly, the number of days equal to or above 31°C between June-August (JJA). Yield loss due to irreversible grain sterility begins at this temperature (Webber et al., 2015, Porter and Gawith, 1999). Aside from being the warmest months in the year, the JJA season is important because it coincides with the anthesis (flowering) stage, which occurs 130 days after emergence or typically around June, as well as overlapping with the grain-filling stage (Acevedo et al., 2002, Kristensen et al., 2011).

The anthesis stage has been shown to be particularly sensitive to high temperatures, and the grain-filling stage that follows it is also sensitive to heat stress. High temperatures at anthesis can reduce the grain number, while heat stress after flowering can reduce the grain size (Semenov and Shewry, 2011). These reductions can reduce grain yield (Semenov and Shewry, 2011). While increased temperatures increase the rate of grain-filling, they also reduce the period for grain-filling; although it could be thought that an increase in the grain-filling rate could compensate for the shorter period, this does not occur at temperatures above 30°C (Farooq et al., 2011).

A second predictor is mean summer (JJA) precipitation ($\bar{P_S}$), which is also averaged over both countries and the four German regions. In the work of Lobell and Burke (2010), this term is averaged over the entire growing season; in the work of Hawkins et al. (2013a), this is a climate index based on a long-term rainfall average over JJA. The third predictor is an interaction term $T_H$ and $\bar{P_S}$, as hot days and precipitation have been shown to interact with each other (i.e. more hot days, less rain) (Hawkins et al., 2013a).

#### 3.3.4.3 General SCCM for evaluation

The full general model is shown in Equation 3.2, where $Y$ is wheat yield, $f(t)$ is the yield time series trend, $\beta_n$ represents the coefficients of the different parameters for the climate indices and their interaction term, and $\varepsilon$ is an error term at time $t$. Although this general model is adopted, the trend and each climate predictor is also evaluated and validated for each site (UK, Germany, German regions); non-significant predictors ($p$>0.05) are not included in the final model of the country or region based on linear regression analysis with ordinary least squares.

$$Y(t) = f(t) + \beta_1 T_H(t) + \beta_2 \bar{P}_S(t) + (\beta_3 T_H(t) \times \bar{P}_S(t)) + \varepsilon(t) \tag{3.2}$$

*Equation 3.2: General statistical crop-climate model for wheat.*

#### 3.3.4.4 Analyzing yield time series trends

Aside from the linear and LPP models, wheat yield data are also fitted with polynomial models, to account for non-linear yield trends (See Table 3.2 for a summary of models). Different models are fitted to the yield time series from the UK, Germany, and German regions. The model with the best fit, as determined by several statistical metrics described in the following section, is used to represent the yield trend, inclusive of changes in yield due to technology and management. In the case of the LPP model, the point of transition or breakpoint when yields begin to decline or stagnate is estimated with the $R\ segmented$ package (Muggeo, 2008), which uses a grid-search type algorithm that estimates a model breakpoint by fitting a linear model iteratively with a linear predictor. Provided with an initial estimated breakpoint, calculations from the algorithm update the breakpoint estimate through 'gap' and 'difference-in-slope' coefficients (Muggeo, 2008). In the final model, the time trend yield is fitted simultaneously with the climate predictors in order to avoid overfitting.

*Table 3.2: Models used to determine yield evolution trends.*

| Model | Description |
|---|---|
| LM0 | Linear model |
| LM1 | LPP before estimated breakpoint |
| LM2 | LPP after estimated breakpoint |
| QM | Quadratic model (QM) |
| CM | Cubic model (CM) |

### 3.3.5 Process-based model: CERES-Wheat

#### 3.3.5.1 Introduction and basis of selection

While considering the limitations of PCMs, they remain the most current tools for understanding the impacts of climate change on crop yield. There are many PCMs that are capable of simulating the effects of climate change on various aspects of wheat production, and many of them have been used and evaluated in many different locations all across the world. In the work of this chapter, the CERES-Wheat PCM (Originally Ritchie and Otter, 1985, now part of DSSAT Jones et al., 2003) is used. CERES has a long development history (Dettori et al., 2011). It has been compared to other crop models in various studies (e.g. Eitzinger et al., 2013, Palosuo et al., 2011, Singh et al., 2008, Eitzinger et al., 2004, Jamieson et al., 1998). It is also included in the Agricultural Model Intercomparison and Improvement Project (AgMIP) (Rosenzweig et al., 2013) and in a crop modeling method comparison study (Liu et al., 2016).

In a review of studies that use crop models in their methodology, CERES was the most used crop model in the majority of papers (White et al., 2011). These reasons of wide usage, evaluation and accessibility are reasons for the selection of CERES-Wheat for the research. Additionally, CERES-Wheat was found to have satisfactory performance at a regional scale in the UK and in Germany (Cho et al., 2012, Nain and Kersebaum, 2007), but also critically evaluated in Northern Germany (Langensiepen et al., 2008). In this chapter, CERES-Wheat is tested and validated at the field-level in Germany prior to its use in crop hindcasts.

#### 3.3.5.2 Modeled processes and mechanisms

CERES-Wheat is able to simulate crop phenological development. This means that it is able to simulate the growth of grains, leaves, stems and roots and biomass accumulation based on thermal time accumulation (Cho et al., 2012). It is also able to account for soil water balance and soil processes important to crops such as nitrogen uptake, which eventually lead to yield (Cho et al., 2012).

In CERES-Wheat, the plant life cycle is divided into several phases. The rate of crop development is governed by thermal time, or growing degree-days (GDD), which is computed based on the daily maximum and minimum temperatures (Jones et al., 2003). Daily plant growth is computed by converting daily intercepted photosynthetically active radiation into plant dry matter using a crop-specific radiation use efficiency (RUE) parameter (Jones et al., 2003). Light interception is computed as a function of leaf area index, plant population, and row spacing (Jones et al., 2003). Information on field management such as planting, harvesting, application of both organic and inorganic fertilizer, and irrigation are also considered.

Abiotic stresses such as water, nitrogen, temperature or atmospheric $CO_2$ modify the amount of new dry matter available for growth each day (Jones et al., 2003). In CERES-Wheat simulations, the wheat crop is allowed to grow and reach physiological maturity. However, growth is terminated if the plant runs out of resources or if the grain growth rate is reduced below a threshold value for several days (Jones et al., 2003). In order to model these processes, the CERES-Wheat PCM has a set of minimum data requirements for the simulation to run (Jones et al. (2003), Table 3.3).

#### 3.3.5.3 Cultivar-related plant processes and responses

Many of the simulated crop metrics like yield or phenology are based on the genetic coefficients for the cultivar (variety) simulated. For example, the grain numbers are based on the cultivar characteristics, which determine its genetic potential, canopy weight, average rate of carbohydrate accumulation during flowering, and temperature, water and nitrogen stresses (Jones et al., 2003). Other genetic coefficients include the cultivar's daylength and vernalization sensitivity. Failed or insufficient vernalization in winter wheat can delay dormancy, which then delays the onset of the reproductive stage of winter wheat (Wang et al., 2015). In CERES-Wheat, these numerous processes are described by cultivar coefficients (Table 3.4).

*Table 3.3: CERES-Wheat minimum data requirements (Jones et al., 2003).*

| **Module** | **Input requirements** |
|---|---|
| Site | Latitude, longitude, elevation, average annual temperature and amplitude, slope, topography, drainage, surface stones |
| Weather | Daily solar radiation, maximum and minimum temperatures, precipitation |
| Soil | Classification and characteristics by layer (e.g. water release curve characteristics, bulk density, organic carbon, pH, root growth factor, drainage) |
| Initial conditions | Previous crop, root and nodule amounts, rhizobia characteristics, water, ammonium and nitrate by layer |
| Management | Cultivar: name and type (genotypic coefficients) Planting: date, depth, method, row spacing, direction, population<br>Water: irrigation and water management (dates, method, amounts, depth)<br>Inputs: Inorganic and organic fertilizer (material, dates, method, amount)<br>Others: tillage, environmental adjustments (e.g. $CO_2$, harvest schedule) |

*Table 3.4: CERES-Wheat (DSSAT) experimental values (Jones et al., 2003) and regional German cultivar coefficients (Nain and Kersebaum, 2007).*

| **Cultivar coefficient** | **Controlled crop process** | **Coefficient values** |
|---|---|---|
| P1D | Photoperiod sensitivity coefficient | 5.0 |
| P1V | Vernalization sensitivity coefficient | 5.0 |
| P5 | Thermal time from the onset of filling to maturity | 8.0 |
| G1 | Grain (kernel) number per unit stem | 3.9 |
| G2 | Potential kernel growth rate | 3.0 |
| G3 | Tiller death coefficient | 3.0 |
| PHINT | Thermal time between the appearance of leaf tips | 95 |

#### 3.3.5.4 Experimental design for the PCM evaluation and simulations

The sensitivity of the CERES-Wheat is tested at the field scale, with BL data from 1978-2014. A simple sensitivity experimental design is adopted to test the PCM responses to increases in temperature and precipitation in both well-fertilized and poorly-fertilized wheat fields. Modifications are daily increases of 1 and 2°C to daily temperature and +1 and 2 standard deviations from the daily mean precipitation for the observed time period. In addition, in order to understand the influence of genetic coefficients, the regional coefficients from Nain and Kersebaum (2007) are used together with the default genetic coefficients for wheat to compare yield responses.

After this sensitivity analysis, the output of the PCM is evaluated with regional genetic coefficients (Table 3.4) so that its performance can be compared with the SCCM. The use of standard genetic coefficients is a commonly applied practice when practical restrictions prevent site-specific model calibrations (Langensiepen et al., 2008), althugh the limitations of this typical approach have already been explained (See Section 3.2.1.1). The following experimental set-up is thus adopted: Optimal fertilizer based on the BL experiments is provided. Fine-scale soil information is used to provide the soil profile and type. One kilometer-grid resolution soil data is taken from a fine-scale soil grid developed for DSSAT (IRI et al., 2015).

Regional yields are simulated with the same experimental setup (input, management, cultivar) for all four German regions. The simulations are performed for each year from 1981-2010 (30 years). Evapotranspiration is calculated using the Priestley-Taylor method (Priestley and Taylor, 1972), which is based on radiation and soil heat flux. Hydrology follows the Ritchie water balance model (Ritchie and Otter, 1985), which uses the upper limit and drained lower limit of the soil as basis of the available soil water. The Godwin model for soil organic matter (Godwin and Jones, 1991) is used, and this models the transport of nitrogen through the soil to deeper layers based on water flux values obtained from the soil water module of the crop model (Jones et al., 2003).

#### 3.3.5.5 Evaluating crop model performance

To evaluate model performance, goodness-of-fit is determined based on the adjusted coefficient of determination ($R^2$) with $p$ <0.05 considered statistically significant, denoted with ($*$) when indicated. Additional validation such as the root mean square error ($RMSE$), Akaike Information Criterion ($AIC$), and Leave-One-Out Cross Validation ($LOOCV$) error statistics are also calculated. $RMSE$ is the square root of the mean square error, which is a measure of how close a fitted line is to data points.

The $AIC$ is a means of measuring relative model quality where a smaller number means the model is closer to the true (unknown) model, and the $LOOCV$ statistics are from a resampling method that estimate test error based on $n-1$ training observations with a prediction made for the excluded observation (James et al., 2013). $R^2$, $RMSE$, $AIC$ and $LOOCV$ error are also calculated between hindcasted and observed yields for each country and region. If the SCCM or PCM do poor jobs of representing crop yield responses to climate, this will be reflected in its validation statistics, such as a low $R^2$, and high error estimates between hindcasted and observed yields.

## 3.4 Results

The results section first reports the analysis of yield trends, followed by the evaluation of the general SCCM, the site sensitivity analysis of a PCM, and finally the yield hindcast comparison between the SCCM and PCM.

### 3.4.1 Yield trend analysis

#### 3.4.1.1 National wheat yield trends in the UK and Germany

Time series analysis of FAO data shows that wheat yields in the UK and Germany since 1961 have increased from around 3-4 t/ha to around 6 t/ha at present. However, after the 1990s, the rate of yield increase slows down, and more yield variability (scatter) in annual yields can be observed, for example for Germany after the year 2000. Figures 3.2A-D and Table 3.5 show the results of fitting different models for the yields in UK and Germany to describe national yield trends.

The linear model LM0 shows that yields in both countries have increased significantly since the 1960s ($R^2$=0.83). The estimated breakpoint between increasing and stagnating yields in the UK and Germany is identified as the year 1999. Two separate linear models are thus calculated for the LPP model: one before the year 1999 and another one after. For the UK, LM1 improves the $R^2$ value for the years 1961-1999 (LM0 $R^2$=0.83, LM1 $R^2$=0.89) when annual yields from 1999 onwards are removed from the time series. Similarly, a separate linear model for Germany before the year 1999 explains yield trends better (LM0 $R^2$=0.92, LM1 $R^2$=0.95). Yields after 1999 for both countries show no significant trend ($R^2$=0.02 and -0.05 for the UK and Germany, respectively) when fitted with another linear model (LM2) after the breakpoint, which is evidence that yields have been stagnating in the UK and Germany in the most recent decade.

For the UK, polynomial models improve $R^2$ values compared to LM0 (QM $R^2$=0.88, CM $R^2$=0.92). The best fit is given by a cubic model, which has the smallest $RMSE$, $AIC$ and $LOOCV$ error estimates. For Germany, QM and CM also improve $R^2$ values compared to LM0 (QM $R^2$=0.93, CM $R^2$=0.95) and result in smaller $RMSE$ values. The $AIC$ value is also smallest for the cubic model, although the $LOOCV$ statistics are similar between the QM and

CM models. These smaller error values using QM and CM show that yield trends for both countries over the past 50 years are not necessarily linear, with the rate of yield increase slowing down in the mid-90s.

#### 3.4.1.2 Regional wheat yield trends

Winter wheat yield data from 1979-2014 shows different levels of productivity across Germany. The highest median yields were in Northern Germany (DEF) at 8 t/ha and the lowest median yields are observed in East Germany (DED) at 6 t/ha (Fig. 3.3). Regions all show significant positive (linear) trends ($R^2$>0.35). Similar to the analysis at the national level, a cubic model best describes yield trends in the time series, with the highest $R^2$ compared to other fitted models, and smaller $RMSE, AIC$ and $LOOCV$ error (Table 3.6). For example, the $RMSE$ is smallest for DE2, DEA, DED, and DEF with the cubic model ($RMSE$=0.42, 0.52, 0.59, 0.52 t/ha respectively) compared to the linear model ($RMSE$=0.58, 0.66, 0.63, 0.68 respectively).

#### 3.4.1.3 Field level yield trends

At the field level, BL wheat yields treated without fertilizer, or only farmyard manure (FYM), show no significant trend (Fig. 3.4A-C). Yields in BL had increasing yields ($p$ <0.05) for all treatments with NPK fertilizers (Fig. 3.4D-F), although the quadratic and cubic models do not show significant improvements compared to a fitted linear time trend (Table 3.7). The effect of the fertilization scheme on the yield trend is not analyzed in this work, but it has been investigated extensively in the work of Merbach and Schulz (2013).

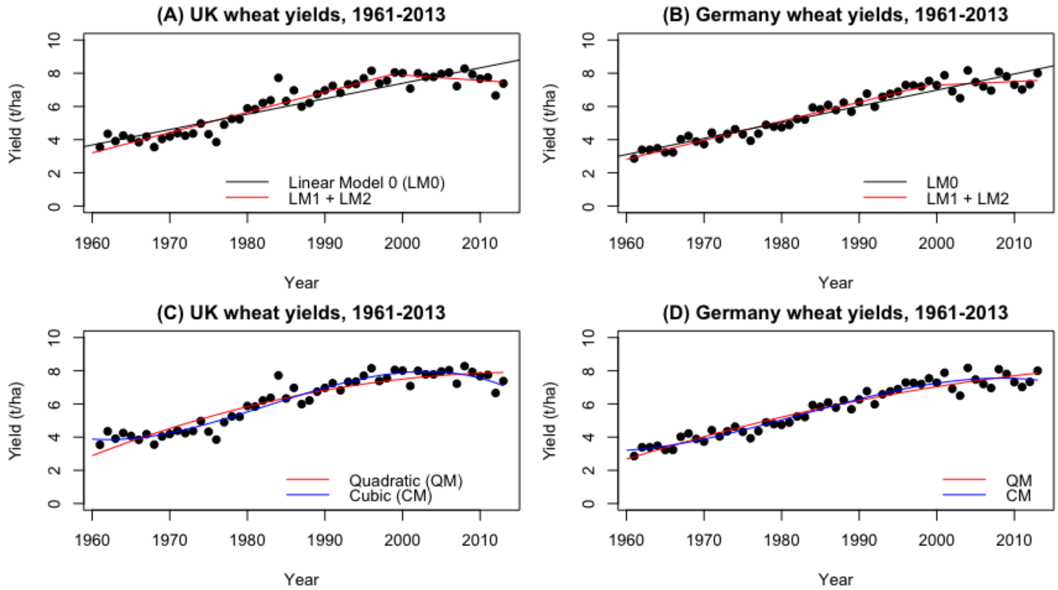



*Figure 3.2: Different models for UK and Germany wheat yield trends, 1961-2013. (A) and (B) show linear models, (C) and (D) are non-linear models.*

*Table 3.5: Summary evaluation statistics for national yield trend analysis.*

| **Country** | **Statistic** | **LM0** | **LM1 (pre-1999)** | **LM2 (post-1999)** | **QM** | **CM** |
|---|---|---|---|---|---|---|
| UK | $R^2$ | 0.83* | 0.89* | 0.02 | 0.88* | 0.92* |
| | $RMSE$ | 0.63 | 0.48 | 0.41 | 0.54 | 0.42 |
| | $AIC$ | 107.7 | 59.8 | 20.6 | 92.4 | 67.9 |
| | $LOOCV$ | 0.43 | 0.25 | 0.24 | 0.33 | 0.20 |
| Germany | $R^2$ | 0.92* | 0.95* | -0.05 | 0.93* | 0.95* |
| | $RMSE$ | 0.43 | 0.29 | 0.48 | 0.39 | 0.35 |
| | $AIC$ | 67.6 | 19.4 | 25 | 59.4 | 49 |
| | $LOOCV$ | 0.20 | 0.09 | 0.3 | 0.17 | 0.14 |

*LM0 is the linear model; LM1 is the linear model to the estimated breakpoint year 1999 and LM2 is the linear model for the years after 1999; QM and CM are the quadratic and cubic models. ($*$) indicates statistical significance ($p$ <0.05).*

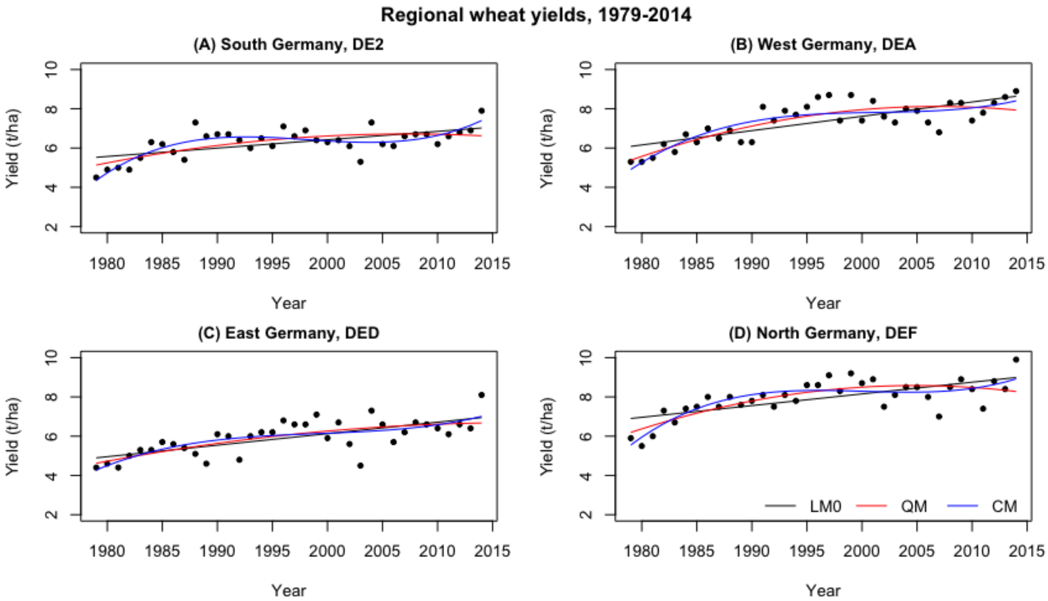



*Figure 3.3: Wheat yields for German regions (A-D), 1979-2014, including different fitted models (linear, quadratic and cubic) to describe the trend.*

*Table 3.6: Summary evaluation statistics for regional yield trend analysis.*

| Region | Statistic | LM0 | QM | CM |
|---|---|---|---|---|
| South Germany, DE2 | $R^2$ | 0.35* | 0.41* | 0.63* |
| | $RMSE$ | 0.58 | 0.54 | 0.42 |
| | $AIC$ | 68.5 | 66.2 | 49.6 |
| | $LOOCV$ | 0.38 | 0.36 | 0.22 |
| West Germany, DEA | $R^2$ | 0.55* | 0.67* | 0.71* |
| | $RMSE$ | 0.66 | 0.56 | 0.52 |
| | $AIC$ | 77.8 | 66.2 | 65.0 |
| | $LOOCV$ | 0.48 | 0.37 | 0.33 |
| East Germany, DED | $R^2$ | 0.47* | 0.48* | 0.5* |
| | $RMSE$ | 0.63 | 0.61 | 0.59 |
| | $AIC$ | 74.2 | 74.5 | 74.2 |
| | $LOOCV$ | 0.43 | 0.45 | 0.46 |
| North Germany, DEF | $R^2$ | 0.43* | 0.57* | 0.66* |
| | $RMSE$ | 0.68 | 0.59 | 0.52 |
| | $AIC$ | 80.8 | 72.1 | 64.5 |
| | $LOOCV$ | 0.53 | 0.45 | 0.36 |

*LM0 is the linear model; QM and CM are the quadratic and cubic models. ($*$) indicates statistical significance ($p$ <0.05).*

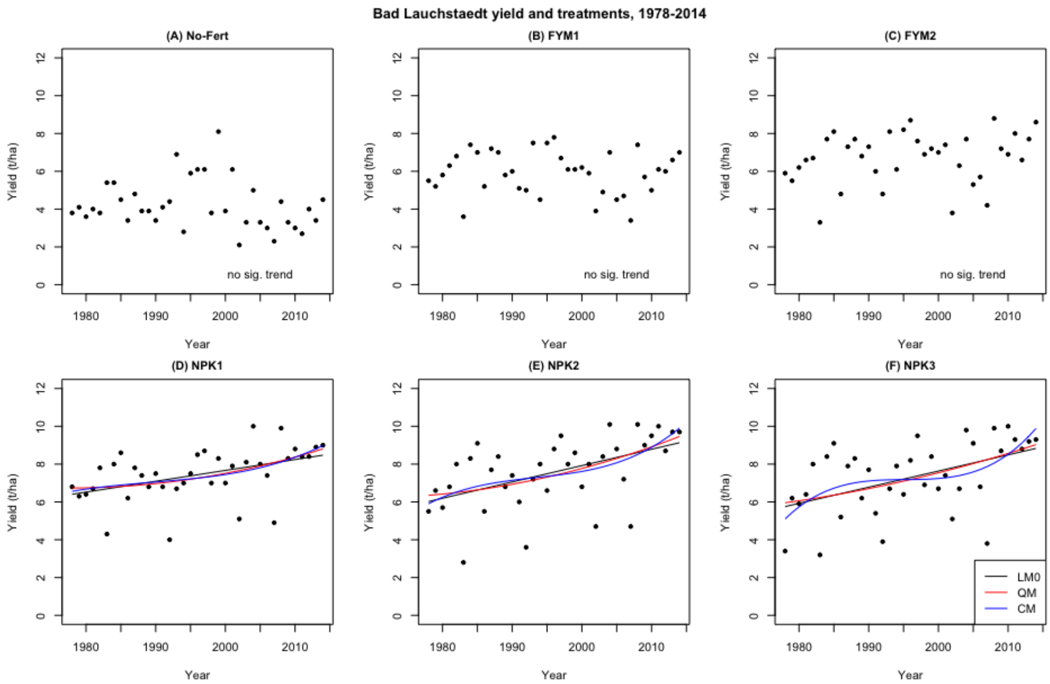



*Figure 3.4: Wheat yields for different fertilizer treatments (A-F) from the Bad Lauchstädt long-term field experiment, 1978-2014. Significant yield trends (linear, quadratic and cubic) are also shown.*

*Table 3.7: Summary evaluation statistics for field-level yield trend analysis.*

| Treatment | Statistic | LM0 | QM | CM |
|---|---|---|---|---|
| NPK only (NPK1) | $R^2$ | 0.18* | 0.17* | 0.15* |
| | $RMSE$ | 1.21 | 1.2 | 1.2 |
| | $AIC$ | 125.3 | 126.6 | 128.5 |
| | $LOOCV$ | 1.62 | 1.63 | 1.7 |
| NPK + 20 t/ha FYM (NPK2) | $R^2$ | 0.24* | 0.22* | 0.21* |
| | $RMSE$ | 1.56 | 1.55 | 1.54 |
| | $AIC$ | 143.7 | 145.3 | 146.7 |
| | $LOOCV$ | 2.68 | 2.73 | 2.8 |
| NPK + 30 t/ha FYM (NPK3) | $R^2$ | 0.22* | 0.2* | 0.22* |
| | $RMSE$ | 1.63 | 1.63 | 1.58 |
| | $AIC$ | 147.2 | 149.1 | 149 |
| | $LOOCV$ | 2.9 | 3.1 | 3.07 |

*Treatments without NPK fertilizers did not show any significant ($p<0.05$) trends and are thus not shown. LM0 is the linear model; QM and CM are the quadratic and cubic models. ($*$) indicates statistical significance ($p<0.05$).*

### 3.4.2 Evaluating the statistical crop-climate model

In this section, the results from the SCCM evaluation are reported. Yield trend fitting at the national and regional level showed that a cubic (non-linear) trend most accurately captures how yields have changed and evolved from the past decades (See Figures 3.2, 3.3), so this is included in the generalized additive model.

#### 3.4.2.1 National-level yield and climate models

The results of using the general wheat SCCM (Eqn. 3.2) with ordinary least squares (multiple linear regression) on climate and wheat data results in distinct SCCMs for the UK and Germany. At the country level, only summer predictors – $T_H$, $\bar{P}_S$ and their interaction term – are statistically significant predictors and the interaction term was significant only for Germany. Equations 3.3 and 3.4 show the country level SCCMs, where $Y$ is wheat yield, $f(t)$ is the cubic time trend, $\beta_n$ represents the coefficients of the different parameters, and $\varepsilon$ is an error term at time $t$:

$$\text{UK:} \quad Y(t) = f(t) + \beta_1 T_H(t) + \beta_2 \bar{P}_S(t) + \varepsilon(t) \tag{3.3}$$

$$\text{Germany:} \quad Y(t) = f(t) + \beta_1 T_H(t) + \beta_2 \bar{P}_S(t) + (\beta_3 T_H(t) \times \bar{P}_S(t)) + \varepsilon(t) \tag{3.4}$$

*Equations 3.3 and 3.4: National statistical crop-climate models (SCCMs).*

These models, based on summer climate predictors, can account for 85% and 94% of yield variability for the UK and Germany from 1961-2013 based on their $R^2$ value (Table 3.8). Comparing the coefficients shows that for the UK and Germany, $T_H$ has a negative effect on yields while the coefficients of $\bar{P}_S$ and the interaction term between $T_H$ and $\bar{P}_S$ are significant and negative, however these effects are small.

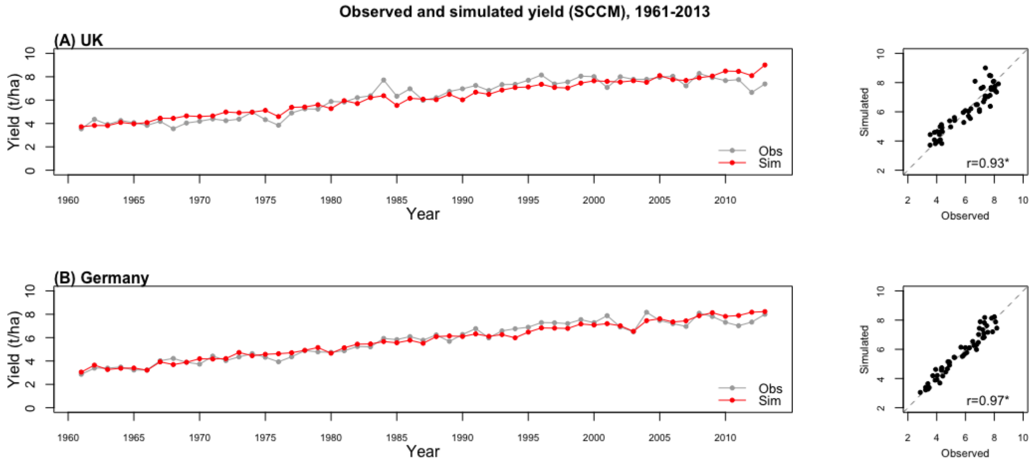



*Figure 3.5: Observed and SCCM simulated yields, (A) UK and (B) Germany, 1961-2013. (C) and (D) compare observed and simulated yields.*

*Table 3.8: Summary statistics for national SCCMs.*

| **Country** | $T_H$ | $\bar{P_S}$ | $T_H \times \bar{P_S}$ | $R^2$ | $RMSE$ | $AIC$ | $LOOCV$ |
|---|---|---|---|---|---|---|---|
| UK | -0.64* | -0.005* | - | 0.85* | 0.59 | 104.5 | 0.47 |
| DE | -0.25* | -0.007* | 0.001* | 0.94* | 0.38 | 59.14 | 0.18 |

*($*$) indicates statistical significance ($p$ <0.05).*

#### 3.4.2.2 Regional-level yield and climate models

At the regional level, the cubic yield trend and significant climate predictors are evaluated through multiple regression which result in distinct regional SCCMs (Equations 3.5-3.7). Significant climate predictors for regions vary: hot days and summer precipitation are significant predictors for yield for East, West and Southern Germany (DED, DEA, DE2). The only significant climate predictor for Northern Germany (DEF) is JJA precipitation (Fig. 3.6). Similar to the development of the country-level SCCMs, summary statistics are calculated per region, including cross-validation error ($LOOCV$) for a hindcast using the regional SCCMs. The calculated coefficients for each region are in Table 3.9.

The SCCM regional yield hindcasts show high and significant correlation ($r$>0.7) to yield observations, for all regions (Fig. 3.7A-D, Table 3.9).

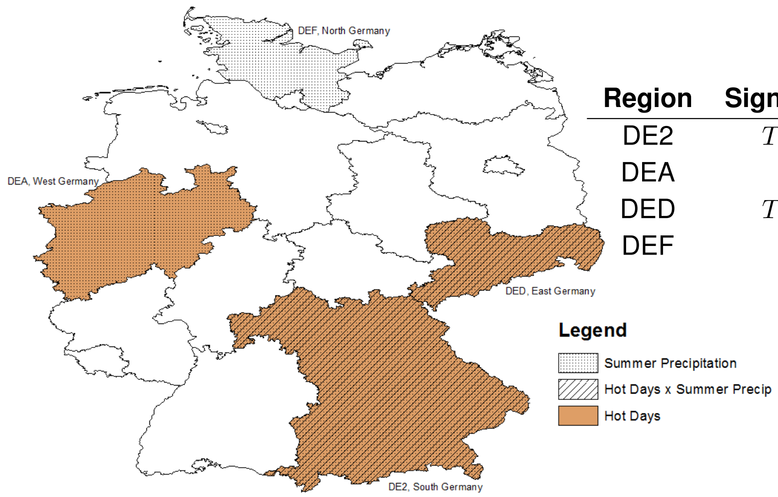
Evaluation statistics between observed and simulated yields also show relatively good agreement, as evidenced by $RMSE$ under 1 t/ha, significant $R^2$ values, and relatively small $LOOCV$ error.

$$\text{DE2, DED:} \quad Y(t) = f(t) + \beta_1 T_H(t) + \beta_2 \bar{P}_S(t) + (\beta_3 T_H(t) \times \bar{P}_S(t)) + \varepsilon(t) \quad (3.5)$$

$$\text{DEA:} \quad Y(t) = f(t) + \beta_1 T_H(t) + \beta_2 \bar{P}_S(t) + \varepsilon(t) \quad (3.6)$$

$$\text{DEF:} \quad Y(t) = f(t) + \beta_1 \bar{P}_S(t) + \varepsilon(t) \quad (3.7)$$

*Equations 3.5-3.7: Regional statistical crop-climate models (SCCMs).*




| **Region** | **Significant predictors** |
|---|---|
| DE2 | $T_H$, $\bar{P}_S$, $T_H \times \bar{P}_S$ |
| DEA | $T_H$, $\bar{P}_S$ |
| DED | $T_H$, $\bar{P}_S$, $T_H \times \bar{P}_S$ |
| DEF | $\bar{P}_S$ |

*Figure 3.6: Significant climate predictors for yield in German states.*

*Table 3.9: Coefficient values for regional German SCCMs.*

| **State** | $T_H$ | $\bar{P}_S$ | $T_H \times \bar{P}_S$ | $R^2$ | $RMSE$ | $AIC$ | $LOOCV$ |
|---|---|---|---|---|---|---|---|
| DE2 | -0.34* | -0.01* | 0.001* | 0.47* | 0.46 | 57.1 | 0.69 |
| DEA | -0.08* | -0.01* | NS | 0.6* | 0.58 | 71 | 1.48 |
| DED | -0.37* | -0.01* | 0.001* | 0.56* | 0.52 | 66.1 | 1.15 |
| DEF | NS | -0.004* | NS | 0.5* | 0.6 | 71.2 | 1.15 |

*($*$) indicates statistical significance ($p$ <0.05).*

Simulated yields from all regional SCCMs show good year-to-year agreement, although they generally underestimate yields between the years 1995-2000. The SCCMs for DE2 and DED, which both contain a significant interaction term ($T_H \times \bar{P}_S$), show their ability to simulate yields during hot summers, for example the large European heat waves in 2003 and 2006. The results of these SCCM hindcasts are compared PCM hindcasts in the next section, after the sensitivity analysis for the PCM.

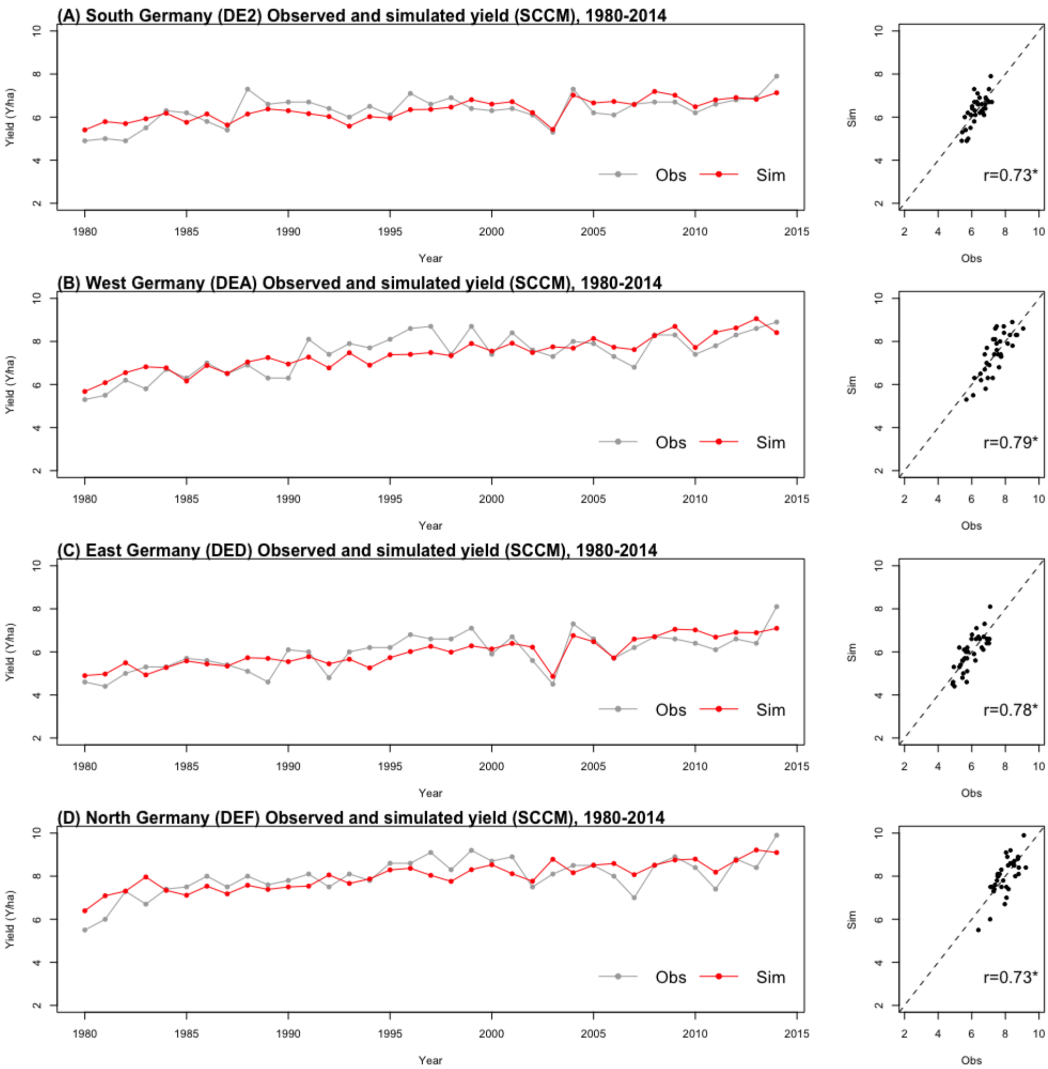



*Figure 3.7: Observed and SCCM simulated yields, (A) DE2, (B) DEA, (C) DED, and (D) DEF, 1978-2013, with a comparison and correlation of observed and simulated yields, ($*$) indicates significance ($p$ <0.05).*

### 3.4.3 Testing the sensitivity of the PCM to environmental modifications

The responses of CERES-Wheat are tested at the site level, using wheat yield and weather data from 1978-2014 from the BL LTFE (Merbach and Schulz, 2013). Two treatments, namely the no-fertilizer and the high-yielding mineral fertilizer with FYM, are chosen for validation to represent the effect under different fertilization schemes.

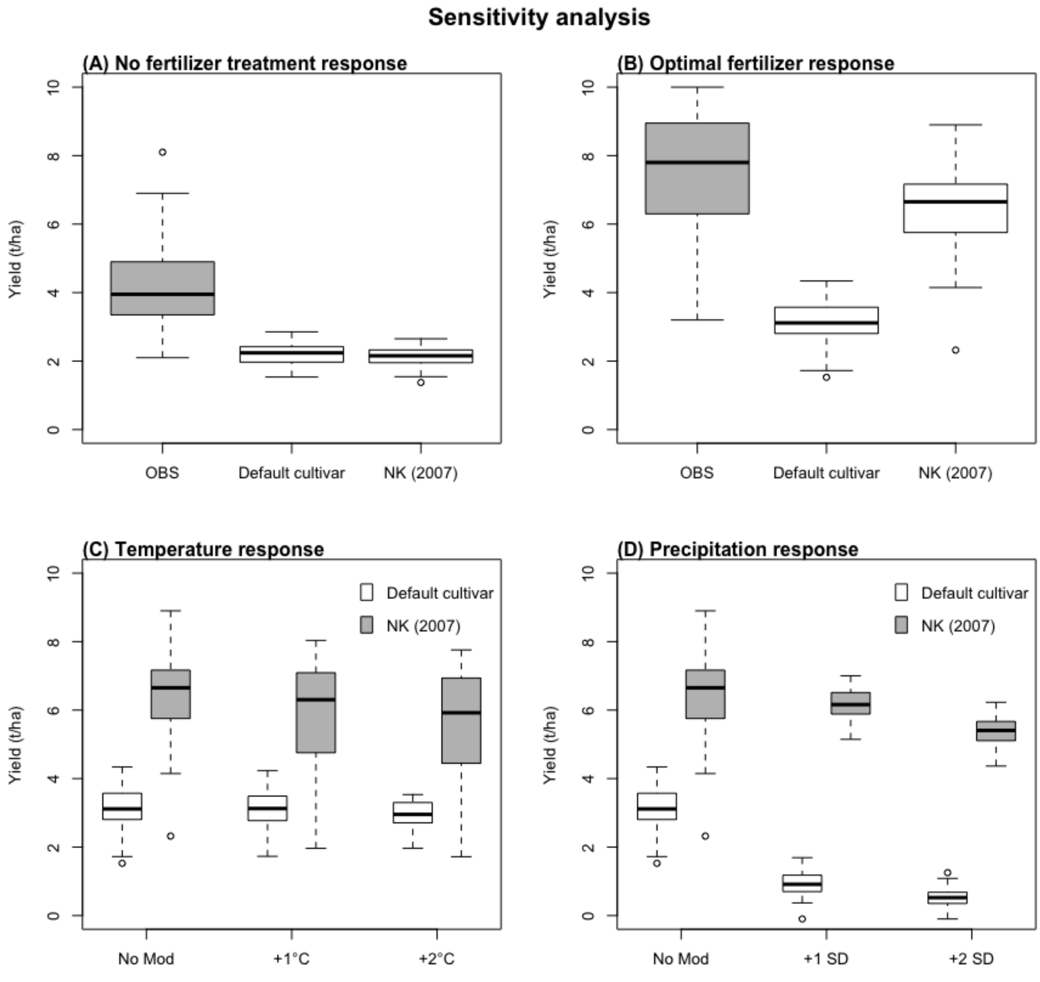



*Figure 3.8: Simple climate sensitivity validation experiment with BL data (1979-2014) and CERES-Wheat, using default genetic coefficients and those from Nain and Kersebaum (2007) (NK 2007). The boxplots show: (A) the yield response to no fertilizer and (B) optimal fertilizer, both compared to observations; (C) changes in response to daily increases in temperature (+1 and 2$^{\circ}C$) and (D) precipitation (+1 and 2 standard deviations.)*

Yields are shown to be sensitive to fertilization, temperature, (increases of 1 and 2 degrees Celsius to daily temperature) and precipitation (+1 and +2 standard deviations to daily precipitation), with the PCM simulating increases to yield with increased fertilization, and decreases to mean yields with increased temperature and precipitation (Tables 3.10A-B). In particular, the yield responses are shown to be very sensitive to the genetic coefficients used to describe the cultivar responses. Given the large number of plant responses that these coefficients control (See Table 3.4), it is not surprising that a default cultivar performs more poorly compared to the Germany genetic coefficients from (Nain and Kersebaum, 2007, or NK2007) when attempting to recreate observed yields (Fig 3.8A-B).

*Table 3.10: Yield responses to environmental modifications in experimental validation.*

| **Experimental modification** | **Yield difference and RMSE relative to respective BL yield observations** | |
|---|---|---|
| | **No fertilizer** | **Optimal fertilizer** |
| Default cultivar | -48%, 2.4 t/ha | -58%, 4.7 t/ha |
| Nain and Kersebaum (2007) cultivar | -50%, 2.5 t/ha | -14%, 2.2 t/ha |

*Table 3.10 continued.*

| **Weather modification** | **Yield change with default cultivar** | **Yield change with NK2007 cultivar** |
|---|---|---|
| +1 C Temperature | -1% | -10% |
| +2 C Temperature | -5% | -14% |
| +1 SD Precipitation | -70% | -3% |
| +2 SD Precipitation | -82% | -16% |

Comparison of the simulated yields with the default and the NK2007 genetic coefficients (Table 3.10A) shows that simulated wheat yields are not significantly different when they are not fertilized, regardless of the cultivar used. In the case of well-fertilized wheat experiments, the NK2007 yield simulations are significantly higher than those simulated with the default cultivar; in addition, the NK2007 yield simulations are closer to observations. Yield simulations with the default cultivar for the well-fertilized experiment have a large $RMSE$ of nearly 5 t/ha with respect to yield observations.

In addition, the genetic coefficients influenced how yields respond to environmental modifications: for example, yield losses with increased precipitation are drastic, with 70-80% simulated yield losses with the default wheat cultivar in contrast to 3-16% with the NK2007 wheat cultivar coefficients.

### 3.4.4 Crop model comparison: regional yield hindcasts

The last section of results is the comparison of regional yield hindcasts using wheat yield observations, simulations from regional SCCMs and simulations from the PCM experiments designed with E-OBS climate, regional genetic coefficients and fine-scale soil data. The results are shown for each region in Fig. 3.9A-D. As reported in Section 3.4.2.2, the yield simulations from the SCCM are generally well-correlated to yield observations for each region (National: $r$>0.9, regional Germany: $r$>0.7). In contrast, yield simulations from the PCM have poor (non-significant) correlation to observations, apart from DED (East Germany), which had $r$=0.5. Although median yields from DEA and DEF (West and North Germany) are within the range of values of yield observations, generally yields from DE2 and DED (South and East Germany) overestimated regional wheat yields, with large $RMSE$ values of approximately 2 t/ha (Table 3.11).

Although these are generally poor results for the PCM, it can be observed the yield impacts of the 2003 heat wave were observed in the climate analysis (See Appendix) are captured by PCM at this point in time. Comparatively, however, the SCCM evaluation performance is much closer to observations in terms of the year-to-year accuracy and smaller errors.

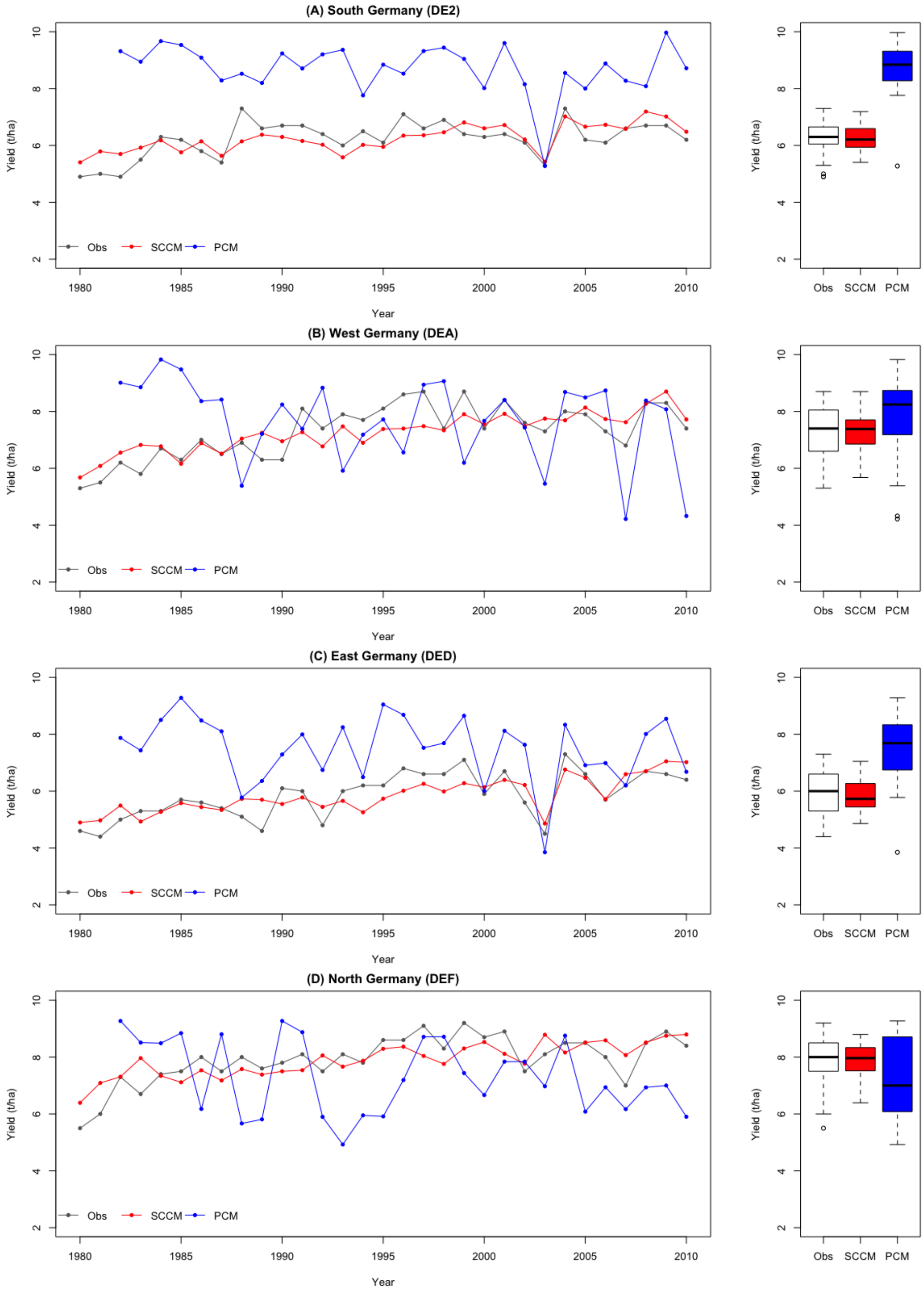



*Figure 3.9: Crop modeling method comparison: SCCM and PCM simulations of regional German wheat yields, 1981-2010.*

*Table 3.11: Summary statistics, CERES-Wheat (PCM) simulations compared to yield observations.*

| **State** | **Correlation ($r$)** | $R^2$ | $RMSE$ |
|---|---|---|---|
| DE2 | NS | NS | 2.5 |
| DEA | NS | NS | 1.8 |
| DED | 0.47* | 0.2* | 1.9 |
| DEF | NS | NS | 1.6 |

*($*$) indicates statistical significance ($p$ <0.05).*

However, as discussed previously, upscaling from fine-scale PCMs to a regional scale is challenging and the input parameters are considerable sources of uncertainty. The context of these multi-method comparisons and their implications for future yield projections are discussed in the following section.

# 3.5 Discussion

In this discussion, the results from the SCCM and PCM evaluation are addressed in terms of the main research question: how well does each crop modeling method capture past yield observations?

## 3.5.1 Using statistical crop modeling approaches

In this chapter, statistical approaches were used to answer research questions on: (1) how yields have changed over the study areas and period of yield observations, (2) what the best model to describe yield trends is, and (3) how models based on empirical data perform with temperature and precipitation indices to hindcast past yields. National and regional wheat data showed that yields have changed significantly in recent decades, with evidence of stagnation in both the UK and Germany at the national level (shown by the LPP and non-linear yield trend) as well as in German regions. Other findings include that hot days above 31°C have already affected wheat yields in the UK and Germany.

These results are supported by the literature review in Chapter 2 where it was reported that wheat yields are negatively affected by periods of high temperatures. The results of the SCCM evaluation showed that the effect of precipitation is fairly unclear because of its small coefficient, but the interaction term between the two climate indices showed that periods of high hot days also typically had low rainfall (e.g. the 2003 heatwave, see Brisson et al., 2010). In Germany, this interaction term was significant at the national level and for several of the four examined regions. Statistical metrics such as high correlation, low $RMSE$ and improved $AIC$ are used as evidence of the good performance at the national and regional level of the statistical approaches used in this chapter. This means that the SCCM has a credible performance, making its usage for the future reasonably justified. However, there are also limitations that are associated with statistical approaches to crop modeling.

#### 3.5.1.1 Limitations of the statistical model approach

While the SCCMs resulted in yield simulations that were significantly correlated to observations, the approach of statistically linking empirical data on climate and crops and using the resulting SCCM for projections is still heavily criticized. For example, while it is frequently reported that SCCMs are useful, their usefulness and predictive power are also frequently diminished in scientific literature because they are limited to responses inside the range of conditions used to develop them (Ewert et al., 2011). Other criticisms are the fact that in the real world, plants respond to highly local weather conditions in complex and non-linear ways (Glotter et al., 2014), in contrast to the relatively simple regression used here in the chapter. In addition, it has been argued that the complexity of crop growth and development makes it impossible to define a general relationship between temperature and rate of development for all phases and varieties of wheat (Porter and Gawith, 1999).

This is where PCMs may have significant advantages by being able to represent the complexity of plant development. PCMs attempt to encompass knowledge of crop physiology and responses to environmental factors, and have been used for decades to gain knowledge on how crops develop, grow, and yield (Semenov et al., 2012, Chenu et al., 2017). They are also able to integrate simulations of external factors not commonly included in statistical models. For example, they can include fertilization management, $CO_2$ concentrations, and wheat variety (Rosenzweig et al., 2014). These factors are difficult to include in SCCMs. However, based on the results in this chapter, the relatively simple approach of the regional SCCMs significantly outperformed the yield simulations generated by the CERES-Wheat PCM.

The PCM evaluation is discussed in the following section, but here it is valuable to point out that despite significant criticism of statistical models (e.g. Semenov et al., 2012, White, 2009) it is argued that there are still merits with the use of statistical approaches, especially when applied to larger areas where fine-scale information cannot be obtained, as is the case in this chapter for the country- and regional-scales. They remain useful, if imperfect tools, for projecting future yield responses and are likely to continue to play an important role in anticipating future impacts of climate change (Lobell and Burke, 2010).

### 3.5.2 Evaluation of the PCM approach

The results of the PCM evaluation were not strong on a stand-alone basis, and this was even more evident when compared to the SCCM: most regions did not have well-correlated PCM yield simulations, and all PCM yield simulations had larger $RMSE$ values relative to observations compared to the SCCM. There are clearly limitations in directly applying the field-based PCM at the regional scale, in addition to existing limitations and differences during calibration between these approaches that may have contributed to this disparity in the measured skill (here, with metrics like $r$ and $RMSE$) of the two crop modeling methods used – it is argued that these are mostly issues connected to scale. Scale issues brought about by the use of plot- or field-scale models at coarser resolutions are well-known within the discipline (e.g. Ewert et al., 2011, Hansen and Jones, 2000, see also Chapter 2 literature review).

Because of this, the research questions posed at the beginning are revisited: what can be expected when using PCMs at a scale that they were not originally designed for? Can the PCM still be used in the future for yield projections, and are there opportunities to improve their usability at larger scales? In the following sections, the evaluation of the PCM performance is discussed in the context of these scale issues to attempt to provide initial options to address these critical questions.

#### 3.5.2.1 Limitations of the PCM approach: scale and data aggregation

Scale is a notable issue in many crop-climate studies in both SCCMs and PCMs, but in opposite directions. SCCMs are generally challenging to use at the field-level scale, where climate model output is hard to obtain and extremely uncertain (Lobell and Burke, 2010) but have reasonable performance at larger scales, as shown in this chapter. In contrast, when PCMs are used at larger scales than the field or plot, the aggregation of input data from finer to coarser resolution will inevitably lead to losses of spatial variability of the dataset (Ewert et al., 2011). At the field scale, there is variation in crop development due to small-scale factors like micro-climates and soil variation (Barlow et al., 2015) which are lost when data is aggregated. Aggregation errors are also amplified as the resolution of soil and climate data decreases (Maharjan et al., 2019, Hoffmann et al., 2016, Folberth et al., 2016). In particular for climate data, given the importance of radiation and its significant effects on crop model output and error (e.g. Trnka et al., 2007), it should be acknowledged that future radiation projections are highly uncertain, making these also contribute to errors (in future yield projections).

However, despite this scale limitation being well-known, field-scale models are used at larger scales roughly 50% of the time (Challinor et al., 2017). Given the awareness and discourse on scale and aggregation error, why does this scale mismatch persist in practice, and how can it be addressed? One method would be to approach crop modeling at the regional scale. This would be advantageous given the reliability and positive evaluation of regionally downscaled climate information, such as over Europe through EURO-CORDEX (e.g. Kotlarski et al., 2014). Some regional-level PCMs exist, for example the General Large Area Model (GLAM, Challinor et al., 2004), which represents a number of plant growth processes applied over a range of environments (Challinor et al., 2004). GLAM has been shown to be successful at reproducing crop metrics like yield from a variety of crops, and it has been used extensively in other crop-climate studies (e.g. Wang et al., 2017, Elliott et al., 2015, Watson and Challinor, 2013, Vermeulen et al., 2013, Challinor et al., 2010). GLAM was also recently compared to a similar generalized additive SCCM used in this chapter (Watson et al., 2015).

However, it should be noted that criticism of regional yield simulations also exists, because regional yields are often poorly correlated with the yields of individual farms and are less valuable in local decision support (Lawless and Semenov, 2005). Using regional scales also does not eliminate the aggregation error from other fine-scale input data on soil. Thus, bridging the gap in the scale and resolution at which climate models, crop models and other local-scale processes operate remains a considerable problem for the impact assessment of climate change (Fowler et al., 2007).

#### 3.5.2.2 Calibration with regional parameters

Another potential solution and common method in the literature of addressing the scale mismatch is by using more generic or regionalized crop model parameters in fine-scale PCMs, which was the approach attempted in this chapter. To do so, firstly, the sensitivity of the PCM was tested using regional crop calibration parameters and generic other parameters whenever possible (e.g. following Palosuo et al., 2011). A review of the methods included in the PCM also showed sufficient representation of plant processes relevant to yield (Section 3.3.5, Table 3.4). Results showed that yields responded to changes in the fertilization scheme, temperature, and precipitation in the sensitivity test. However, it was also shown that these results were highly influenced by the genetic parameters which determine the magnitude of the yield responses.

Cultivar coefficients for PCMs like CERES-Wheat are important in the field of crop-climate research because of the powerful simulation effects that they have over crop growth and development which cascade into yield simulations (See Table 3.4). The sensitivity of yields to crop parameterization indicates the importance of the initial calibration prior to simulation, and this is argued to have a large influence on the bias in PCM yield hindcasts in this chapter. PCMs such as CERES-Wheat require extensive input data and parameters. However, despite its widespread use, it was challenging to find regional genetic coefficients for CERES-Wheat Germany in crop-climate studies, as many studies that use CERES-Wheat frequently report 'iterative' or 'trial-and-error' calibration and parameterization procedures (e.g. Li et al., 2015, Dettori et al., 2011, Wang et al., 2009) and do not often report these coefficients nor have detailed procedures about how they were obtained (e.g. Li et al., 2015, Thaler et al., 2012, Lobell et al., 2012, Ruiz-Ramos and Mínguez, 2010, Wang et al., 2009). It is argued that this limited reporting hinders the reproducibility of results from evaluation.

In addition to the iterative nature of determining genetic coefficients for CERES-Wheat, there is still a remaining large number of input parameters that are dependent on fine-scale validation data, which are not always available or accessible, even in LTFE. As discussed in the introduction, in the typical PCM set-up, it is usually deemed sufficient or even appropriate that before it is applied to the regional scale, the PCM is highly tuned to a field-level site (where calibration data can be more easily monitored or obtained, Maharjan et al., 2019, Hoffmann et al., 2016). These data include fine-scale information such as phenology, the harvest index, and tiller/leaf growth which are important variables that measure the stages of wheat growth and development.

However, it is argued that the 'typical' method of field-level calibration (whether the results are satisfactory or not) and its subsequent direct application at the regional scale does not fully address the scale issue. In particular for the regional genetic parameters, cultivars are often very specific to a region or locality so using regionalized parameters will inevitably result in simulation biases relative to local yields. Cultivars can also vary widely as they are grown under different conditions of soil and climate, even within the same state or region (Curtis, 2012). As mentioned in the introduction to the chapter, a study that performed extensive calibration to determine genetic coefficients through built-in functions in CERES-Wheat, Langensiepen et al. (2008) also found large ranges of error above 2 t/ha using CERES-Wheat to simulate wheat yields in Germany. These errors were deemed too large to permit the practical application of CERES-Wheat for optimizing fertilizer management in North Germany (Langensiepen et al., 2008). This is in line with the idea that if a model needs to be skillful if its assessment of risk is to be correct (Challinor et al., 2017), which underlines the importance of validation.

#### 3.5.2.3 Continuing mismatch between PCM application and its scale

In spite of these challenges, the use of field-scale PCMs beyond its original scale is widespread in crop-climate modeling studies (e.g. Challinor et al., 2017), where CERES (DSSAT) is one of the most widely used PCMs based on an extensive review (e.g. White et al., 2011). In contrast to the findings of Langensiepen et al. (2008) and this chapter, CERES-Wheat has been shown to be able to feasibly simulate regional yields in the UK and Germany (e.g. Cho et al., 2012, Nain and Kersebaum, 2007). While a globally gridded version of the larger DSSAT suite, of which CERES crop models are a part of, is included in the large-scale multi-crop model comparison project, the Agricultural Model Intercomparison and Improvement Project (AgMIP, Rosenzweig et al., 2013), the end-user is more likely to utilize the publicly accessible field-based model to answer smaller-scale research questions, and it is argued that this is where issues of the continuing mismatch between the scale design of PCMs and their application are likely to persist.

Recently released guidelines to crop modeling emphasize ’good practices’ in crop modeling (Challinor et al., 2017), including better transparency and measures to enhance reproducibility, but it is argued that the scale gap – and the methods that attempt to address it but not solve it – remain difficult to make feasible guidelines for due to limited available data and resources to run fully parameterized, well-calibrated PCMs. However, these guidelines are important to the credibility and robustness of crop yield projections. In the final section of the discussion, other potential solutions to the issue are also suggested.

### 3.5.3 Setting guidelines for crop modeling practices

Both the statistical and process-based approaches to crop modeling have powerful simulation capability, but also significant limitations that contribute to uncertainty. While progress continues to be made with the use of both SCCMs and PCMs, and comparing their results to each other in ensemble approaches (e.g. Lobell and Asseng, 2017, Liu et al., 2016, Moore et al., 2017), the results in this chapter show that yield simulations from contrasting crop modeling approaches can differ significantly, due to reasons like the scale mismatch and calibration/validation practices that do not wholly address the gap between design and application, particularly for PCMs.

In their work, Challinor et al. (2017) discuss the scale mismatch and offer the following suggestions for good crop modeling practice, including appropriate complexity, creating model ensembles based on skill, correcting biases in climate model output, including uncertainty estimates in simulations, and performing evaluation as a continuous process (and over a broad range of contexts), among others. Specifically for crop model calibration and ongoing evaluation, heavy emphasis is put on adequate input and validation data throughout the growth of the simulated crop, as well as thorough documentation for transparency and reproducible results (Challinor et al., 2017). In this regard, the work of AgMIP has made significant progress in its management of different crop model protocols, evaluation, and its implications (Müller et al., 2017, Elliott et al., 2015).

In the meantime, how can progress be made in answering the research questions of the study when working with limited data to carry out thorough evaluation, which can result in less than ideal results (as evidenced by the PCM evaluation results for the PCM in this chapter)? Reasonable performance in the context of this chapter and study is therefore reliant on the reported favorable evaluation of CERES-Wheat in the past, its represented processes that are important to wheat production (focusing on yield), and plausible performance (e.g. with the DED region) even with limited parameterization. Here, the research focus is reiterated: to comparatively assess the contributed uncertainty of the two different crop modeling approaches, as well as 'pay attention' to the otherwise standardized intermediate steps of downscaling and bias correction, and their impacts on the cascade of uncertainty. While a more thorough evaluation would perhaps provide an improvement of the statistical metrics between PCM yields and observations, the results point to the need for more thorough regional calibration, rather than casting doubt on the knowledge that can be gleaned from the PCM simulation results, which includes information on the amount of dry matter, and estimates of dates of anthesis and maturity (See Jones et al., 2003).

While the input-intensive nature of PCMs may not change soon, and may in fact increase as more knowledge is gained on crop growth and development, there are only a limited amount of ways to make field-level PCMs more usable at the regional scale for which they are frequently applied. It is suggested that results here could be improved greatly by using available regional data (apart from yield) for the 'ongoing evaluation' (e.g. Challinor et al., 2017) recommended for good practices in crop modeling. Apart from the further development of both field- and regional-scale crop models (both process-based and statistical approaches), publicly accessible databases of the PCM calibration parameters, such as CERES-Wheat coefficients for well-used wheat cultivars, should be developed and disseminated in scientific literature to promote transparent and reproducible methods in PCMs. In addition, greater availability of regional information on wheat growth and development characteristics would be valuable resources to aid in more fairly evaluating PCMs applied at larger scales.

#### 3.5.3.1 Characterizing uncertainty through multi-model and multi-method comparisons

Other recommendations that can be followed from the Challinor et al. (2017) paper are the use of ensembles to better evaluate skill. Although this is a single-SCCM and single-PCM study, the results of the chapter connect to the larger discourse on crop modeling method uncertainty. Within the broad classifications of these two crop modeling methods, there exist numerous types of statistical approaches and PCMs of various scales. Numerous institutions develop crop models, each with their own formulation, requirements, and implementation procedures. In an effort to characterize how different impact models compare to each other, multi-model ensembles (MMEs) are used for many different types of models. For example, MMEs can be composed of crop models, such as those in AgMIP (Rosenzweig et al., 2013), climate models in the Coupled Model Intercomparison Project (CMIP) (Eyring et al., 2016) or other sectoral impact models in the ISI-MIP (Warszawski et al., 2014). MMEs are a useful method for characterizing uncertainty due to different model structure. Ensembles allow for direct comparison of simulations in order to quantify and explore uncertainty (Challinor et al., 2013).

While the value of multiple models for impact assessment in quantifying uncertainty is increasingly well-documented, it also has been argued that there are underlying conceptual differences between MMEs of impact models and MMEs of climate models, and ultimately should have very different objectives (Challinor et al., 2014). Climate models can be assessed on a number of physical properties, such as how well they can represent precipitation and temperature. In contrast, impact models like crop models require calibration towards a small subset of variables towards simulating one measurable value, such as yield (Challinor et al., 2014). This means that comparing impact models is less advanced than comparing climate models due to constraints in comparable properties, and that significant difficulties exist in obtaining adequate data, particularly at regional scales (Challinor et al., 2014).

It is often the objective in climate MMEs to narrow the range of uncertainty to seek consensus between models. However, it has been argued that this objective may be too limited (Knutti and Sedláček, 2012, Challinor et al., 2014). As models improve and represent more processes in greater detail, there is greater confidence in their projections. Despite these improvements, agreement or convergence between model simulations may remain slow (Knutti and Sedláček, 2012). In contrast to climate model ensembles, rather than focusing on a single objective of narrowing uncertainty, exploring the differences between crop models matters, and is actually valuable (Challinor et al., 2014). It is thus argued that despite varying results between SCCMs and PCMs, applying multi-method ensembles can further improve the understanding behind projecting the impacts of climate change.

Therefore, due to the fundamentally different approaches between PCMs and SCCMs – for example, significant differences in parameter inputs, calibration needs, and included processes, as shown in this chapter – single-PCM or single-SCCM crop modeling method approaches contain significant uncertainty, and more studies which use both approaches comparatively will help to characterize these significant sources of uncertainty and error in yield simulations.

## 3.6 Conclusion

Important food crops like wheat are sensitive to climate. Climate largely determines where and when crops can be cultivated and how they are managed, and also has an influence on their harvest and yield. Extremes of climate and climate change, already observed in Europe, have present-day impacts on wheat yields in the UK and Germany. In particular, heat stress can have significant negative effects on wheat. Achieving a science-based understanding of these impacts and making insights into the future, in the context of adaptation, is largely dependent on crop models. The fundamentally different approaches of crop modeling – PCMs and SCCMs – have different input requirements and associated limitations, which can lead to very different yield simulation results, as shown in this chapter. However, it is important to continue an approach that investigates and communicates these differences, as they help elucidate the limitations and needed improvements to the typical impact assessment method often described in crop-climate studies.

In this context, continued work on transparently communicating calibration and validation procedures is particularly important when using PCMs, which have high input demands at a fine scale for which they were originally designed. Although their regional application is possible, it is not without challenges due to scale and aggregation error. Thus, the reproducibility of previous research that employs these methods is of extreme significance and studies that use PCM should report methods with detail, and consider suggestions made here and elsewhere in the scientific literature on good crop modeling practice. Following these guideliness will aid in not only the evaluation of the model or method, but in the robustness of results into the future.

# Chapter 4

# Evaluating the added value of downscaled GCM output

## 4.1 Introduction

Global climate models (also general circulation models, or GCMs) and earth system models (ESMs) are the most current scientific tools that are used to understand the atmosphere, oceans, and their feedback with land systems. As discussed in the literature review in Chapter 2, GCMs still have numerous limitations which stem from, *inter alia*, limited abilities to represent complex physical processes and resolve key climate features (Section 2.2.2). Combined with these limitations, climate change projections are also uncertain when combined with plausible future greenhouse gas scenarios and natural variability (e.g. Hawkins and Sutton, 2011, 2009). Therefore, the errors from GCMs can have large impacts on the robustness of climate change impact assessments.

Based on the literature review, the coarse scale of GCMs is often criticized because impact assessments often require fine-scale climate information. Therefore, the downscaling of GCMs for impact assessments is now common practice, or even considered 'necessary' (Glotter et al., 2014, p.8776) particularly for climate change studies. However, the benefits of using regional climate models (RCMs) for downscaling are still debated in the scientific community. Because RCMs are driven by GCMs, there are implicit assumptions on the skill of GCMs, as well as the added value provided by RCMs (e.g. Feser et al., 2011). In this chapter, these implicit claims are investigated by analyzing historical GCM and GCM-RCM (dynamically downscaled GCMs) simulations in terms of their closeness to observations, in order to address the uncertainty from climate models in crop yield simulations and projections. While the evaluation of climate models and the added value of RCMs is widespread in the climate modeling discipline, it is a relatively uncommon component in impact assessment, so this chapter provides information on how important this step is in projecting future crop yield under climate change.

### 4.1.1 Simulations of the climate system with the CMIP ensemble of GCMs

Exploring where GCMs are skillful in representing key features of the atmosphere and oceans is typically done through multi-model ensembles (MME). MMEs of GCMs are an effective way of comparing different GCMs under a common framework and structured experiments. The Coupled Model Intercomparison Project (CMIP) is a promiment example of an MME, with more than 20 modeling centers and 50 models (GCMs/ESMs) participating in its fifth phase (CMIP5) (Taylor et al., 2012). CMIP5 provides a multi-model context for determining why similarly forced GCMs produce a range of responses, and they are also a way of understanding the climate system, inclusive of its feedback to the carbon cycle, and exploring climate predictability (Taylor et al., 2012).

CMIP5 models have shown significant changes and improvements in the performance of participating GCMs relative to its predecessor CMIP3, which was used extensively in the Intergovernmental Panel on Climate Change Fourth Assessment Report (AR4) (Kumar et al., 2014). CMIP5 has also been shown generally capable of simulating climate extremes and their trend patterns (Sillmann et al., 2013). However, there has also been some instances of degradation in skill and only little changes to uncertainty in CMIP5 compared to CMIP3 (Kumar et al., 2014). While current models are generally able to simulate European climate with considerable skill, they are still affected by common errors, such as a tendency to underestimate blocking frequencies (McSweeney et al., 2015, Woollings, 2010). Blocking describes a weather pattern in which the prevailing westerly winds and storms are blocked by a persistent and stationary anomaly, generally an anticyclone (area of high pressure) (Woollings, 2010).

Biases in the representation of blocking such as the Greenland and summer Pacific blocking frequencies are associated with errors in the representation of storm tracks, while biases in winter European blocking frequency are related to the North Atlantic storm track tilt and Mediterranean cyclone density (Zappa et al., 2014). The North Atlantic has an important influence on European climate: the single most important factor for year to year fluctuations in the seasonal climate around the Atlantic Basin is the state of the North Atlantic Oscillation (NAO) (Scaife et al., 2014), so biases in this basin are crucial to the accurate representation of European climate. Other biases in the surface storm track north of the Gulf Stream have been found to be connected to biases in sea surface temperatures (Booth et al., 2017).

Given the existing biases in GCMs, the use of dynamical downscaling methods that make use of a 'nested' RCM within a driving GCM has been met with scientific criticism and debate because the output of an RCM is heavily influenced by the lateral boundary conditions of the driving GCM. If the large-scale climatology of the driving GCM has large systematic errors, these will be transmitted to the nested RCM (Giorgi and Gutowski, 2015). This has been previously discussed with the example of biased storm tracks and the resulting biased precipitation simulations in Sections 2.2.2 and 2.2.4 of the Literature Review. Therefore, the usefulness of RCMs, in terms of their "added value" to existing GCM simulations, has been well-debated in the field of climate modeling. Added value is the term used to describe additional knowledge gained from RCMs (Feser et al., 2011) compared to the information available from GCMs alone.

### 4.1.2 Added value of regional climate models

Comparing GCM and GCM-RCM output is relevant in the context of the scientific discussion in the climate modeling discipline community on the added value of RCMs. RCMs target regional (sub-continental to sub-national) scales, and have approximate spatial resolution ranges from 1 to 50 km, in contrast to GCM resolutions that are about 100 km and coarser (Rummukainen, 2016). In principle, several improvements to GCM simulations can be expected when RCMs are used, due to their higher resolution: for example, numerical truncation error in the discretization of field equations is automatically reduced with the use of finer computational grids, and these finer grids also permit the explicit representation of small-scale processes that are precluded in low-resolution simulations (Di Luca et al., 2015).

These small-scale processes such as local orography, land-sea contrast, and atmospheric features such as convective cells are important influences to regional climate, in addition to the prevailing large-scale conditions (Feser et al., 2011). Therefore, it is both well-established and unsurprising that regional climate modeling adds "detail" to the driving GCM results. Although a number of key studies have already demonstrated that RCMs can realistically simulate general climate patterns in comparison to observations (e.g. in Europe, Kotlarski et al., 2014), the added value of the information provided by RCMs remains a long-standing and central issue in the climate modeling literature and community (Rummukainen, 2016, Giorgi and Gutowski, 2015).

For example, many RCM studies implicitly assume a superiority of the RCM output over the driving global data (Feser et al., 2011), or simply have claims that the added value of RCMs consists of "more spatial detail" (Takayabu et al., 2016). These are claims that should be examined, but are not usually explicitly proven (Takayabu et al., 2016, Feser et al., 2011). Thus, this chapter focuses on evaluating the simulations of past climate from both GCMs and GCM-RCMs, and this is considered an important step in the process of assessing the impacts of climate change on wheat production in the study sites in Europe.

## 4.2 Chapter approach and objectives

Given the increasing availability and resolution of climate model simulations, including downscaled climate model output, there have been numerous studies with different emphases that have evaluated the performance of both GCMs and RCMs for specific climate phenomena and variables over varying timescales and geographical locations. In this chapter, historical GCM and GCM-RCM simulations are compared to past observations of climate to identify the skill of the selected GCMs and GCM-RCM combinations in representing temperature and precipitation in the UK and Germany, in the interest of utilizing these climate models to project future climate change and future crop yields.

Through this comparison, the research aims to go beyond the implicit claim that using RCMs makes climate model output more skillful, a step that is mostly out of the scope of climate impact studies.

### 4.2.1 Chapter research questions

In this chapter, the main research question addressed is: how do GCMs and GCM-RCMs compare to each other when representing climate variables that are relevant to crop growth and development? In addition, the following questions guide the data and methods for the analysis in the chapter:

> **(1)** How well do historical GCMs and GCM-RCM simulations capture temperature and precipitation in the past for the UK, Germany and its selected regions?
>
> **(2)** How does GCM-RCM output compare to the output of their coarser driving GCMs, and also to observations?
>
> **(3)** Based on the comparisons to answer questions (1) and (2), do RCMs add value in the context of the climate needs for the crop models used in the study?

Based on the literature review, it is hypothesized that GCM-RCMs will reproduce past observations of temperature and precipitation better compared to GCM-only simulations, particularly for temperature. However, despite this better performance, it is also anticipated that these uncorrected simulations contain significant biases relative to observations.

## 4.3 Data and methods

### 4.3.1 Overview of chapter experimental design

In this chapter, output from GCMs and GCM-driven RCMs are processed and compared to the climate observations from E-OBS (Haylock et al., 2008) that were analyzed in terms of annual and seasonal timescales in Chapter 3. The approach of the chapter is to use simple but challenging tests to determine whether added value of GCM-RCMs exists over GCMs in the context of the climate output needed for the crop models. This means that the respective outputs of GCM-RCMs and GCMs alone are compared relative to observations in order to answer the research questions. In the following section, the basis for selecting the different GCMs and RCMs that are used in the chapter are discussed and reviewed.

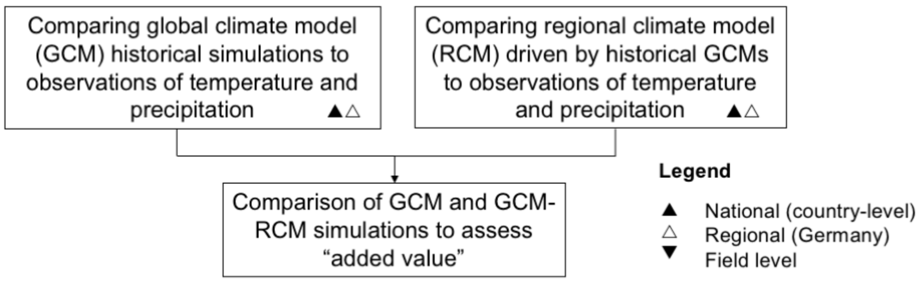



*Figure 4.1: Overview of Chapter 4 research design.*

## 4.3.2 Selection of GCMs, RCMs, GCM-RCM combinations

### 4.3.2.1 Multi-model ensembles of GCMs and RCMs

Simulations of global climate from GCMs in this chapter are taken from CMIP5 (Taylor et al., 2012). Although a more recent sixth phase of CMIP also exists (CMIP6, Eyring et al., 2016), the CMIP5 database has more readily available simulations for the research. CMIP5 has several core simulations, including a "historical" run, which is forced by observed atmospheric composition changes (reflecting both anthropogenic and natural forcings) and land cover (Taylor et al., 2012). CMIP5 also has available future climate projections forced with specified emission concentrations from the representative concentration pathways (RCPs). For CMIP5, four RCPs have been formulated that are based on a range of projections of future population growth, technological development, and societal response: a high emissions scenario (RCP8.5), midrange emissions scenarios (RCP6.0 and RCP4.5) and a low emissions scenario (RCP2.6) (Taylor et al., 2012, van Vuuren et al., 2011, Moss et al., 2010).

Simulations of regional climate over the European domain are taken from the Coordinated Regional Downscaling Experiment (CORDEX), in particular EURO-CORDEX (Jacob et al., 2014, Giorgi and Gutowski, 2015), which is considered the state-of-the-art RCM intercomparison project over Europe. EURO-CORDEX has been well-studied in the literature, especially in the context of CMIP5, for instance CMIP5 GCM ensemble evaluations (e.g. Jury et al., 2015), a thorough evaluation of temperature and precipitation simulations from ensembles of RCMs (e.g. Kotlarski et al., 2014), and it has been used in recent agricultural impact studies in Europe (e.g. Balkovič et al., 2018).

In this chapter, the closeness of climate simulations to observations is important to the research objectives as an impact assessment study demands realistic, or at least plausible, climate representation for the crop models. Apart from the evaluation of GCMs and GCM-RCMs, the RCMs themselves also need to be evaluated, which is undertaken in the next chapter (Chapter 5).

Because of the design of the study and the limited availability of simulations within the same experimental ensemble in public databases of GCM and GCM-RCM output, the selection of GCMs has to consider which GCM-RCM combinations are available under EURO-CORDEX. It is also important that the chosen GCM-RCM combinations are available into the future in order to create yield projections until the end of the century. In the following section, the basis for selection of RCMs and GCMs is discussed.

#### 4.3.2.2 RCM review and selection

Although there are numerous RCMs available from EURO-CORDEX, three RCMs are chosen due to their availability at the time of analysis for the needed climate variables from the historical (driven by GCMs) and evaluation simulations (for the evaluation of RCMs in Chapter 5) on a daily time step, a simulation period of sufficient length for the past (30 years or more), and if they are a member of the same experimental ensemble (i.e. r1p1i1). The three chosen RCMs are CCLM4-8-17, RACMO22E, and RCA4 (hereafter referred to as CCLM, RACMO and RCA, Table 4.1a), which are available at both 0.11 and 0.44°, which is a resolution of approximately 12 and 50 km, respectively. However, not all the chosen RCMs have simulations of a sufficient length (> 30 years) for the 0.11° resolution, so only the 0.44° resolution (EUR-44) simulations are used. In addition, it has been found that for seasonal mean quantities averaged over large European sub-domains (i.e. the research design of this chapter), no clear benefit of an increased spatial resolution was identified (Kotlarski et al., 2014).

These three RCMs are also chosen because of their common availability for future climate simulations forced by the RCP8.5 and RCP2.6 scenarios (the highest and lowest emission scenarios from Moss et al., 2010, van Vuuren et al., 2011) and in the newly-released (2017) EURO-CORDEX-Adjust, which is a database of bias-corrected future climate simulations performed by climate research centers within the CORDEX framework over the European domain. The three chosen RCMs have been well-used and evaluated in climate modeling studies over Europe, where they have generally shown satisfactory performances.

A key paper on the evaluation of RCMs over the European domain (Kotlarski et al., 2014) evaluated 17 simulations of temperature and precipitation from 6 RCMs (including CCLM, RACMO and RCA) at two different resolutions against E-OBS data. For temperature, their analysis revealed a cold bias of up to -2°C for most models, most seasons and most subdomains (Kotlarski et al., 2014). Over the British Isles, models typically had a dry bias in most seasons – in contrast to the generally positive precipitation biases over Europe, which were significantly larger in the higher-resolution experiments (Kotlarski et al., 2014).

Other evaluation studies have focused on individual GCM-RCM pairs. For example, RACMO nested within the EC-EARTH GCM (Hazeleger et al., 2010), showed significant improvements to climate simulations of fine-scale precipitation maxima and minima forced by the Alpine topography and the Italian coastlines, which were both well-captured by the RACMO RCM (Giorgi and Gutowski, 2015). A study that evaluated mean and extreme precipitation regimes over Spain using an ensemble of RCMs – including CCLM, RACMO, and RCA – showed good representation of the mean regimes and the annual cycle, but an overestimation of rainfall frequency that led to a wrong estimation of wet and dry spells (Herrera et al., 2010).

This means that while RCM performance may be acceptable, there are still significant biases that vary from one sub-domain to the next. Biases from RCMs were also found in a study that evaluated climate change indices of temperature and precipitation derived from the output of chosen RCMs, again including CCLM, RACMO and RCA (Dosio, 2016). The results of their study showed that, in general, the chosen RCMs underestimated maximum temperatures, performed relatively better in simulating minimum temperatures, but overestimated precipitation. Their study reports that RCA driven by GCM HadGEM2-ES, (Collins et al., 2011) showed the smallest biases (Dosio, 2016).

A different study that investigated added value found that improved skill from RCMs was not clear when downscaled output with RACMO and CCLM was compared to bias-corrected ECHAM5 GCM (Eden et al., 2014). While the comparison of directly bias-corrected GCMs compared to bias-corrected downscaled GCMs is not covered by the chapter research questions, and the research focuses instead on the evaluation of the chosen GCMs, GCM-RCMs, and RCMs themselves (the lattermost in the subsequent Chapter 5), it is argued that this important step of evaluating the climate models in the context of agricultural impact assessment is needed to better characterize uncertainty. In addition, this evaluation of driving GCMs and downscaling RCMs is argued to be crucial in building confidence in crop yield projections.

#### 4.3.2.3 GCM review and selection

Similar to RCM selection, driving GCMs are selected for the study based on their availability on several levels: on a daily timestep over a sufficient time period (>30 years) for maximum and minimum temperature as well as precipitation, within the same ensemble member from CMIP5 (r1i1p1), and lastly for future climate simulations forced by the RCPs 8.5 and 2.6. The subset of available CMIP5 GCM simulations that suited these criteria are also reviewed in terms of available scientific literature on their evaluation. The five chosen GCMs are CNRM-CM5.1, EC-EARTH, HadGEM2-ES, IPSL-CM5A-MR, and MPI-ESM-LR (referred to as CC, EC-EARTH, HadGEM, IPSL, and MPI, Table 4.1b) which all have varying resolutions of approximately 150 km. Historical runs for the chosen GCMs (and GCM-RCMs) are available from 1976-2005.

Several studies which have evaluated the chosen GCMs are discussed in the following paragraphs. In a review of CMIP5 GCM performance in the context of EURO-CORDEX, a model performance index was used to rank and evaluate surface temperature and precipitation for several GCMs, including the GCMs selected for this study, apart from EC-EARTH (Jury et al., 2015).

*Table 4.1: GCM and RCM selection and combinations.*

*Table 4.1a. Selected regional climate models (RCMs) from EURO-CORDEX for evaluation simulations.*

| **RCM, abbreviation** | **Downscaled resolution** | **Reference and institutes** |
|---|---|---|
| CCLM4-8-17 (CCLM) | 0.44° | Jaeger et al. (2008), CLM Community |
| RACMO22E (RACMO) | 0.44° | van Meijgaard et al. (2008), Royal Netherlands Meteorological Institute (KNMI) |
| RCA4 (RCA) | 0.44° | Kjellström et al. (2016), Rossby Centre, Swedish Meteorological and Hydrological Institute |

*Table 4.1b. Selected global climate models (GCMs) from CMIP5 for historical simulations.*

| **GCM/ESM, abbreviation** | **Resolution** | **Reference and institutes** |
|---|---|---|
| CNRM-CM5.1 (CC) | 1.406° x 1.406° | Voldoire et al. (2013), Centre National de Recherches Météorologiques (CNRM) and Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique |
| EC-EARTH | 1.125° x 1.125° | Hazeleger et al. (2010), Royal Netherlands Meteorological Institute (KNMI) |
| HadGEM2-ES (HadGEM) | 1.250° x 1.875° | Collins et al. (2011), Martin et al. (2011), UK Met Office Hadley Centre |
| IPSL-CM5A-MR (IPSL) | 1.25° x 1.875° | Dufresne et al. (2013), Institut Pierre Simon Laplace (IPSL) |
| MPI-ESM-LR (MPI) | 1.875° x 1.875° | Giorgetta et al. (2013), Max Planck Institute for Meteorology |

*Table 4.1c. Paired GCM-RCM combinations for evaluation.*

| **No.** | **Global climate model and regional climate model combination** |
|---|---|
| 1 | CCLM-MPI |
| 2 | RACMO-ECEARTH |
| 3 | RCA-CC |
| 4 | RCA-HadGEM |
| 5 | RCA-IPSL |
| 6 | RCA-MPI |

Based on their evaluation, the GCM MPI-ESM-LR had a model performance index higher than the multi-model mean, in addition CNRM-CM5 and HadGEM2-ES had above average performance compared to other GCMs in the study (Jury et al., 2015). In the same evaluation study, IPSL-CM5A-MR and the lower-resolution IPSL-CM5A-LR had lower average MPI with respect to the other GCMs, as it performed well for near-surface variables but poorly for upper-air variables (Jury et al., 2015). A different version of an IPSL GCM, IPSL-CM5B-LR, was found to have 'implausible' projections and all three IPSL models demonstrated poor realism of the annual cycle of rainfall in most regions in Europe (McSweeney et al., 2015, p.3237).

In a study that compared the precipitation output of 34 GCMs from CMIP5 to high-resolution satellite gauge-adjusted observations, selected climate models in the study were outperformed by the multi-model ensemble mean and median, although biases over Europe were generally lower than other regions (Mehran et al., 2014). Their sample of GCMs included the GCMs selected for the study, apart from EC-EARTH. The quantile bias analyses in the study indicated that CMIP5 simulations are particularly biased at high quantiles (extremes) of precipitation (Mehran et al., 2014). Following their method of using a 'volumetric hit index' – which is the volume of precipitation detected correctly by GCMs above a set threshold – the chosen GCMs were thus ranked (from best-performing): MPI, CC, HadGEM, IPSL. However, after the removal of mean-field bias, these GCMs had fairly even index scores. Output apart from simulated precipitation extremes were improved after correction (Mehran et al., 2014).

Other studies which assess GCM performance have found that IPSL-CM5A-LR (the lower-resolution version of the IPSL model that is used in this chapter) performed relatively well compared to observational data (Yoo and Cho, 2018), based on simple statistical measures, namely root mean square error ($RMSE$) and correlation coefficient of empirical orthogonal functions (EOF). This is in contrast to the aforementioned studies which have found IPSL to have poor to mixed results (e.g. Jury et al., 2015, McSweeney et al., 2015). In the Yoo and Cho (2018) study, EC-EARTH was included in the 20-member MME, and it showed relatively small normalized $RMSE$ for gridded data within the ensemble. In addition, the first EOF of CC, IPSL and EC-EARTH explained approximately 97.5% of the pattern variance of output, which was very close to the variance explained by observations (97.23%) (Yoo and Cho, 2018).

Therefore, the selected GCMs for the study that meet the availability criteria show relatively fair performances based on a review of the literature; however they are also reported to have several biases, and these are investigated in the following results section.

#### 4.3.2.4 Processing the GCM-RCM combinations

After the literature review and data availability checks, the final GCM-RCM combinations which are used for the past climate evaluation are the following: CCLM-MPI, RACMO-ECEARTH, RCA-CC, RCA-HadGEM, RCA-IPSL, and RCA-MPI (Table 4.1c). Although this is a relatively small ensemble of 5 GCMs and 6 GCM-RCM combinations, it is found suitable for the research questions of the study.

CMIP5 and EURO-CORDEX use netCDF as the file format for climate model output. The rotated lat-lon grid of climate model simulations are transformed to a regular lat-lon grid to match the E-OBS (Haylock et al., 2008) and NUTS grids using the $remapcon$ and $setgrid$ functions from the Climate Data Operators ($cdo$) toolkit (CDO, 2018). Further processing of netCDF files is completed with the R package $ncdf4$ (Pierce, 2017). Country and regional land-based grid cells for the UK, Germany, and German regions (See Table 3.1) are selected based on geographical boundaries defined by the European NUTS gridding system. Similar to the selection of grid cells for observations, the grid cells that matched the lat-lon grid coordinates of GCM and RCM simulations are extracted.

Selected grid cells are then aggregated to represent country and regional averages of climate simulations for the needed variables of maximum and minimum temperature, and precipitation for 1976-2005 for the historical GCM and GCM-RCMs. The processed GCM and GCM-RCM output are then compared to observations and to each other with the statistical methods discussed at the end of this data and methods section, recalling that it is the hypothesis that uncorrected GCM-RCM output has a better performance compared to uncorrected GCM output.

### 4.3.3 Statistical analyses and evaluation

#### 4.3.3.1 Comparing the output of GCMs and RCMs to observations

Building on the work from the previous chapter, daily values of maximum and minimum temperature as well as precipitation for the UK, Germany and four German regions, are taken from E-OBS, which is a gridded dataset of land-only gridded daily high-resolution estimates of these climate variables in Europe (Haylock et al., 2008). An advantage of using E-OBS is its spatial and temporal coverage, which makes it ideal for an approximate evaluation of RCM-simulated temperature and precipitation characteristics over Europe (Kotlarski et al., 2014).

In the analysis of climate simulations and BC methods, several statistical metrics are used to evaluate climate model performances, such as bias (simulations minus observations), correlation ($r$), and $RMSE$. Mean bias is calculated for annual and seasonal maximum and minimum temperature (Tmax, Tmin) for the period between 1976-2005, where a negative (positive) mean bias indicates that simulated temperatures are cooler or have fewer hot days (warmer or more hot days) than observations. For precipitation, a negative (positive) mean bias indicates that total annual or seasonal precipitation is underestimated (overestimated). To answer the research question on added value of RCMs, "simple" or conventional statistical measures of $RMSE$ and correlation are adopted, similar to the approach of Yoo and Cho (2018). This straightforward approach is a step to show the relative performance of RCMs compared to GCMs.

#### 4.3.3.2 Challenging tests of added value

As reported earlier in this chapter (Section 4.1.2), the concept and proof of added value from RCMs is strongly debated in the scientific community. There are many key studies that are on opposite ends of the debate: for example, some studies have confirmed that there is added value from RCMs, particularly for climate projections (e.g. Rummukainen, 2016), while other studies have found that in a setup where GCMs and GCM-RCMs are both directly post-processed (bias corrected) there is no clear added value by RCMs (e.g. Eden et al., 2014). Therefore, it must be acknowledged that the design of the chapter to find this added value in uncorrected projections is a difficult test, as is any added value test that is largely dependent on the context (e.g. model, variable, scale, region, experiment set-up) (Rummukainen, 2016).

Furthermore, correlations are a particularly challenging test for precipitation compared to precipitation as the link between temperature and external forcings is more clear. For instance, it has been shown that the addition of anthropogenic forcings to climate model simulations produces better agreement with the evolution of observed temperatures (Meehl et al., 2012). However challenging, it is argued, after the review of literature in Chapter 2, that finding and characterizing added value from RCMs is a method that is not common in disciplinary impact assessment studies (e.g. crop modeling studies), as assumptions are often made that climate model output is automatically 'better' using finer-scale data. However, given the growing dependence and demand of impact studies for high-resolution climate projections, issues such as the value of downscaling must also be taken into account when considering the cascade of uncertainty in climate impact assessment.

## 4.4 Results

In this section, the results of the comparison of (1) uncorrected historical GCM simulations to the same output downscaled by an RCM, and (2) the comparison of both the raw and downscaled output to observations are reported.

### 4.4.1 Comparing climate model output to observations

The results of the comparison of climate model output to observations is reported here in order to assess how well-represented temperature and precipitation are in both historical GCM and GCM-RCM (downscaled) output.

#### 4.4.1.1 Biases and error in simulated maximum and minimum temperature (country level)

At the country level, the 5 chosen GCMs and 6 GCM-RCM combinations show biases relative to observations for annual averages of maximum and minimum temperature. Generally, it can be observed that both GCMs and GCM-RCMs underestimate maximum temperatures (Tmax, Figs. 4.2A and 4.3A), where bias of the ensemble mean of GCMs is smaller than GCM-RCMs. There are smaller observed biases for minimum temperature (Tmin, Figs. 4.2B and 4.3B) from both GCMs and GCM-RCMs, and the ensemble mean of the latter has a smaller/similar bias compared to GCMs.

These biases are also reflected in the calculated $RMSE$ relative to observations over the same period (Tables 4.2 and 4.3). GCM-RCMs have larger $RMSE$ than GCMs for Tmax: 1.89 and 1.4°C for the UK and Germany, compared to $RMSE$=1.37 and 1.24°C using the ensemble of GCMs. For the UK, Tmin is overestimated by historical GCMs, but this is improved using RCMs, and the multi-model mean is closer to observations: $RMSE$=1.05 (0.43°C) for the ensemble mean of historical GCMs (GCM-RCMs). For Germany, the ensemble mean of Tmin simulations by both GCMs and GCM-RCMs have similar error relative to observations ($RMSE$=0.7 for GCMs, $RMSE$=0.6 for GCM-RCMs).

In terms of individual climate model performances, historical simulations from IPSL significantly underestimate Tmax over the UK, even when downscaled with RCA ($RMSE$=2.88°C and $RMSE$=2.68°C with RCA). Historical simulations from MPI for Tmin over the UK and Germany have a large positive bias and $RMSE$ (2.7°C for the UK and 1.72°C for Germany); using the RCMs like CCLM and RCA reduce this bias, for example CCLM-MPI brings Tmin simulations closer to observations ($RMSE$=0.85°C for the UK and 0.94°C for Germany).

It can be observed that RACMO has different effects on output from driving GCM EC-EARTH: downscaled RACMO-ECEARTH Tmin has reduced error and bias relative to the historical GCM simulations for the UK ($RMSE$=1.41 to 0.71°). However, RACMO increases the $RMSE$ for Germany Tmin compared to its historical GCM simulations and observations ($RMSE$=0.95 to 1.79°C).

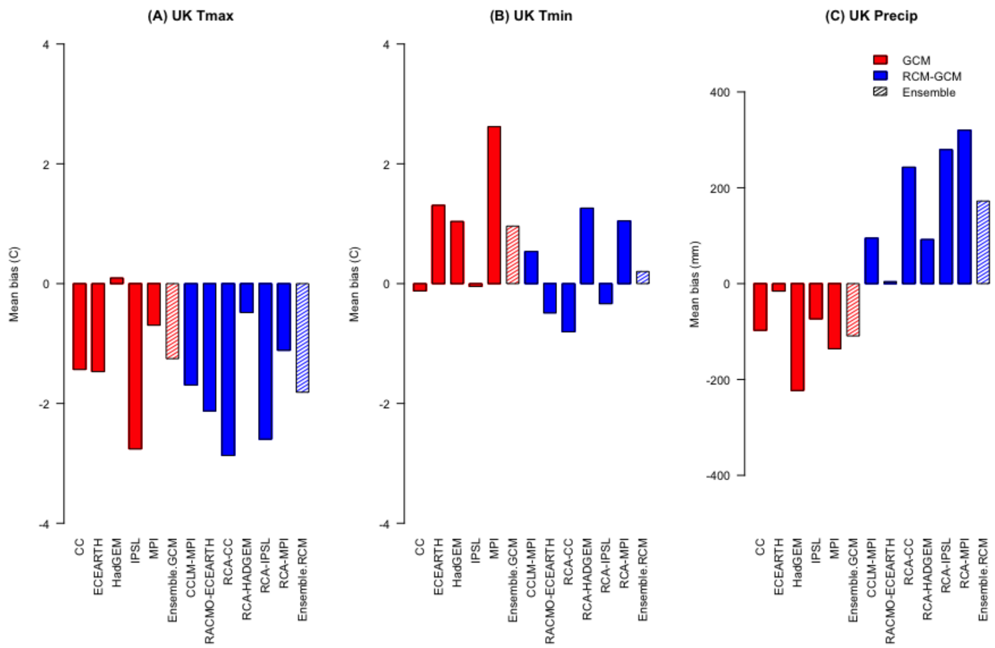



*Figure 4.2: Mean bias of GCM and GCM-RCM simulations: A) annual average maximum temperature, b): minimum temperature, and C) total annual precipitation for 1976-2005 in the UK.*

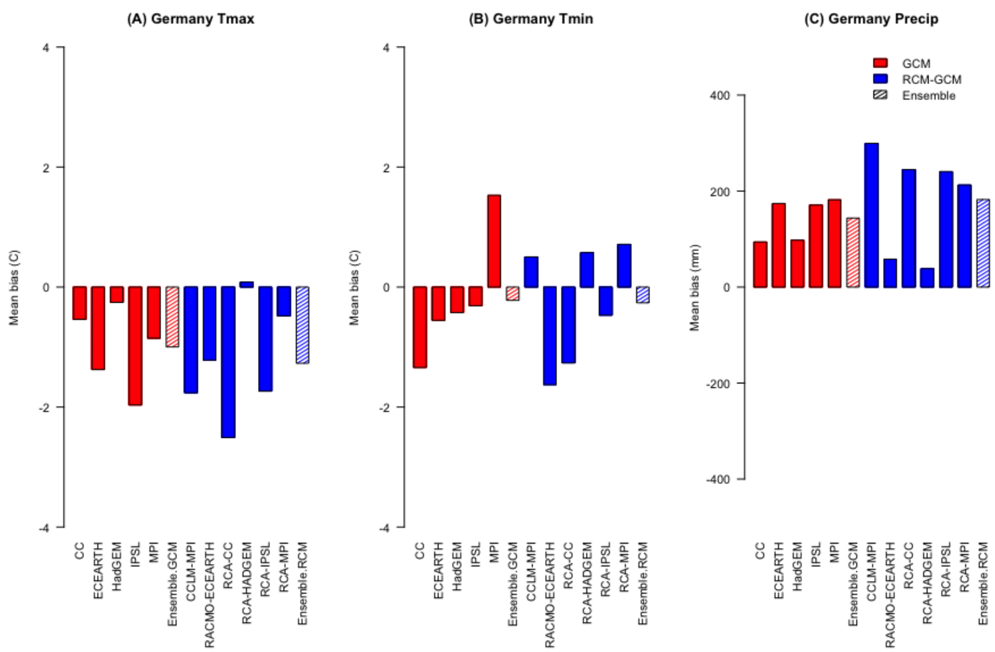



*Figure 4.3: Mean bias of GCM and GCM-RCM simulations: A) annual average maximum temperature, b): minimum temperature, and C) total annual precipitation for 1976-2005 in Germany.*

#### 4.4.1.2 Biases and error in simulated precipitation (country level)

For total annual precipitation, historical GCMs are shown to generally underestimate rainfall for the UK, and positive biases generally increase when using RCMs over the UK and Germany (Figs. 4.2C and 4.3C), which is also reflected in the large $RMSE$ values for GCM-RCMs compared to GCMs alone. For example, RCA-downscaled MPI has a large $RMSE$ of 347mm compared to the 193mm from MPI alone (Table 4.3).

*Table 4.2: $RMSE$ between historical GCM Tmax, Tmin and Precip to observations.*

| GCM | Annual Tmax | | Annual Tmin | | Annual Precip | |
|---|---|---|---|---|---|---|
| | UK | Germany | UK | Germany | UK | Germany |
| **CC** | 1.6 | 1.11 | 0.58 | 1.62 | 170.42 | 172.88 |
| **ECEARTH** | 1.59 | 1.55 | 1.41 | 0.95 | 156.42 | 220.68 |
| **HadGEM** | 0.7 | 0.98 | 1.15 | 0.97 | 261.92 | 177.1 |
| **IPSL** | 2.88 | 2.2 | 0.64 | 1.04 | 156.64 | 210.06 |
| **MPI** | 1.09 | 1.39 | 2.7 | 1.72 | 193.69 | 245.14 |
| **Ens.mean** | 1.37 | 1.24 | 1.05 | 0.72 | 164.7 | 184.72 |

*Table 4.3: $RMSE$ between historical GCM-RCM simulations Tmax, Tmin and Precip to observations.*

| GCM-RCM | Annual Tmax | | Annual Tmin | | Annual Precip | |
|---|---|---|---|---|---|---|
| | UK | Germany | UK | Germany | UK | Germany |
| **CCLM-MPI** | 1.89▲ | 2.04▲ | 0.85▼ | 0.94▼ | 161.33▼ | 324.77▲ |
| **RACMO-ECEARTH** | 2.2▲ | 1.4▼ | 0.71▼ | 1.79▲ | 144.69▼ | 139.05▼ |
| **RCA-CC** | 2.94▲ | 2.68▲ | 0.99▲ | 1.51▼ | 286.18▲ | 271.89▲ |
| **RCA-HADGEM** | 0.77▲ | 0.76▼ | 1.35▲ | 0.94▼ | 175.59▼ | 152.88▼ |
| **RCA-IPSL** | 2.68▼ | 2▼ | 0.59▼ | 0.92▼ | 305.39▲ | 274▲ |
| **RCA-MPI** | 1.36▲ | 1.05▼ | 1.23▼ | 1.05▼ | 347.28▲ | 257.11▲ |
| **Ens.mean** | 1.89▲ | 1.4▲ | 0.42▼ | 0.6▼ | 206.62▲ | 213.03▲ |

*A ▲(▼) indicates a relative increase (decrease) in $RMSE$ relative to the driving GCM.*

#### 4.4.1.3 Regional-level annual average temperatures and total precipitation

At the German regional level, both historical GCMs and GCM-RCM simulations of annual average Tmax, Tmin and total annual precipitation show biases relative to observations, although the size and sign of the bias varies from region to region as well as for the particular climate variable analyzed. For example, similar to the national level, GCMs and GCM-RCMs generally underestimate annual average Tmax in the German regions (Figs. 4.4A-4.7A). In contrast, for all the four regions analyzed, the multi-model mean of annual average Tmin has small biases and small $RMSE$ relative to observations for both GCMs and GCM-RCMs, typically under 1°C (Figs. 4.4B-4.7B, Tables 4.4 and 4.5). The use of RCMs has mixed effects on the ensemble means: it only reduces $RMSE$ in DE2 (Tmax), DED (Tmin), and DEF (Tmin).

At the individual model level, more reductions of $RMSE$ can be observed: for example, RACMO-ECEARTH (Tmax DE2), RCA-MPI (reductions in $RMSE$ across all regions for Tmax and Tmin), RCA-IPSL (Tmax DE2, DED, DEF, Tmin DEA, DED and DEF), RCA-HadGEM (Tmax for all regions, Tmin DE2, DEA), RCA-CC (Tmin DE2, DEF), and CCLM-MPI (Tmin across all regions). Similar to the country-level analysis, total annual precipitation is generally overestimated by both GCMs and GCM-RCMs (Figs. 4.4C-4.7C), and some of these biases increase when using RCMs, for example with CCLM-MPI in DE2, DEA and DED; RCA-CC in DEA and DED; and RCA-MPI in DEA, DED, and DEF.

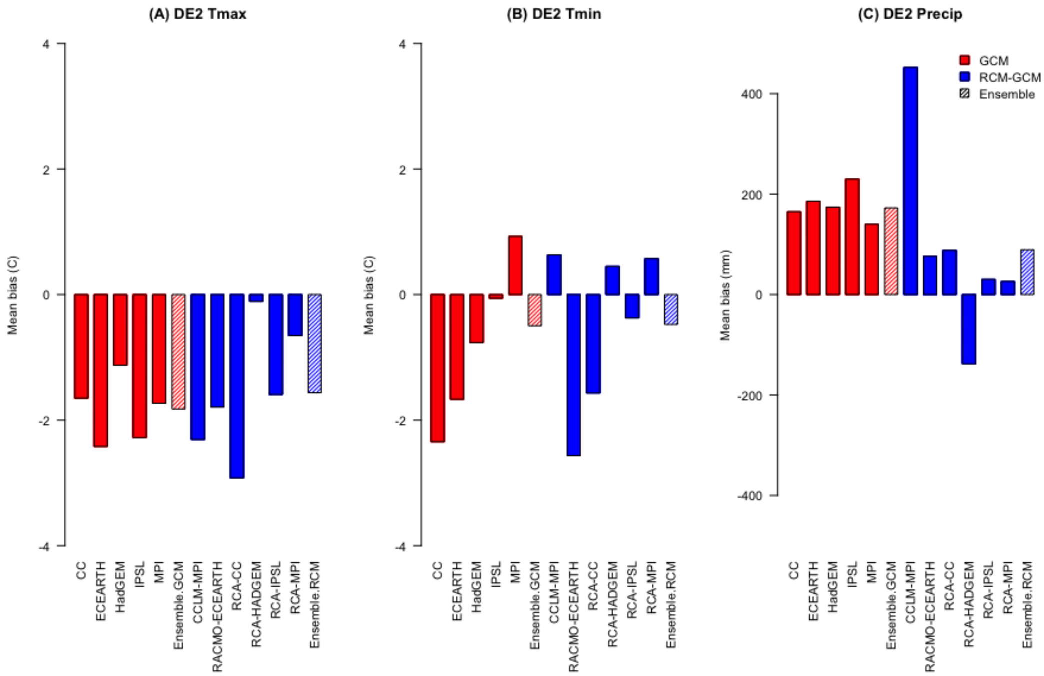



*Figure 4.4: Mean bias of GCM and GCM-RCM simulations: A) annual average maximum temperature, B): minimum temperature, and C) total annual precipitation for 1976-2005 in DE2 (South Germany).*

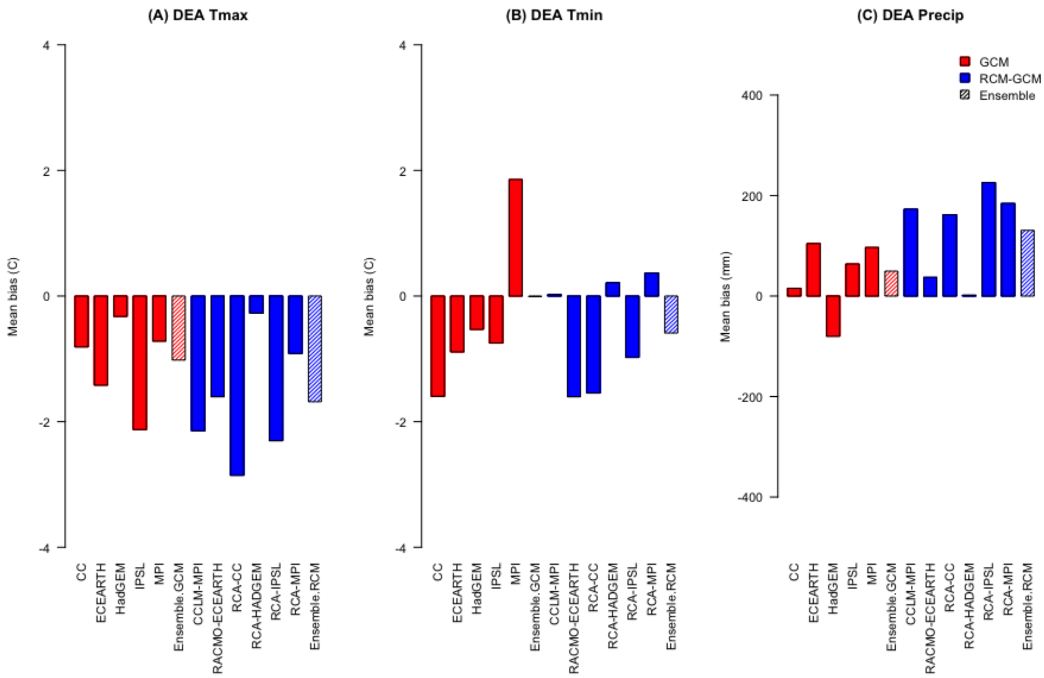



*Figure 4.5: Mean bias of GCM and GCM-RCM simulations: A) annual average maximum temperature, B): minimum temperature, and C) total annual precipitation for 1976-2005 in DEA (West Germany).*

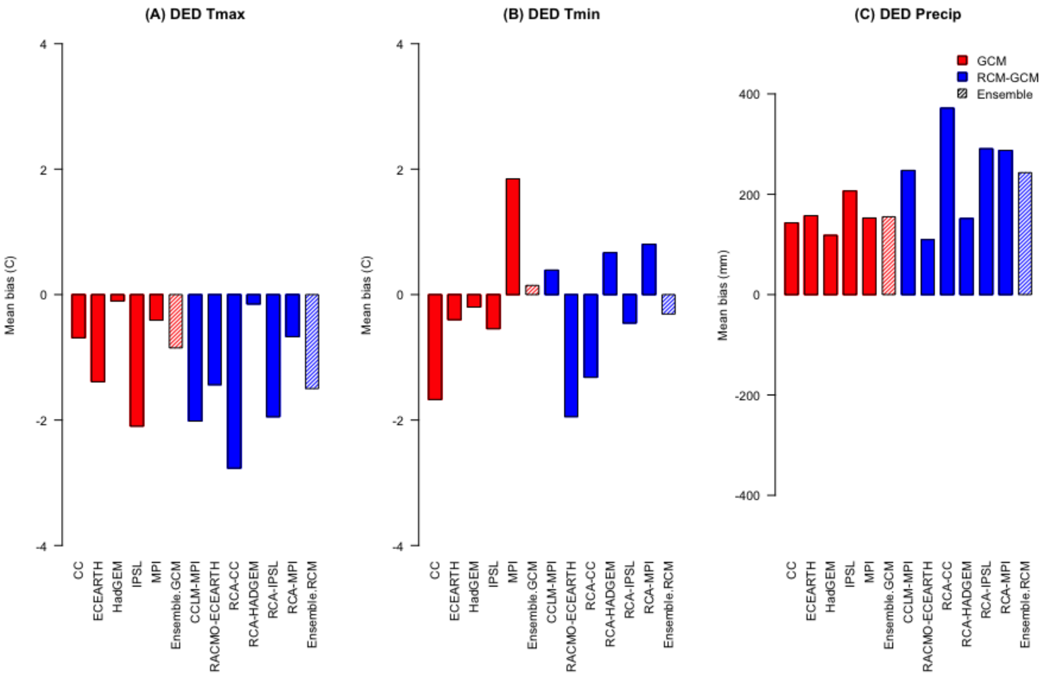



*Figure 4.6: Mean bias of GCM and GCM-RCM simulations: A) annual average maximum temperature, B): minimum temperature, and C) total annual precipitation for 1976-2005 in DED (East Germany).*

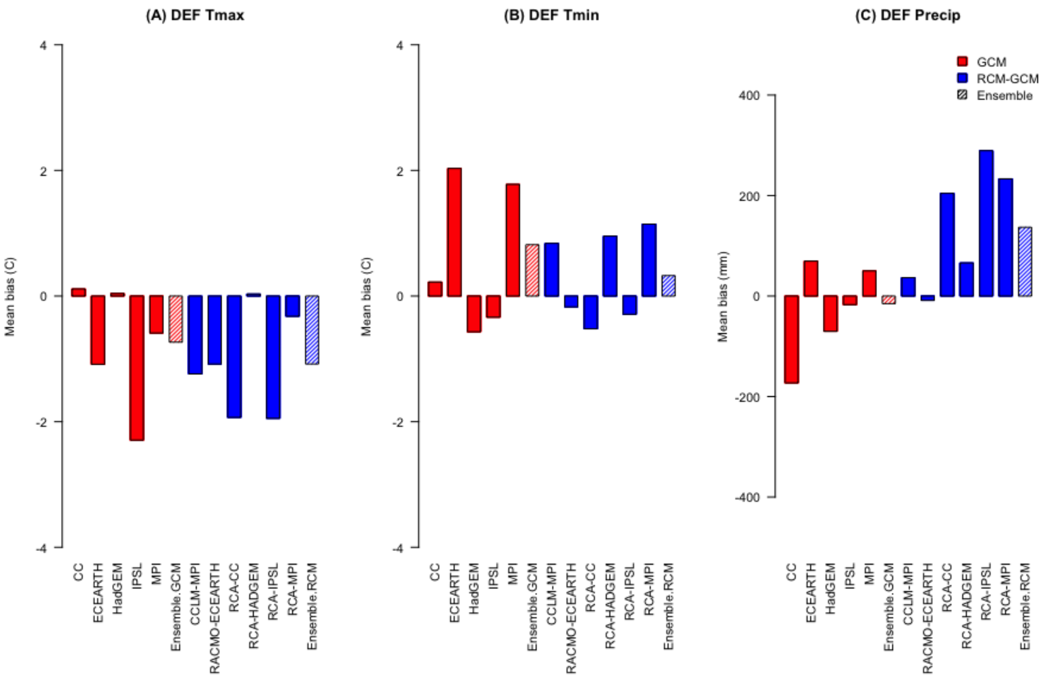



*Figure 4.7: Mean bias of GCM and GCM-RCM simulations: A) annual average maximum temperature, b): minimum temperature, and C) total annual precipitation for 1976-2005 in DEF (North Germany).*

*Table 4.4: $RMSE$ between historical GCM Tmax, Tmin and Precip to observations, German regions, 1976-2005.*

| GCM-RCM | Annual Tmax | | | | Annual Tmin | | | | Annual Precip | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **DE2** | **DEA** | **DED** | **DEF** | **DE2** | **DEA** | **DED** | **DEF** | **DE2** | **DEA** | **DED** | **DEF** |
| **CC** | 1.96 | 1.37 | 1.43 | 1.03 | 2.52 | 1.85 | 1.97 | 0.93 | 235.25 | 164.76 | 211.04 | 230.98 |
| **ECEARTH** | 2.57 | 1.64 | 1.69 | 1.35 | 1.85 | 1.18 | 0.95 | 2.17 | 235.42 | 191 | 214.99 | 188.4 |
| **HadGEM** | 1.51 | 1.1 | 1.18 | 0.99 | 1.19 | 0.99 | 0.99 | 1.06 | 248.73 | 178.02 | 186.9 | 178.39 |
| **IPSL** | 2.52 | 2.4 | 2.44 | 2.55 | 1 | 1.22 | 1.18 | 1.13 | 272.54 | 158.82 | 248.88 | 145.07 |
| **MPI** | 2.05 | 1.38 | 1.26 | 1.35 | 1.2 | 2.03 | 2 | 2.03 | 242.27 | 204.68 | 228.44 | 164.05 |
| **Ens.mean** | 2.01 | 1.35 | 1.28 | 1.15 | 0.85 | 0.69 | 0.72 | 1.11 | 220.29 | 146.4 | 199.68 | 137.56 |

*Table 4.5: $RMSE$ between historical GCM-RCM Tmax, Tmin and Precip to observations, German regions, 1976-2005.*

| GCM-RCM | Annual Tmax | | | | Annual Tmin | | | | Annual Precip | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **DE2** | **DEA** | **DED** | **DEF** | **DE2** | **DEA** | **DED** | **DEF** | **DE2** | **DEA** | **DED** | **DEF** |
| **CCLM-MPI** | 2.52▲ | 2.41▲ | 2.32▲ | 1.7▲ | 0.99▼ | 0.79▼ | 0.89▼ | 1.25▼ | 495.97▲ | 252.7▲ | 287.99▲ | 160.37▼ |
| **RACMO-ECEARTH** | 2▼ | 1.77▲ | 1.7▲ | 1.34▼ | 2.68▲ | 1.78▲ | 2.1▲ | 0.74▼ | 167.49▼ | 155.96▼ | 176.87▼ | 161.85▼ |
| **RCA-CC** | 3.12▲ | 3.02▲ | 3▲ | 2.15▲ | 1.81▼ | 1.74▼ | 1.64▼ | 0.96▲ | 181.27▼ | 209.33▲ | 398.12▲ | 259.24▲ |
| **RCA-HADGEM** | 0.82▼ | 0.9▼ | 0.93▼ | 0.81▼ | 0.91▼ | 0.75▼ | 1.07▲ | 1.2▲ | 218.81▼ | 162.21▼ | 235.15▲ | 171.38▼ |
| **RCA-IPSL** | 1.95▼ | 2.52▲ | 2.3▼ | 2.15▼ | 0.91▼ | 1.23▲ | 1.04▼ | 0.87▼ | 150.49▼ | 274.69▲ | 324▲ | 322.4▲ |
| **RCA-MPI** | 1.13▼ | 1.36▼ | 1.25▼ | 1.1▼ | 0.93▼ | 0.87▼ | 1.1▼ | 1.46▼ | 159.48▼ | 249.72▲ | 328.32▲ | 291.7▲ |
| **Ens.mean** | 1.7▼ | 1.82▲ | 1.69▲ | 1.3▲ | 0.72▼ | 0.8▲ | 0.64▼ | 0.69▼ | 159.35▼ | 183.67▲ | 269.86▲ | 188.55▲ |

*A ▲(▼) indicates a relative increase (decrease) in $RMSE$ relative to the driving GCM.*

### 4.4.2 Comparison of relative GCM and RCM performance: correlation

In the previous section, the relative bias and error of GCM and GCM-RCM output was compared to observations. This section continues with the comparison of GCMs and downscaled GCM output in order to answer research questions on added value. In this section, the correlation between GCM and GCM-RCM output is reported. It is important for climate model output to match closely with observations because the realism of climate input data is crucial to plausible yield projections, so a higher correlation statistic is considered favorable.

The performance of GCMs and RCMs driven by GCMs are compared to how closely they are correlated to observations (Figs. 4.8 and 4.9), as well as the four German regions used in the study (Figs. 4.10 and 4.13). Overall, it can be observed that the correlation statistic of the multi-model mean of GCM-RCMs exceeds that of the multi-GCM mean in the UK and Germany for Tmax and Tmin. For example, the correlation of the GCM-RCM ensemble mean for Tmax and Tmin is over $r$=0.6 for the UK, Germany, DE2 (Tmax only), DEA, and DEF (Tmin only). In contrast, the ensemble mean correlation of historical simulations from GCMs is not significant for Tmax nor Tmin in any of the German regions as well as at the country level in Germany.

For precipitation, both GCMs and downscaled GCMs typically do not have any significant correlation for the UK and Germany and at the regional level. This indicates that precipitation simulation from climate models needs significant improvement, a known limitation of climate models and their ability to simulate precipitation (See Chapter 2, Section 2.2.2).

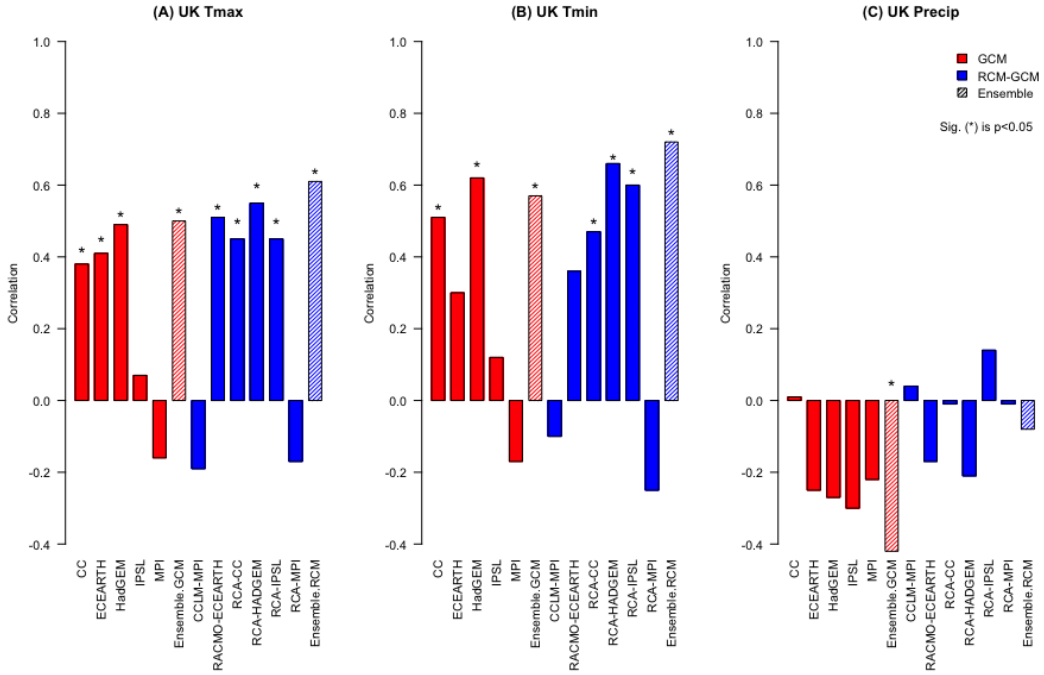



*Figure 4.8: Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for the UK.*

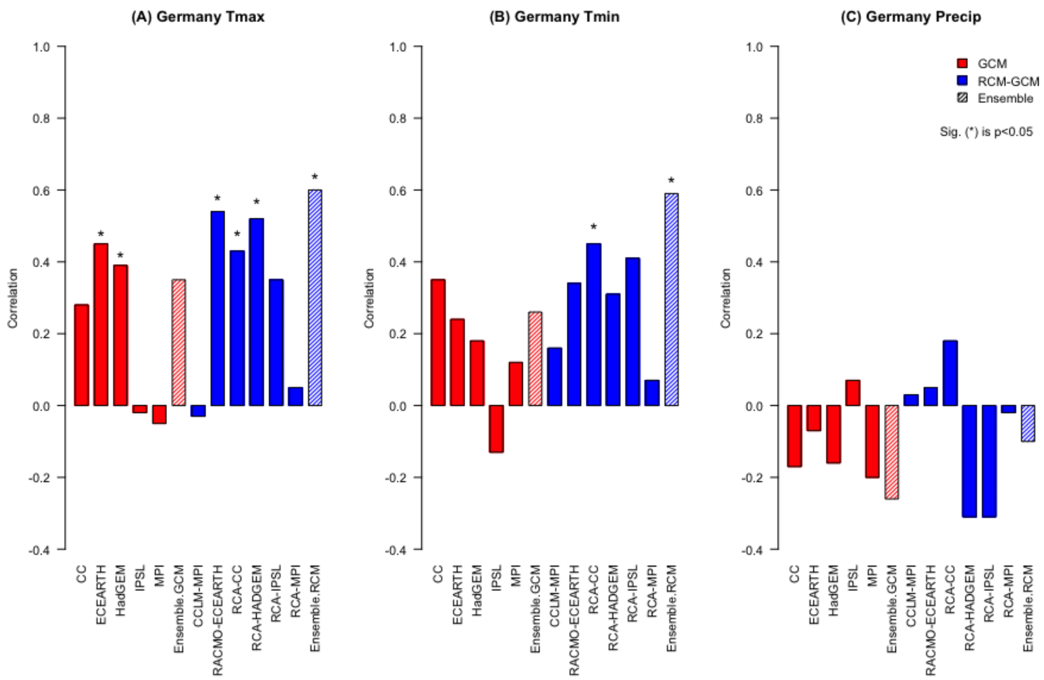



*Figure 4.9: Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for Germany.*

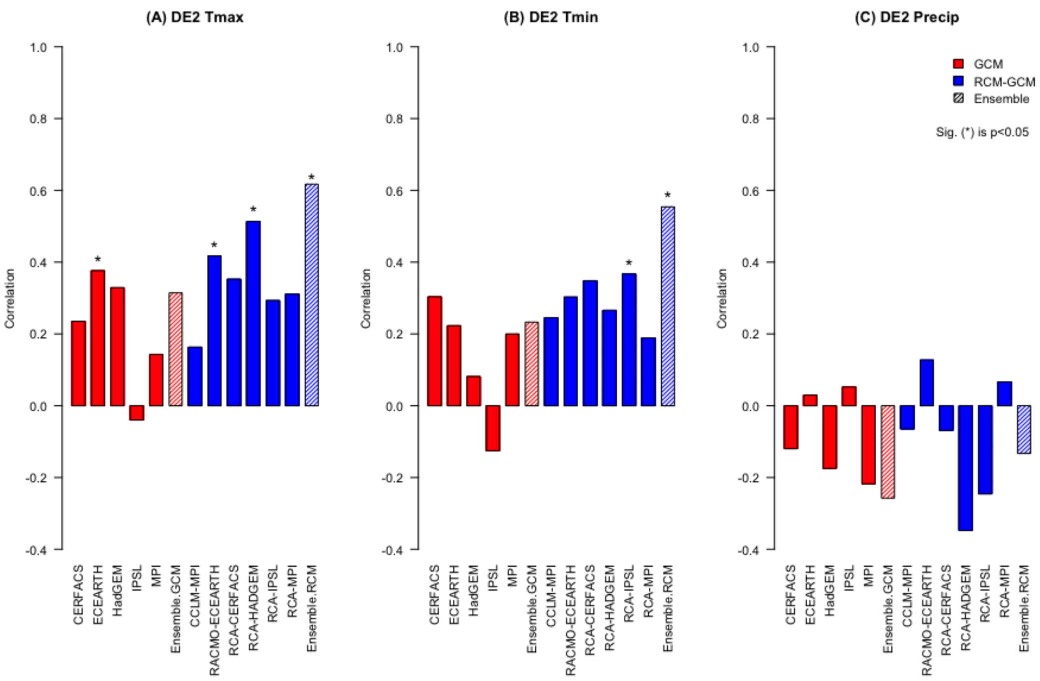



*Figure 4.10: Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for DE2.*

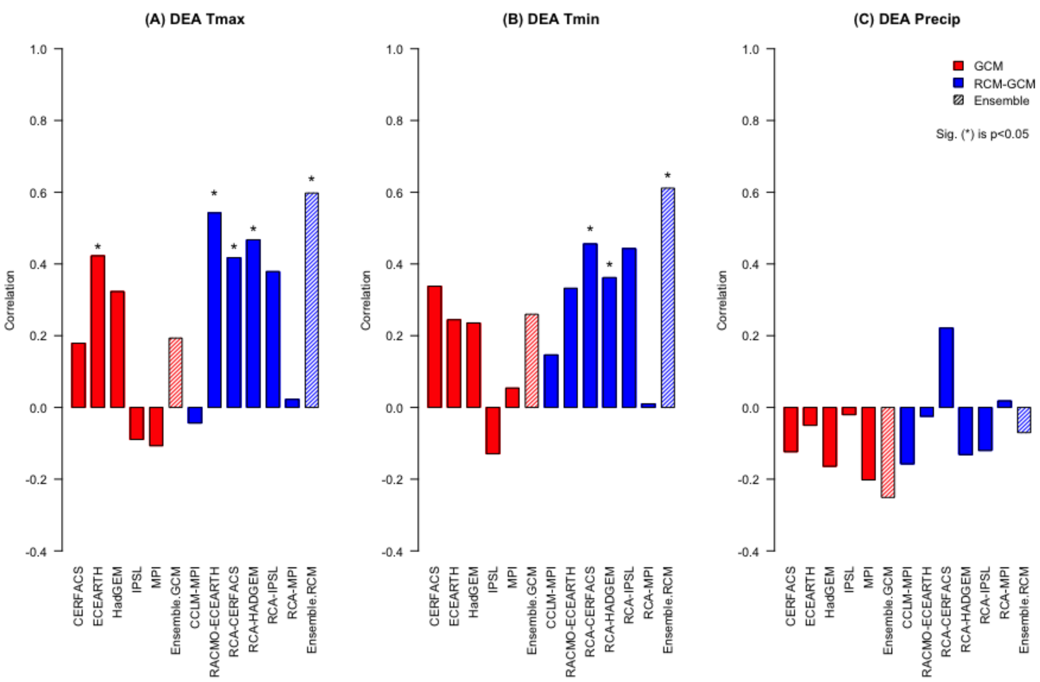



*Figure 4.11: Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for DEA.*

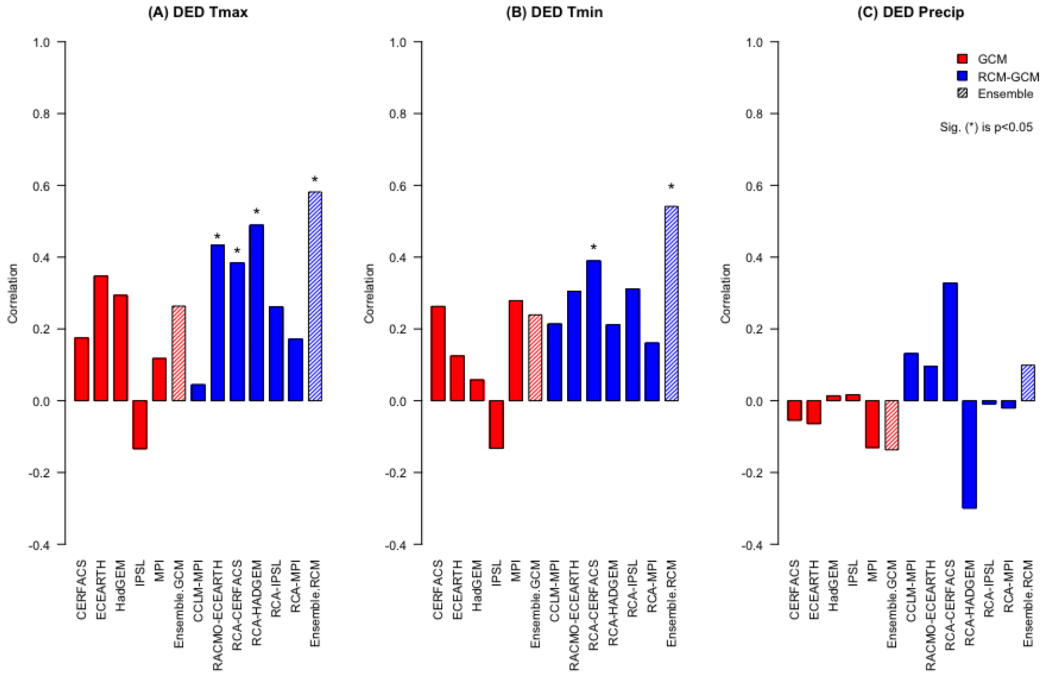



*Figure 4.12: Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for DED.*

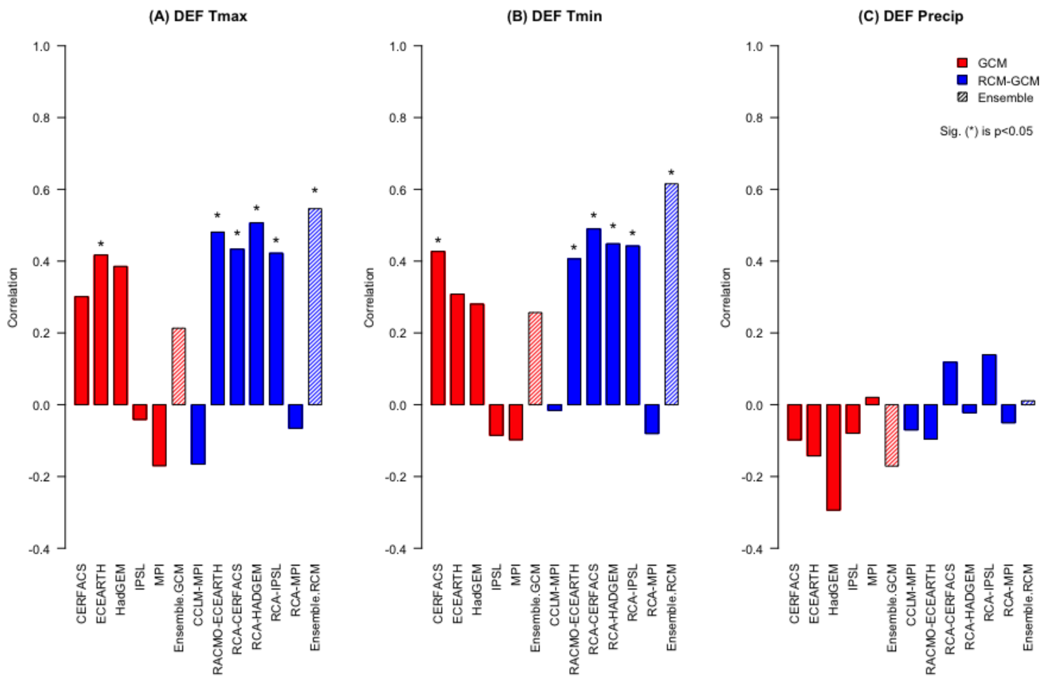



*Figure 4.13: Correlation of historical GCM and RCM-downscaled GCM simulations of total annual precipitation, 1976-2005 for DEF.*

## 4.5 Discussion

The chapter was designed to answer questions on how well GCMs and GCM-RCMs can capture past climate trends and how simulations from GCMs and GCM-RCMs compare to each other, in the context of the debated added value of RCMs. The main findings of the chapter, contextualized by the research questions (RQs) are as follows:

### 4.5.1 Summary of findings

- Historical simulations of GCMs and their dynamically downscaled simulations (GCM-RCMs) contain significant biases relative to observations on annual timescales, but the multi-model mean of the analyzed climate variables (Tmax, Tmin and precipitation) has smaller biases relative to observations and outperforms individual models. **(RQ 1)**

- In general, both GCMs and GCM-RCMs tended to underestimate Tmax, overestimate Tmin, and have a wide range of biases for precipitation.

- Comparing GCMs and GCM-RCMs with simple statistical metrics such as correlation and $RMSE$ showed that the multi-model mean of GCM-RCMs typically outperformed the multi-model mean of GCMs, as well as individual GCM-RCM pairs, and individual GCMs in terms of correlation. **(RQ 2)**

- Total annual precipitation from historical simulations was not well-correlated to observations. It also had large positive biases relative to observations; although the former is is not unexpected given that they were unforced by observations, and the latter considering known challenges in the simulation of precipitation by climate models. Therefore, it was observed that the performance of GCMs in simulating precipitation over the UK, Germany and German regions needs significant improvement.

- Overall, the relatively better performance of GCM-RCMs in the study design (better correlation, some smaller errors, and benefits associated with higher resolution) was considered sufficent evidence for their use in projecting the impacts of climate change on crops, at least for the context of the study. A more thorough characterization of added value beyond the chosen metrics is out of the scope of the research but it is clear that this GCM and GCM-RCM comparison is necessary to avoid overstating the 'benefits' of RCMs. **(RQ 3)**

In the following section, these key results are discussed, focusing on how these results relate to the main research objectives.

### 4.5.2 Comparing of GCMs and GCM-RCM simulations

The major research question addressed by this chapter is how much do RCMs add to the GCM simulations – do they improve them, make them worse, or make no difference? In terms of $RMSE$, the advantages of using downscaled simulations are shown to be mixed: in many cases the use of an RCM reduces $RMSE$, particularly for Tmin (See Tables 4.2 and 4.4) and in other cases, using RCMs increases the measured error.

For example, using downscaled simulations versus historical simulations reduces $RMSE$ for Tmax in the UK and Germany for RACMO-ECEARTH, RCA-HadGEM, RCA-IPSL (Tables 4.2 and 4.3). Downscaled climate output also has smaller error for Tmin, with the multi-model mean ensemble error having significantly smaller $RMSE$. In contrast however, the multi-model mean of downscaled Tmax and precipitation simulations for the UK and Germany had larger in $RMSE$ relative to GCM-only simulations.

In this chapter, it was also shown that correlations were often better, and some biases relative to observations were reduced, when using GCM-RCMs compared to uncorrected historical GCM output. For example: the more well-correlated ensemble mean of Tmin for the UK and Germany, and across all chosen German regions, individual GCMs and RCMs (e.g. Tmax with RCA-IPSL over IPSL alone in the UK and Germany, RACMO-ECEARTH over EC-EARTH alone in the UK, among others). However, the use of RCMs did not always result in smaller errors. In fact, when using RCMs for precipitation, some of these errors increased substantially (e.g. CCLM-MPI for Germany).

These imbalanced advantages of using RCMs have been reported in other studies and in the review (See Sections 4.1.2, 4.3.3.2). It is argued, therefore, based on these results, that the proof of added value from RCMs is challenging – in fact, many studies that have attempted to find this added value from RCMs have not shown unequivocal gains (Di Luca et al., 2016). Therefore, although some reduction of $RMSE$ is observed in precipitation when using RCMs, the choice of using GCM-RCMs over GCMs only does not appear to be unilaterally in favor of RCMs. Rather than this being a discouraging result in favor of the additional spatial and temporal information already known to be provided by RCMs, it provides an interesting result, considering that RCMs are often accepted as 'better' than GCMs simply because of the more spatial resolution they provide. The results reported here are a reminder that claims of added value should always be verified.

However, in order to move forward with the analysis and provide usable climate model output for the impact assessment, it should be recalled that the tests used here are relatively 'tough' tests for added value – particularly for precipitation– given that climate models still need significant improvement to represent the complexity of the atmospheric processes that drive the representation of precipitation in climate models. In addition, this larger bias and error could be due to the drizzle effect associated with climate models. This tendency of climate models, particularly RCMs, to overestimate the occurrence of wet days and underestimate heavy precipitation is well known (Maraun, 2013, Maraun et al., 2010, Piani et al., 2010). Although the results show that not all RCMs could not improve the accuracy of precipitation simulations, it remains that even un-downscaled GCM simulations of precipitation contain considerable error, therefore RCMs that are driven by the same GCMs cannot be expected to completely compensate for these errors.

Therefore, it can be reported that the selected RCMs in this study are able to result in improvements in the multi-model ensemble mean's correlation to observations for maximum and minimum temperatures, and are also able to reduce some bias and error – particularly for minimum temperatures – relative to the historical driving GCM simulations. These reasons of improved correlation, and some reduction of biases, are deemed sufficient justification for the use of RCMs in the study and RCMs are found to be appropriate to answer the research questions of the study. However, errors remain, and are introduced by RCMs themselves: because of this, the selected RCMs are evaluated and bias-corrected in the next chapter.

### 4.5.3 Connecting to the added value debate: advantages of regional climate and impact modeling

The debate on added value of RCMs was an important question to address in the chapter because of the 'garbage-in-garbage-out' discussion which questions the utility of RCMs (e.g. Maraun, 2016, Jury et al., 2015, Giorgi and Gutowski, 2015), that is, if RCMs will produce anything better if they still remain driven by GCMs that contain significant errors. Despite the advantages of modeling climate and impacts at regional scales, it cannot be discounted that RCMs are still highly dependent on the skill of GCMs. Additionally, it is not clear nor obvious whether RCMs can improve the larger-scale climate properties, such as the continental to sub-continental surface temperature distribution (Sørland et al., 2018).

A critical study which investigated downscaling jointly with bias correction showed that after correction, GCM- and RCM-driven US maize yields were essentially indistinguishable (Glotter et al., 2014). Although the comparison of bias-corrected GCM-only output used to simulate yields is not in the scope of the research, this points to the importance of taking a critical stance on the utility of RCMs. RCMs should not be used like 'black boxes' for producing regional climate information (Giorgi and Gutowski, 2015, p.471) with the automatic assumption that their output is better.

With multiple criticisms on the utility of regional climate modeling, why advocate for regional scales or RCMs? It is argued that these more local scales will continue to be significant in impact assessment, and that this significance may continue to increase. The development of adaptation and response strategies to climate change depend on information and geographical features (e.g. coastlines, mountain ranges) at smaller scales (Sørland et al., 2018). Compared with the coarser GCM output, downscaling (both through RCMs and statistical methods) has been found to add value in several ways. For example, RCMs can often better capture meso-scale phenomena and climate dynamics, which the results of statistical downscaling can complement because of its different set of information and assumptions (Takayabu et al., 2016).

In addition, [climate and impact] model accuracy and data quality are often better at local to regional scales, which has led to questions on whether projections of climate change impacts are better made by ensembles that are global or regional in scope (e.g. Challinor et al., 2014, 2017). Furthermore, adaptation is important at regional and local scales: if effective climate change adaptation measures, particularly for food security, are to happen in the next few decades, the uncertainties and lack of skill in simulated regional climates need to be communicated and understood by both researchers and policymakers (Ramirez-Villegas et al., 2013).

To conclude, recent studies report that RCMs in general outperform GCMs in many aspects, and that there are benefits to the use of GCM-RCM model chains in regional climate change assessments (Sørland et al., 2018, Olsson et al., 2016). A recent review also concluded that the answer to whether RCMs provide added value is "yes" – added value is meaningfully underlined when there is a clear physical context for it to appear in – and that it is more important to ask where this added value can be found rather than whether it exists or not (Rummukainen, 2016). Added value has been found to depend on many factors, including the general setup of the experiment (model, scale, region, boundary conditions, and intended application) as well as the specific climate variables being analyzed (Di Luca et al., 2016, Rummukainen, 2016). Therefore, more than ever, careful framing of research questions and development of targeted and appropriate methods are important to make the most of the full range of models (both climate- and impact-oriented) and communities of researchers, rather than deepen the existing divide (Challinor et al., 2017, 2014).

In this context, considering the results found in this chapter, and given the many perceived advantages of using RCMs, particularly EURO-CORDEX, which has been well-evaluated over Europe alongside the GCMs that drive many of its experiments (Jacob et al., 2014, Kotlarski et al., 2014, Jury et al., 2015), the results of the chapter point to RCMs as being appropriate for the study. In future chapters, reanalysis-driven RCMs themselves are evaluated compared to observations, and the error of GCM-RCMs and RCMs in climate projections are also analyzed to further examine how biases in climate models can affect future yield projections.

## 4.6 Conclusion

Although GCMs are powerful tools and continue to increase in resolution and complexity, their coarse scale and biases limit their direct application in impact studies. RCMs are seen as a method to generate higher-resolution simulations from existing GCMs. They are considered useful particularly for crop studies which rely on high-quality and realistic simulations of temperature and precipitation. However, the added value of RCMs is not always clear as they are themselves driven by GCMs.

To justify the selection of using downscaled GCM output for impact analyses, the work of this chapter examined differences between the output of historical GCM simulations and those that were downscaled by RCMs. It was found that RCMs have some advantages over GCMs in terms of better correlation and some improvements to simulations of temperature. However, these findings are limited to the design and context of the chapter, and the debate on added value remains an active point of discussion in many scientific communities. It is therefore important that the limitations of these downscaling methods, alongside the limitations of GCMs, are communicated in order to provide a better understanding of uncertainty.

# Chapter 5

# Evaluation of RCMs and the effect of bias correction on climate and yield simulations

## 5.1 Introduction

The assessment of added value in the previous chapter showed that downscaling global climate models (GCMs) with regional climate models (RCMs) shows some added value over GCMs alone. As previously discussed, there are benefits associated with using downscaled simulations, along with their increasing availability and access. However, RCMs themselves can introduce error, therefore affecting impact projections that utilize their output. Therefore, RCM evaluation in a perfect boundary setting is an important piece of information that can reveal RCM deficiencies (Kotlarski et al., 2014). In this chapter, the biases of selected reanalysis-driven RCMs are evaluated, and these are corrected with a number of bias correction (BC) methods of varying complexity. Finally, this BC RCM output is used to simulate yield in order to investigate the effects of BC on RCM output and yield simulations.

### 5.1.1 RCMs and bias correction in impact assessment

As previously discussed in the literature review (Chapter 2, Section 2.2.4), the use of downscaling methods – either statistical downscaling or through RCMs (dynamical downscaling) – is an attempt to resolve the scale discrepancy between GCM grid cell resolution and the fine-scale resolution required for local and regional impact assessment (Maraun et al., 2010), such as modeling the impacts of climate change on agriculture. There are many reasons for utilizing downscaling to bridge the scale gap between GCM and impact models. A primary reason is that it is not considered 'sensible' to use hourly and/or daily GCM outputs directly for agriculture impact assessment because of the challenges in interpreting the higher temporal resolution data over large grid boxes (Luo and Yu, 2012, p.560).

Based on the literature review, progress in regional climate modeling has made the use of RCMs more attractive in a number of climate impact analyses, such as hydrological studies (e.g. Pasten Zapata, 2017, Olsson et al., 2016, Teutschbein and Seibert, 2012) and agriculture-oriented studies (e.g. Balkovič et al., 2018, Wilcke and Bärring, 2016, Waongo et al., 2015, Glotter et al., 2014, Oettli et al., 2011. RCMs have a history within the field of agricultural impact assessment, as they have been applied to agricultural impacts assessments as early as 1998 (Luo and Yu, 2012). Therefore, while their added value is still an active point of discussion in the scientific community (See Chapter 4), it is argued that downscaling methods do, and will continue to, play a significant role in impact assessment.

Similar to the use of RCMs, the use of bias correction (BC, see Section 2.2.5), has become somewhat standardized in impact assessment methods. Previous studies have found that precipitation output from GCMs cannot be used to force hydrological or other impact models without some form of prior correction if realistic output is sought (Piani et al., 2010), so it is often applied, in varying forms of complexity, to post-process climate model output.

However, BC is also seen as a debatable step if it is applied without careful design and thought to its implications on the climate change signal and climate physics (e.g. avoiding 'naive application', Maraun et al., 2017, see Section 2.4.2.1). The choices of RCMs, their driving GCMs, and BC methods are therefore important to evaluate and not take simply as standard, especially considering research findings that showed that GCM and RCM simulations that were used as input to crop modeling showed no significant differences after BC (Glotter et al., 2014), and that biases in wheat yield simulations have been caused by biases in rainfall inherited from GCMs (Macadam et al., 2016).

The choices of RCMs and BC methods are also relevant in the context of the impact modeling methods they are used as input to. The large differences between statistical and process-based approaches to crop modeling mean that climate model output is utilized in different ways.

### 5.1.2 Comparing crop model approaches and past yield simulations

#### 5.1.2.1 Differences in crop modeling method

As previously discussed in Chapters 2 and 3, the impacts of climate change on crop yield are typically assessed through the use of crop models, which are categorized into process-based models (PCMs) or statistical crop-climate models (SCCMs). These crop modeling approaches are methodologically distinct: PCMs consider plant growth processes, genetics and field management. SCCMs, on the other hand, are built considering empirical data, and rely on a much smaller set of input data for calibration. The issues between the choice of using PCMs or SCCMs have been summarized as calibration parameter differences, scale mismatch, challenges in upscaling parameters, aggregation error, stationarity, among other differences (See Chapter 2). These numerous issues are argued to cause serious limitations in the resulting confidence in crop yield projections when they are generated using only a single crop model, or more importantly, with only a single crop modeling approach. The numerous advantages of the PCM approach – namely being able to consider field management and genetics, which have a larger effect on yields compared to climate (Porter and Semenov, 2005) – are also matched by the advantages of ease of use and transparency of statistical approaches (e.g. as reviewed by Lobell and Burke (2010)).

These fundamental differences in crop modeling method are important, particularly when recalling the cascade of uncertainty (e.g. Wilby and Dessai, 2010) in agricultural impact projections. For example, PCMs are linked to uncertainties that are introduced by crop modeling parameters (e.g. Watson et al., 2015, Koehler et al., 2013), and related to factors such as soil and genetics (e.g. Folberth et al., 2016, Langensiepen et al., 2008). A major issue is the assumed stationarity in impact responses (Lobell and Burke, 2010, Lobell and Field, 2007). These differences between SCCMs and PCMs have been investigated in previous chapters, which have focused on evaluating simulations of wheat yields in the UK and Germany, as well as four German regions (See Chapter 3).

#### 5.1.2.2 Review of previous chapter findings

In Chapter 3, SCCMs and PCMs were compared using climate observations (E-OBS, Haylock et al., 2008) as input in order to assess their skill in producing yield "hindcasts". A hindcast is the term used in the study to describe recreations of past yield. The PCM used for the comparison is CERES-Wheat (part of DSSAT, Jones et al., 2003). It was shown to be highly sensitive to input parameters such as the genetic coefficients. The SCCMs for the study are generalized additive models based on the work of

Hawkins et al. (2013a), Lobell and Burke (2010). They were evaluated for the UK, Germany, and the German regions that are the focus of the study (See Equations 3.3-3.7). These SCCMs consider a non-linear time trend and climate indices that are relevant for heat stress, which is thought to be the current and future major cause of climate-driven yield loss (e.g. Semenov et al., 2014).

A comparison of the yield hindcasts generated by the PCM and SCCMs showed that, in the study, SCCMs out-performed PCMs based on smaller root mean square error ($RMSE$), higher correlation, and generally better agreement (smaller biases) with the yield observations in wheat yield hindcasts from German regions between 1981-2010. Despite this difference in skill, it is argued that there are still many benefits to the use of systems-based understanding that decision support systems like PCMs can provide. In addition, the important outcomes resulting from the evaluation of the PCM and SCCM in Chapter 3 – particularly what can be considered a 'reasonable' performance from the PCM in particular – have implications for the work in this chapter.

Both crop modeling approaches have advantages and disadvantages, and shared limitations (e.g. considering extremes and the influence of pests). It is argued that PCMs and SCCMs remain difficult to compare to each other, but these differences that are valuable to be explored further (e.g. Challinor et al., 2014, see Chapter 3 discussion) to better characterize how decisions in the impact simulation cascade affect the resulting projections. Therefore, this chapter adopts this method comparison approach while investigating the effect of BC on yield simulations.

## 5.2 Chapter approach and objectives

Based on the overall research objectives of investigating uncertainty from the ‘intermediate’ steps needed to link climate and crop models (See Chapter 1, Fig. 1.4), this chapter takes on the objectives of evaluating the skill of RCMs, comparing how different approaches to BC affect RCM output, and whether these BC methods also have impacts on yield simulations generated by two different crop modeling methods. By doing so, the study aims to contribute to knowledge in the interdisciplinary field of crop-climate research.

### 5.2.1 Chapter research questions

**(1)** How do reanalysis-driven RCM simulations (evaluation RCM simulations or perfect boundary setting RCM) of temperature and precipitation compare to observations? What are their biases?

**(2)** How do different BC methods, chosen for their varying complexity, affect these climate simulations relative to observations? Are they effective in reducing biases?

**(3)** How does BC of RCM evaluation simulations used for crop models affect yield simulations: (a) relative to observations and the E-OBS hindcast, and (b) relative to yield simulations generated by uncorrected RCM output?

**(4)** How do yield simulations compare when they are generated with the same bias-corrected climate data but with a different crop modeling approach?

**(5)** By reducing the biases climate model simulations, do simulations from the PCM or SCCM improve?

In order to answer these questions, after the evaluation and correction of reanalysis-driven RCMs, the resulting BC RCM output are used in two different crop modeling methods in order to assess the impact of BC on yield simulations. Of particular interest is how these yield hindcasts differ because the two different crop modeling methods utilize climate inputs in distinct ways. Because the RCMs are driven by reanalysis data, it is expected that their outputs are reasonably correlated to observations and their biases are relatively small compared to observations. Where present, these biases are assumed to be caused by the inherent differences between RCMs. Despite projected small biases, it is also hypothesized that BC is still needed to overcome significant errors relative to observations, due to inherent RCM error. It is also expected that yields driven with bias-corrected RCM output will have better agreement to reference yields compared to yields driven by uncorrected RCM output. Through this analysis, it is the objective of the chapter to form a cohesive simulation approach that includes selection of a BC method for the future yield projections.

## 5.3 Data and methods

### 5.3.1 Overview of chapter experimental design

The structure of the chapter (Fig. 5.1) is framed around the objective of assessing RCM evaluation simulations for their skill in reproducing E-OBS observations of temperature and precipitation. After this evaluation, RCM output is then subjected to BC using three different methods: linear scaling, variance correction, and quantile-quantile mapping. The simulations are then re-evaluated relative to observations to test the effectiveness of the BC step. After BC, the RCM output are used as input to the SCCMs (Eqns. 3.5-3.7) and the PCM so that the regional yield simulations can be compared.

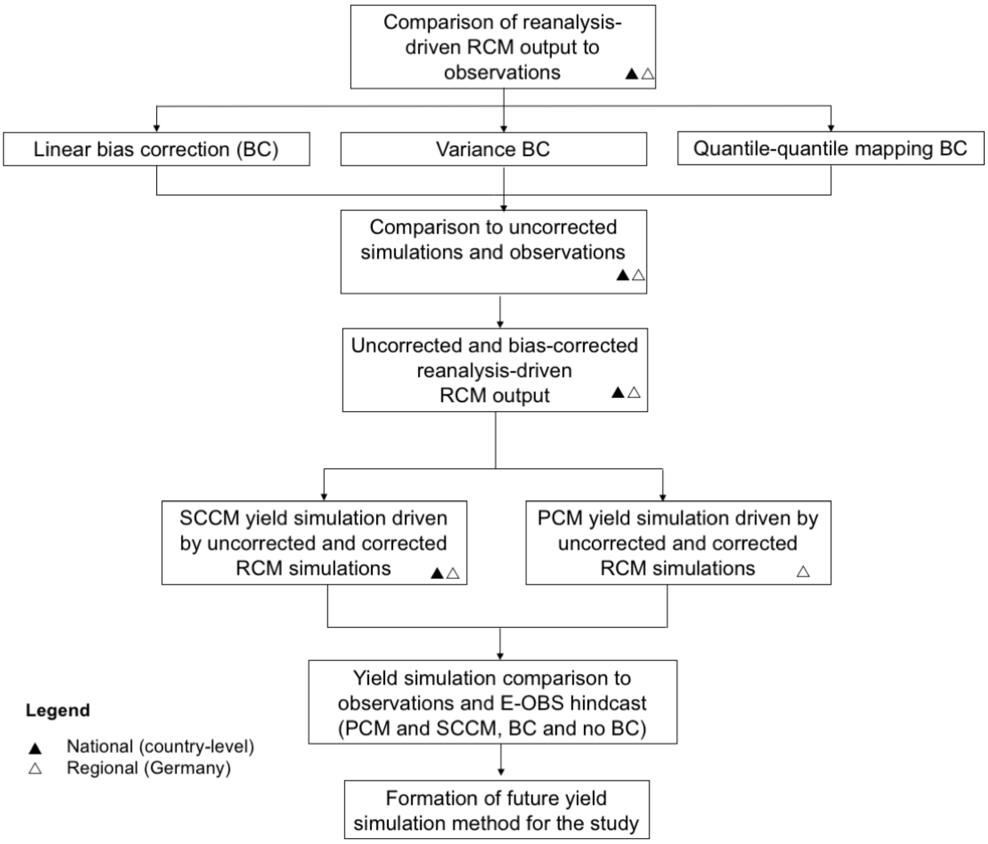



*Figure 5.1: Overview of Chapter 5 research design.*

### 5.3.2 Observations of climate and yield

Daily observations of precipitation, maximum and minimum temperature over the UK, Germany, and four German regions are taken from the gridded dataset E-OBS (Haylock et al., 2008) from 1981-2010. The German regions are identified by their NUTS code: DE2 (South Germany), DEA (West Germany), DED (East Germany) and DEF (North Germany) (Table 3.1). Yield data from the Food and Agriculture Organization (FAO (2014), for the country level) and German Statistical Office (Destatis, 2018) for the regional level are used in this chapter, also from 1981-2010.

### 5.3.3 Simulations of temperature and precipitation from regional climate models

The three RCMs in this chapter are the RCMs used in the previous chapter which evaluated added value, and are taken from the the Coordinated Downscaling Experiment over the European domain (EURO-CORDEX, Jacob et al., 2014). The three RCMs are CCLM, RACMO and RCA (Table 5.1), which are available from 1981-2010 for the evaluation simulation at 0.44° resolution. Daily maximum temperature (Tmax), minimum temperature (Tmin) and precipitation are derived from RCM evaluation simulations. The relevant climate indices for the SCCMs are also derived from the daily timestep using uncorrected and corrected RCM output.

*Table 5.1: Selected regional climate models (RCMs) from EURO-CORDEX for evaluation simulations.*

| RCM | Reference and institutes |
|---|---|
| CCLM4-8-17 (CCLM) | Jaeger et al. (2008), CLM Community |
| RACMO22E (RACMO) | van Meijgaard et al. (2008), Royal Netherlands Meteorological Institute (KNMI) |
| RCA4 (RCA) | Kjellström et al. (2016), Rossby Centre, Swedish Meteorological and Hydrological Institute |

An RCM evaluation simulation is when GCM-scale reanalysis data (taken as observations) are used as initial and boundary conditions to drive the downscaling RCM. In the case of EURO-CORDEX, reanalysis data is from ERA-Interim (Dee et al., 2011). RCM evaluation simulations from EURO-CORDEX are useful for the study, as using reanalysis data downscaled by an RCM allows for the evaluation of the RCM itself, and provides an objective measure of the skill of the RCM downscaling accuracy by approximately synchronizing the sequence of simulated and observed time series (Eden et al., 2014, Hwang et al., 2014, Menut et al., 2013). How well RCMs perform is a major research question of the thesis, and this method of RCM evaluation has been used in several impact assessment studies (e.g. Hwang et al., 2014, Menut et al., 2013, Oettli et al., 2011). Several statistical metrics, explained later in this section are calculated to assess the performance RCMs relative to observations in order to determine their biases. After this analysis, several BC methods are applied to RCM output.

### 5.3.4 Bias correction methods

The three BC approaches selected for the study can be applied directly to correct climate model output: (1) Simple scaling with linear transformation, (2) Variance scaling, which corrects variation as well as the mean of the simulations through power transformation (for precipitation) and variance scaling (for temperature), and (3) Distribution mapping through quantile-quantile mapping. These methods have been chosen to show varying complexity in the adjustment method. Stochastic weather generators (WGs) are another valuable method of BC and have been recently revisited as an alternative to 'classical' BC methods (e.g. Maraun et al., 2017, Maraun, 2016). WGs which generate random weather data with the properties of the observation or calibration dataset; however they are not examined in this research because the realism (i.e. year-to-year values) of annual and seasonal climate is important in the RCM evaluation in order to generate yield hindcasts.

The equations for the chosen BC methods from Teutschbein and Seibert (2012) are as follows, where $T$ and $P$ are daily $(d)$ temperature and precipitation. An asterisk ($^*$) signifies the corrected RCM output. Monthly mean and standard deviation are denoted by $\mu_m$ and $\sigma_m$, respectively. The coefficient of variation ($CV$) is standard deviation divided by the mean ($\sigma/\mu$). The time period for the EURO-CORDEX RCM evaluation simulations ($eval$) and E-OBS observations ($obs$) is from 1981-2010.

Although daily climate values are also analyzed in this chapter, the results and analyses are explained primarily in terms of annual and seasonal summer climate for clarity. All observations and RCM simulations are corrected on a daily basis before obtaining seasonal means for use in the statistical crop-climate model, while daily values are needed for the CERES-Wheat model.

**(1) Simple scaling: linear correction for temperature and precipitation**

In this BC method, daily precipitation from the RCM evaluation simulations $P_{eval}(d)$ is corrected with a factor based on the ratio of the long-term monthly mean $\mu_m$ of observed and the RCM-downscaled reanalysis data (Eqn. 5.1). For temperature, daily Tmax from the RCM evaluation simulation, $T_{eval}(d)$, is corrected with an additive term based on the difference of the long-term monthly mean between the observations $T_{obs}(d)$ and RCM evaluation simulation $T_{eval}(d)$ simulation (Eqn. 5.2):

$$P^*_{eval}(d) = P_{eval}(d) \cdot \left[\frac{\mu_m(P_{obs}(d))}{\mu_m(P_{eval}(d))}\right] \tag{5.1}$$

$$T^*_{eval}(d) = T_{eval}(d) + \mu_m(T_{obs}(d)) - \mu_m(T_{eval}(d)) \tag{5.2}$$

**(2a) Variance scaling: power transformation of daily precipitation**

While linear scaling accounts for a bias in the mean, it does not allow differences in the variance to be corrected. In the first step of variance scaling (Eqn. 5.3), the coefficient of variation $CV$ of daily precipitation from observations and simulations is equated, such that $P^{b_m}_{eval}(d)$ is scaled by a non-linear correction by a parameter $b_m$. The estimate of parameter $b_m$ is found using the root-finding Brent-Dekker algorithm ($R$ library $pracma$), an algorithm for real, univariate, continuous functions (Brent, 1971).

$$f(b_m) = 0 = CV_m(P_{obs}(d)) - CV_m(P^{b_m}_{eval}(d)) \tag{5.3}$$

After estimating $b_m$, the parameter is used as an exponent to adjust the variance statistics of the daily RCM evaluation simulations for precipitation $P_{eval}(d)$ (Eqn. 5.4):

$$P^{*1}_{eval}(d) = P^{b_m}_{eval}(d) \tag{5.4}$$

In a final step, the long-term monthly mean of observed precipitation is matched to the monthly mean $\mu_m$ of the intermediary precipitation time series obtained from step two $P^{*1}_{eval}(d)$ (Eqn. 5.5):

$$P^{*}_{eval}(d) = P^{*1}_{eval}(d) \cdot \left[\frac{\mu_m(P_{obs}(d))}{\mu_m(P^{*1}_{eval}(d))}\right] \tag{5.5}$$

**(2b) Variance scaling for temperature**

This is a corresponding approach to power transformation for precipitation to correct temperature. It results in a corrected time series with the same mean and standard deviation as the observed time series. In a first step, the mean of the RCM-simulated time series are adjusted by linear scaling, where $T^{*1}_{eval}(d)$ follows from Eqn. 5.2:

$$T^{*2}_{eval}(d) = T^{*1}_{eval}(d) - \mu_m(T^{*1}_{eval}(d)) \tag{5.6}$$

The standard deviations are scaled based on the ratio of observed and RCM evaluation run output on a daily timestep:

$$T^{*3}_{eval}(d) = T^{*2}_{eval}(d) \cdot \left[\frac{\sigma_m(T_{obs}(d))}{\sigma_m(T^{*2}_{eval}(d))}\right] \tag{5.7}$$

In a final step, the corrected time-series is shifted:

$$T^{*}_{eval}(d) = T^{*3}_{eval}(d) + \mu(T^{*1}_{eval}(d)) \tag{5.8}$$

**(3) Quantile-quantile mapping**

Realistic representation of precipitation is crucial for any impact and vulnerability assessment. Hence, crop modelers use BC methods that can correct the intensity histogram of simulated precipitation (Piani et al., 2010).

These methods are known as distribution mapping, also known as quantile-quantile (QQ) mapping. QQ mapping is a method for BC where a Gamma distribution ($\gamma$, with shape parameters $\alpha$ and $\beta$) is assumed to be suitable for distributions of precipitation events and that a Gaussian distribution ($N$ for normal) is assumed to fit best for temperatures (Teutschbein and Seibert, 2012, Piani et al., 2010, Eqns. 5.9, 5.10).

$$P^{*}_{eval}(d) = F^{-1}_{\gamma}(F_{\gamma}(P_{eval}(d)|\alpha_{contr,m}, \beta_{contr,m})|\alpha_{obs,m}, \beta_{obs,m}) \tag{5.9}$$

$$T^{*}_{eval}(d) = F^{-1}_{N}(F_{N}(T_{eval}(d)|\mu_{contr,m}, \sigma^{2}_{contr,m})|\mu_{obs,m}, \sigma^{2}_{obs,m}) \tag{5.10}$$

**(3a) QQ mapping: dealing with the drizzle effect and cross validation**

Because it has been shown that RCMs simulate too many days with very low precipitation (Chen et al., 2013) – the so-called drizzle effect – an initial step prior to QQ mapping is to adjust the number of dry days, by matching them with the number of observed dry days using a wet day threshold below which all simulated values are changed to zero, a method used in a recent hydrological impact study (Pasten Zapata, 2017).

After this correction, the steps described by Teutschbein and Seibert (2012) are followed: cumulative distribution functions (CDFs) are constructed for observations and RCM simulations for all days within a certain month. The value of the RCM-simulated precipitation/temperature of day $d$ within month $m$ is then searched on the observational CDF of the RCM simulations together to find its corresponding cumulative probability. The value of precipitation $P$ or temperature $T$ of the same cumulative probability is then located on the empirical CDF of observations, and this value is used as the corrected value for the RCM simulation.

An additional cross-validation step called five-fold cross validation, based on the work of Maraun and Widmann (2015), is used for QQ mapping. In this approach, conceptually similar to the leave-one-out cross validation described in Chapter 3, the calibration period is divided into five sections of the same length (in this study, using daily data from 30-year periods). The correction is calibrated using four of the sections and these are used to correct the remaining section not used for calibration. This is repeated until all sections have been corrected in this cross-validation method.

### 5.3.5 Crop modeling methods

Following their evaluation in Chapter 3, the two crop modeling methods that are used in this chapter are CERES-Wheat (originally Ritchie and Otter, 1985, now part of DSSAT, Jones et al., 2003) and the regional SCCMs (Eqns. 3.5-3.7). Because of the limited regional yield data for the UK, the crop modeling method comparison is only used for the four German regions for which data is available. A full description of the PCM and SCCMs is found in Chapter 3, Section 3.3.5. For the basic PCM experimental setup, the same regional parameters, including the genetic coefficients from Nain and Kersebaum (2007) and fine-scale soil data (IRI et al., 2015) are used, and $CO_2$ is not considered in the PCM for simplicity and better comparability with the SCCM which cannot consider $CO_2$.

### 5.3.6 Methods of statistical evaluation

#### 5.3.6.1 Evaluation of RCM output relative to observations and uncorrected RCM output

Similar to the previous chapter, RCM outputs are compared to observations of Tmax, Tmin and precipitation by using statistical metrics such as the bias, $RMSE$, and correlation ($r$). A negative (positive) mean bias indicates that simulated temperatures are cooler or have fewer hot days (warmer or more hot days) than observations. For precipitation, a negative (positive) mean bias indicates that total annual or seasonal precipitation is underestimated (overestimated). The Kolmogorov-Smirnov (KS) test statistic is also calculated to show characteristics of distribution of simulations relative to observations. The KS test statistic provides the maximum distance between the cumulative distributions of the observation and simulation dataset, and can be used to evaluate the similarity or dissimilarity of two datasets (Dobor et al., 2016). Significant $p$-values ($p$ <0.05) indicate rejection of the null hypothesis that the samples have identical distribution.

As indicated previously, the results in this chapter focus on annual and seasonal timescales for clarity in the discussion, particularly the JJA climate indices that are used for the SCCMs. The JJA season is also crucial as it is when when heat stress on crucial wheat development stages is possible (e.g. anthesis, grain-filling). However, supplementary information on simulations on a daily timestep can be found in the Appendix (Section 8.4.1), including Taylor diagrams and empirical CDFs and probability density functions (PDFs). Climate model output (uncorrected and bias-corrected) are additionally represented in Taylor diagrams (Taylor, 2001) with the $R$ package $plotrix$ (J, 2006) in the Appendix.

In a Taylor diagram, the radial distance from the origin is proportional to the standard deviation. The centered root mean square difference is proportional to the distance between the observations (denoted on the x-axis) and the simulation. The correlation is given by the azimuthal position of the point (Taylor, 2001), see sample diagram that shows both negative and positive correlation, Fig. 5.2). Empirical CDFs and PDFs are also found in the Appendix to show the distribution of daily climate model output and observed values, as well as daily bias-corrected output.

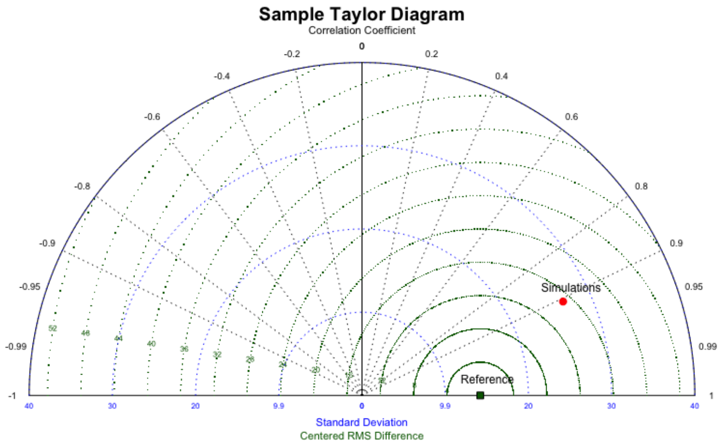



*Figure 5.2: Sample Taylor diagram for displaying pattern statistics, based on Taylor (2001). The radial distance from the origin of is proportional to the standard deviation (blue lines). The centered RMS difference between 'Observations' and 'Simulations' is proportional to their distance apart (green lines). The correlation between the two fields is given by the azimuthal position, for example, $r$=0.9-0.95 for Simulations.*

#### 5.3.6.2 Yield evaluation relative to observations and uncorrected RCM output-driven yields

The results of the yield simulations at the regional level which are generated using the PCM and SCCM are compared to yield observations and to the yield hindcast that uses E-OBS as input data. The simulations are evaluated for how close they are to these reference yields using $r$, $RMSE$, and bias. A 'good' performance is defined here as significant correlation, small $RMSE$, and minimal bias (positive/negative) compared to the reference yields. Statistical significance is marked at $p$ <0.05 and significant differences are tested with a Student's t-test. Following this statistical evaluation, the effects of BC on yield simulations using the two crop modeling methods are discussed in the context of developing a unified approach for projecting yield into the future.

## 5.4 Results

The results section is organized as follows: Firstly, the results of the RCM evaluation are reported. This is followed by the effect of BC on RCM output, and finally the effect of BC on yield simulations generated by the SCCM and PCM. Because there is no country-level PCM yield hindcast, the yield hindcasts from SCCMs and PCMs are reported and compared for each German region, relative to the reference yields (observations and E-OBS hindcast with the respective crop model approach) before and after BC. The ranges of yield simulations with each correction are also reported. Supplementary information on individual RCM-driven yield performances is shown in the Appendix (Tables A11-A14).

### 5.4.1 Evaluating reanalysis-driven RCMs

The results from the analysis of the error contributed by RCMs are performed by comparing the output of reanalysis-driven RCMs to observations of annual and seasonal Tmax, Tmin and precipitation (1981-2010). The results show that uncorrected RCM evaluation simulations can capture the year-to-year values and range of observations relatively well: for example, there are significant and high correlations between raw RCM output and observations: $R^2$=0.8-0.9 for Tmax, Tmin for the UK, Germany and German regions (Tables 5.2-5.7), and for precipitation this ranges between $R^2$=0.5-0.8 (Tables 5.4-5.8).

While well-correlated, RCM evaluation simulations still contain biases. For example, it can be observed that CCLM, RACMO, and RCA generally underestimate annual Tmax, overestimate Tmin and have biases for annual and seasonal precipitation in the UK and Germany (Figs. 5.3 and 5.4), as well as the four different German regions (Figs. 5.5-5.8) – similar to the general findings of RCM bias reported by Kotlarski et al. (2014).

The negative biases of uncorrected RCM evaluation simulations range between -1.13 to -1.48°C for UK Tmax, 0.37-0.74 °C for UK Tmin, and -113 to 132 mm for UK precipitation, -0.9 to -1.49°C for Germany Tmax, -0.86 to 0.27 for Germany Tmin, and 38-113 mm for Germany precipitation (Tables 5.2-5.4). German regions also show similar ranges of bias to Germany for Tmax, Tmin precipitation (Tables 5.6-5.8), with RACMO usually showing the largest negative biases for Tmin in all four German regions.

The summer climate indices used for the SCCM are derived from daily Tmax and precipitation in JJA. The results show that for the UK, RCM evaluation simulations from CCLM, RACMO and RCA simulate zero days above 31°C, while over Germany (including all four German regions) RCA tends to overestimate the number of hot days while RACMO and CCLM have negative biases for the number of hot days (Figs. 5.3D- 5.8D).

JJA precipitation shows relatively small biases for the UK and Germany (-40 to 40mm, UK and -16 to 0.7mm, Germany). In German regions, the biases from RCMs are also relatively small, with the largest negative bias from raw RCA simulations (-87mm for DE2). Additional statistics on the KS test values (Tables 5.5 and 5.9) and the daily empirical cumulative distribution function and probability density functions, (available in the Appendix, 8.4.1, Figs. A10-A15 and Tables A6 and A10) show that the uncorrected RCM simulations have significantly different distributions relative to observations. Additional information from the analysis of the seasonal climate indices is also available in Appendix (Section 8.4.1B.).

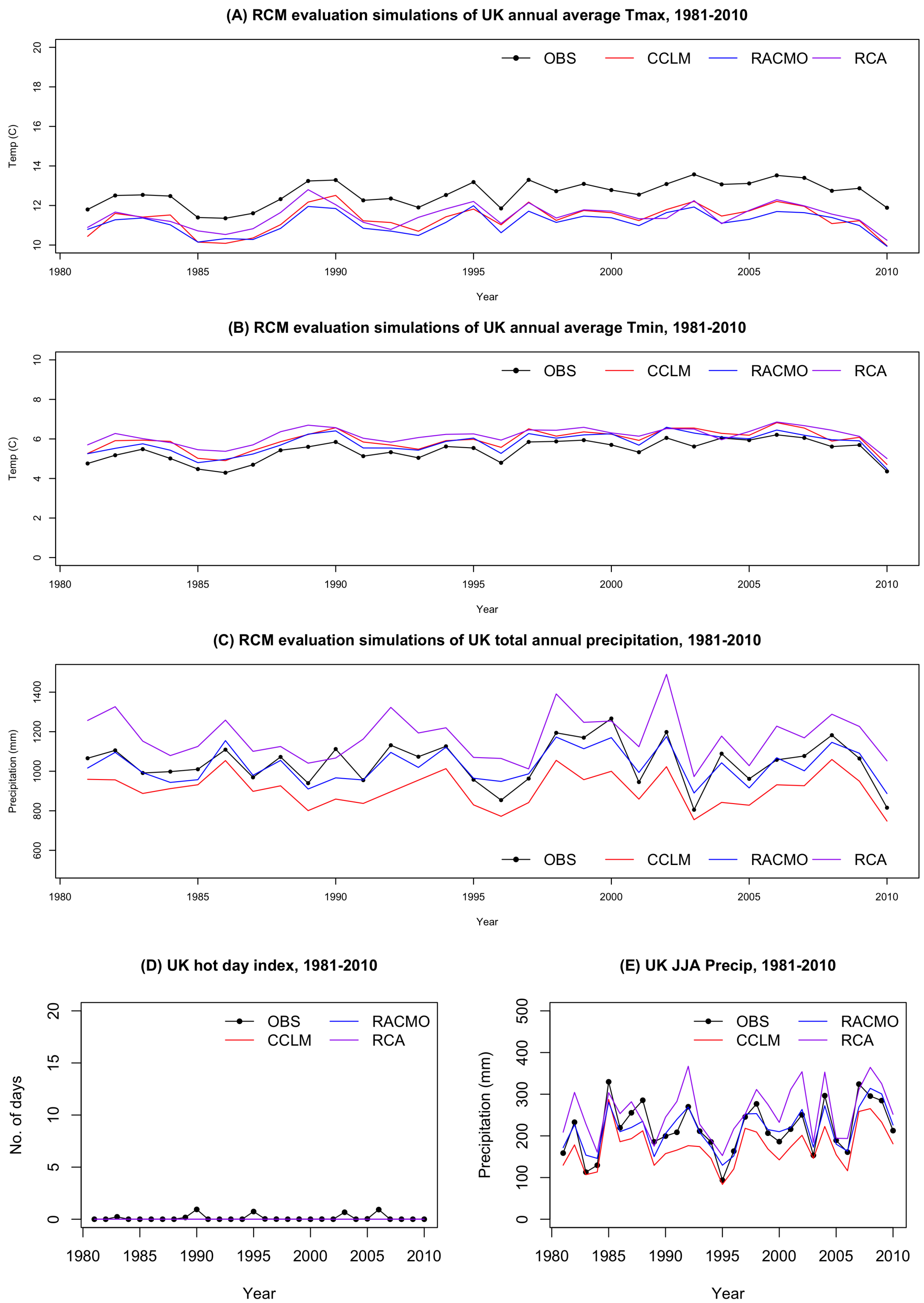



*Figure 5.3: Uncorrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31° C), and (E) total JJA precipitation for the UK, 1981-2010.*

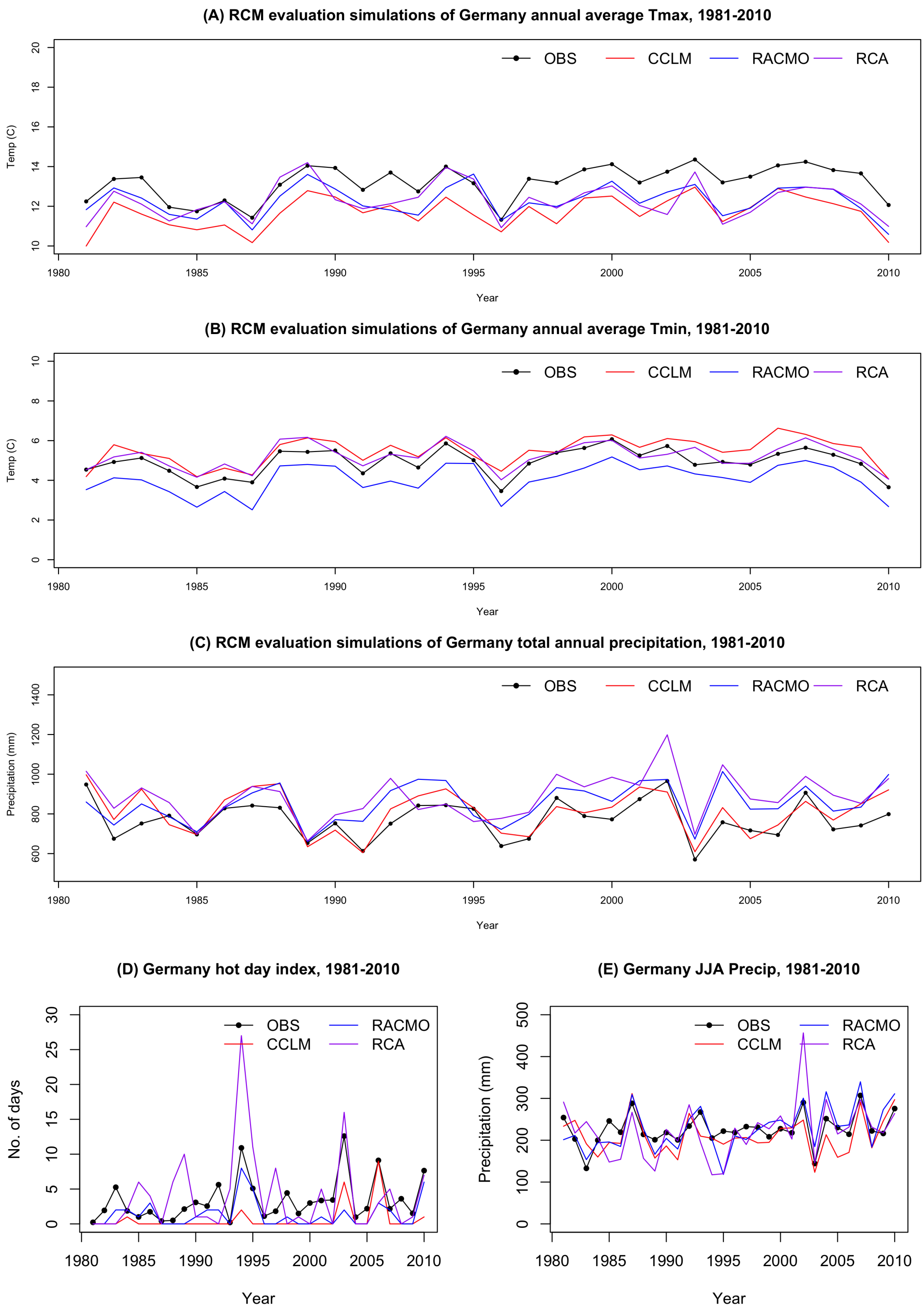



*Figure 5.4: Uncorrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31° C), and (E) total JJA precipitation for Germany, 1981-2010.*

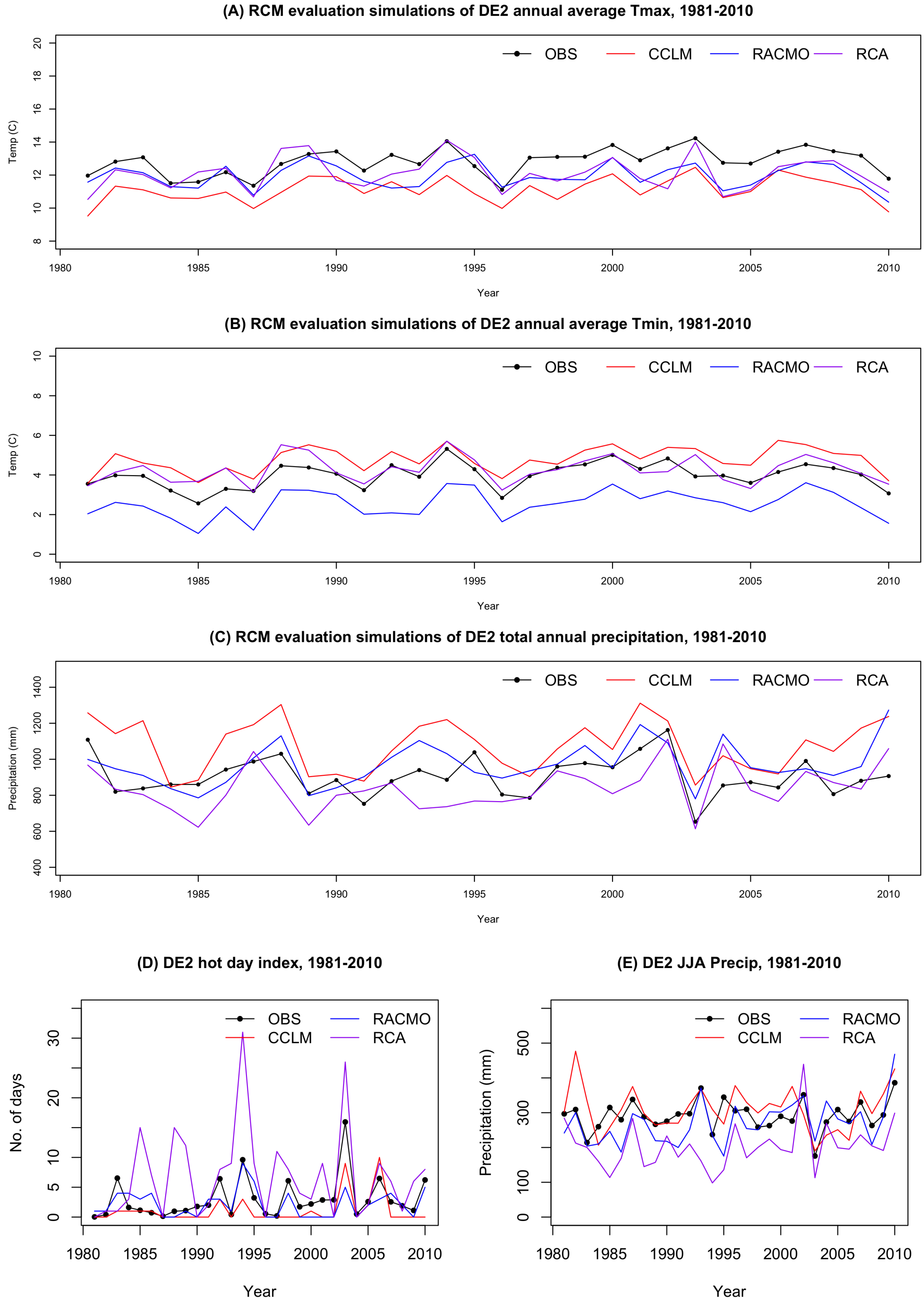



*Figure 5.5: Uncorrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31° C), and (E) total JJA precipitation for DE2 (South Germany), 1981-2010.*

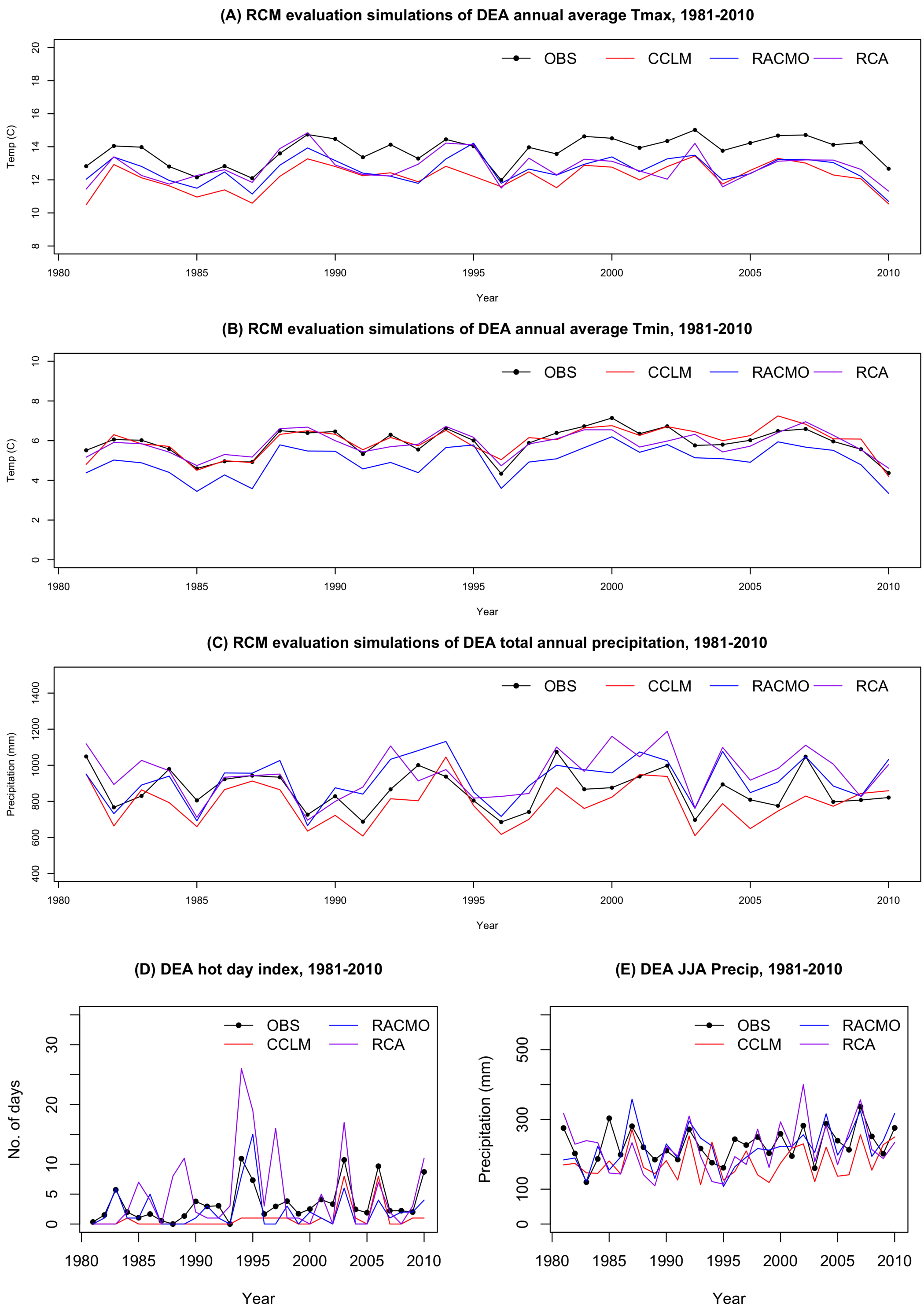



*Figure 5.6: Uncorrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31°C), and (E) total JJA precipitation for DEA (West Germany), 1981-2010.*

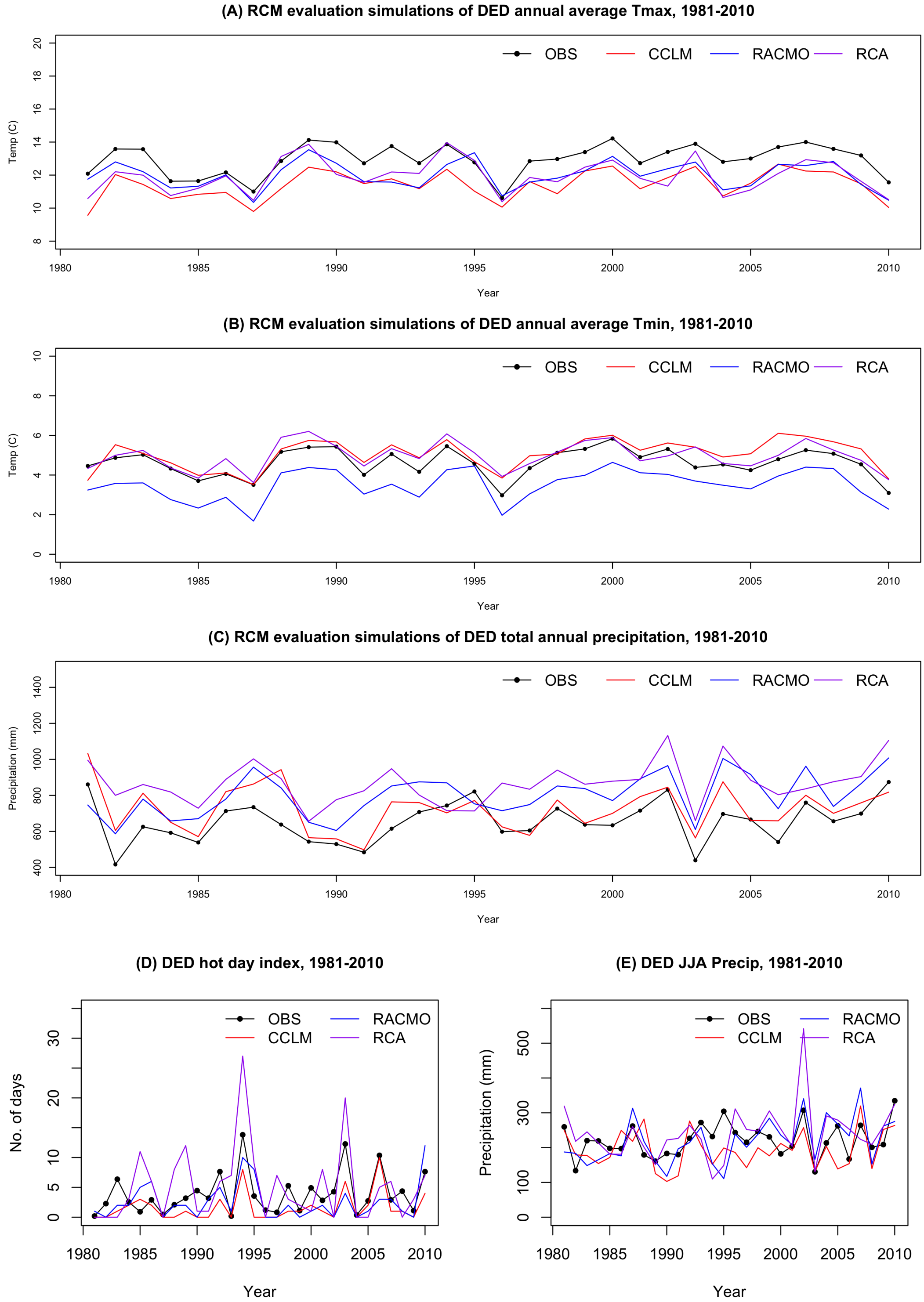



*Figure 5.7: Uncorrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31°C), and (E) total JJA precipitation for DED (East Germany), 1981-2010.*

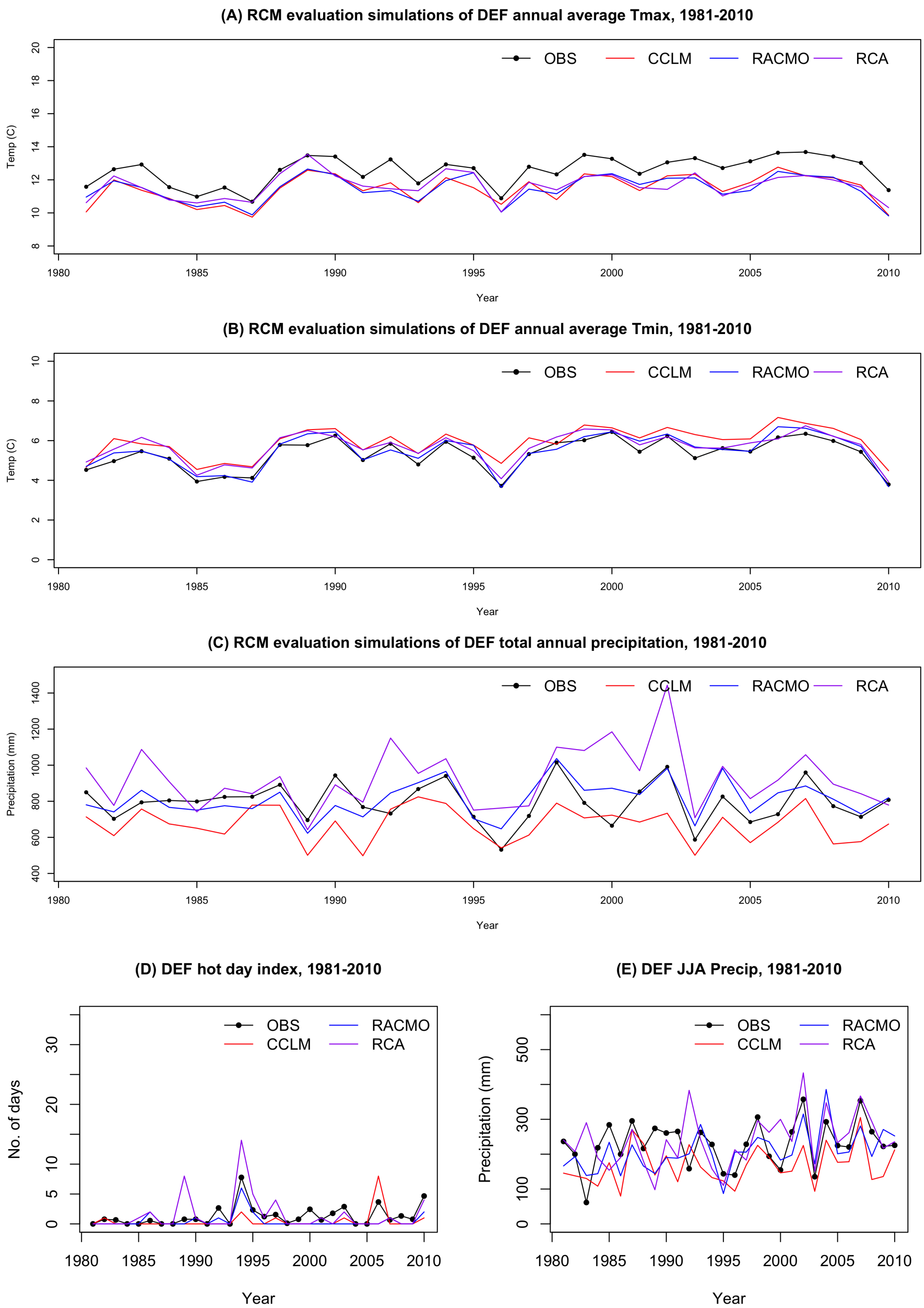



*Figure 5.8: Uncorrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31°C), and (E) total JJA precipitation for DEF (North Germany), 1981-2010.*

### 5.4.2 Results of the BC of RCM output

The results from the evaluation of RCMs shows that while biases can be considered relatively small, and that correlations are significant and high – particularly when recalling the results from uncorrected historical simulations of temperature and precipitation (Chapter 4) – RCMs still have error relative to observations. Therefore, they are bias corrected with different methods of varying complexity; the results of these corrections are reported in this section.

#### 5.4.2.1 Effects of BC on national-level climate simulations

The different BC methods of simple scaling, variance and power transformation, and quantile-quantile mapping are applied to the RCM evaluation simulations. The results of the correction are shown in Figs. 5.9 and 5.10 for the UK and Germany. It can be observed that the BC methods are effective in reducing the range of error of the uncorrected simulations relative to observations. Although some biases remain, for example, in the seasonal climate indices (See Section 8.4.1B.), all BC methods are effective to bring the simulations closer to observations.

For example, biases in annual Tmax and Tmin in the UK and Germany are reduced to zero from previous negative biases (Tables 5.2-5.3). BC methods are also shown to be effective for precipitation: after BC, QQ-mapping corrected simulations have biases of less than 15mm (11mm) compared to biases ranging from -130 to 130mm (38 to 113 mm) for the UK (Germany) and total annual precipitation (Table 5.4).

While the statistical analysis shows that correlation is not typically improved by BC, uncorrected simulations (on the annual timescale from 1981-2010) are already highly correlated to observations with $r$>0.9 for the UK and $r$>0.7 for Germany for all RCMs for annual average Tmax (Table

5.2), Tmin (Table 5.3), and total annual precipitation (Table 5.4). There are some observed increases to correlation using variance transformation and QQ mapping for Tmax, and the linear and variance methods for precipitation at the country level.

$RMSE$ and bias are greatly reduced across all climate variables in the UK and Germany after BC. Because of the nature of the computation of the power transformation correction method, annual precipitation totals are in perfect agreement with observations. However, the seasonal and daily analysis (in the Appendix 8.4.1) shows that after correction, distributions of daily precipitation are no longer significantly different to observations, based on empirical CDFs and PDFs, along with the KS test statistic for the UK and Germany (Table 5.5, Appendix Figs. A10 and A11). The Taylor diagrams also show that corrected RCM evaluation simulations have properties closer to that of observations – with high correlation and similar standard deviation; however, a small spread remains with precipitation simulations (Appendix Fig. A16).

Overall, BC methods show the capability to improve correlation, reduce the $RMSE$ and biases of the regional climate simulations of daily, seasonal and annual temperatures and precipitation, and the resulting climate indices.

*Table 5.2: Statistical comparison of annual averages of maximum temperature from RCM evaluation simulations and observations, UK and Germany, 1981-2010.*

| Model and BC method | UK | | | Germany | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.93, * | 1.29 | -1.27 | 0.91, * | 1.54 | -1.49 |
| **Raw RACMO** | 0.92, * | 1.5 | -1.48 | 0.78, * | 1.07 | -0.92 |
| **Raw RCA** | 0.81, * | 1.19 | -1.13 | 0.7, * | 1.12 | -0.9 |
| **Linear BC CCLM** | 0.93, * | 0.25 | 0 | 0.91, * | 0.36 | 0 |
| **Linear BC RACMO** | 0.92, * | 0.25 | -0.01 | 0.78, * | 0.54 | -0.02 |
| **Linear BC RCA** | 0.81, * | 0.37 | -0.01 | 0.7, * | 0.68 | -0.02 |
| **Variance BC CCLM** | 0.93, * | 0.23 | 0 | 0.92, * | 0.34 | 0 |
| **Variance BC RACMO** | 0.92, * | 0.25 | -0.01 | 0.8, * | 0.52 | -0.02 |
| **Variance BC RCA** | 0.82, * | 0.36 | -0.01 | 0.73, * | 0.63 | -0.02 |
| **QQ BC CCLM** | 0.94, * | 0.23 | -0.08 | 0.92, * | 0.35 | -0.09 |
| **QQ BC RACMO** | 0.93, * | 0.24 | -0.08 | 0.81, * | 0.51 | -0.09 |
| **QQ BC RCA** | 0.83, * | 0.36 | -0.08 | 0.74, * | 0.64 | -0.09 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $R$, smaller $RMSE$ or bias) relative to the uncorrected RCM simulation. $RMSE$ and bias are in Celsius.*

*Table 5.3: Statistical comparison of annual averages of minimum temperature from RCM evaluation simulations and observations, UK and Germany, 1981-2010.*

| Model and BC method | UK | | | Germany | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| Raw CCLM | 0.94, * | 0.56 | 0.53 | 0.89, * | 0.62 | 0.53 |
| Raw RACMO | 0.94, * | 0.4 | 0.37 | 0.94, * | 0.9 | -0.86 |
| Raw RCA | 0.87, * | 0.79 | 0.74 | 0.9, * | 0.4 | 0.27 |
| Linear BC CCLM | 0.94, * | 0.18 | 0 | 0.89, * | 0.32 | 0.01 |
| Linear BC RACMO | 0.94, * | 0.17 | -0.01 | 0.94, * | 0.25 | -0.01 |
| Linear BC RCA | 0.87, * | 0.26 | 0 | 0.9, * | 0.29 | 0 |
| Variance BC CCLM | 0.94, * | 0.19 | 0 | 0.91, * | 0.3 | 0.01 |
| Variance BC RACMO | 0.94, * | 0.18 | 0 | 0.94, * | 0.25 | -0.01 |
| Variance BC RCA | 0.87, * | 0.26 | 0 | 0.91, * | 0.27 | 0 |
| QQ BC CCLM | 0.94, * | 0.2 | -0.06 | 0.89, * | 0.35 | 0 |
| QQ BC RACMO | 0.94, * | 0.21 | -0.07 | 0.93, * | 0.29 | -0.01 |
| QQ BC RCA | 0.89, * | 0.25 | -0.06 | 0.92, * | 0.26 | 0 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $R$, smaller $RMSE$ or bias) relative to the uncorrected RCM simulation. $RMSE$ and bias are in Celsius.*

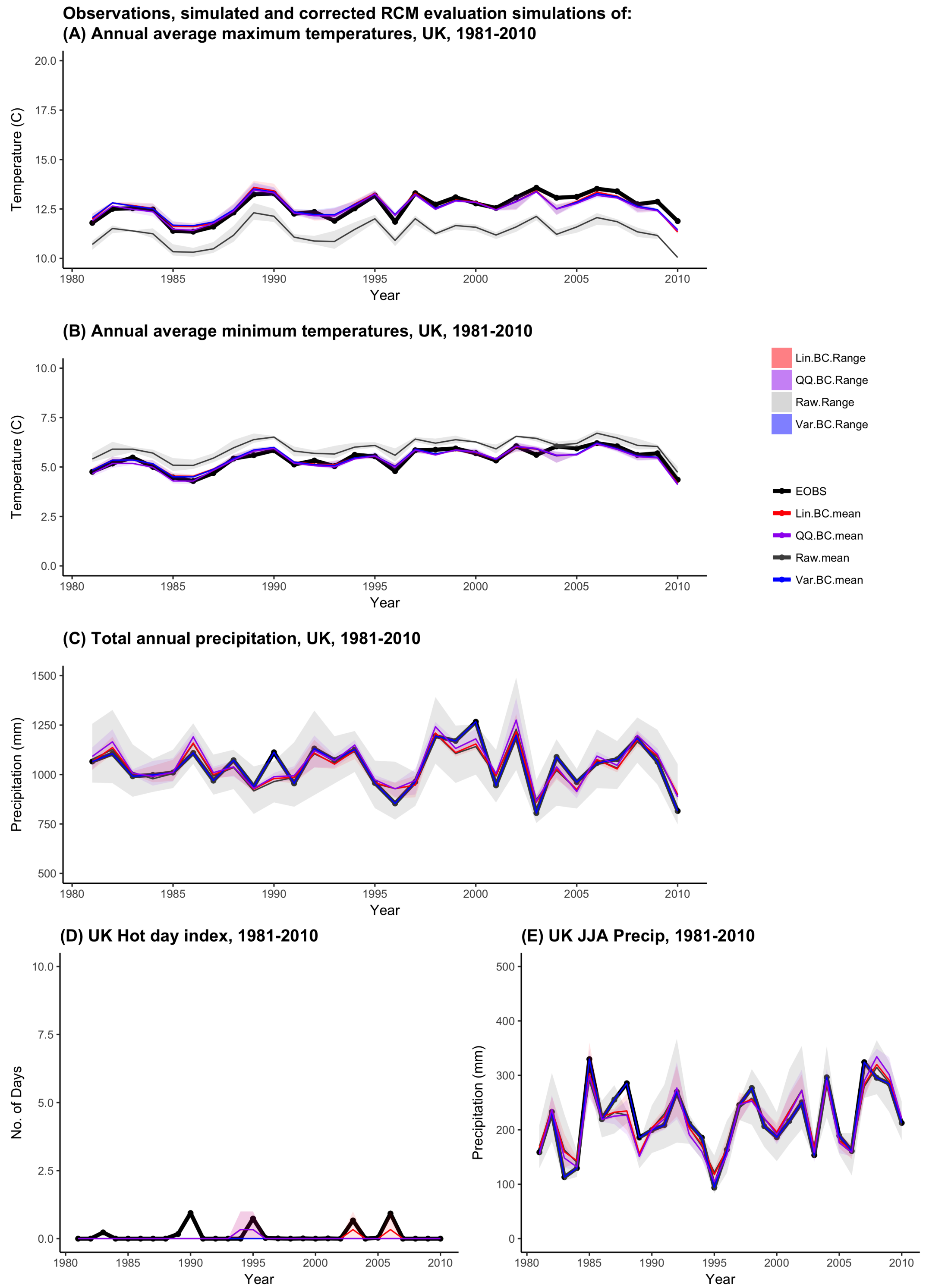



*Figure 5.9: Bias-corrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31°C), and (E) total JJA precipitation for the UK, 1981-2010. Three different BC methods are used and their ranges are shown relative to E-OBS climate observations.*

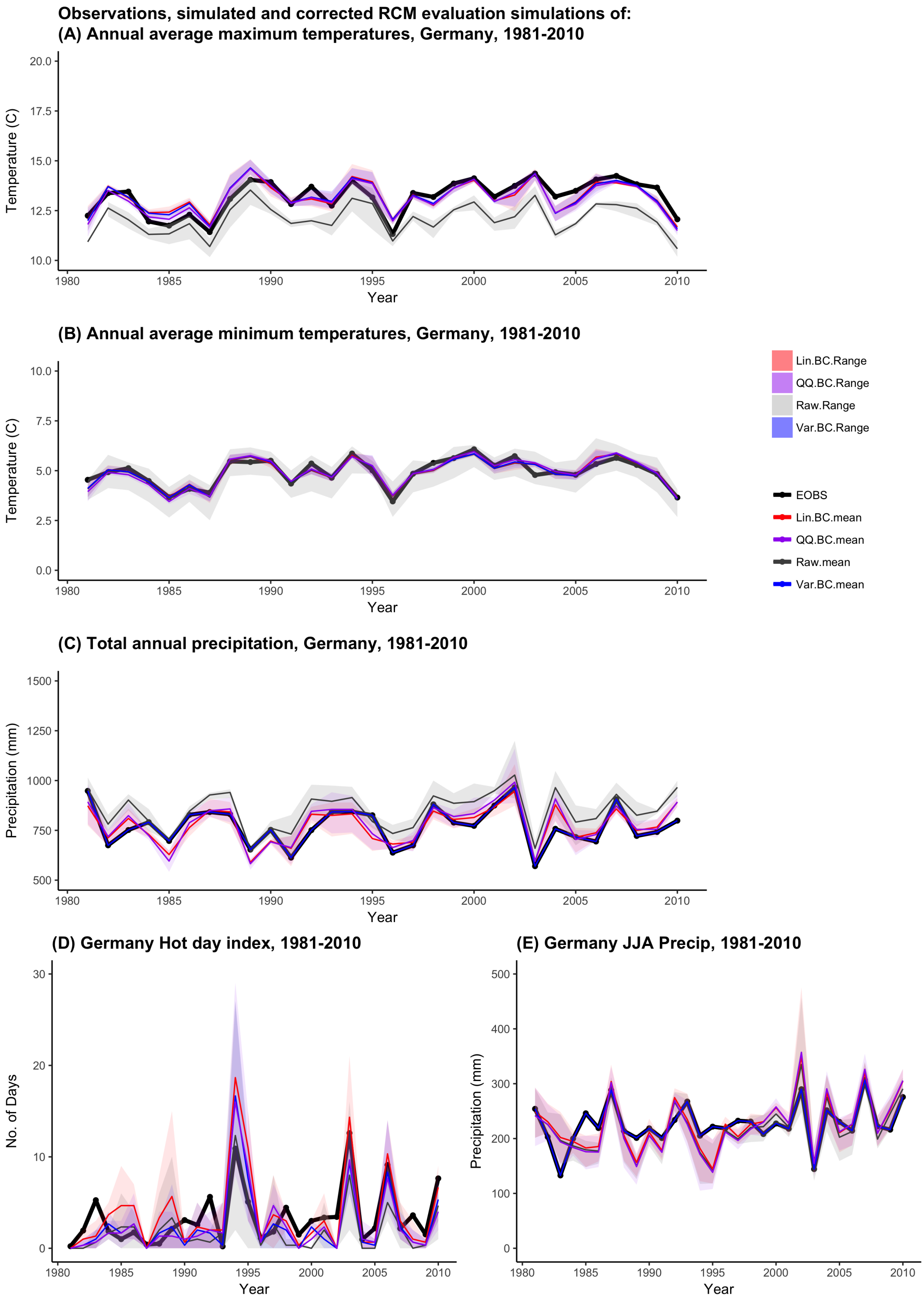



*Figure 5.10: Bias-corrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31° C), and (E) total JJA precipitation for Germany, 1981-2010. Three different BC methods are used and their ranges are shown relative to E-OBS climate observations.*

*Table 5.4: Statistical comparison of total annual precipitation from RCM evaluation simulations and observations, UK and Germany, 1981-2010.*

| Model and BC method | UK | | | Germany | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.86, * | 145.02 | -133.19 | 0.84, * | 69.57 | 38.41 |
| **Raw RACMO** | 0.89, * | 53.82 | -14.28 | 0.73, * | 107.61 | 81.16 |
| **Raw RCA** | 0.8, * | 150.79 | 132.24 | 0.69, * | 140.77 | 113.53 |
| **Linear BC CCLM** | 0.87, * | 54.39 | -4.75 | 0.85, * | 55.7 | 0.36 |
| **Linear BC RACMO** | 0.9, * | 50.03 | -0.96 | 0.72, * | 68.97 | 1.92 |
| **Linear BC RCA** | 0.82, * | 65.45 | -2.04 | 0.67, * | 81.87 | 2.4 |
| **Variance BC CCLM** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **Variance BC RACMO** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **Variance BC RCA** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **QQ BC CCLM** | 0.87, * | 55.51 | 9.42 | 0.85, * | 57.9 | 7.9 |
| **QQ BC RACMO** | 0.89, * | 54.8 | 15.46 | 0.73, * | 74.69 | 12.5 |
| **QQ BC RCA** | 0.81, * | 73.47 | 9.89 | 0.68, * | 93.86 | 10.87 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $R$, smaller $RMSE$ or bias) relative to the uncorrected RCM simulation. $RMSE$ and bias are in mm.*

*Table 5.5: Kolmogorov-Smirnov (KS) test statistics on the distribution of annual averages of maximum and minimum temperature, and total annual precipitation from RCM evaluation simulations and observations, UK and Germany, 1981-2010.*

| Model and BC method | UK | | | Germany | | |
|---|---|---|---|---|---|---|
| | Tmax | Tmin | Precip | Tmax | Tmin | Precip |
| **Raw CCLM** | 0.73, * | 0.47, * | 0.57, * | 0.7, * | 0.4, * | 0.23 |
| **Raw RACMO** | 0.77, * | 0.33 | 0.2 | 0.57, * | 0.53, * | 0.33 |
| **Raw RCA** | 0.7, * | 0.63, * | 0.43, * | 0.53, * | 0.2 | 0.47, * |
| **Linear BC CCLM** | 0.13 | 0.13 | 0.13 | 0.13 | 0.07 | 0.13 |
| **Linear BC RACMO** | 0.1 | 0.1 | 0.17 | 0.2 | 0.17 | 0.13 |
| **Linear BC RCA** | 0.13 | 0.2 | 0.17 | 0.2 | 0.13 | 0.1 |
| **Variance BC CCLM** | 0.17 | 0.13 | 0.03 | 0.13 | 0.1 | 0.03 |
| **Variance BC RACMO** | 0.1 | 0.1 | 0.03 | 0.2 | 0.17 | 0.03 |
| **Variance BC RCA** | 0.13 | 0.17 | 0.03 | 0.2 | 0.13 | 0.03 |
| **QQ BC CCLM** | 0.2 | 0.13 | 0.13 | 0.2 | 0.1 | 0.2 |
| **QQ BC RACMO** | 0.13 | 0.13 | 0.13 | 0.23 | 0.23 | 0.17 |
| **QQ BC RCA** | 0.17 | 0.2 | 0.13 | 0.2 | 0.1 | 0.17 |

*($*$) indicates a difference to the distribution of observations with statistical significance ($p$ <0.05).*

#### 5.4.2.2 Effects of BC on regional climate simulations

After BC, errors are also significantly reduced regardless of the BC method used at the regional level (Figs. 5.11-5.14). Several improvements to RCM simulations occur after BC: although correlation is not always improved, biases and $RMSE$ are greatly reduced (Tables 5.6-5.7). In all four regions, variance and QQ mapping methods are additionally able to improve correlation of Tmax. For Tmin, this improvement is less consistent, but variance and QQ mapping methods are able to improve correlation of Tmin for CCLM and RCA in all four regions. Biases are also shown to be reduced, although not completely eliminated, in precipitation simulations for all regions (Tables 5.8). $RMSE$ is not reduced using QQ mapping for RACMO and RCA simulations of precipitation in DE2 (RCA), DEA (RACMO), and DEF (RACMO). The annual KS test statistic (Table 5.9) shows that after BC, distributions of annual Tmax, Tmin and total annual precipitation are not significantly different to observations.

However, on a daily level, some RCM simulations still have significantly different distributions, particularly for precipitation. QQ mapping is shown to be most effective in correcting distributions for Tmax and Tmin (Appendix Table A10). The QQ mapping method also removes the drizzle days that RCMs introduce to simulations of precipitation (Appendix Figs. A12-A15). Taylor diagrams show that raw temperature simulations have small biases relative to observations, with high and significant correlation across the four regions ($r$>0.6). BC is able to further improve these simulations and adjust their variability, but regional precipitation still shows some errors, apart from those corrected with power transformation (by nature of the computation) (Appendix Fig. A17).

Similar to the national level, BC proves to be a useful step in reducing the introduced error from the RCMs. In the next section, these results are used in the SCCM and PCM to determine how BC affects yield simulations.

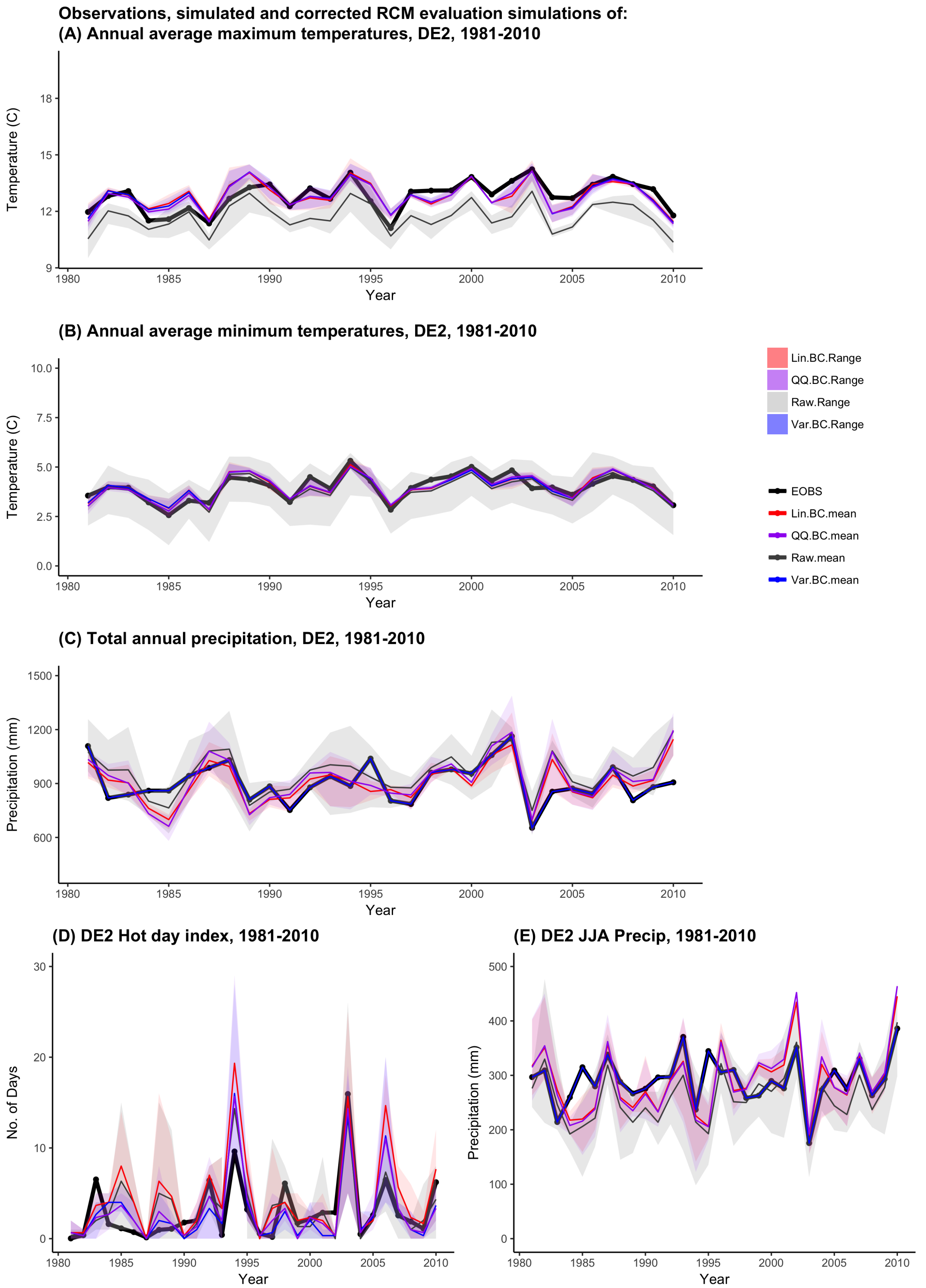



*Figure 5.11: Bias-corrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31°C), and (E) total JJA precipitation for DE2 (South Germany), 1981-2010. Three different BC methods are used and their ranges are shown relative to E-OBS climate observations.*

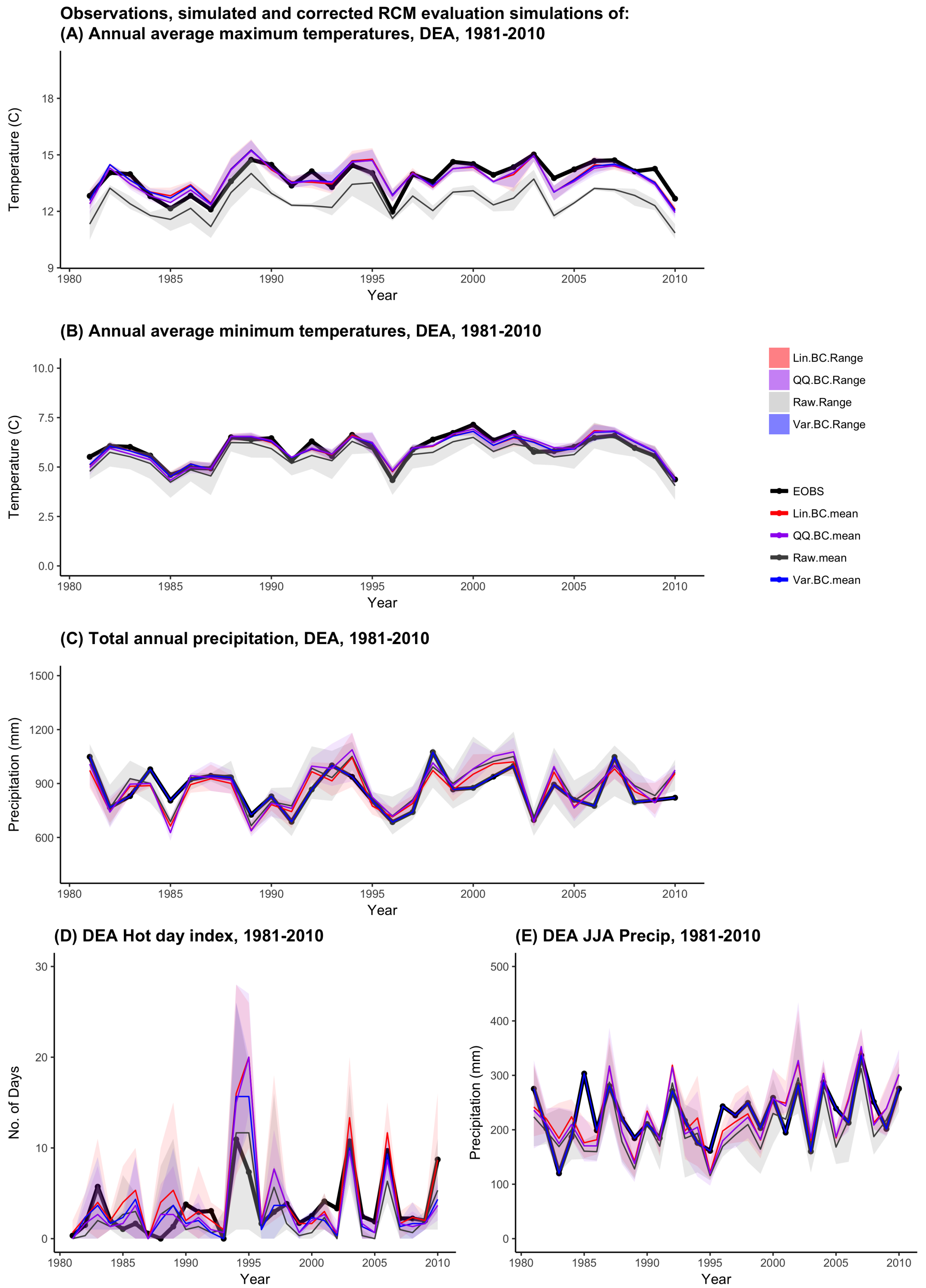



*Figure 5.12: Bias-corrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31°C), and (E) total JJA precipitation for DEA (West Germany), 1981-2010. Three different BC methods are used and their ranges are shown relative to E-OBS climate observations.*

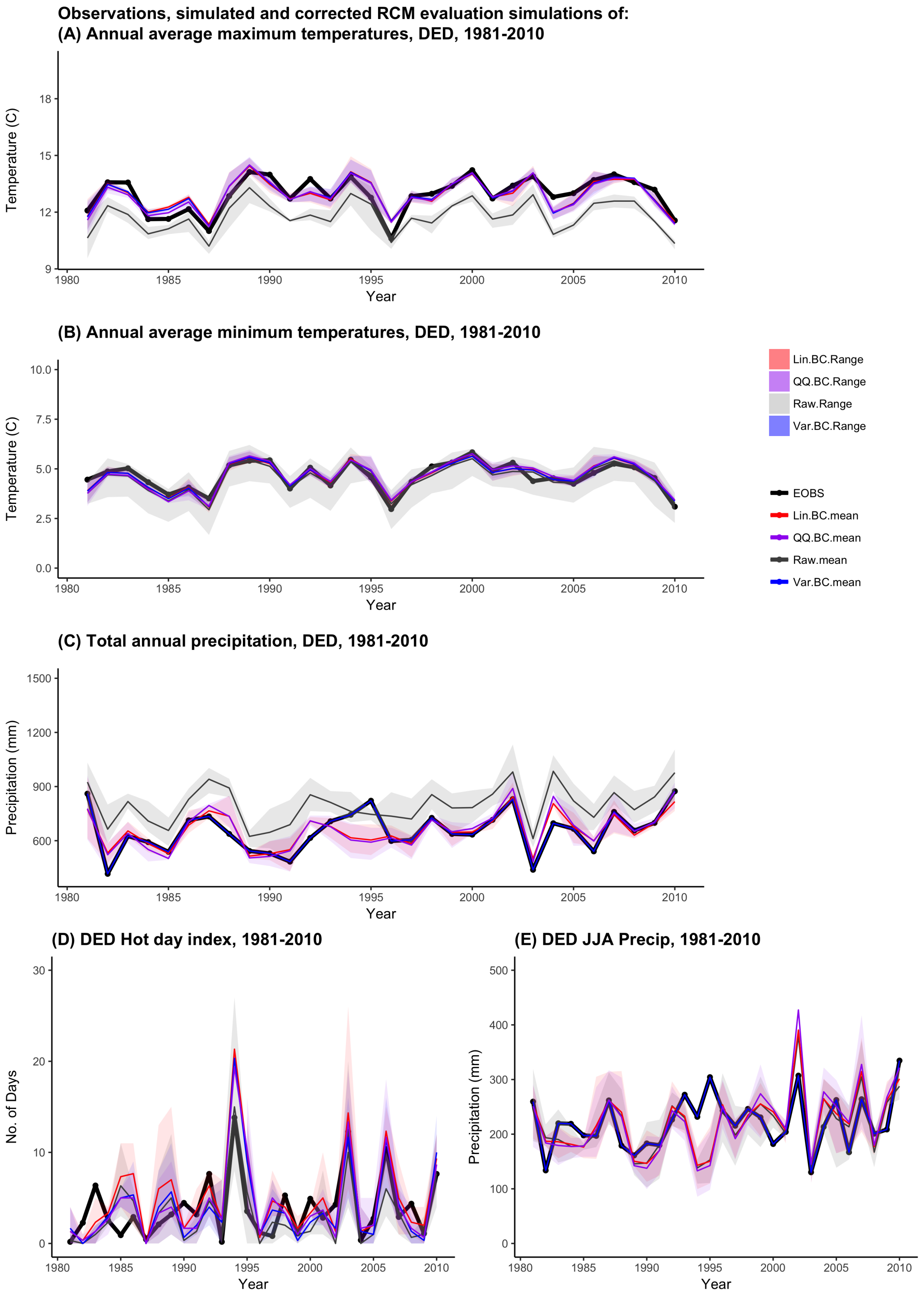



*Figure 5.13: Bias-corrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31°C), and (E) total JJA precipitation for DED (East Germany), 1981-2010. Three different BC methods are used and their ranges are shown relative to E-OBS climate observations.*

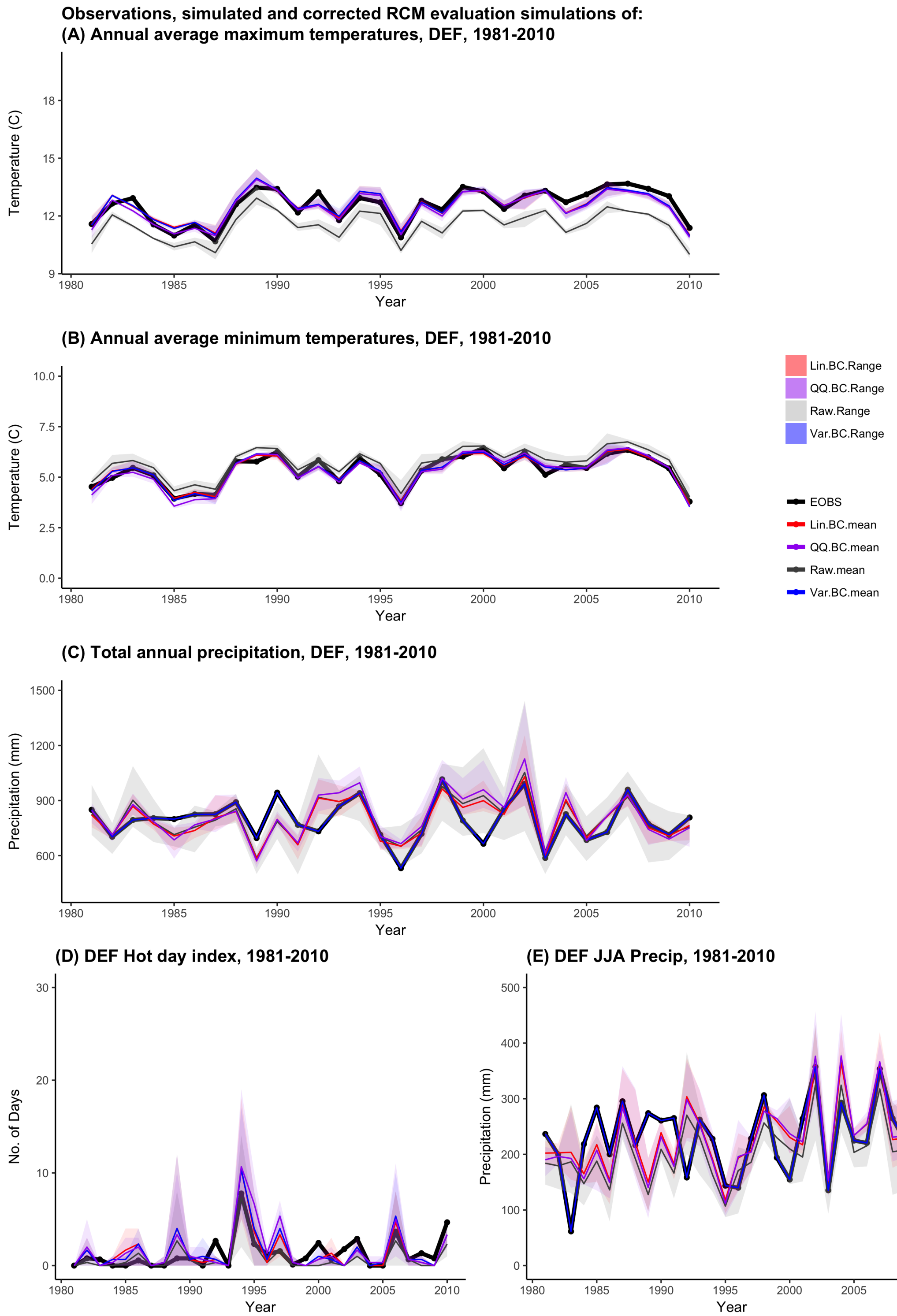



*Figure 5.14: Bias-corrected RCM evaluation simulations of annual (A) average maximum temperature, (B) average minimum temperature, (C) total precipitation; seasonal (D) hot days (above 31°C), and (E) total JJA precipitation for DEF (North Germany), 1981-2010. Three different BC methods are used and their ranges are shown relative to E-OBS climate observations.*

*Table 5.6: Statistical comparison of annual averages of maximum temperature from RCM evaluation simulations and observations, German regions, 1981-2010.*

| Model and BC method | DE2 (South Germany) | | | DEA (West Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.87, * | 1.75 | -1.7 | 0.89, * | 1.65 | -1.61 |
| **Raw RACMO** | 0.66, * | 1.08 | -0.87 | 0.79, * | 1.28 | -1.16 |
| **Raw RCA** | 0.63, * | 1.07 | -0.72 | 0.66, * | 1.24 | -1.01 |
| **Linear BC CCLM** | 0.87, * | 0.4 | 0 | 0.89, * | 0.38 | 0.01 |
| **Linear BC RACMO** | 0.66, * | 0.64 | -0.02 | 0.79, * | 0.53 | -0.02 |
| **Linear BC RCA** | 0.63, * | 0.79 | -0.01 | 0.66, * | 0.71 | -0.02 |
| **Variance BC CCLM** | 0.88, * | 0.38 | 0 | 0.9, * | 0.37 | 0.01 |
| **Variance BC RACMO** | 0.69, * | 0.61 | -0.02 | 0.81, * | 0.51 | -0.02 |
| **Variance BC RCA** | 0.67, * | 0.71 | -0.01 | 0.7, * | 0.66 | -0.02 |
| **QQ BC CCLM** | 0.89, * | 0.38 | -0.05 | 0.91, * | 0.38 | -0.08 |
| **QQ BC RACMO** | 0.7, * | 0.61 | -0.06 | 0.81, * | 0.51 | -0.08 |
| **QQ BC RCA** | 0.68, * | 0.71 | -0.06 | 0.7, * | 0.68 | -0.08 |

*Table 5.6 continued.*

| Model and BC method | DED (East Germany) | | | DEF (North Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.9, * | 1.58 | -1.53 | 0.94, * | 1.13 | -1.09 |
| **Raw RACMO** | 0.79, * | 1.12 | -0.96 | 0.9, * | 1.15 | -1.09 |
| **Raw RCA** | 0.76, * | 1.2 | -1 | 0.81, * | 1.05 | -0.92 |
| **Linear BC CCLM** | 0.9, * | 0.41 | 0 | 0.94, * | 0.3 | 0 |
| **Linear BC RACMO** | 0.79, * | 0.58 | -0.02 | 0.9, * | 0.38 | -0.02 |
| **Linear BC RCA** | 0.76, * | 0.66 | -0.01 | 0.81, * | 0.51 | -0.02 |
| **Variance BC CCLM** | 0.91, * | 0.38 | 0 | 0.94, * | 0.29 | 0 |
| **Variance BC RACMO** | 0.81, * | 0.56 | -0.02 | 0.9, * | 0.38 | -0.02 |
| **Variance BC RCA** | 0.79, * | 0.63 | -0.01 | 0.82, * | 0.5 | -0.02 |
| **QQ BC CCLM** | 0.91, * | 0.4 | -0.05 | 0.95, * | 0.32 | -0.12 |
| **QQ BC RACMO** | 0.82, * | 0.55 | -0.06 | 0.91, * | 0.39 | -0.12 |
| **QQ BC RCA** | 0.79, * | 0.64 | -0.05 | 0.82, * | 0.52 | -0.12 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $R$, smaller $RMSE$ or bias) relative to the uncorrected RCM simulation. $RMSE$ and bias are in Celsius.*

*Table 5.7: Statistical comparison of annual averages of minimum temperature from RCM evaluation simulations and observations, German regions, 1981-2010.*

| Model and BC method | DE2 (South Germany) | | | DEA (West Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.85, * | 0.86 | 0.79 | 0.9, * | 0.33 | 0.06 |
| **Raw RACMO** | 0.87, * | 1.48 | -1.44 | 0.94, * | 0.96 | -0.93 |
| **Raw RCA** | 0.78, * | 0.51 | 0.28 | 0.87, * | 0.35 | -0.05 |
| **Linear BC CCLM** | 0.85, * | 0.34 | 0.01 | 0.9, * | 0.33 | 0.01 |
| **Linear BC RACMO** | 0.87, * | 0.34 | 0 | 0.94, * | 0.26 | -0.01 |
| **Linear BC RCA** | 0.78, * | 0.43 | 0.01 | 0.87, * | 0.35 | -0.01 |
| **Variance BC CCLM** | 0.87, * | 0.33 | 0.01 | 0.91, * | 0.32 | 0.01 |
| **Variance BC RACMO** | 0.86, * | 0.33 | 0 | 0.93, * | 0.26 | -0.01 |
| **Variance BC RCA** | 0.8, * | 0.41 | 0 | 0.89, * | 0.32 | -0.01 |
| **QQ BC CCLM** | 0.87, * | 0.35 | 0 | 0.89, * | 0.36 | 0.01 |
| **QQ BC RACMO** | 0.87, * | 0.33 | 0 | 0.92, * | 0.3 | 0.01 |
| **QQ BC RCA** | 0.82, * | 0.39 | 0 | 0.89, * | 0.32 | 0.02 |

*Table 5.7 continued.*

| Model and BC method | DED (East Germany) | | | DEF (North Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.86, * | 0.57 | 0.42 | 0.93, * | 0.65 | 0.59 |
| **Raw RACMO** | 0.91, * | 1.17 | -1.13 | 0.96, * | 0.29 | 0.14 |
| **Raw RCA** | 0.88, * | 0.47 | 0.32 | 0.96, * | 0.41 | 0.34 |
| **Linear BC CCLM** | 0.86, * | 0.38 | 0.01 | 0.93, * | 0.28 | 0 |
| **Linear BC RACMO** | 0.91, * | 0.32 | 0 | 0.96, * | 0.25 | -0.02 |
| **Linear BC RCA** | 0.88, * | 0.34 | 0 | 0.96, * | 0.22 | -0.02 |
| **Variance BC CCLM** | 0.88, * | 0.37 | 0.01 | 0.94, * | 0.28 | 0 |
| **Variance BC RACMO** | 0.9, * | 0.32 | 0 | 0.95, * | 0.28 | -0.01 |
| **Variance BC RCA** | 0.89, * | 0.33 | 0 | 0.96, * | 0.22 | -0.02 |
| **QQ BC CCLM** | 0.85, * | 0.43 | 0.03 | 0.93, * | 0.34 | -0.06 |
| **QQ BC RACMO** | 0.89, * | 0.36 | 0.02 | 0.95, * | 0.33 | -0.06 |
| **QQ BC RCA** | 0.88, * | 0.34 | 0.03 | 0.97, * | 0.21 | -0.06 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $R$, smaller $RMSE$ or bias) relative to the uncorrected RCM simulation. $RMSE$ and bias are in Celsius.*

*Table 5.8: Statistical comparison of total annual precipitation from RCM evaluation simulations and observations, German regions, 1981-2010.*

| Model and BC method | DE2 (South Germany) | | | DEA (West Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.7, * | 197.34 | 169.55 | 0.77, * | 104.16 | -72.27 |
| **Raw RACMO** | 0.53, * | 127.19 | 65.39 | 0.74, * | 102.65 | 57.9 |
| **Raw RCA** | 0.55, * | 129.25 | -66.2 | 0.67, * | 132.53 | 89.01 |
| **Linear BC CCLM** | 0.7, * | 89.7 | -1.15 | 0.77, * | 82.46 | -0.39 |
| **Linear BC RACMO** | 0.51, * | 109.75 | 1.08 | 0.73, * | 83.5 | 1.11 |
| **Linear BC RCA** | 0.55, * | 127.64 | 2.85 | 0.66, * | 96.35 | 3.27 |
| **Variance BC CCLM** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **Variance BC RACMO** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **Variance BC RCA** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **QQ BC CCLM** | 0.71, * | 94.75 | 13.33 | 0.8, * | 78.93 | 13.63 |
| **QQ BC RACMO** | 0.51, * | 122.55 | 18.61 | 0.71, * | 105.72 | 26.4 |
| **QQ BC RCA** | 0.55, * | 156.43 | 23.38 | 0.65, * | 115.01 | 23.8 |

*Table 5.8 continued.*

| Model and BC method | DED (East Germany) | | | DEF (North Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.77, * | 106.79 | 69.04 | 0.7, * | 145.37 | -120.77 |
| **Raw RACMO** | 0.73, * | 167.81 | 144.63 | 0.71, * | 82.47 | 18.77 |
| **Raw RCA** | 0.59, * | 235.1 | 211.01 | 0.55, * | 191.64 | 129.91 |
| **Linear BC CCLM** | 0.79, * | 76.06 | 3.57 | 0.71, * | 86.26 | -2.21 |
| **Linear BC RACMO** | 0.72, * | 81.16 | 2.94 | 0.71, * | 80.41 | 2.36 |
| **Linear BC RCA** | 0.59, * | 97.03 | 0.77 | 0.56, * | 125.73 | 5.25 |
| **Variance BC CCLM** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **Variance BC RACMO** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **Variance BC RCA** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **QQ BC CCLM** | 0.78, * | 80.65 | -3.57 | 0.72, * | 86.54 | 12.15 |
| **QQ BC RACMO** | 0.74, * | 88.15 | 11.04 | 0.66, * | 103.93 | 25.48 |
| **QQ BC RCA** | 0.58, * | 108.42 | 8.72 | 0.53, * | 158.84 | 23.1 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $R$, smaller $RMSE$ or bias) relative to the uncorrected RCM simulation. $RMSE$ and bias are mm.*

*Table 5.9: Kolmogorov-Smirnov (KS) test statistics on the distribution of annual averages of maximum and minimum temperature, and total annual precipitation from RCM evaluation simulations and observations, German regions, 1981-2010.*

| Model and | DE2 (South Germany) | | | DEA (West Germany) | | |
|---|---|---|---|---|---|---|
| BC method | Tmax | Tmin | Precip | Tmax | Tmin | Precip |
| **Raw CCLM** | 0.73, * | 0.57, * | 0.53, * | 0.7, * | 0.17 | 0.3 |
| **Raw RACMO** | 0.5, * | 0.7, * | 0.37, * | 0.63, * | 0.53, * | 0.3 |
| **Raw RCA** | 0.47, * | 0.23 | 0.37, * | 0.53, * | 0.17 | 0.33 |
| **Linear BC CCLM** | 0.13 | 0.17 | 0.23 | 0.17 | 0.13 | 0.13 |
| **Linear BC RACMO** | 0.2 | 0.17 | 0.1 | 0.13 | 0.1 | 0.13 |
| **Linear BC RCA** | 0.13 | 0.23 | 0.13 | 0.17 | 0.13 | 0.13 |
| **Variance BC CCLM** | 0.17 | 0.17 | 0.03 | 0.13 | 0.13 | 0.03 |
| **Variance BC RACMO** | 0.17 | 0.17 | 0.03 | 0.17 | 0.1 | 0.03 |
| **Variance BC RCA** | 0.17 | 0.17 | 0.03 | 0.2 | 0.17 | 0.03 |
| **QQ BC CCLM** | 0.23 | 0.17 | 0.23 | 0.17 | 0.17 | 0.17 |
| **QQ BC RACMO** | 0.17 | 0.17 | 0.17 | 0.17 | 0.13 | 0.2 |
| **QQ BC RCA** | 0.2 | 0.13 | 0.13 | 0.23 | 0.13 | 0.17 |

*Table 5.9 continued.*

| Model and | DED (East Germany) | | | DEF (North Germany) | | |
|---|---|---|---|---|---|---|
| BC method | Tmax | Tmin | Precip | Tmax | Tmin | Precip |
| **Raw CCLM** | 0.77, * | 0.33 | 0.33 | 0.6, * | 0.43, * | 0.5, * |
| **Raw RACMO** | 0.53, * | 0.57, * | 0.53, * | 0.6, * | 0.17 | 0.17 |
| **Raw RCA** | 0.5, * | 0.23 | 0.7, * | 0.57, * | 0.33 | 0.4, * |
| **Linear BC CCLM** | 0.2 | 0.1 | 0.13 | 0.1 | 0.13 | 0.13 |
| **Linear BC RACMO** | 0.2 | 0.1 | 0.1 | 0.13 | 0.1 | 0.1 |
| **Linear BC RCA** | 0.17 | 0.13 | 0.2 | 0.2 | 0.17 | 0.17 |
| **Variance BC CCLM** | 0.13 | 0.13 | 0.03 | 0.1 | 0.1 | 0.03 |
| **Variance BC RACMO** | 0.2 | 0.1 | 0.03 | 0.13 | 0.1 | 0.03 |
| **Variance BC RCA** | 0.17 | 0.13 | 0.03 | 0.2 | 0.17 | 0.03 |
| **QQ BC CCLM** | 0.13 | 0.17 | 0.17 | 0.13 | 0.1 | 0.17 |
| **QQ BC RACMO** | 0.2 | 0.13 | 0.1 | 0.2 | 0.17 | 0.13 |
| **QQ BC RCA** | 0.2 | 0.17 | 0.17 | 0.23 | 0.13 | 0.23 |

*($*$) indicates statistical significance ($p$ <0.05).*

### 5.4.3 Comparing the effect of BC on past yield simulations generated by the SCCM

The results of using uncorrected and BC evaluation-run RCM output show different effects on regional yield hindcasts, and the magnitude of these effects depends on the BC method, region and crop modeling method (Figs. 5.15A-5.18A). After BC, it can be observed that only variance methods reduce biases in SCCM yield projections (Tables 5.10-5.13): for example, in DE2 bias (and $RMSE$) are reduced to -0.06 (0.26 t/ha) over the simulation period 1981-2010 compared to the initial bias of -0.14 (0.52 t/ha).

Among the BC methods, variance and QQ mapping show the ability to reduce $RMSE$ compared to the simple linear method, reduce the mean biases over the period of simulations, and improve or maintain high positive correlation to the E-OBS yield hindcast and to observations. The use of variance BC is also able to improve the significant correlation of DE2, DEA, and DED ($r$=0.8.4, 0.94, 0.9, respectively).

It can also be observed that variance-corrected yields from DEF perfectly agree with the hindcast as the only significant climate predictor is total summer precipitation, which, by definition of the correction and the derivation of the climate index, perfectly equates with climate observations (Table 5.13).

The use of QQ mapping brought small improvements to yield hindcasts, for example in DE2 where it increased correlation, albeit marginally, and reduced $RMSE$ relative to observations ($RMSE$=0.7 t/ha compared to 0.78 t/ha for the raw simulations). Significant differences to the E-OBS hindcast are found between linear BC RCA yield simulations in DE2, raw CCLM yields in DEA, and linear BC RCA simulations in DED (Figs. 5.15C-5.15C). After BC, the raw CCLM yields in DEA are no longer significantly different to the E-OBS hindcast.

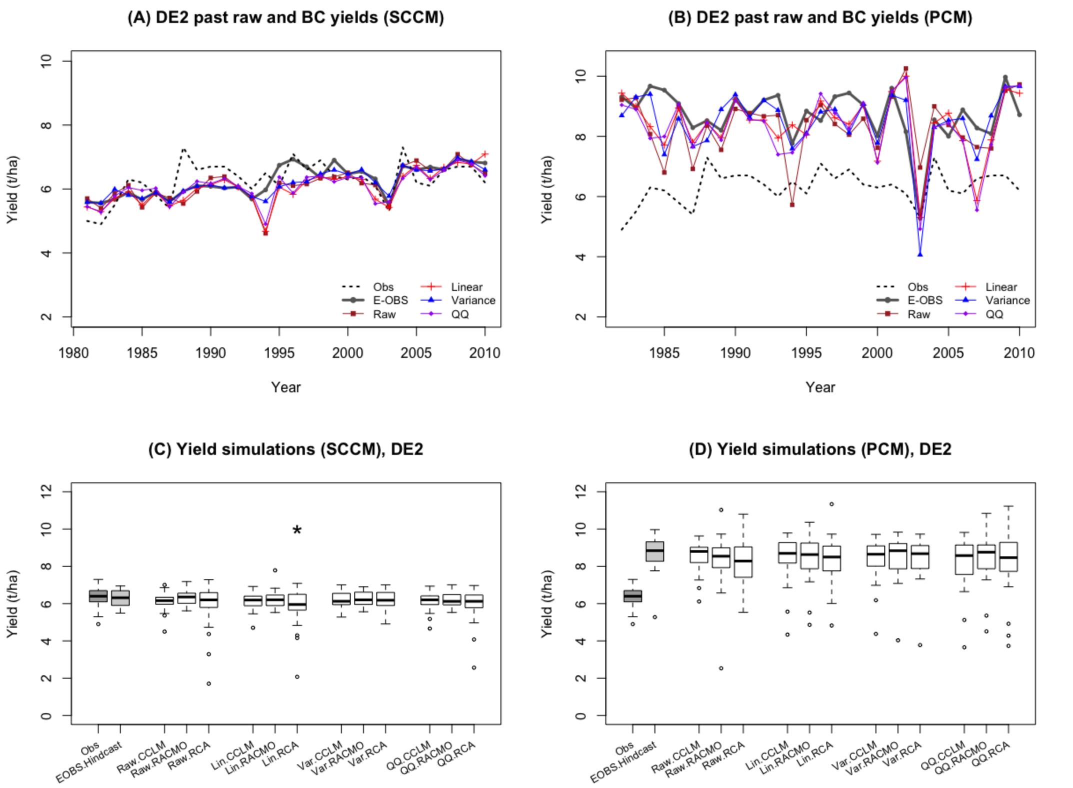



*Figure 5.15: DE2 (South Germany) SCCM yield simulation with uncorrected and bias-corrected climate model output. A * indicates a significant difference between the yield simulation and the E-OBS yield hindcast based on a t-test.*

*Table 5.10: Statistical evaluation between yield simulations, a yield hindcast generated with E-OBS, and yield observations, DE2 (South Germany).*

| BC method | SCCM | | | | | | PCM | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Correl. | | RMSE | | Mean bias | | Correl. | | RMSE | | Mean bias | |
| | EOBS | Obs | EOBS | Obs | EOBS | Obs | EOBS | Obs | EOBS | Obs | EOBS | Obs |
| **Raw** | 0.66* | 0.24 | 0.52 | 0.78 | -0.14 | -0.19 | 0.36 | -0.01 | 1.28 | 2.37 | -0.43 | 1.96 |
| **Linear** | 0.63* | 0.22 | 0.56 | 0.81 | -0.25 | -0.3 | 0.52* | 0.18 | 1.02 | 2.35 | -0.32 | 2.07 |
| **Variance** | 0.84* | 0.44* | 0.26 | 0.53 | -0.06 | -0.11 | 0.75* | 0.26 | 0.76 | 2.42 | -0.22 | 2.17 |
| **QQ** | 0.53* | 0.26 | 0.52 | 0.7 | -0.18 | -0.23 | 0.55* | 0.18 | 1.1 | 2.36 | -0.38 | 2.01 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $r$, smaller $RMSE$ or bias) relative to the yield hindcast (EOBS) or observations (OBS). $RMSE$ and bias are in t/ha.*

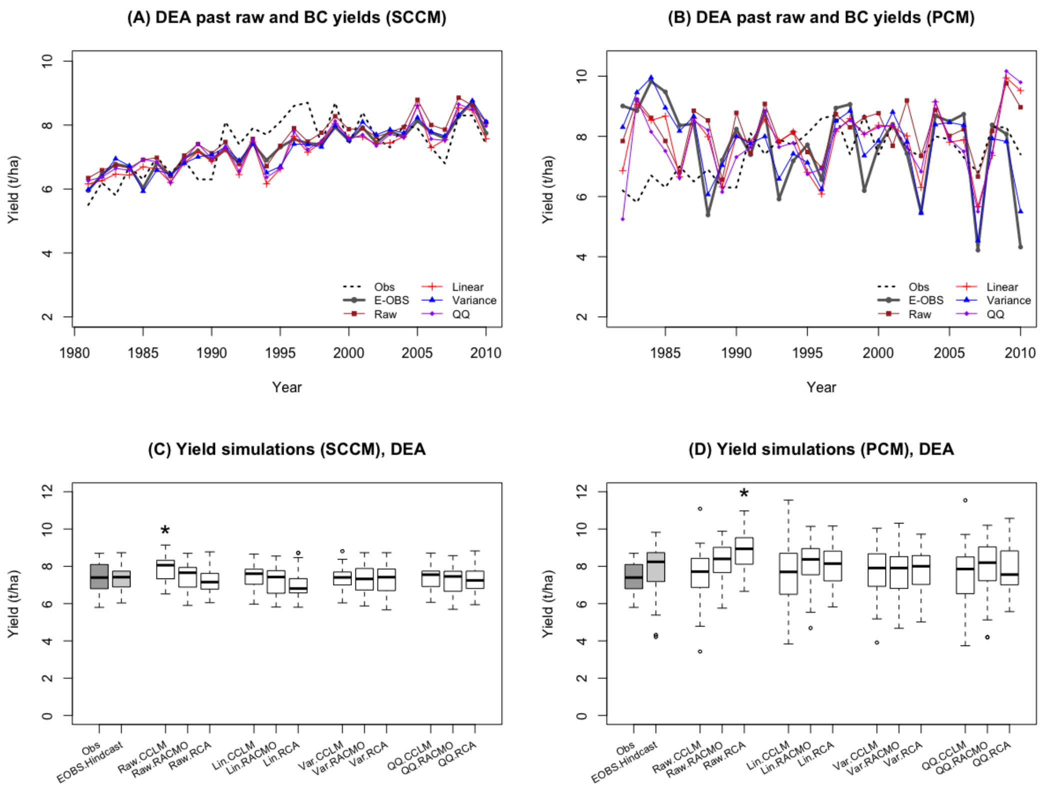



*Figure 5.16: DEA (West Germany) SCCM yield simulation with uncorrected and bias-corrected climate model output. A * indicates a significant difference between the yield simulation and the E-OBS yield hindcast based on a t-test.*

*Table 5.11: Statistical evaluation between yield simulations, a yield hindcast generated with E-OBS, and yield observations, DEA (West Germany).*

| Model and BC method | SCCM | | | | | | PCM | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Correl. | | RMSE | | Mean bias | | Correl. | | RMSE | | Mean bias | |
| | EOBS | Obs | EOBS | Obs | EOBS | Obs | EOBS | Obs | EOBS | Obs | EOBS | Obs |
| **Raw** | 0.92* | 0.62* | 0.26 | 0.66 | 0.02 | -0.06 | 0.06 | 0.01 | 1.85 | 1.61 | 0.86 | 1.1 |
| **Linear** | 0.87* | 0.59* | 0.39 | 0.74 | -0.2 | -0.28 | 0.23 | 0.1 | 1.63 | 1.4 | 0.39 | 0.63 |
| **Variance** | 0.94* | 0.59* | 0.25 | 0.71 | -0.03 | -0.12 | 0.93* | -0.12 | 0.57 | 1.55 | 0.08 | 0.33 |
| **QQ** | 0.87* | 0.57* | 0.33 | 0.71 | -0.03 | -0.12 | 0 | 0.34 | 1.9 | 1.31 | 0.29 | 0.53 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $r$, smaller $RMSE$ or bias) relative to the yield hindcast (EOBS) or observations (OBS). $RMSE$ and bias are in t/ha.*

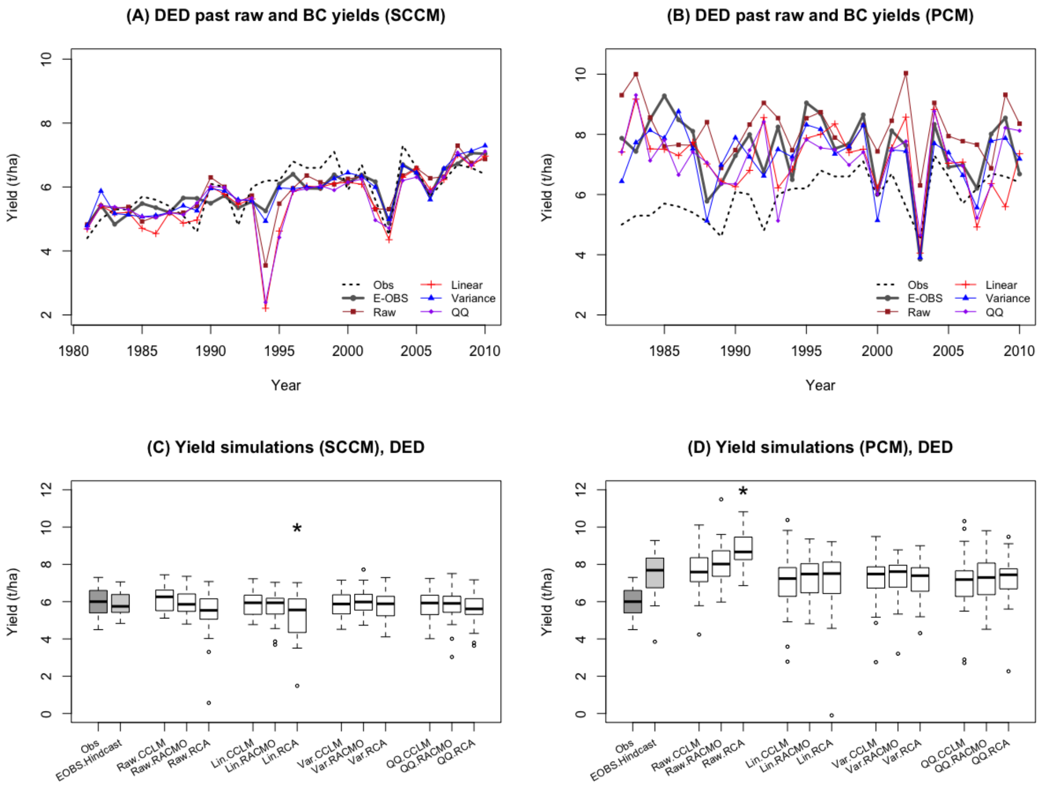



*Figure 5.17: DED (East Germany) SCCM yield simulation with uncorrected and bias-corrected climate model output. A * indicates a significant difference between the yield simulation and the E-OBS yield hindcast based on a t-test.*

*Table 5.12: Statistical evaluation between yield simulations, a yield hindcast generated with E-OBS, and yield observations, DED (East Germany).*

| **Model and BC method** | **SCCM** | | | | | | **PCM** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Correl.** | | **RMSE** | | **Mean bias** | | **Correl.** | | **RMSE** | | **Mean bias** | |
| | **EOBS** | **Obs** | **EOBS** | **Obs** | **EOBS** | **Obs** | **EOBS** | **Obs** | **EOBS** | **Obs** | **EOBS** | **Obs** |
| **Raw** | 0.67* | 0.45* | 0.68 | 0.89 | -0.24 | -0.31 | 0.24, | 0.08 | 1.61 | 2.75 | 0.96 | 2.51 |
| **Linear** | 0.63* | 0.41* | 1.04 | 1.23 | -0.48 | -0.54 | 0.42* | 0.23 | 1.25 | 1.72 | -0.29 | 1.25 |
| **Variance** | 0.9* | 0.66* | 0.31 | 0.59 | 0.01 | -0.05 | 0.82* | 0.47* | 0.71 | 1.58 | -0.27 | 1.27 |
| **QQ** | 0.52* | 0.27 | 1 | 1.21 | -0.33 | -0.39 | 0.42* | 0.38* | 1.17 | 1.57 | -0.29 | 1.25 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $r$, smaller $RMSE$ or bias) relative to the yield hindcast (EOBS) or observations (OBS). $RMSE$ and bias are in t/ha.*

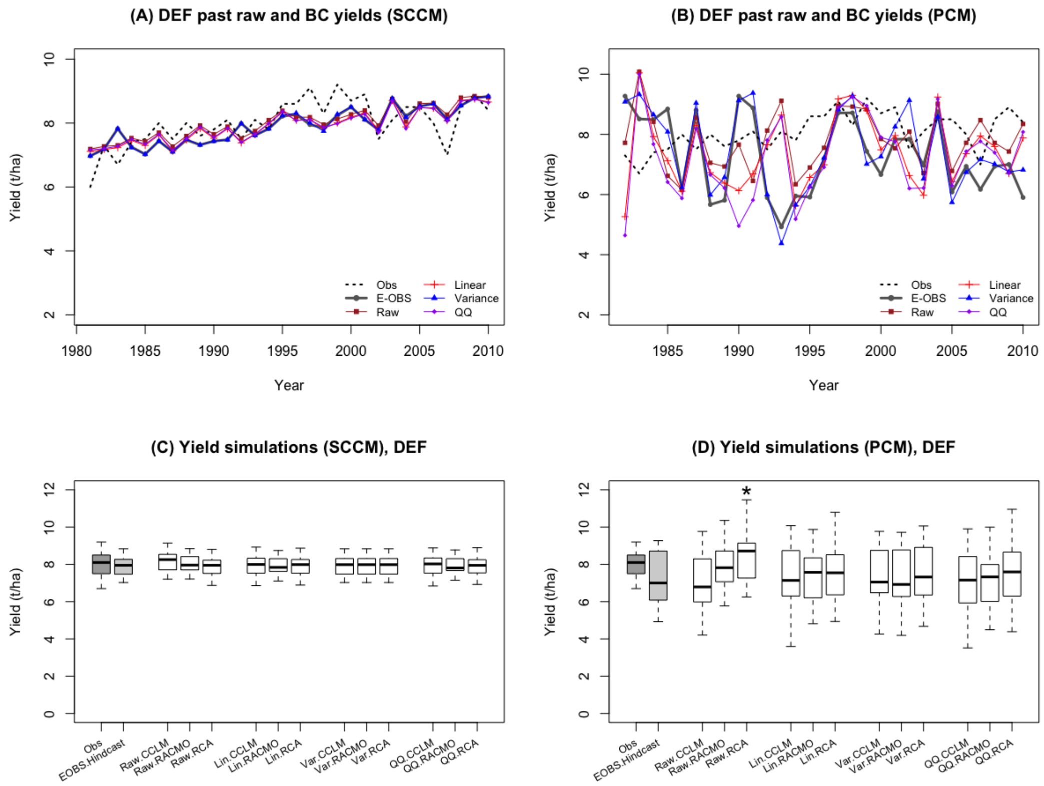



*Figure 5.18: DEF (North Germany) SCCM yield simulation with uncorrected and bias-corrected climate model output. A * indicates a significant difference between the yield simulation and the E-OBS yield hindcast based on a t-test.*

*Table 5.13: Statistical evaluation between yield simulations, a yield hindcast generated with E-OBS, and yield observations, DEF (North Germany).*

| Model and BC method | SCCM | | | | | | PCM | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Correl. | | RMSE | | Mean bias | | Correl. | | RMSE | | Mean bias | |
| | EOBS | Obs | EOBS | Obs | EOBS | Obs | EOBS | Obs | EOBS | Obs | EOBS | Obs |
| **Raw** | 0.86* | 0.63* | 0.27 | 0.5 | 0.02 | -0.11 | 0.3 | -0.22 | 1.62 | 1.32 | 0.84 | 0.07 |
| **Linear** | 0.86* | 0.63* | 0.27 | 0.51 | -0.02 | -0.15 | 0.33 | 0 | 1.42 | 1.43 | 0.2 | -0.57 |
| **Variance** | 1* | 0.57* | 0.03 | 0.56 | 0 | -0.13 | 0.92* | -0.21 | 0.58 | 1.75 | 0.15 | -0.62 |
| **QQ** | 0.83* | 0.63* | 0.29 | 0.51 | -0.01 | -0.14 | 0.15 | 0.08 | 1.7 | 1.63 | 0 | -0.78 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $r$, smaller $RMSE$ or bias) relative to the yield hindcast (EOBS) or observations (OBS). $RMSE$ and bias are in t/ha.*

Although the mean bias over the simulation period 1981-2010 between actual, hindcasted, and SCCM simulated yields is already quite small (under 1 t/ha), what these results show is that apart from the ability of BC to successfully bring RCM output more similar to E-OBS climate observations, BC is also able to improve the yield simulations. BC is able to improve SCCM yields relative to the E-OBS SCCM hindcast, and in some cases, relative to actual yield observations as well (e.g. DE2 and DED with variance correction).

However, yield simulations are also affected by the computational artifacts of the correction, for instance that since the DEF SCCM only has summer precipitation as a significant predictor, and that variance correction results in perfectly equal simulations to observations. Generally, it can also be observed that, similar to their evaluation in Chapter 3, the $RMSE$ and biases of SCCM yields are much smaller than yield simulations from the PCM, which is discussed in the next section.

### 5.4.4 Comparing the effect of BC on past yield simulations generated by the PCM

The application of BC also results in improvements to yield generated by the PCM using individual RCM output (Figs. 5.15B-5.18B, Appendix Tables A15-A18). For example, all BC methods reduce $RMSE$, improve correlation, and reduce bias in DE2 and DED relative to both observations and E-OBS hindcast yields (Tables 5.10 and 5.12). For DEA, all methods reduce the PCM yield bias relative to both the E-OBS hindcast and observations: from 0.86 t/ha bias to 0.39, 0.08, and 0.29 t/ha for linear, variance and QQ mapping methods respectively. However, only variance methods in DEA are able to improve correlation to $r$=0.93 (Table 5.11).

In DEF (North Germany), all BC methods significantly reduce bias from 0.84 t/ha relative to the E-OBS hindcast to 0.2, 0.15 and 0 t/ha for linear, variance and QQ correction methods, respectively (Table 5.13). It can be observed that median yields are more similar to observations and the yield hindcast after using BC. For example, in DEA, DED, and DEF, yield simulations with raw RCA output are significantly different to their respective E-OBS hindcast based on t-tests. After correction, yields are no longer significantly different (Figs. 5.16D-5.18D).

After BC, the size of the bias (in t/ha) relative to the respective E-OBS hindcast of both PCMs and SCCMs is fairly comparable, typically under 0.5 t/ha. These results indicate that BC is also capable of improving PCM yield simulations because it is able to improve the climate model output (daily Tmax, Tmin, Precipitation) that is used as input. However, BC does not necessarily improve the yield simulations relative to actual yield, as seen in the remaining large difference between yield simulations and observations in DE2. This discrepancy is therefore still connected to the crop modeling method itself, and this is discussed later in the chapter.

#### 5.4.4.1 Comparison of ensemble SCCM and PCM yields with bias-corrected climate input

A comparison of the ensemble SCCM and PCM yields, averaged per BC method, generated using raw and BC RCM output is shown in Fig. 5.19. It can be observed in DEA, DED, and DEF that ensemble median yields and ranges are more similar to both observations and the yield hindcast generated with E-OBS after BC (Fig. 5.19B-D). For example, in DE2, the range of yield simulations using the PCM and uncorrected RCM output is large, but after BC this is significantly reduced by linear, variance and QQ methods (Fig. 5.19A), although the bias between simulated and observed yields in this region is still high.

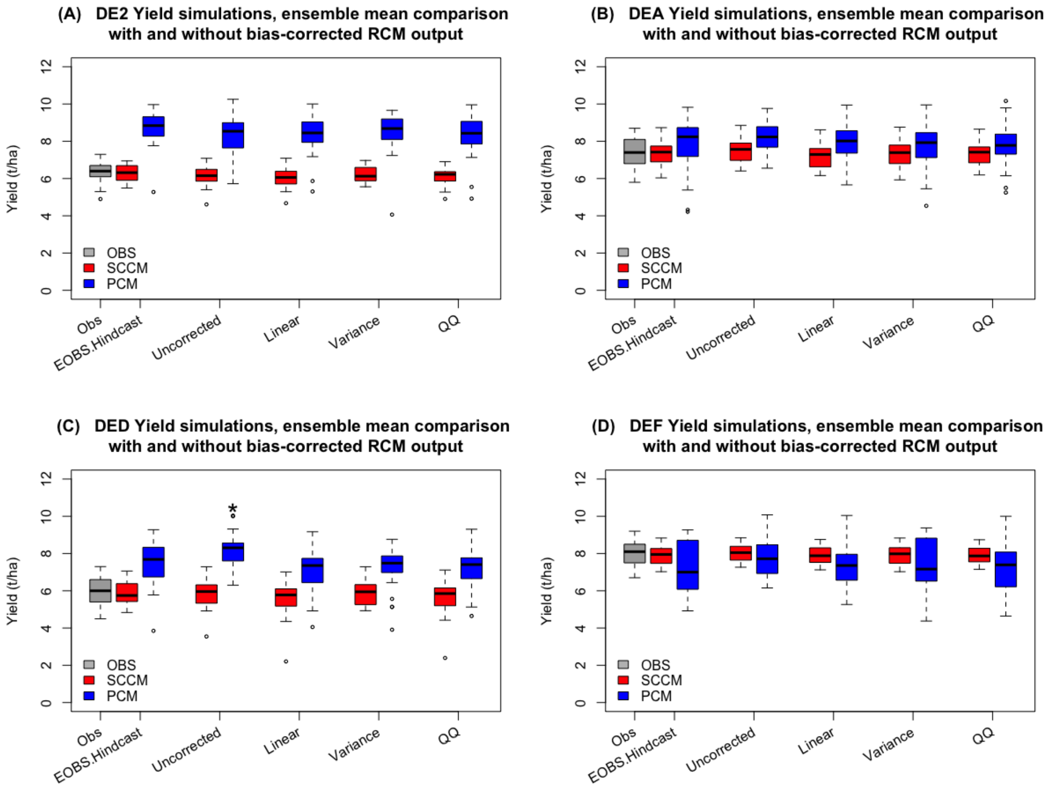



*Figure 5.19: Comparison of yield observations, SCCM and PCM yield hindcasts, and ensemble mean of yields generated by bias-corrected RCM output for German regions, 1982-2010. A * indicates a significant difference between the yield simulation and the E-OBS yield hindcast based on a t-test.*

Overall, the analysis of the effect of BC on yield simulations from both the SCCM and PCM approaches shows that the while all BC methods are generally effective in minimizing bias in PCM yields, variance and QQ correction are more useful in the SCCM yields compared to linear methods (Tables 5.10-5.13). Ensemble means of PCM yields with uncorrected RCM output in this study have generally poor and non-significant correlation to reference yields, but this is improved by BC, particularly in the DE2 and DED regions. The results of the RCM evaluation, correction, and their application into crop models are discussed in the following section.

## 5.5 Discussion

In this chapter, the research questions are focused on how well RCMs perform relative to observations and whether BC improves the correlation and reduces bias/$RMSE$ when using different methods of varying complexity. Another research question for the chapter is whether improved climate model output with reduced biases from the correction can reduce input error, thereby also improving yield simulations from both statistical and process-based approaches to crop modeling. The results presented are discussed in this section to connect the results to the research questions and to previous studies.

Additionally, the results in this chapter are used in the formulation of an approach for use in the following chapters which aim to utilize future climate projections to generate yield simulations. To begin the discussion, firstly, the results of the RCM evaluation and bias correction are addressed.

### 5.5.1 Error in reanalysis-driven RCM simulations

#### 5.5.1.1 Results of the bias analysis for climate

Statistical evaluation shows that climate simulations from reanalysis-driven RCMs contained biases when used directly (uncorrected) in a comparison to observations. Typically, Tmax was underestimated, Tmin was overestimated, and individual RCMs had positive or negative biases in simulating precipitation, but this varied from region to region. These biases from RCM evaluation simulations were addressed using different BC methods of varying complexity, which were all effective in reducing the biases in RCM output, $RMSE$, and also led to some improvements in correlation relative to climate observations.

Because of its design, QQ mapping was also able to correct the distribution of daily Tmax and Tmin at the regional level (not achieved by the more simple methods), as evidenced by the KS test statistic. Although some RCMs still had better performances than others – for example, the CCLM model typically had the highest correlation among other RCMs for Tmax and precipitation – all bias corrected simulations showed significant improvements, and their corrected output were closer to observations regardless of the BC method used.

#### 5.5.1.2 Intrinsic error in RCMs

It is evident from the RCM evaluation that the error from RCMs should be assessed before their use in impact studies, regardless of the BC method used, or if BC is applied at all. Biases in the evaluation RCM simulations are significant, because they are intrinsic to the RCMs (Ruiz-Ramos et al., 2016, Maraun et al., 2010, Fowler et al., 2007), as RCM evaluation simulations are all driven by the same ERA-Interim data. When the boundary conditions of RCMs are reanalysis output, correlations in terms of the time evolution of RCM output and observations are expected (de Elía et al., 2017).

In contrast, when using historical GCM simulations as boundary conditions to downscaling RCMs (i.e. ‘free-running’ GCM simulations), downscaling skill depends strongly on the biases inherited from the driving GCM (See Section 2.2.2). In the results presented, errors and differences between climate observations and simulated temperature and precipitation from RCM evaluation simulations were therefore a result of the differences between RCMs chosen for the experiment.

It should be noted, however, that atmospheric reanalyses such as the ERA-Interim data which was used as boundary conditions for the RCMs, are also based on imperfect models, meaning there may still some differences between reanalysis datasets and actual climate observations which can affect the evaluation of RCMs (Kotlarski et al., 2014, de Elía et al., 2017). While the use of simulations from EURO-CORDEX and E-OBS is a common and suitable pairing over the European domain (e.g. Kotlarski et al., 2014) and matches the purposes of the study, a comparison of reanalysis datasets is also a point that can be explored in future studies.

Although the type of RCM evaluation used in the chapter does not necessarily allow for uncovering the physical reasons for the found biases (similar to e.g. Kotlarski et al., 2014), the results of this chapter indicate that RCMs can contribute to error in impact projections.

#### 5.5.1.3 Utilizing bias corrected simulations for impact assessment

The choice of RCM and BC method paired with impact assessment models is also important. For example, it has been shown that the choice of downscaling is significant in reproducing past yields (e.g. Ramarohetra et al., 2015). Similarly, the large influence of the choice of RCM was found in a hydrological impact study, where it was shown that uncertainty in regional climate projections due to different RCMs is greater than the uncertainty stemming from different BC methods (Hwang et al., 2014, Teutschbein and Seibert, 2012).

Other examples which highlight the careful pairing of climate and impact models include studies which show that errors in precipitation were found to propagate through, and even enhanced by, non-linear processes that simulated stream flow (Hwang et al., 2014), so it is argued that climate model outputs with minimal biases are intuitively ideal. In a crop modeling study, the performances of RCMs in reproducing crucial climate variables for crop production were shown to be extremely variable, which led to a large range of crop yield projections (Oettli et al., 2011). Climate simulations that retain large errors, as shown in the uncorrected temperature and precipitation simulations, could therefore result in over- or underestimation- of projected yield changes and impacts.

Crops are sensitive to the timing of extreme temperature and precipitation events (Glotter et al., 2014), and certain temperatures can trigger different developmental stages. For example, a cool bias that underestimates Tmax during sensitive growth stages could underestimate yield impacts in a future climate with more hot days. Heat stress is predicted to be a more significant stressor than drought for wheat production in the future (Semenov and Shewry, 2011), so a realistic representation of Tmax is critical.

In summary, this cascade of errors in using RCMs and BC has been shown to affect projections from impact models, and this has been observed in the results of the yield hindcasts of the chapter. In the following section, the effects of the intrinsic error in RCMs on yield simulations is discussed.

## 5.5.2 Improvement of crop yield simulations through BC

### 5.5.2.1 Comparing the effect of different BC methods on yield simulations

In this chapter, all BC methods were generally able to reduce bias and $RMSE$ in yield simulations (driven by E-OBS), although their correlation to actual yield observations was not always improved. While variance correction was shown to be effective in reducing bias across both crop modeling methods, as well as for individual and ensemble yield simulations, the computational artifact of perfectly equal total precipitation will be problematic, particularly for the future yield simulations generated by the SCCM. For example, the North Germany state (DEF) only has summer precipitation as a significant climate predictor, and after BC of daily values the sum total equated to the total over the observations period. This means that any projections for this region, using the SCCM with JJA precipitation, will continue to reproduce these computational artifacts.

Another BC method which was fairly consistent in reducing input error and thus improving yield simulations was QQ mapping, which shows promise for use into the future. QQ mapping has been well-used and evaluated in climate impact studies (e.g. Macadam et al., 2016, Staffell and Pfenninger, 2016, Teutschbein and Seibert, 2012, Eisner et al., 2012, Oettli et al., 2011). QQ mapping brought improvements to SCCM and PCM yield simulations through the reduction of bias in most of the yield simulations for German regions in the study. These results were also reported in a recent study that investigated the effect of QQ mapping on wheat yield simulations for Australia (Macadam et al., 2016). Their study concluded that BC is a necessary step in yield simulation, in addition to the initial reduction of bias because of the use of RCMs. Biases in rainfall were found to be inherited from the forcing GCMs; therefore BC would have still been necessary even at finer climate model resolutions (Macadam et al., 2016).

However, there are exceptions: for example, in DE2 (South Germany) it can be observed that there is a large bias between PCM simulations and observations even after BC. The discrepancy between the yields from the PCM and actual yields was discussed extensively in Chapter 3 as a result of applying a field-based, input-intensive crop model to a regional scale, with implications for aggregation error due to the scaling-up of processes. This is likely to have contributed to the large bias in DE2. Because a regional genetic coefficient is used in the research to facilitate comparison between the SCCM and PCM, it is argued that the coefficients do not necessarily reflect the actual cultivars grown in DE2. The regional coefficients from the study of Nain and Kersebaum (2007) are derived from experimental work completed in North-Central Germany.

Apart from adopting more rigorous evaluation, this also emphasizes what was discussed in Chapter 3, which was the need for better reporting of regional data that can be used for calibration. This underscores the need for more regional-level yield data and (related to phenology and other developmental stages of wheat) to aid in better calibrating the CERES/DSSAT parameters. In addition, better reporting of input parameters such as regional genetic coefficients often used in in the 'iterative' process of crop model calibration could improve yield hindcasts performed with climate observations as input.

#### 5.5.2.2 Comparing BC RCM yields simulated with different methods

The results in this chapter further add evidence that the impact projections are affected by the cascade of error and uncertainty from climate and crop models. The results also affirm the that the uncertainty in crop model projections is also a result of the linkages between climate and crop models. Given the relative acceptance and 'standardization of BC' (e.g. Chen et al., 2015, p.1123) as a necessary step in impact assessment, it was important to address how different BC methods affect yield simulations, giving particular focus to the crop modeling method as well.

Similar to the work of Macadam et al. (2016), BC was shown to be mostly effective in reducing large errors, for example with the PCM. By reducing the input error of the RCM simulations, BC was also able to improve the statistics between PCM yields and observations, which is research question 5 of the chapter. However, BC was less effective in improving small errors in the yield simulations which were generated by the SCCM. However, in the results presented here, BC with QQ mapping was not always effective in reducing biases for the SCCM, and in some cases made these errors slightly larger. In their work, this is theorized to be due to how QQ mapping uses a cumulative probability distribution function and how the BC procedure corrects biases on a daily, rather than seasonal, timestep (Macadam et al., 2016).

This problem of over-correction both through variance correction and QQ mapping, which re-scale the simulated time series in an attempt to explain unexplained small-scale variability results in the ‘inflation’ of the simulated time series (Maraun et al., 2015, p.2138): the drizzle effect for area means is over-corrected, area-mean extremes are overestimated, and trends are affected by the correction method.

The results in this chapter show that BC of the climate model output that is used as input to crop models can generally provide improvements to wheat yield simulations, both in SCCMs and PCMs, and relative to observations and a observation-driven yield hindcast. This is not surprising, because in order to realistically simulate changes in crop yields, crop models must be forced with climate data that closely represent relevant aspects of climate, especially considering that crops like wheat have non-linear responses to climate and other environmental factors (Macadam et al., 2016, Glotter et al., 2014, Hawkins et al., 2013a). Therefore, having climate model output that is closer in its mean and distribution to climate observations will inevitably bring improvements to yield simulations, at least relative to a hindcast driven by climate observations. This means that the choice of RCM (or driving GCM, and GCM-RCM combination) – and its evaluation – remains an important decision in the formulation of an impact assessment methodology.

Whether BC improves the yield simulations relative to actual yield observations themselves is argued to be more dependent on how well the crop modeling method can capture the reality it seeks to represent, partnered with minimal input errors, and even more specific to the the case of PCMs, the inclusion of factors apart from climate. In contrast to studies which have focused on climate model or crop model contributions to uncertainty, there have been relatively few studies that have assessed the effects of BC on wheat yield simulations in a manner that considers the crop modeling method. In this regard, and considering past results in previous chapters (e.g. See Chapter 3, crop model evaluation) how do these results inform the development of a methodology for future climate and crop yield projections?

### 5.5.3 Formulating a method for future yield projections

#### 5.5.3.1 An ‘optimal’ BC method?

As shown in the results, all BC methods used in this chapter were capable of improving uncorrected RCM simulations of daily Tmax, Tmin, and precipitation, which are important in winter wheat growth and development. However, there are differences in the ability of different BC methods to improve not only the mean values of climate variables but also their other properties such as distribution and variation, which are better accounted for in more sophisticated methods such as QQ mapping.

Among the BC methods presented here, QQ mapping was able to effectively correct the daily distribution of Tmax and Tmin in German regions. While both linear and variance correction were shown to be also effective, the latter’s computational artifacts (i.e. perfect monthly or annual total precipitation) may make it challenging to use in future simulations. In addition, because linear approaches only consider changes in the mean, extreme values in future climate scenarios are often leveled off (Supit et al., 2012). Furthermore, changes in climate variability are likely to be hard to correct using a mean (linear) BC only (Challinor et al., 2005).

It is argued that while it can be considered reasonable to limit the choice of BC to the best performing method, (e.g. here, QQ mapping), for simulations of future climate a singular choice is more difficult to justify. Should a single BC method be considered ‘optimal’ for the study? Ideally, rather than a single choice of BC method, an ensemble of bias-corrected climate model (GCM, GCM-RCM, or RCM) simulations can better characterize uncertainty from both the choice of climate model and the methods used to bring them closer to observations. However, based on practicality, and considering the overarching research questions on the study, it has been shown that the BC method of QQ mapping is not only effective in reducing bias in climate model output, it is also effective in reducing large biases in the yield simulations generated by an upscaled PCM. It can also bring some improvements to yields from SCCMs because it is effective in reducing biases in the driving RCMs. QQ mapping is also able to correct the distribution of daily temperatures and precipitation which are important for the PCM, and does not contain computational artifacts when used with the SCCM, as does the variance correction demonstrated here for the North Germany region (DEF).

Its frequent use and evaluation in climate impact studies make QQ mapping a suitable choice for performing BC on simulations of future climate from GCMs and RCMs. Therefore, the following chapters use only QQ mapping, although as shown in the results, linear and variance mapping are also efficient correction methods, depending on the context and research questions. However, it should be noted that QQ mapping, and BC methods in general, are highly criticized for several reasons.

### 5.5.4 Revisiting the issues and limitations of BC

BC may be a feasible way of reducing RCM error in climate simulations and this leads to reductions ‘downstream’ the impact assessment cascade. However, it should be recognized that there are also significant criticisms of the use of BC, and there are limitations in the method itself. This means that the methodology of simulating future yields needs to be reflective of the changes that BC can impose on the direction and magnitude of projected yield changes.

#### 5.5.4.1 Physical consistency and limited applicability

While useful in reducing error in model output, BC may also disrupt the physical consistency of climate simulations (Piani et al., 2010). In addition, while BC methods are capable of reducing biases, there is no BC approach that can completely remove biases over a validation period (Chen et al., 2015). For example, in the results of this chapter, while BC significantly improved the mean, variation and distribution of climate simulations, some biases remained after BC (e.g. DE2 PCM simulations).

Limitations of BC are also dependent on the chosen methods for impact projection or assessment. It is usually not possible to correct all relevant errors for which observations are available (Hawkins et al., 2013b). For instance, in the work of this chapter, temperature and precipitation were corrected as these are the needs of the SCCM and PCM. However, should more climate variables beyond temperature and precipitation be needed for the various crop modeling methods that exist, (e.g. the CERES-Wheat model also takes non-essential input information on humidity, wind speed, and other variables) it can be a challenge to find observational data to verify and evaluate simulations, limiting the value of the BC step.

#### 5.5.4.2 The issue of stationarity and bias correction

A significant limitation of BC is the issue of stationarity. Although BC of climate model output has emerged as a standard procedure in most recent climate change impact studies, there is a problematic assumption that climate model biases remain constant over time (Chen et al., 2015). This issue of stationarity is one of the major criticisms in the usage of BC, and in projecting future climate using BC. This is because the stationarity in the design of BC methods results in the calibrated BC coefficients, whether for correcting evaluation simulations or historical GCM-driven simulations, also being used to correct future climate change simulations. A major limitation of this approach is that it is impossible to say whether the variability of future climate will bear any resemblance to the variability of past climate used as the 'control' or calibration period for bias correction. This inherent assumption of stationarity is the weak point of any BC method (Teutschbein and Seibert, 2012).

This assumption of stationarity is also a major issue in the use of model output statistics (MOS, i.e. the combination of downscaling and correction approaches (Maraun et al., 2010)) for climate output. In numerical weather prediction, MOS-applied predictions can be verified by forecast verification scores. In contrast, MOS in climate studies is almost exclusively bias correction. For climate change studies, evaluation and assessment of MOS for future climate is essentially impossible (Maraun, 2016).

#### 5.5.4.3 Climate change signal modification

One of the most widely-discussed drawbacks of using BC is that while its use may be justified (as observed in the results of this chapter with RCM errors), BC may change the climate signal or trend that arises from the climate simulations (Hempel et al., 2013). Because of this – since many impact studies explicitly seek to investigate the impacts of climate change – whether or not the BC is appropriate for future use remains a topic of discussion (Hempel et al., 2013). There have been some recommendations of abandoning statistical BC approaches toward a more stochastic approach, meaning randomly adding small-scale variability (e.g. Maraun, 2013).

Therefore, while BC has obvious benefits, its use needs to be carefully considered in impact studies. The validity of the assumption of stationarity in bias has important implications for impact studies and needs to be verified to properly address uncertainty in future climate projections (Chen et al., 2015).

#### 5.5.4.4 Negotiating the BC issue for impact studies

Other approaches to negotiate the BC issue, particularly to manage the ‘destruction of physical consistency’ are to use a combination of dynamical downscaling and BC in a ‘trend-preserving approach’ (Hempel et al., 2013, p.220), which is used in the large Inter-Sectoral Impacts Model Intercomparison Project (ISI-MIP, Warszawski et al., 2014). In this approach, physical consistency is ensured by correcting low-resolution model data (e.g. sea surface temperatures) in order to provide correct boundaries for an RCM. However, this does not solve the problem completely, because the RCM itself introduces bias into simulations (Hempel et al., 2013). This non-trivial introduction of error by RCMs is also discussed in Chapter 2 and is the primary reason for the evaluation of RCMs in this chapter.

While this trend-preserving BC approach is expected to reduce the deviation between high-resolution simulations and observations while ensuring physical consistency of different climate variables (Hempel et al., 2013), it is evident that there are no clear-cut solutions to resolving the BC issue, nor the use of RCMs versus GCMs alone, which mean that devising a singular approach for future yield simulations is challenging and very contextual to the research questions being investigated.

Apart from more sophisticated BC methods, directly addressing and reducing RCM error means that there need to be improvements in the physical realism of processes within both GCMs and RCMs. In a recent review of climate change experiments performed at very high resolution (approximately 1.5-km RCM resolution), Kendon et al. (2017) discuss several high-resolution experiments that are now available to provide potential added value to future projections for convective precipitation, wind gusts, hail, fog, and lightning (Kendon et al., 2017). However, despite promising results and potential contributions from these high-resolution experiments, it remains essential that driving GCMs capture the key large-scale processes driving future changes before any regional downscaling is attempted (Kendon et al., 2017).

Independent of the improvement of climate and impact models, what is clear in moving forward with the BC issue is argued to be that as the use of RCMs and BC methods become more popular, standard, and accessible, bias-corrected climate model data may serve as the basis for real-world adaptation actions. In fact, this is the objective of many studies: that climate model simulations can provide plausible representations of future reality, giving society time to develop and implement climate change adaptation strategies. Climate model simulations thus contain an undeniable ethical dimension and should thus be plausible, defensible and actionable (Maraun, 2016). Therefore, the uncertainties of climate models, and the methods that are used to produce and utilize them should be communicated transparently.

#### 5.5.4.5 Remaining limitations

In the simulations with the PCM, daily solar radiation is an essential climate variable that is used by DSSAT/CERES-Wheat to simulate radiation use efficiency. In this case, solar radiation data from the ERA-Interim reanalysis dataset (Dee et al., 2011) was used as observations. As described in Macadam et al. (2016), it may be necessary to correct simulated solar radiation towards values derived from satellite products and/or observations of sunshine hours or other variables, however this may make the BC less reliable than for temperature and rainfall. Although solar radiation is not bias-corrected because in this study due to a larger focus on temperature and precipitation, and incomplete data for direct observations of solar radiation, improved input data could also result in improvements to yield simulations

## 5.6 Conclusion

The realism or plausibility of climate model simulations is important for assessing or projecting the impacts of climate change. While climate models continue to improve, downscaling and using BC remain an accessible ways of processing climate simulations for impact assessment by bridging the scale gap, and reducing errors in climate model output.

The use of BC on RCM output that is used as input to crop models has been shown to improve yield simulations to be closer to an observation-driven yield simulation because of the reduction of input error. However, each BC method used here affected yield simulations differently, particularly when using either the SCCM or PCM approach, which handle and utilize climate data in distinct ways. While QQ mapping was chosen to be a suitable method for future yield projections, regardless of this choice, it is clear that there are still gaps in effectively linking the output of climate models as input to crop models through downscaling and BC. It is also important to be conscious of the criticisms of RCMs and BC, which are methods that may add to error and uncertainty in the yield simulations.

It is recommended that prior to the use of climate model simulations in impacts or projection studies, the skill of RCMs and the effect of BC methods used to correct them need to be evaluated. Significant work needs to be done in order to negotiate the benefits provided by BC vis-à-vis its limitations. In order to move forward to project the impacts of climate change on yields, the work of the subsequent chapters continues the comparative crop modeling method approach, as well as a method that considers error introduced by RCMs individually as well as from joint GCM-RCM error.

# Chapter 6

# Projections of future temperature and precipitation

## 6.1 Introduction

According to the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report, climate change caused by increased greenhouse gas (GHG) emissions is anticipated to bring further warming to the earth's atmosphere and oceans, leading to changes in all components of the climate system (Summary for Policymakers (SPM), IPCC, 2014). Projections of future climate in the latest IPCC report are based on simulations of climate models forced with future emission scenarios, the Representative Concentration Pathways (RCPs, Moss et al., 2010, van Vuuren et al., 2011). Simulations from climate and earth system models project that global surface temperature change is likely to exceed 2°C for the higher emission scenarios RCP6.0 and RCP8.5 (SPM IPCC, 2014). Although there is less certainty on regional precipitation changes, increases in global mean surface temperature are expected to change normal water cycle patterns. Extreme precipitation events are very likely to become more intense and frequent by the end of this century (SPM IPCC, 2014).

These changes are anticipated to be detrimental to crop production because of the sensitivity of many crops, including wheat, to rising and extreme temperatures and variable precipitation (Asseng et al., 2014, and see Chapter 1 for more on wheat physiology). In this chapter, the projected changes in temperature and precipitation from dynamically downscaled climate model simulations from the Coordinated Regional Downscaling Experiment over Europe (EURO-CORDEX, Jacob et al., 2014) are investigated in order to use these simulations as input into crop models that can assess the potential impacts of climate change on future food security. In addition, previously discussed issues such as the effect of bias correction to handle errors in climate model output (See Chapter 5 discussion) are addressed in the chapter questions and design.

### 6.1.1 Bias correction and its contribution to uncertainty

While global climate models and earth system models (also general circulation models, GCMs and ESMs) are powerful simulation tools that use physics- and mathematics-based approaches to represent the atmosphere, ocean and land systems, the complexity of the earth's systems make it challenging to generate climate simulations without some uncertainty (See Chapter 2, Section 2.2.2). This uncertainty is linked to, among other sources, the numerous parameterizations that need to be made and challenges in representing complex processes. These modeled and parameterized processes also affect other simulated climate processes, making error propagation an issue. Projecting into the future is also challenging, considering natural variability and scenario uncertainty – i.e. how society develops or regulates greenhouse gas emissions. Error is also introduced when using regional climate models (RCMs) for downscaling because of the different structures and parameters of RCMs, and the possible pairings with a diverse number of driving GCMs.

RCM and GCM errors are typically addressed by bias correction (BC) methods, which attempt to minimize or eliminate biases in climate model output. It has been argued that utilizing GCM or GCM-RCM output without some form of BC will not accurately represent potential changes to climate (Hawkins et al., 2013b, Piani et al., 2010). Therefore it is 'unavoidable' that climate risk and impact assessment studies see the need to utilize BC methods (Iizumi et al., 2017, p.7800). However, in principle, BC is a post-processing step that does not address underlying error embedded in GCMs/ESMs or RCMs, only the error in their output.

BC is an additional source of uncertainty in climate risk assessments because different BC methods and reference daily weather data sets often lead to different impact outcomes (Iizumi et al., 2017). BC has also been criticized heavily because of its potential to introduce uncertainty into climate model output (e.g. the results from Chapter 5), potential modifications to the climate signal or to the embedded climate physics (e.g. Hempel et al., 2013), and also because in principle, there is no certainty that future climate will behave like past climate (e.g. Maraun, 2016). Because of the scientific discussion around BC, as well as the importance of BC in the impact assessment chain, the investigation of how it contributes to uncertainty is likewise a key objective of the research. While the decomposition of uncertainty linked to BC is addressed in Chapter 7, to prepare the climate simulations for yield projections and the uncertainty analysis, this chapter focuses on investigating how temperature and precipitation are projected to change until the end of the century. Additionally, how BC modifies these future projections is also investigated.

### 6.1.2 Chapter approach and objectives

In this chapter, dynamically downscaled GCM output (paired GCM-RCMs), both uncorrected and bias corrected, are used to investigate changes to climate, particularly for temperature and precipitation. These climate variables which are important to wheat growth, development, and yield. It is the objective of the chapter to compare uncorrected and BC climate projections under two future emission scenarios, and using two different approaches to calibrate the BC. Two different calibrations for the BC are used in order to separately address and characterize joint GCM-RCM and RCM-only error. These BC outputs are generated for their use in the process-based model (PCM) and statistical crop-climate model (SCCM) that are used to generate future yield simulations in the subsequent chapter.

The hypothesis is that BC climate simulations will project different changes to temperature and precipitation relative to uncorrected simulations, and that there may be modifications to the robustness of the climate change signal after BC. It is also hypothesized that the different GCM-RCM combinations will respond differently to the evaluation- or historical-based calibration, depending on the contributed error of the choice of GCM or RCM. The details of these different calibrations are described in the following data and methods section.

#### 6.1.2.1 Chapter research questions

**(1)** What are the trends and changes projected for temperature and precipitation from different GCM-RCM output? How do projections from different GCM-RCMs compare to each other?

**(2)** How do future climate projections from the chosen GCM-RCM combinations compare under the high and low emission (RCP8.5 and 2.6) scenarios?

**(3)** How are projected climate changes affected by the use of BC? Does BC change future climate projections relative to changes projected by uncorrected simulations?

**(4)** How do the results of two BC approaches, one that corrects GCM-RCM error and the other RCM error, compare?

**(5)** How can the results of the chapter inform the selection of GCM-RCM pairs for use in impact assessment?

## 6.2 Data and methods

Daily simulations of future precipitation, maximum and minimum temperature (Tmax and Tmin) from EURO-CORDEX over the UK, Germany and four German states are selected, using the highest and lowest RCP emission scenarios (RCP8.5 and RCP2.6, Moss et al., 2010, van Vuuren et al., 2011) from 2011 until the end of the century, in 30-year periods. These variables are crucial input data for the PCM and SCCMs that are used to generate yield projections. General changes and trends in these climate variables are investigated for the UK and Germany relative to the baseline period of historical simulations from 1976-2005. Figure 6.1 shows an outline of the general methods used in this chapter.

### 6.2.1 Climate models and future emission scenarios

Six combinations of GCM-RCMs – the same used in previous chapters – are used for comparison of their output, and to identify how each climate model projects future climate. These GCM-RCMs are: CCLM-MPI, RACMO-ECEARTH, RCA-CC, RCA-HadGEM, RCA-IPSL and RCA-MPI (Table 6.1). These GCMs and RCMs are chosen based on their inclusion in large model intercomparison projects (e.g. CMIP5 and CMIP6 (Eyring et al., 2016, Taylor et al., 2012), EURO-CORDEX and EURO-CORDEX-Adjust (Jacob et al., 2014)), in comparison and evaluation studies over Europe (e.g. Jury et al., 2015, Kotlarski et al., 2014), and generally satisfactory performances over the UK and Germany (See Chapter 4).

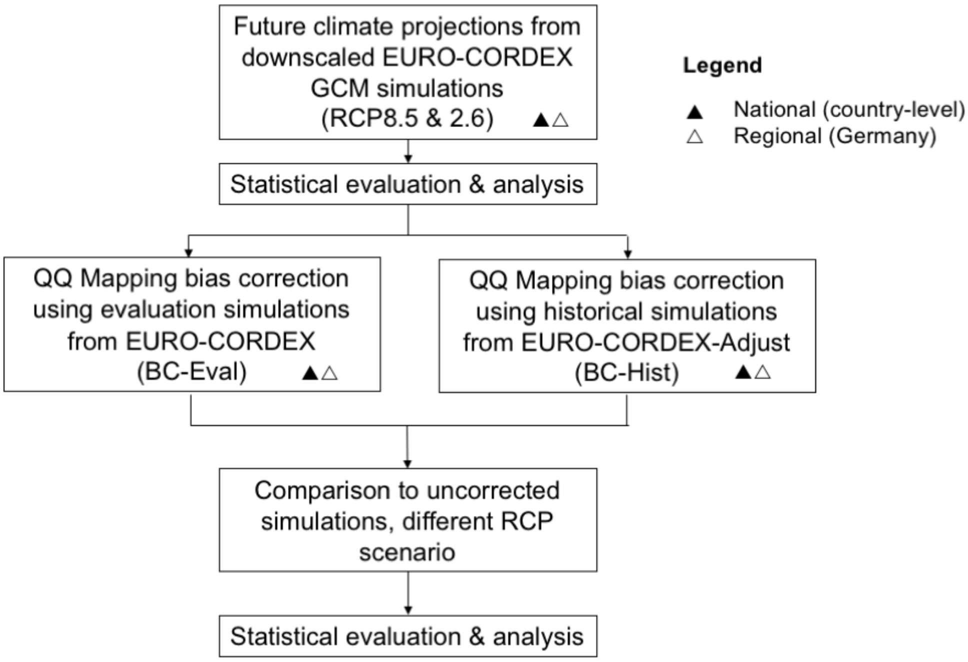



*Figure 6.1: Overview of Chapter 6 research design.*

*Table 6.1: Paired GCM-RCM combinations for future climate projections with available RCP scenarios.*

| GCM-RCM combination | Available RCP scenario |
|---|---|
| CCLM-MPI | RCP8.5 |
| RACMO-ECEARTH | RCP8.5 |
| RCA-CC | RCP8.5 |
| RCA-HadGEM | RCP8.5 and 2.6 |
| RCA-IPSL | RCP8.5 |
| RCA-MPI | RCP8.5 and 2.6 |

### 6.2.2 Bias correction method and approaches

BC methods are operationally used to post-process regional climate projections (Maraun et al., 2017). The motivation for the BC of climate model output is primarily to correct biases, because if the historical or evaluation time series are biased compared to the observations, then logically, future projections will also be biased. BC in the form of quantile-quantile (QQ) mapping (See Chapter 5 methods and discussion, Teutschbein and Seibert, 2012, Piani et al., 2010) is applied to these simulations in this chapter in order to correct the mean and distribution of climate model output. Although it was shown in Chapter 3 that a variety of BC methods are effective in bringing climate simulations closer to observations, QQ mapping is chosen because of its design to correct the distribution of simulations, and includes a cross-validation step and management of drizzle days (See Chapter 5 Methods).

#### 6.2.2.1 Calibration approaches to investigate GCM-RCM errors

Two approaches to calibrate the distribution parameters are used: (1) calibration using past RCM evaluation simulations (BC-Eval) and (2) calibration using past GCM-RCM historical simulations (BC-Hist). These approaches are used in order to identify the error coming from RCMs alone, and that coming from the use of downscaled GCM simulations which have both GCM and RCM error. As previously discussed in Chapters 4 and 5, RCM evaluation simulations are driven by reanalysis data at their boundaries, with EURO-CORDEX utilizing ERA-Interim to drive RCM evaluation runs (Jacob et al., 2014, Dee et al., 2011). In contrast, historical RCM simulations are driven by GCM simulations that have been driven by with time-varying external forcings (e.g. GHGs and other radiative forcings) at their boundaries. Thus historical simulations do not assimilate observations (or reanalysis data) and do not match the temporal evolution of atmospheric states in the real world (Eden et al., 2014). The comparison of historical GCM-RCMs and GCM-only simulations was performed in Chapter 4, where some "added value" from downscaling was found, and the error introduced by RCMs alone was evaluated in Chapter 5, where different choices of RCMs were found to introduce their own error to climate output.

The primary reason for using these two different calibration approaches is that it allows for the careful comparison and selection of GCM-RCM combinations with small biases (Pasten Zapata, 2017), so the results may be useful in future work for selecting climate models for impact studies. This approach to comparing calibration for BC also allows for the identification or isolation of error from GCMs and their downscaling RCM. The results of these two different BC approaches are compared to each other and also to uncorrected simulations. The low/high emission scenarios represented by the RCPs are also compared. These recent approaches to understanding the effects of BC on projections through these different calibrations are aimed to provide novel results to better understand climate model error.

### 6.2.3 Future emission scenarios

What the future world will look like in terms of the sustainability of global development, including the resulting GHG emissions, and any policies or regulatory measures to control GHGs is another significant source of uncertainty in climate and crop projections. There are extensive uncertainties in future forcings and responses to climate change, necessitating the use of scenarios to explore the potential consequences of different response options (Moss et al., 2010) To do this, RCPs are used (See Chapter 2, Section 2.2.2). The RCPs span the range of radiative forcing values from 2.6 to 8.5 W/m$^2$ in four scenarios (van Vuuren et al., 2011).

However, although four RCPs have been developed, not all future simulations are available for the selected GCM-RCMs in the EURO-CORDEX database at the time of analysis. For the purpose of the chapter, RCP8.5 and RCP2.6, the highest and lowest emission pathways, are chosen to represent unabated global warming and a peak and decline in emissions, respectively. In addition, RCP2.6 is only available for the RCA-HadGEM and RCA-MPI simulations. The $CO_2$ equivalent of RCP8.5 is 1370 parts per million (ppm) by the year 2100, and RCP2.6 would peak at 490 ppm before the year 2100, followed by a decline (Moss et al., 2010).

### 6.2.4 Statistical analyses and evaluation

As in previous chapters, the analysis of the chapter is for the UK and Germany, including four German states, because of the high productivity and importance of wheat in these locales. The four German states are identified by their EU NUTS code: DE2 (South), DEA (West), DED (East), and DEF (North). The climate model outputs are used to compare GCM-RCM performances before and after correction, in particular measuring projected changes.

The projected changes in the uncorrected future projections are compared to uncorrected past historical simulations (1976-2005, 30 years). Projections corrected considering RCM error only (BC-Eval) are compared to the calibration period of BC past RCM evaluation simulations (available for 1981-2010, also 30 years). For BC-Hist, the calibration period is BC past historical simulations. Early century refers to 2011-2040, mid-century 2041-2070, and late century is from 2071-2100.

Linear regression is also used to determine whether significant linear trends in temperature and precipitation exist. Trend analysis is also extended to climate indices such as the number of days above 31°C between June and August (JJA), a critical temperature for wheat (Porter and Gawith, 1999) and the total JJA precipitation, a period of potential heat stress for wheat (See Chapter 1 for a review of wheat physiology). The analysis in the chapter is limited to annual and seasonal trends in climate model output, and a more in-depth analysis of extremes such as their duration are recommended for future work. Additional t-tests are performed to test for significance between time-series.

## 6.3 Results

In this results section, climate change projections for temperature and precipitation over the chosen wheat-growing countries and regions are analyzed: firstly, national and regional uncorrected projections from the high and low emissions scenarios that are taken directly from the EURO-CORDEX simulations, followed by an analysis of how BC (with BC-Eval and and BC-Hist) affects the temperature and precipitaton simulations and their projected changes. The summer climate index projections are also reported. The results of the uncorrected projections are shown as the GCM-RCM ensemble for the purpose of clarity; however additional information on the effect of BC-Eval and BC-Hist on individual GCM-RCMs is also provided.

### 6.3.1 Projected temperature changes

#### 6.3.1.1 Uncorrected projections of annual Tmax and Tmin

At the national level, uncorrected climate projections show warming temperatures for the UK and Germany (Figs. 6.2 I-II and 6.3 I-II, Tables 6.2 and 6.3). The ensemble of GCM-RCMs show increases relative to the past historical simulation baseline, and have significant increasing trends until the end of the century for the UK and Germany. Future projections of both Tmax and Tmin show increases of over 2°C for the UK and over 3°C for Germany under RCP2.6 and RCP8.5 for the period 2071-2100. Differences between individual GCM-RCM projections show that the largest projected increases in temperature are from RCA-HadGEM in both the UK and Germany (Tables 6.4-6.5 and Tables 6.6-6.7). RCA-HadGEM projections are typically up to a degree warmer than the RCP8.5 ensemble mean, but cooler than the RCP2.6 ensemble mean for both Tmax and Tmin.

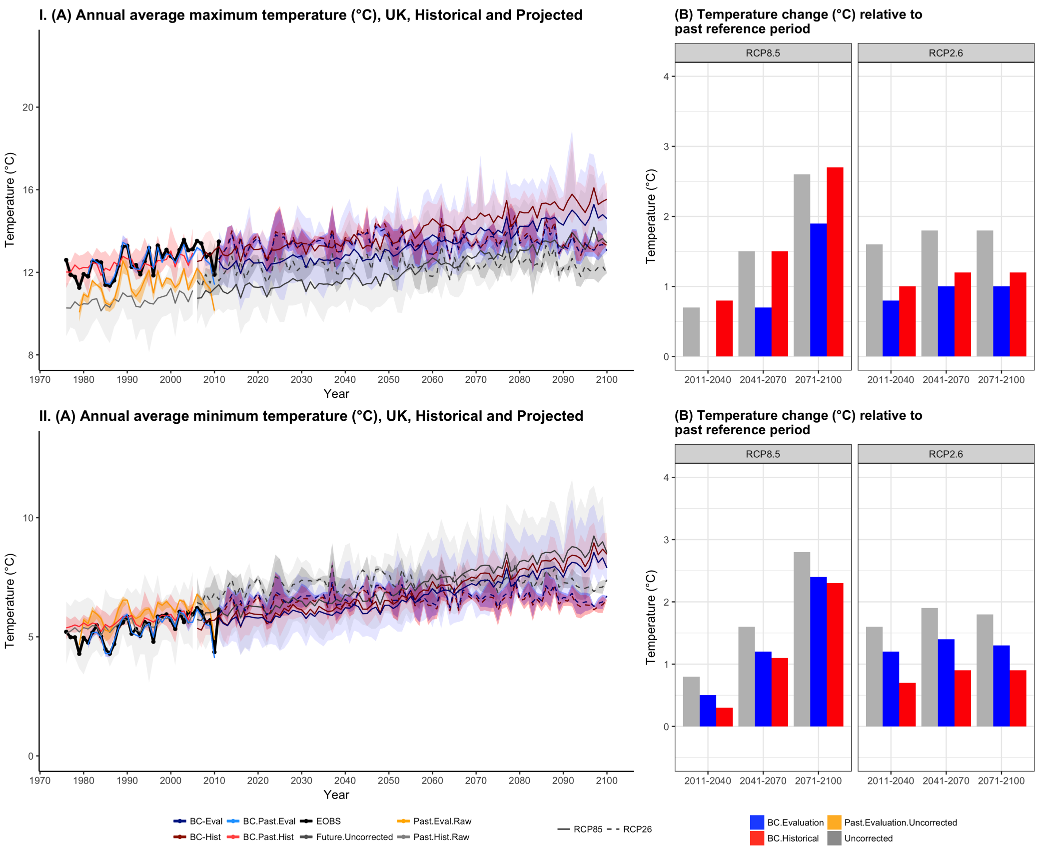



*Figure 6.2: Ensemble GCM-RCM historical and projected changes to annual averages of I. Maximum temperature and II. Minimum temperature for the UK until the end of the century using different calibration for the bias correction (BC-Eval and BC-Hist). I-II (A) shows the range of simulations under RCP8.5 and RCP2.6. I-II (B) shows changes in these climate variables relative to the respective past calibration period.*

*Table 6.2: Summary of ensemble projected temperature changes (UK).*

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **Tmax RCP85 Mean** | 0.7 | 1.5 | 2.6 | +, 0.9* | 0▼ | 0.7▼ | 1.9▼ | +, 0.9* | 0.8▲ | 1.5 | 2.7▲ | +, 0.88* |
| | 6.60 | 14.10 | 24.50 | | 0 | 5.6 | 15.2 | | 6.4 | 12.1 | 21.8 | |
| **Tmax RCP26 Mean** | 2.7 | 2.8 | 2.9 | +, 0.01 | 0.8▼ | 1▼ | 1▼ | +, 0.01 | 0.9▼ | 1.1▼ | 1.2▼ | +, 0.01 |
| | 28.20 | 29.20 | 30.30 | | 6.4 | 8 | 8 | | 7.2 | 8.8 | 9.6 | |
| **Tmin RCP85 Mean** | 0.8 | 1.6 | 2.8 | +, 0.94* | 0.5▼ | 1.2▼ | 2.3▼ | +, 0.93* | 0.3▼ | 1.1▼ | 2.3▼ | +, 0.93* |
| | 14.50 | 29.10 | 50.90 | | 9.3 | 22.4 | 43 | | 5.3 | 19.4 | 40.6 | |
| **Tmin RCP26 Mean** | 2.6 | 2.9 | 2.8 | +, 0.02 | 1.2▼ | 1.4▼ | 1.3▼ | +, 0.02 | 0.6▼ | 0.8▼ | 0.7▼ | +, 0.02 |
| | 57.80 | 64.50 | 62.30 | | 22.4 | 26.1 | 24.3 | | 10.4 | 13.8 | 12.1 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective calibration period. (∗) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

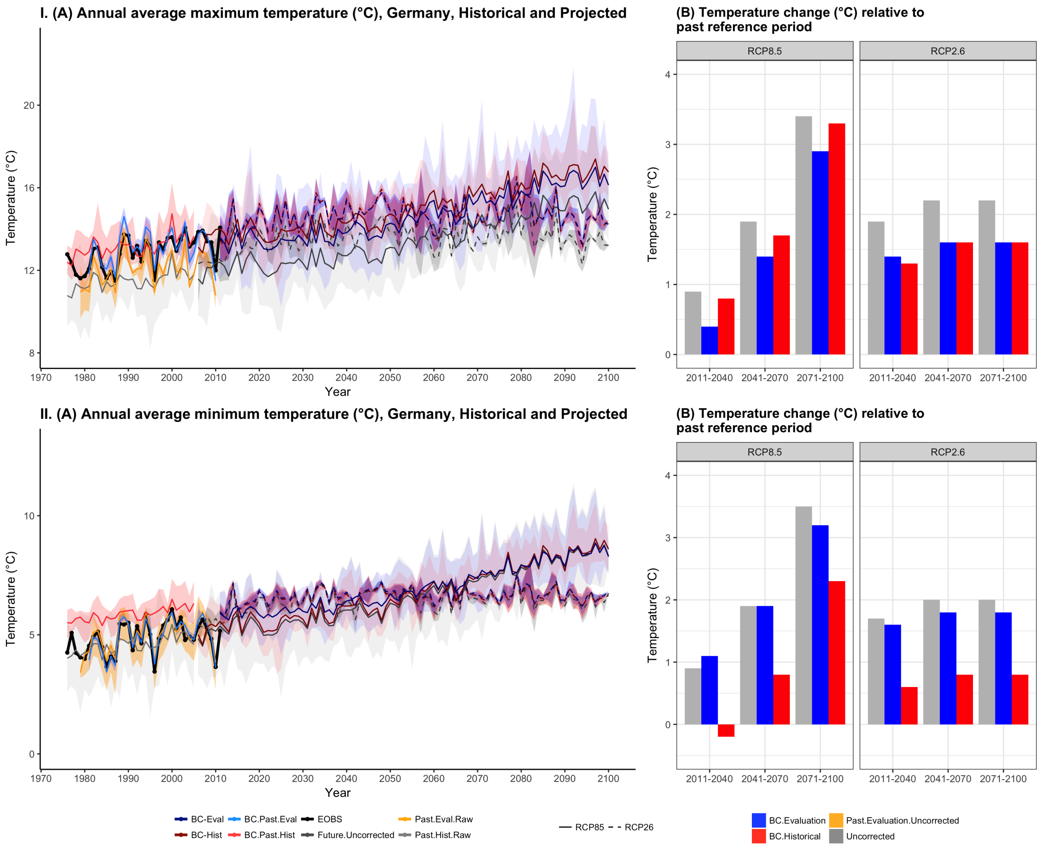



*Figure 6.3: Ensemble BC GCM-RCM historical and projected changes to annual averages of I. Maximum temperature and II. Minimum temperature for Germany until the end of the century using different calibration for the bias correction (BC-Eval and BC-Hist). I-II (A) shows the range of simulations under RCP8.5 and RCP2.6. I-II (B) shows changes in these climate variables relative to the respective past calibration period.*

*Table 6.3: Summary of ensemble projected temperature changes (Germany).*

| | Uncorrected 2011-2040 | Uncorrected 2041-2070 | Uncorrected 2071-2100 | Uncorrected Trend ($R^2$) | BC-Eval 2011-2040 | BC-Eval 2041-2070 | BC-Eval 2071-2100 | BC-Eval Trend ($R^2$) | BC-Hist 2011-2040 | BC-Hist 2041-2070 | BC-Hist 2071-2100 | BC-Hist Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Tmax RCP85 Mean** | 0.9 | 1.9 | 3.4 | +, 0.87* | 0.4▼ | 1.4▼ | 2.9▼ | +, 0.86* | 0.8▼ | 1.7▼ | 3.3▼ | +, 0.85* |
| | 7.80 | 16.50 | 29.60 | | 3.1 | 10.7 | 22.1 | | 6.1 | 12.9 | 25.1 | |
| **Tmax RCP26 Mean** | 3.2 | 3.4 | 3.4 | +, 0.02 | 1.4▼ | 1.6▼ | 1.7▼ | +, 0.01 | 1.3▼ | 1.5▼ | 1.5▼ | +, 0.02 |
| | 31.20 | 33.20 | 33.20 | | 10.7 | 12.2 | 13 | | 9.8 | 11.3 | 11.3 | |
| **Tmin RCP85 Mean** | 0.9 | 1.9 | 3.5 | +, 0.92* | 1.1▲ | 1.9 | 3.2▼ | +, 0.92* | -0.2▼ | 0.8▼ | 2.3▼ | +, 0.91* |
| | 19.70 | 41.60 | 76.60 | | 22.3 | 38.5 | 64.9 | | -3.4 | 13.6 | 39.2 | |
| **Tmin RCP26 Mean** | 2.7 | 3 | 3 | +, 0.06* | 1.6▼ | 1.8▼ | 1.8▼ | +, 0.04* | 0.3▼ | 0.5▼ | 0.4▼ | +, 0.02 |
| | 75.80 | 84.20 | 84.20 | | 32.5 | 36.5 | 36.5 | | 4.8 | 8.1 | 6.5 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective calibration period. (∗) indicates a significant ($p < 0.05$) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

As expected, the difference in the trajectories of Tmax and Tmin is observed to be dependent on the RCP scenario. The RCP2.6 projected changes of temperature start out high – projected changes are approximately 2°C higher relative to the calibration period (here uncorrected historical simulations) – early into the century, but these changes level off by the mid- and late century. For example, for both the UK and Germany, the uncorrected projected change between 2011-2040 is approximately 0.8°C for the ensemble mean of projected Tmax and Tmin under RCP8.5. In contrast, the projected change for the RCP2.6 scenario is larger for the same time period: above 2.7 (2.6)°C for Tmax (Tmin) for the UK and above 3.2 (2.7)°C for Germany under RCP2.6. Towards the mid- and late-century, the projected changes to temperature under RCP2.6 remain the same.

In contrast, it can be observed in the results that projected changes in Tmax and Tmin under RCP8.5 continue to increase over time. Indicating a stabilization of emissions under RCP2.6, the ensemble mean of Tmax and Tmin do not show any significant trends for both countries under the RCP2.6 scenario (Tables 6.2 and 6.3), with the exception of RCA-MPI and the RCP2.6 ensemble mean for Germany Tmax – although $R^2$ values are small ($R^2$<0.06). The ensemble mean and all individual GCM-RCMs show significant positive trends for future projections of annual Tmax and Tmin under the RCP8.5 scenario for the UK and Germany (Tables 6.4 and 6.7).

Projections also show that that the beginning of the uncorrected future climate starts from the end period of uncorrected historical simulations of Tmax and Tmin, which are often below the mean of observations. In Chapter 4, historical simulated temperatures from GCM-RCMs showed significant negative biases relative to observations, meaning that these projected changes toward the end of the century could be significantly underestimating the magnitude and range of future temperatures. For this reason, the use of BC on future climate projections is investigated in the following section.

#### 6.3.1.2 The effect of bias correction on temperature projections

A summary of the effects of BC is that ensemble means of BC-Eval and BC-Hist temperature projections generally show reductions in projected changes after BC, but the effect of BC on individual projections varies. Regardless of these modifications, simulations under both BC calibration approaches continue to project significant future warming. Between BC-Eval and BC-Hist, how big the 'jump' from uncorrected projections also changes: this depends on whether the removal of the RCM bias can cause significant changes to the projection or whether it is the larger error from the joint GCM-RCM choice that contributes more error. These different 'cases' are further explained in the discussion; here, the effect of BC is discussed in more detail.

**(1) Effect of BC on ranges and scenario differences of simulations**

The use of the two differently-calibrated BC approaches – BC-Eval, to correct RCM error and BC-Hist to correct GCM-RCM error – results in projections that 'jump' from the uncorrected range to begin at the same magnitude as the end of the 30-year period of BC past evaluation and BC past historical simulations. This means that projections shift upward compared to the uncorrected projections which are below the observational mean. Both BC-Eval and BC-Hist simulations show increases (upward shifts) in the range of temperatures under both scenarios and for both Tmax and Tmin for the UK and Germany, with significant increasing trends for both the ensemble means and the individual GCM-RCMs.

The range of BC projections relative to the uncorrected projections varies depending on the variable and location. For example, for Tmax, the range of BC-Hist is higher than BC-Eval in the UK (RCP8.5), but not in Germany (Figs. 6.2-I and 6.3-I). For Tmin, the ranges of the ensemble means of BC-Eval and BC-Hist are smaller than uncorrected projections for the UK, but they are relatively similar to each other in Germany (Figs. 6.2-II and 6.3-II).

In terms of differences between scenarios, the RCP8.5 and 2.6 pathways remain distinct even after BC: projected changes in Tmax under RCP8.5 remain low in the early century, but steadily increase towards the end of the century. The RCP2.6 projections project a large positive temperature change, but this again levels off in the mid- and late-century for both BC-Eval and BC-Hist. BC-Hist corrected projections show significant increasing trends for both the ensemble means and the individual GCM-RCMs.

**(2) Effect of BC on projected changes in temperature**

Between raw, BC-Hist and BC-Eval projections, the largest projected changes (relative to the respective calibration period) are typically from the uncorrected projections. BC-Eval and BC-Hist generally reduce the projected changes in Tmax and Tmin in the UK (Figs. 6.2-I and II (B)) and Germany (Figs. 6.3-I and II (B)) for both scenarios. In addition to these smaller projected relative changes after BC, significant trends are only observed for projections forced by RCP8.5 (Tables 6.4-6.7).

In the case of individual GCM-RCMs, the use of BC-Eval results in smaller projected changes relative to the changes from uncorrected projections for all GCM-RCMs apart from RCA-HadGEM. RCA-HadGEM was noted in the previous section as having the largest projected changes among the GCM-RCMs, so this increase due to BC-Eval brings the projected change by the end of the century to 4.1$^{\circ}$C in the UK, for example. BC-Eval Tmin also generally has smaller projected changes compared to uncorrected Tmin, apart from RCA-HadGEM and RCA-MPI. Relative to past BC historical simulations, the projected changes in BC-Hist are 2.7 (2.3)$^{\circ}$C for Tmax (Tmin) under RCP8.5 and 1.2 (1.2)$^{\circ}$C for RCP2.6 by the end of the century for the UK (Tables 6.2-6.5). For Germany, these projected changes are: 3.3 (2.3)$^{\circ}$C for Tmax (Tmin) for RCP8.5 and 1.2 (0.4) for RCP2.6.

*Table 6.4: GCM-RCM annual projected Tmax changes for the UK, in °C and in percentage.*

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **CCLM-MPI** | 0.5 | 1.2 | 2.1 | +, 0.58* | 0▼ | 0.6▼ | 1.3▼ | +, 0.59* | 0.8▲ | 1.5▲ | 2.3▲ | +, 0.56* |
| | 4.70 | 11.20 | 19.50 | | 0 | 4.8 | 10.4 | | 6.5 | 12.1 | 18.6 | |
| **RACMO-ECEARTH** | 0.7 | 1.4 | 2.6 | +, 0.79* | -0.1▼ | 0.7▼ | 2▼ | +, 0.79* | 0.7 | 1.4 | 2.7▲ | +, 0.74* |
| | 6.80 | 13.60 | 25.20 | | -0.8 | 5.6 | 16 | | 5.6 | 11.2 | 21.7 | |
| **RCA-CC** | 0.6 | 1.4 | 2.8 | +, 0.81* | -1.3▼ | -0.5▼ | 0.9▼ | +, 0.81* | 0.5▼ | 1.3▼ | 2.8 | +, 0.79* |
| | 6.30 | 14.60 | 29.20 | | -10.4 | -4 | 7.2 | | 4 | 10.4 | 22.5 | |
| **RCA-HADGEM** | 1.1 | 2.1 | 3.6 | +, 0.71* | 1.6▲ | 2.6▲ | 4.1▲ | +, 0.71* | 0.9▼ | 1.8▼ | 3.3▼ | +, 0.67* |
| | 9.20 | 17.60 | 30.10 | | 12.8 | 20.7 | 32.7 | | 7.1 | 14.3 | 26.2 | |
| **RCA-IPSL** | 0.8 | 1.5 | 2.6 | +, 0.68* | -0.9▼ | -0.1▼ | 1▼ | +, 0.68* | 1▲ | 1.7▲ | 2.8▲ | +, 0.65* |
| | 8.10 | 15.20 | 26.40 | | -7.2 | -0.8 | 8 | | 8.2 | 14 | 23 | |
| **RCA-MPI** | 0.6 | 1.2 | 2.1 | +, 0.59* | 0.4▼ | 1.1▼ | 2▼ | +, 0.59* | 0.7▲ | 1.4▲ | 2.3▲ | +, 0.51* |
| | 5.30 | 10.60 | 18.50 | | 3.2 | 8.8 | 16 | | 5.6 | 11.3 | 18.6 | |
| ***RCP85_Mean*** | 0.7 | 1.5 | 2.6 | +, 0.9* | 0▼ | 0.7▼ | 1.9▼ | +, 0.9* | 0.8▲ | 1.5 | 2.7▲ | +, 0.88* |
| | 6.60 | 14.10 | 24.50 | | 0 | 5.6 | 15.2 | | 6.4 | 12.1 | 21.8 | |
| **RCA-HADGEM_RCP26** | 2.1 | 2.1 | 2.3 | +, 0 | 1.3▼ | 1.4▼ | 1.5▼ | +, 0 | 1.4▼ | 1.5▼ | 1.6▼ | +, 0.02 |
| | 19.50 | 19.50 | 21.40 | | 10.4 | 11.2 | 12 | | 11.3 | 12.1 | 12.9 | |
| **RCA-MPI_RCP26** | 1.4 | 1.6 | 1.6 | +, 0.01 | 0.2▼ | 0.5▼ | 0.4▼ | +, 0.01 | 0.5▼ | 0.9▼ | 0.8▼ | +, 0.01 |
| | 13.60 | 15.50 | 15.50 | | 1.6 | 4 | 3.2 | | 4 | 7.2 | 6.4 | |
| ***RCP26_Mean*** | 2.7 | 2.8 | 2.9 | +, 0.01 | 0.8▼ | 1▼ | 1▼ | +, 0.01 | 0.9▼ | 1.1▼ | 1.2▼ | +, 0.01 |
| | 28.20 | 29.20 | 30.30 | | 6.4 | 8 | 8 | | 7.2 | 8.8 | 9.6 | |

*Table 6.5: GCM-RCM annual projected minimum temperature changes for the UK, in °C and in percentage.*

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **CCLM-MPI** | 0.6 | 1.4 | 2.3 | +, 0.72* | 0.7▲ | 1.3▼ | 2.1▼ | +, 0.72* | 0.4▼ | 1.1▼ | 2.1▼ | +, 0.71* |
| | 10.30 | 24 | 39.40 | | 13.1 | 24.3 | 39.2 | | 7.2 | 19.7 | 37.6 | |
| **RACMO-ECEARTH** | 0.9 | 1.7 | 3 | +, 0.85* | 0.2▼ | 0.9▼ | 2▼ | +, 0.86* | 0.5▼ | 1.2▼ | 2.5▼ | +, 0.85* |
| | 18.70 | 35.30 | 62.40 | | 3.7 | 16.8 | 37.4 | | 8.8 | 21.1 | 44 | |
| **RCA-CC** | 0.7 | 1.6 | 3 | +, 0.86* | -0.5▼ | 0.2▼ | 1.5▼ | +, 0.86* | 0▼ | 0.9▼ | 2.3▼ | +, 0.84* |
| | 15.60 | 35.60 | 66.70 | | -9.3 | 3.7 | 28 | | 0 | 15.5 | 39.7 | |
| **RCA-HADGEM** | 1.1 | 2.1 | 3.7 | +, 0.83* | 1.7▲ | 2.6▲ | 4.1▲ | +, 0.83* | 0.3▼ | 1.3▼ | 2.8▼ | +, 0.79* |
| | 16.80 | 32 | 56.40 | | 31.7 | 48.5 | 76.6 | | 5.1 | 22.3 | 48.1 | |
| **RCA-IPSL** | 0.9 | 1.7 | 2.9 | +, 0.79* | 0▼ | 0.7▼ | 1.8▼ | +, 0.78* | 0.3▼ | 1.1▼ | 2.3▼ | +, 0.79* |
| | 18.10 | 34.20 | 58.40 | | 0 | 13.1 | 33.6 | | 5.3 | 19.6 | 41 | |
| **RCA-MPI** | 0.7 | 1.3 | 2.3 | +, 0.72* | 1▲ | 1.7▲ | 2.6▲ | +, 0.73* | 0.4▼ | 1.1▼ | 2.1▼ | +, 0.68* |
| | 11 | 20.50 | 36.20 | | 18.7 | 31.7 | 48.5 | | 7.2 | 19.9 | 38.1 | |
| ***RCP85_Mean*** | 0.8 | 1.6 | 2.8 | +, 0.94* | 0.5▼ | 1.2▼ | 2.3▼ | +, 0.93* | 0.3▼ | 1.1▼ | 2.3▼ | +, 0.93* |
| | 14.50 | 29.10 | 50.90 | | 9.3 | 22.4 | 43 | | 5.3 | 19.4 | 40.6 | |
| **RCA-HADGEM_RCP26** | 1.7 | 1.9 | 1.9 | +, 0.01 | 1.5▼ | 1.7▼ | 1.7▼ | +, 0.01 | 1.2▼ | 1.4▼ | 1.4▼ | +, 0.01 |
| | 29.10 | 32.60 | 32.60 | | 28 | 31.7 | 31.7 | | 21.5 | 25.1 | 25.1 | |
| **RCA-MPI_RCP26** | 1.9 | 2.2 | 2.1 | +, 0.01 | 0.8▼ | 1▼ | 0.9▼ | +, 0.01 | 0.3▼ | 0.6▼ | 0.5▼ | +, 0.01 |
| | 39.50 | 45.70 | 43.70 | | 14.9 | 18.7 | 16.8 | | 5.3 | 10.6 | 8.8 | |
| ***RCP26_Mean*** | 2.6 | 2.9 | 2.8 | +, 0.02 | 1.2▼ | 1.4▼ | 1.3▼ | +, 0.02 | 0.6▼ | 0.8▼ | 0.7▼ | +, 0.02 |
| | 57.80 | 64.50 | 62.30 | | 22.4 | 26.1 | 24.3 | | 10.4 | 13.8 | 12.1 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective calibration period. (∗) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

*Table 6.6: GCM-RCM annual projected Tmax changes for Germany, in °C and in percentage.*

| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CCLM-MPI** | 0.5 | 1.5 | 2.6 | +, 0.58* | 0.1▼ | 0.9▼ | 2▼ | +, 0.59* | 0.6▲ | 1.5 | 2.6 | +, 0.52* |
| | 4.50 | 13.60 | 23.60 | | 0.8 | 6.9 | 15.3 | | 4.6 | 11.4 | 19.8 | |
| **RACMO-ECEARTH** | 0.8 | 1.4 | 3.1 | +, 0.67* | 0.2▼ | 0.8▼ | 2.5▼ | +, 0.67* | 0.8 | 1.4 | 3.1 | +, 0.61* |
| | 6.90 | 12.10 | 26.90 | | 1.5 | 6.1 | 19.1 | | 6.1 | 10.7 | 23.6 | |
| **RCA-CC** | 0.6 | 1.6 | 3.1 | +, 0.74* | -1.2▼ | -0.3▼ | 1.3▼ | +, 0.73* | 0.5▼ | 1.5▼ | 3▼ | +, 0.7* |
| | 5.90 | 15.60 | 30.20 | | -9.2 | -2.3 | 9.9 | | 3.8 | 11.3 | 22.7 | |
| **RCA-HADGEM** | 1.4 | 2.5 | 4.5 | +, 0.6* | 2.3▲ | 3.4▲ | 5.5▲ | +, 0.6* | 1.2▼ | 2.3▼ | 4.2▼ | +, 0.59* |
| | 10.90 | 19.50 | 35 | | 17.6 | 25.9 | 42 | | 9.1 | 17.5 | 31.9 | |
| **RCA-IPSL** | 1.2 | 2.5 | 4 | +, 0.66* | 0.2▼ | 1.5▼ | 3▼ | +, 0.66* | 0.9▼ | 2.2▼ | 3.6▼ | +, 0.62* |
| | 10.90 | 22.70 | 36.30 | | 1.5 | 11.4 | 22.9 | | 6.8 | 16.7 | 27.3 | |
| **RCA-MPI** | 0.6 | 1.7 | 3.2 | +, 0.58* | 0.8▲ | 1.9▲ | 3.5▲ | +, 0.58* | 0.6 | 1.7 | 3.1▼ | +, 0.51* |
| | 4.90 | 13.80 | 26.10 | | 6.1 | 14.5 | 26.7 | | 4.6 | 13 | 23.7 | |
| ***RCP85_Mean*** | 0.9 | 1.9 | 3.4 | +, 0.87* | 0.4▼ | 1.4▼ | 2.9▼ | +, 0.86* | 0.8▼ | 1.7▼ | 3.3▼ | +, 0.85* |
| | 7.80 | 16.50 | 29.60 | | 3.1 | 10.7 | 22.1 | | 6.1 | 12.9 | 25.1 | |
| **RCA-HADGEM_RCP26** | 3 | 3.1 | 3.2 | +, 0 | 2▼ | 2.1▼ | 2.2▼ | +, 0.01 | 1.8▼ | 1.9▼ | 2▼ | +, 0.02 |
| | 27.30 | 28.20 | 29.10 | | 15.3 | 16 | 16.8 | | 13.7 | 14.5 | 15.3 | |
| **RCA-MPI_RCP26** | 1.3 | 1.7 | 1.7 | +, 0.01 | 0.8▼ | 1.2▼ | 1.1▼ | +, 0.01 | 1▼ | 1.4▼ | 1.3▼ | +, 0.01 |
| | 11.30 | 14.70 | 14.70 | | 6.1 | 9.2 | 8.4 | | 7.6 | 10.7 | 9.9 | |
| ***RCP26_Mean*** | 3.2 | 3.4 | 3.4 | +, 0.02 | 1.4▼ | 1.6▼ | 1.7▼ | +, 0.01 | 1.3▼ | 1.5▼ | 1.5▼ | +, 0.02 |
| | 31.20 | 33.20 | 33.20 | | 10.7 | 12.2 | 13 | | 9.8 | 11.3 | 11.3 | |
| | **Uncorrected** | | | | **BC-Eval** | | | | **BC-Hist** | | | |

*Table 6.7: GCM-RCM annual projected minimum temperature changes for Germany, in °C and in percentage.*

| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CCLM-MPI** | 0.6 | 1.5 | 2.8 | +, 0.71* | 1.1▲ | 1.8▲ | 2.7▼ | +, 0.74* | -0.2▼ | 0.7▼ | 2▼ | +, 0.68* |
| | 11.30 | 28.10 | 52.50 | | 22.3 | 36.5 | 54.8 | | -3.5 | 12.4 | 35.4 | |
| **RACMO-ECEARTH** | 0.9 | 1.7 | 3.4 | +, 0.77* | 1▲ | 1.5▼ | 2.7▼ | +, 0.81* | -0.5▼ | 0.2▼ | 1.9▼ | +, 0.76* |
| | 28.10 | 53.10 | 106.20 | | 20.3 | 30.5 | 54.8 | | -8 | 3.2 | 30.4 | |
| **RCA-CC** | 0.7 | 1.9 | 3.5 | +, 0.81* | 0.1▼ | 0.9▼ | 2.2▼ | +, 0.82* | -0.7▼ | 0.4▼ | 2▼ | +, 0.8* |
| | 19.60 | 53.30 | 98.20 | | 2 | 18.2 | 44.6 | | -11.3 | 6.5 | 32.3 | |
| **RCA-HADGEM** | 1.3 | 2.5 | 4.3 | +, 0.8* | 1.9▲ | 3▲ | 4.7▲ | +, 0.81* | 0.2▼ | 1.3▼ | 3.1▼ | +, 0.73* |
| | 24.10 | 46.30 | 79.60 | | 38.5 | 60.8 | 95.3 | | 3.5 | 22.4 | 53.5 | |
| **RCA-IPSL** | 1.2 | 2.3 | 3.9 | +, 0.83* | 0.9▼ | 1.8▼ | 3.2▼ | +, 0.83* | 0▼ | 1.1▼ | 2.6▼ | +, 0.79* |
| | 27.50 | 52.80 | 89.50 | | 18.2 | 36.5 | 64.9 | | 0 | 19.1 | 45.2 | |
| **RCA-MPI** | 0.7 | 1.7 | 3.2 | +, 0.73* | 1.5▲ | 2.4▲ | 3.7▲ | +, 0.75* | -0.1▼ | 0.9▼ | 2.4▼ | +, 0.67* |
| | 12.60 | 30.70 | 57.80 | | 30.4 | 48.7 | 75 | | -1.8 | 16.2 | 43.2 | |
| ***RCP85_Mean*** | 0.9 | 1.9 | 3.5 | +, 0.92* | 1.1▲ | 1.9 | 3.2▼ | +, 0.92* | -0.2▼ | 0.8▼ | 2.3▼ | +, 0.91* |
| | 19.70 | 41.60 | 76.60 | | 22.3 | 38.5 | 64.9 | | -3.4 | 13.6 | 39.2 | |
| **RCA-HADGEM_RCP26** | 1.2 | 1.3 | 1.4 | +, 0.02 | 1.8▲ | 1.9▲ | 1.9▲ | +, 0.02 | 1.1▼ | 1.3 | 1.3▼ | +, 0.02 |
| | 22.50 | 24.40 | 26.30 | | 36.5 | 38.5 | 38.5 | | 19.5 | 23 | 23 | |
| **RCA-MPI_RCP26** | 2.9 | 3.2 | 3.2 | +, 0.04* | 1.5▼ | 1.7▼ | 1.6▼ | +, 0.01 | 0▼ | 0.2▼ | 0.1▼ | +, 0.01 |
| | 90.60 | 100 | 100 | | 30.4 | 34.5 | 32.4 | | 0 | 3.2 | 1.6 | |
| ***RCP26_Mean*** | 2.7 | 3 | 3 | +, 0.06* | 1.6▼ | 1.8▼ | 1.8▼ | +, 0.04* | 0.3▼ | 0.5▼ | 0.4▼ | +, 0.02 |
| | 75.80 | 84.20 | 84.20 | | 32.5 | 36.5 | 36.5 | | 4.8 | 8.1 | 6.5 | |
| | **Uncorrected** | | | | **BC-Eval** | | | | **BC-Hist** | | | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective calibration period. ($*$) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

#### 6.3.1.3 Regional changes to annual temperatures

In terms of regional projected changes to temperature in Germany, the uncorrected projections from the six GCM-RCMs are shown in Figs. 6.4-6.7 (I-II). The ensemble results are shown in Tables 6.8-6.11. Similar to projected changes at the national level, temperatures are projected to increase until the end of the century, albeit with different pathways depending on the emissions scenario, where stabilization of temperatures occurs under RCP2.6.

Ensemble means of annual Tmax and Tmin show significant trends for RCP8.5 in all regions, but not for RCP2.6. Some exceptions are the ensemble mean for DED and DEF Tmin under RCP2.6, where trends are significantly increasing – however, the $R^2$ values are close to zero. Uncorrected projected changes to temperatures (relative to the uncorrected historical baseline) are, on average, 3.4 (3.6)$^{\circ}$C for all regions for Tmax (Tmin) under RCP8.5; this is 3.4 (3)$^{\circ}$C under RCP2.6 by the end of the century.

After BC, the range and magnitude of projected temperatures also jumps or shifts abruptly upward to begin from where past BC simulations end in 2010. After BC, the increasing trends remain significant for all regions, and the projected changes to regional annual Tmax (relative to the respective calibration period) by the end of the century are 3.2, 2.5, 3, 3$^{\circ}$C for BC-Eval and 3.7, 3.1, 3.4, 3.5$^{\circ}$C for BC-Hist for each region respectively (DE2, DEA, DED, DEF) with the RCP8.5 scenario. Similar to the national level, the effect of BC-Eval and BC-Hist on the ensemble and individual GCM-RCM simulations of temperature depends on how large the biases are in the calibration periods that are used to calibrate the correction: large differences in the bias contributed by the RCM and by the driving GCM lead to differences in the BC-Eval and BC-Hist ranges.

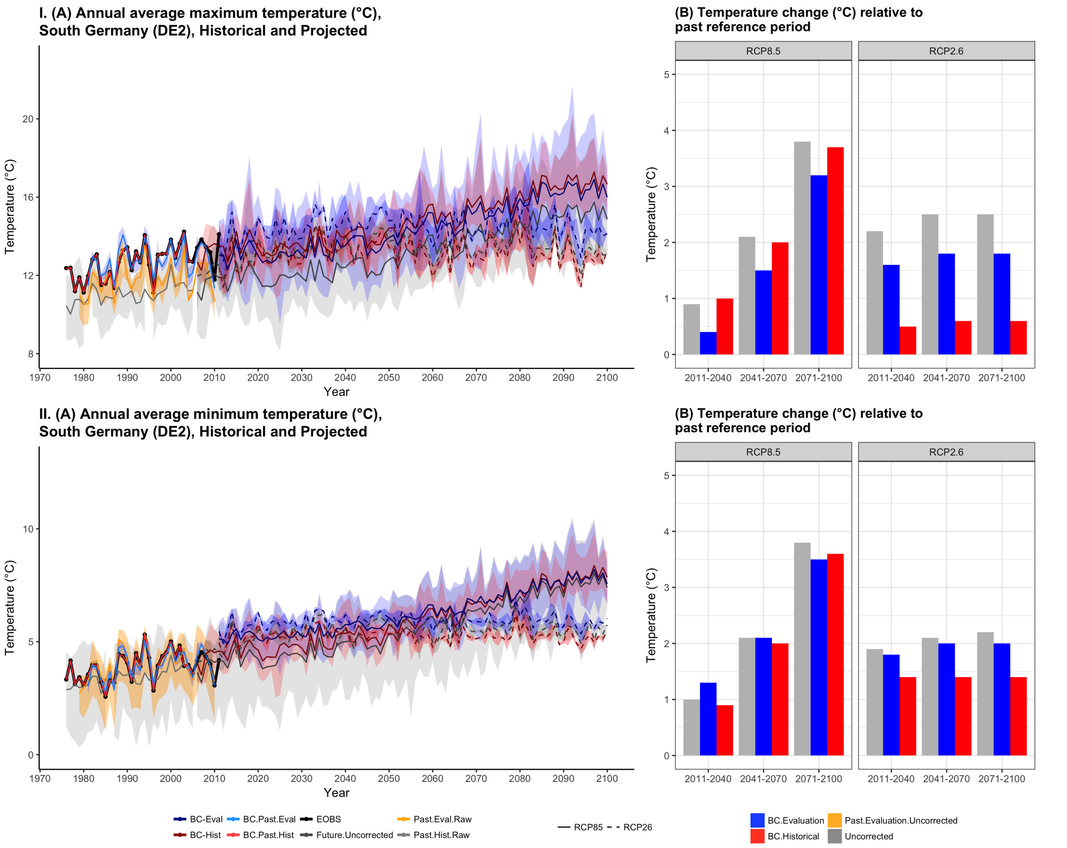



*Figure 6.4: Ensemble BC GCM-RCM historical and projected changes to annual averages of I. Maximum temperature and II. Minimum temperature for South Germany (DE2) until the end of the century using different calibration for the bias correction (BC-Eval and BC-Hist). I-II (A) shows the range of simulations under RCP8.5 and RCP2.6. I-II (B) shows changes in these climate variables relative to the respective past RR.*

*Table 6.8: Summary of ensemble projected temperature changes (DE2).*

| | Uncorrected 2011-2040 | Uncorrected 2041-2070 | Uncorrected 2071-2100 | Uncorrected Trend ($R^2$) | BC-Eval 2011-2040 | BC-Eval 2041-2070 | BC-Eval 2071-2100 | BC-Eval Trend ($R^2$) | BC-Hist 2011-2040 | BC-Hist 2041-2070 | BC-Hist 2071-2100 | BC-Hist Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Tmax RCP85** | 0.9 | 2.1 | 3.8 | +, 0.87* | 0.4▼ | 1.5▼ | 3.2▼ | +, 0.86* | 1▲ | 2▼ | 3.7▼ | +, 0.85* |
| | 8.20 | 19 | 34.40 | | 3.1 | 11.8 | 25.1 | | 7.9 | 15.9 | 29.4 | |
| **Tmax RCP26** | 3.6 | 3.9 | 3.9 | +, 0.02 | 1.6▼ | 1.8▼ | 1.8▼ | +, 0.01 | 0.5▼ | 0.6▼ | 0.6▼ | +, 0.01 |
| | 37.20 | 40.30 | 40.30 | | 12.5 | 14.1 | 14.1 | | 4 | 4.8 | 4.8 | |
| **Tmin RCP85** | 1 | 2.1 | 3.8 | +, 0.92* | 1.3▲ | 2.1 | 3.5▼ | +, 0.92* | 0.9▼ | 2▼ | 3.6▼ | +, 0.91* |
| | 29.40 | 61.70 | 111.60 | | 32.7 | 52.7 | 87.9 | | 23.2 | 51.6 | 92.9 | |
| **Tmin RCP26** | 3 | 3.2 | 3.3 | +, 0.07* | 1.8▼ | 2▼ | 2▼ | +, 0.02 | 1.4▼ | 1.4▼ | 1.4▼ | -, -0.01 |
| | 129.80 | 138.50 | 142.80 | | 45.2 | 50.2 | 50.2 | | 36.1 | 36.1 | 36.1 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective RR. ($*$) indicates a significant ($p < 0.05$) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

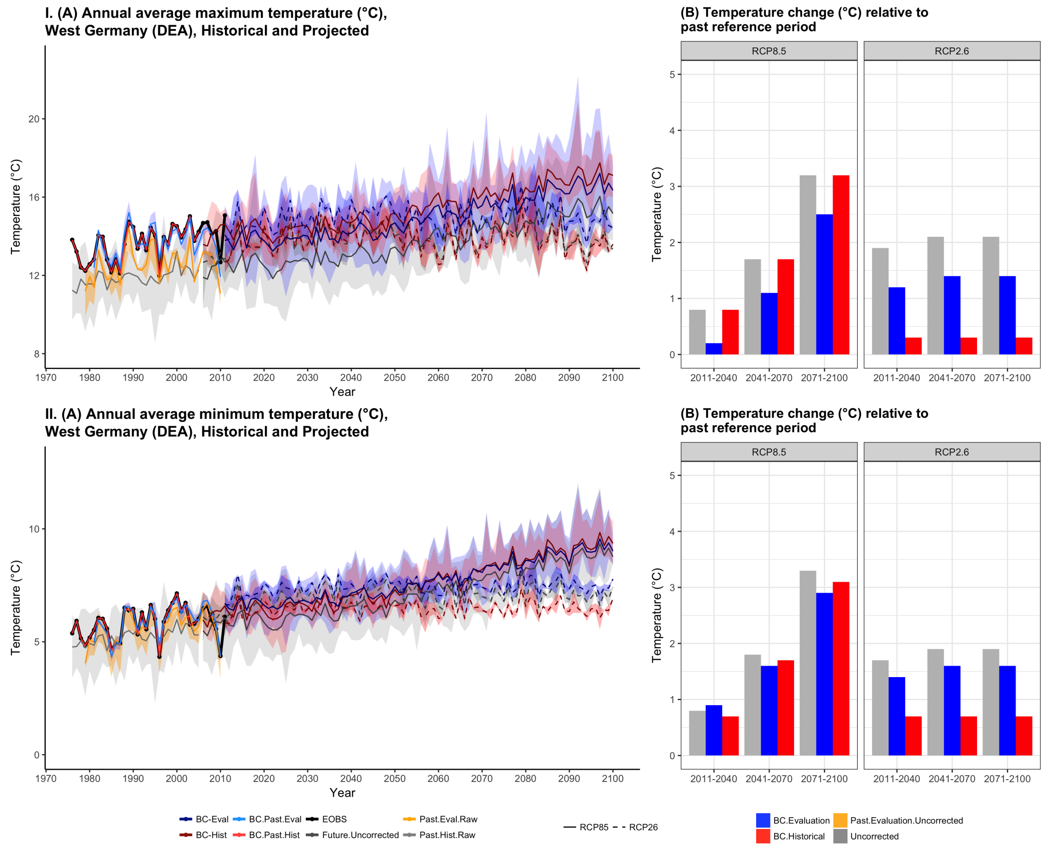



*Figure 6.5: Ensemble BC GCM-RCM historical and projected changes to annual averages of I. Maximum temperature and II. Minimum temperature for West Germany (DEA) until the end of the century using different calibration for the bias correction (BC-Eval and BC-Hist). I-II (A) shows the range of simulations under RCP8.5 and RCP2.6. I-II (B) shows changes in these climate variables relative to the respective past RR.*

*Table 6.9: Summary of ensemble projected temperature changes (DEA).*

| | Uncorrected 2011-2040 | Uncorrected 2041-2070 | Uncorrected 2071-2100 | Uncorrected Trend ($R^2$) | BC-Eval 2011-2040 | BC-Eval 2041-2070 | BC-Eval 2071-2100 | BC-Eval Trend ($R^2$) | BC-Hist 2011-2040 | BC-Hist 2041-2070 | BC-Hist 2071-2100 | BC-Hist Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Tmax RCP85** | 0.8 | 1.7 | 3.2 | +, 0.85* | 0.2▼ | 1.1▼ | 2.5▼ | +, 0.86* | 0.8 | 1.7 | 3.1▼ | +, 0.84* |
| | 6.70 | 14.30 | 26.90 | | 1.5 | 8 | 18.3 | | 5.9 | 12.5 | 22.9 | |
| **Tmax RCP26** | 3 | 3.2 | 3.3 | +, 0.01 | 1.2▼ | 1.4▼ | 1.4▼ | +, 0.01 | 0.3▼ | 0.3▼ | 0.3▼ | +, 0.01 |
| | 28 | 29.90 | 30.80 | | 8.8 | 10.2 | 10.2 | | 2.2 | 2.2 | 2.2 | |
| **Tmin RCP85** | 0.8 | 1.8 | 3.3 | +, 0.91* | 0.9▲ | 1.6▼ | 2.9▼ | +, 0.91* | 0.7▼ | 1.7▼ | 3.1▼ | +, 0.9* |
| | 15.30 | 34.40 | 63.10 | | 15.2 | 27.1 | 49.1 | | 12 | 29.2 | 53.3 | |
| **Tmin RCP26** | 2.6 | 2.8 | 2.9 | +, 0.05* | 1.4▼ | 1.6▼ | 1.6▼ | +, 0.04* | 0.7▼ | 0.7▼ | 0.7▼ | +, -0.01 |
| | 60.90 | 65.60 | 67.90 | | 23.7 | 27.1 | 27.1 | | 12 | 12 | 12 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective RR. ($*$) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

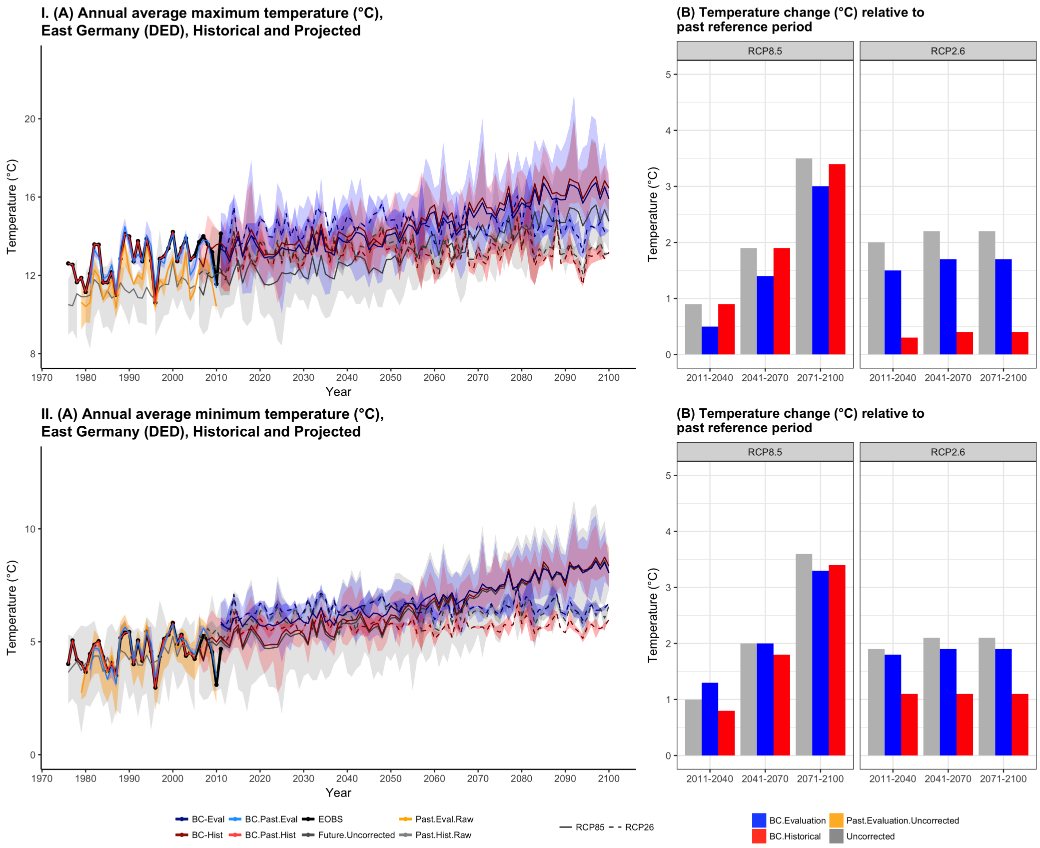



*Figure 6.6: Ensemble BC GCM-RCM historical and projected changes to annual averages of I. Maximum temperature and II. Minimum temperature for East Germany (DED) until the end of the century using different calibration for the bias correction (BC-Eval and BC-Hist). I-II (A) shows the range of simulations under RCP8.5 and RCP2.6. I-II (B) shows changes in these climate variables relative to the respective past RR.*

*Table 6.10: Summary of ensemble projected temperature changes (DED).*

| | Uncorrected 2011-2040 | Uncorrected 2041-2070 | Uncorrected 2071-2100 | Uncorrected Trend ($R^2$) | BC-Eval 2011-2040 | BC-Eval 2041-2070 | BC-Eval 2071-2100 | BC-Eval Trend ($R^2$) | BC-Hist 2011-2040 | BC-Hist 2041-2070 | BC-Hist 2071-2100 | BC-Hist Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Tmax RCP85** | 0.9 | 1.9 | 3.5 | +, 0.85* | 0.5▼ | 1.4▼ | 3▼ | +, 0.85* | 0.9 | 1.9 | 3.4▼ | +, 0.84* |
| | 8 | 16.90 | 31.10 | | 3.9 | 10.9 | 23.3 | | 7.1 | 14.9 | 26.7 | |
| **Tmax RCP26** | 3.3 | 3.5 | 3.5 | +, 0.01 | 1.5▼ | 1.7▼ | 1.7▼ | +, 0.01 | 0.3▼ | 0.4▼ | 0.4▼ | +, -0.01 |
| | 33.10 | 35.10 | 35.10 | | 11.6 | 13.2 | 13.2 | | 2.4 | 3.1 | 3.1 | |
| **Tmin RCP85** | 1 | 2 | 3.6 | +, 0.92* | 1.3▲ | 2 | 3.2▼ | +, 0.92* | 0.8▼ | 1.8▼ | 3.4▼ | +, 0.9* |
| | 23.50 | 46.90 | 84.40 | | 27.9 | 42.9 | 68.7 | | 17.5 | 39.3 | 74.3 | |
| **Tmin RCP26** | 2.9 | 3.1 | 3.1 | +, 0.05* | 1.8▼ | 1.9▼ | 1.9▼ | +, 0.02 | 1.1▼ | 1.1▼ | 1.1▼ | +, -0.01 |
| | 89.1 | 95.3 | 95.3 | | 38.6 | 40.8 | 40.8 | | 24.1 | 24.1 | 24.1 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective RR. (∗) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

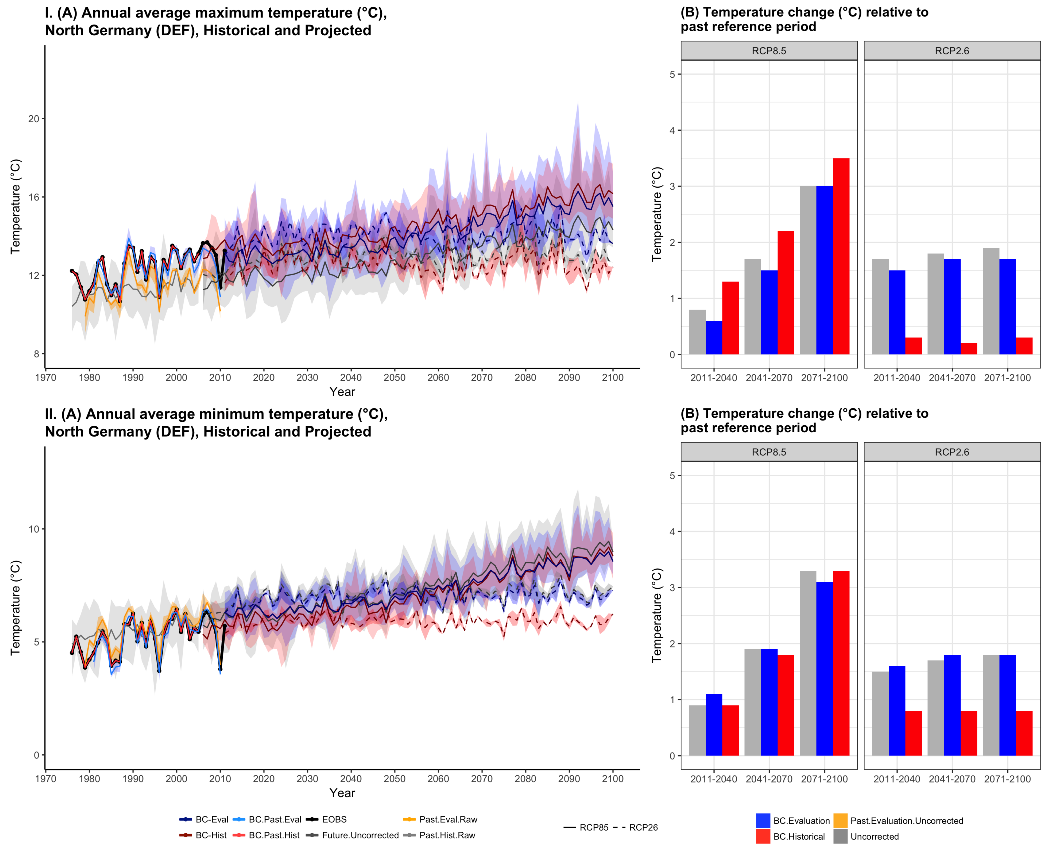



*Figure 6.7: Ensemble BC GCM-RCM historical and projected changes to annual averages of I. Maximum temperature and II. Minimum temperature for North Germany (DEF) until the end of the century using different calibration for the bias correction (BC-Eval and BC-Hist). I-II (A) shows the range of simulations under RCP8.5 and RCP2.6. I-II (B) shows changes in these climate variables relative to the respective past RR.*

*Table 6.11: Summary of ensemble projected temperature changes (DEF).*

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **Tmax RCP85** | 0.8 | 1.7 | 3 | +, 0.87* | 0.6▼ | 1.5▼ | 3 | +, 0.87* | 1.3▲ | 2.2▲ | 3.5▲ | +, 0.84* |
| | 7.10 | 15.10 | 26.70 | | 4.8 | 12.1 | 24.1 | | 10.6 | 17.9 | 28.4 | |
| **Tmax RCP26** | 2.5 | 2.7 | 2.8 | +, 0.02 | 1.5▼ | 1.7▼ | 1.8▼ | +, 0.01 | 0.3▼ | 0.2▼ | 0.3▼ | -, -0.01 |
| | 24.10 | 26 | 27 | | 12.1 | 13.7 | 14.5 | | 2.4 | 1.6 | 2.4 | |
| **Tmin RCP85** | 0.9 | 1.9 | 3.3 | +, 0.9* | 1.1▲ | 1.9 | 3.1▼ | +, 0.92* | 0.9 | 1.8▼ | 3.2▼ | +, 0.9* |
| | 16.40 | 34.70 | 60.30 | | 20.9 | 36.1 | 58.8 | | 17.5 | 35 | 62.1 | |
| **Tmin RCP26** | 2.4 | 2.6 | 2.6 | +, 0.05* | 1.6▼ | 1.8▼ | 1.8▼ | +, 0.04* | 0.8▼ | 0.8▼ | 0.8▼ | +, -0.01 |
| | 51.90 | 56.20 | 56.20 | | 30.4 | 34.2 | 34.2 | | 15.5 | 15.5 | 15.5 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective RR. ($*$) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

### 6.3.2 Changes to total annual precipitation

#### 6.3.2.1 Uncorrected projections of annual precipitation

Future precipitation from GCM-RCMs shows some positive trends over the UK and Germany. For example, over the UK, three of the total six GCM-RCMs show significant trends: RCP8.5 RCA-CC, RCA-IPSL, and RCA-MPI, and the ensemble mean for RCP8.5. For Germany, RCP8.5 RACMO-ECEARTH, RCA-CC, RCA-IPSL, and the RCP8.5 ensemble mean show significant increasing tends. Under RCP2.6 both uncorrected RCA-MPI and RCA-HadGEM show significant increasing trends but the ensemble mean does not for the UK, while under RCP2.6, no significant trends for uncorrected precipitation are observed for Germany (Tables 6.12, 6.13). Unlike temperature, precipitation projections do not show strong divergence between RCP8.5 and RCP2.6 emission scenarios (Figs. 6.8, 6.9).

The past historical simulations of precipitation have a large positive bias over the period 1976-2005 relative to E-OBS data (observations) of annual precipitation. This large bias is retained into the future, as projections remain approximately 200mm away from the mean of observations for both the UK and Germany (Figs. 6.8, 6.9). The largest projected precipitation change over the UK and Germany is from RCA-CC (RCP8.5), which projects a 160mm increase by the end of the century relative to the mean of past uncorrected historical precipitation for both countries; however, other RCP8.5 GCMs project much smaller changes which are typically under 100mm by the end of the century.

Under RCP2.6, projected precipitation changes show large differences between RCA-HadGEM and RCA-MPI. RCP2.6 RCA-HadGEM projects large reductions in total annual precipitation for Germany, approximately 200mm, while RCA-MPI projects increases over 160mm on average in all three 30-year future time periods. In contrast, in the UK, RCA-HadGEM projects small positive increases under 100mm, but RCA-MPI projects large increases in total annual precipitation, over 300mm on average for each 30-year future period. Due to these large contrasts in the value of projected changes, the ensemble mean between these two GCM-RCMs available for RCP2.6 does not show any significant trend for either country. As the number of GCM-RCMs under RCP2.6 is limited, having more members in the ensemble could produce more robust results.

#### 6.3.2.2 The effect of BC on precipitation projections

After BC, it can be observed that precipitation projections generally shift ('jump') downward, closer to the mean of observations and their BC RR (evaluation simulations for BC-Eval and historical simulations for BC-Hist) in both the UK and Germany. This shift is anticipated as the use of BC, in this case quantile-quantile mapping, reduces the large positive biases to bring simulations closer to the mean and distribution of the RR used for calibration, as it was shown in Chapter 5 with RCM simulations.

In the following paragraphs, more specific changes to the range and projected precipitation changes are discussed:

**(1) Effect of BC on ranges and scenario differences of simulations**

In terms of trends, after BC, the significant increasing linear trends are retained for RCP8.5: RCA-CC, RCA-IPSL, RCA-MPI and the RCP8.5 mean for the UK. In Germany, the positive trend of RACMO-ECEARTH is lost for both BC-Eval and BC-Hist. RCA-CC is one of the only GCM-RCMs that has a significant ($p$<0.05) linear increasing trend for precipitation in the UK and in Germany ($R^2$=0.16 and 0.18, respectively) across uncorrected and BC projections.

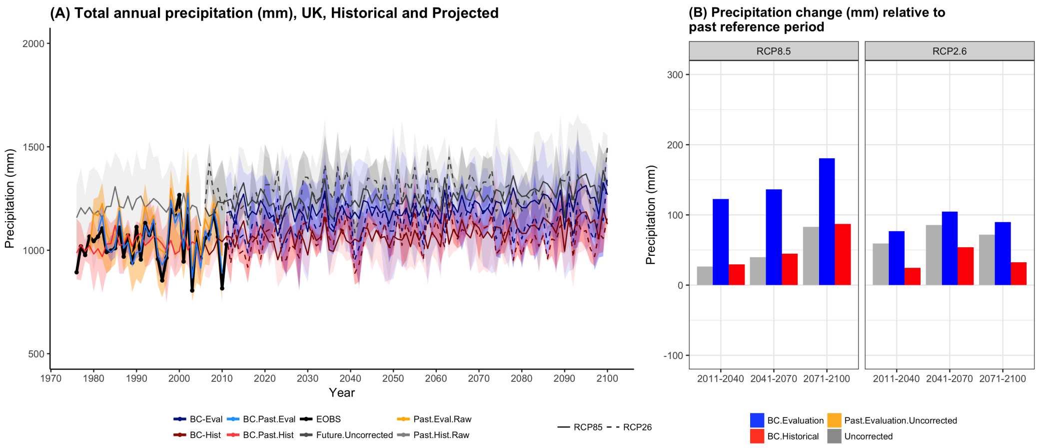



*Figure 6.8: Ensemble BC GCM-RCM historical and projected changes to total annual precipitation for the UK until the end of the century using different calibration for the bias correction (BC-Eval and BC-Hist). (A) shows the range of simulations under RCP8.5 and RCP2.6. (B) shows changes in these climate variables relative to the respective past RR.*

*Table 6.12: GCM-RCM annual projected precipitation changes for the UK, in mm and in percentage.*

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **CCLM-MPI** | -4.4 | 18 | 45.5 | +, 0.01 | 268.1▲ | 281.5▲ | 298.9▲ | +, -0.01 | -3.1▲ | 18.8▲ | 44.5▼ | +, 0.02 |
| | -0.4 | 1.6 | 4 | | 25.5 | 26.8 | 28.4 | | -0.3 | 1.8 | 4.3 | |
| **RACMO-ECEARTH** | 19.6 | 5.8 | 30.2 | +, 0 | 20▲ | 4.3▼ | 28.8▼ | +, 0.01 | 19.2▼ | 4.8▼ | 30.7▲ | +, 0.01 |
| | 1.9 | 0.6 | 2.9 | | 1.9 | 0.4 | 2.7 | | 1.9 | 0.5 | 3 | |
| **RCA-CC** | 43.3 | 106.9 | 159.7 | +, 0.21* | 125.5▼ | 190.8▲ | 245.5▲ | +, 0.23* | 52.7▲ | 117.5▲ | 172▲ | +, 0.26* |
| | 3.40 | 8.40 | 12.50 | | 11.9 | 18.1 | 23.3 | | 5.1 | 11.5 | 16.8 | |
| **RCA-HADGEM** | 58 | 38.7 | 70.4 | +, 0.01 | 2.6▼ | -12.7▼ | 23.1▼ | +, 0.01 | 65.2▲ | 49.7▲ | 84.4▲ | +, 0.01 |
| | 5.10 | 3.40 | 6.20 | | 0.2 | -1.2 | 2.2 | | 6.4 | 4.9 | 8.3 | |
| **RCA-IPSL** | 34.6 | 31.6 | 108.6 | +, 0.1* | 152.8▲ | 153.5▲ | 234.2▲ | +, 0.12* | 26.4▼ | 31.1▼ | 106▼ | +, 0.15* |
| | 2.60 | 2.40 | 8.30 | | 14.5 | 14.6 | 22.3 | | 2.6 | 3 | 10.3 | |
| **RCA-MPI** | 8.6 | 38.9 | 80.4 | +, 0.07* | 172.8▲ | 206.2▲ | 250.9▲ | +, 0.09* | 15.9▼ | 46.1▲ | 84.7▲ | +, 0.1* |
| | 0.60 | 2.90 | 5.90 | | 16.4 | 19.6 | 23.9 | | 1.5 | 4.5 | 8.2 | |
| ***RCP85_Mean*** | 26.6 | 40 | 82.4 | +, 0.19* | 122.8▲ | 136.4▲ | 179.3▲ | +, 0.19* | 29.4▲ | 44.7▲ | 86.8▲ | +, 0.23* |
| | 2.20 | 3.30 | 6.80 | | 11.7 | 12.9 | 17 | | 2.9 | 4.4 | 8.4 | + |
| **RCA-HADGEM_RCP26** | 44 | 96.5 | 42.1 | +, 0.1* | -11.5▼ | 42.4▼ | -11.6▼ | +, 0.12* | 20▼ | 76.1▼ | 20.3▼ | +, 0.15* |
| | 3.90 | 8.50 | 3.70 | | -1.1 | 4 | -1.1 | | 1.9 | 7.3 | 1.9 | |
| **RCA-MPI_RCP26** | 320.3 | 320.3 | 336 | +, 0.19* | 168.9▼ | 170.2▼ | 184▼ | +, 0.19* | 8.2▲ | 11▼ | 21.5▼ | +, 0.23* |
| | 30.80 | 30.80 | 32.30 | | 16.1 | 16.2 | 17.5 | | 0.8 | 1.1 | 2.1 | |
| ***RCP26_Mean*** | -11.3 | 15 | -6.5 | +, 0.02 | 77▲ | 104.6▲ | 82.4▲ | +, 0.01 | 27.9▲ | 57.4▲ | 32.9▲ | +, 0.01 |
| | -0.9 | 1.2 | -0.5 | | 7.3 | 9.9 | 7.8 | | 2.7 | 5.6 | 3.2 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective RR. (∗) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

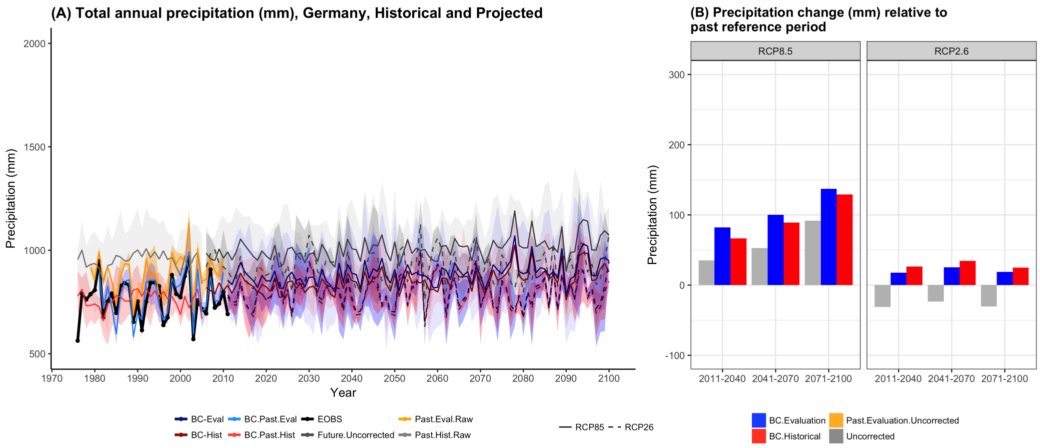



*Figure 6.9: Ensemble BC GCM-RCM historical and projected changes to total annual precipitation for Germany until the end of the century using different calibration for the bias correction (BC-Eval and BC-Hist). (A) shows the range of simulations under RCP8.5 and RCP2.6. (B) shows changes in these climate variables relative to the respective past RR.*

*Table 6.13: GCM-RCM annual projected precipitation changes for Germany, in mm and in percentage.*

| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CCLM-MPI** | -4.7 | 8 | 39.9 | +, 0.01 | 225.8▲ | 223.2▲ | 245.4▲ | +, -0.01 | 83.6▲ | 101.8▲ | 131.4▲ | +, 0.03 |
| | -0.40 | 0.80 | 3.70 | | 29 | 28.6 | 31.5 | | 10.7 | 13 | 16.7 | |
| **RACMO-ECEARTH** | 91.3 | 103.8 | 120 | +, 0.15* | -36.5▼ | -7.9▼ | 3.2▲ | +, 0.01 | 89.6▼ | 109.4▲ | 126.6▲ | +, 0.01 |
| | 11.10 | 12.60 | 14.60 | | -4.7 | -1 | 0.4 | | 11.7 | 14.3 | 16.5 | |
| **RCA-CC** | 43.1 | 81.2 | 154.9 | +, 0.15* | 140.8▲ | 177.2▲ | 249.3▲ | +, 0.16* | 73.6▲ | 106.3▲ | 183.2▲ | +, 0.18* |
| | 4.30 | 8 | 15.30 | | 18 | 22.6 | 31.9 | | 9.8 | 14.2 | 24.4 | |
| **RCA-HADGEM** | 30.2 | 60.4 | 81.2 | +, 0.01 | -59▼ | -28.1▼ | -4.6▼ | +, 0.03* | 59.8▲ | 93.6▲ | 122.7▲ | +, 0.04* |
| | 3.80 | 7.50 | 10.10 | | -7.5 | -3.6 | -0.6 | | 7.9 | 12.4 | 16.3 | |
| **RCA-IPSL** | 22.1 | 28.4 | 78.5 | +, 0.08* | 118.5▲ | 127.1▲ | 177.4▲ | +, 0.1* | 45.3▲ | 62.4▲ | 115.9▲ | +, 0.14* |
| | 2.20 | 2.80 | 7.80 | | 15.1 | 16.2 | 22.7 | | 6 | 8.3 | 15.4 | |
| **RCA-MPI** | 30 | 34.7 | 69.8 | +, 0.01 | 102.4▲ | 108.7▲ | 144▲ | +, -0.01 | 47.6▲ | 60▲ | 94.3▲ | +, 0.03 |
| | 3.10 | 3.50 | 7.10 | | 13.1 | 13.9 | 18.4 | | 6.3 | 7.9 | 12.4 | |
| ***RCP85_Mean*** | 35.3 | 52.7 | 90.6 | +, 0.17* | 82.2▲ | 100.3▲ | 135.9▲ | +, 0.16* | 66.6▲ | 88.9▲ | 128.9▲ | +, 0.24* |
| | 3.70 | 5.60 | 9.50 | | 10.5 | 12.8 | 17.4 | | 8.8 | 11.7 | 16.9 | |
| **RCA-HADGEM_RCP26** | -230 | -223.2 | -227.5 | +, 0.01 | -59.6▲ | -52.3▲ | -56▲ | +, 0.01 | 6.3▲ | 14.1▲ | 9.9▲ | +, 0.01 |
| | -21.60 | -20.90 | -21.30 | | -7.6 | -6.7 | -7.2 | | 0.8 | 1.8 | 1.3 | |
| **RCA-MPI_RCP26** | 175 | 183.2 | 166.2 | +, 0.01 | 93.7▼ | 102▼ | 85.1▼ | +, 0.01 | 17.5▼ | 25.5▼ | 8.9▼ | +, 0.01 |
| | 21.20 | 22.20 | 20.20 | | 12 | 13 | 10.9 | | 2.3 | 3.3 | 1.2 | |
| ***RCP26_Mean*** | -93.5 | -86 | -98.1 | +, 0.01 | 17.5▲ | 25.3▲ | 13.6▲ | +, -0.01 | 36.5▲ | 44.4▲ | 32.8▲ | +, -0.01 |
| | -9.20 | -8.50 | -9.70 | | 2.2 | 3.2 | 1.7 | | 4.9 | 5.9 | 4.4 | |
| | **Uncorrected** | | | | **BC-Eval** | | | | **BC-Hist** | | | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective RR. (∗) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

Significant positive trends are also observed for RCA-MPI and RCA-HadGEM, but not the ensemble mean, for RCP2.6 for both BC-Eval and BC-Hist in the UK only; RCP2.6 precipitation projections still show no trends for Germany. In the UK, the use of BC-Eval and BC-Hist both shift the ensemble mean of projections of total annual precipitation downward relative to the uncorrected projections (Fig. 6.8A). For Germany, BC-Eval and BC-Hist precipitation projections are also shifted downward relative to uncorrected projections, as a result of reducing the positive biases based on the calibration period (Fig. 6.9A).

Similar to uncorrected projections, BC-Eval and BC-Hist precipitation do not show any clear differences in trajectories between RCP8.5 and RCP2.6. Projected changes after BC under RCP2.6 are generally smaller than RCP8.5 for the UK and Germany, at least for the two GCM-RCMs which were available for the comparison. For example, in Germany BC-Hist RCP2.6 RCA-MPI projects relative increases of 21.5mm by the end of the century, which is lower compared to uncorrected projections (336mm) and BC-Eval (180mm) for the same scenario, and to the projected decreases of 85mm under RCP8.5.

**(2) Effect of BC on projected changes in precipitation**

In terms of projected changes, the ensemble mean of BC-Eval simulations shows the largest projected precipitation changes compared to the uncorrected and BC-Hist changes (Figs. 6.8B, 6.9B). For example, the RCP8.5 BC-Eval mean projects 180mm increases by the end of the century versus 80mm from both uncorrected and RCP8.5 BC-Hist projections for the UK. For Germany, the RCP8.5 BC-Eval ensemble projected change is 136mm, compared to BC-Hist (129mm) and uncorrected projections (91mm). For RCP2.6, BC-Eval ensemble is also greater than BC-Hist and uncorrected projections, but for the UK only.

In terms of individual climate models, all six GCM-RCM BC-Eval and BC-Hist simulations for RCP8.5 project precipitation increases in the UK. The largest projected BC-Eval increases are from CCLM-MPI, with over 270mm projected changes (a roughly 25% increase) relative to the RR for the UK. The smallest projected changes are from BC-Eval RACMO-ECEARTH and RCA-HadGEM with just under 30mm (UK) and 5mm (Germany) projected changes by the end of the century (Tables 6.12 and 6.13). Most BC-Eval GCM-RCMs also project increases in rainfall for Germany, apart from RACMO-ECEARTH (RCP8.5) and RCA-HadGEM (both scenarios), which project decreases in total annual precipitation.

BC-Hist precipitation changes from individual GCM-RCMs are all positive changes for the UK and Germany across both emission scenarios. In the UK, all individual BC-Hist GCM-RCMs show projected precipitation changes under 100mm by the end of the century under both scenarios, apart from RCP8.5 BC-Hist RCA-CC which projects a larger change of 172mm relative to the RR. In Germany BC-Hist RCA-MPI (RCP2.6) projects relative increases of 21.5mm by the end of the century, which is lower compared to uncorrected projections (336mm) and BC-Eval (180mm) for the same scenario. In contrast, BC-Hist RCA-MPI under RCP8.5 at the higher emissions scenario projects larger increases of 85mm over Germany. The highest projected precipitation changes by the end of the century are from RCA-CC (180mm) for RCP8.5 for Germany.

In the following section, regional precipitation changes for Germany are discussed, followed by the projected changes in summer climate indices.

#### 6.3.2.3 Projected changes to regional precipitation

At the German regional level, there are significant positive trends for future projected precipitation in all four regions (Table 6.14, Fig. 6.10) under RCP8.5. This means that in general, increasing annual precipitation can be expected in the four German regions: the projected changes are all under 120mm for the ensemble means for RCP8.5: 70mm (DE2), 80mm (DEA), 70mm (DED) and 120mm (DEF).

The shift after BC relative to the uncorrected projections depends on how close the evaluation and historical simulations are initially: for example, in DE2, DEA, and DEF (South, West, and North Germany respectively), while uncorrected past historical simulations have a positive bias relative to observations, i.e. the uncorrected GCM-RCMs were too 'wet' ($RMSE$=176mm on average, see Chapter 4 for historical analysis), the difference between the mean of historical simulations and observations is relatively small (Fig. 6.10 I, II, IV). The difference between uncorrected evaluation simulations and observations is also relatively small for the chosen RCMs.

*Table 6.14: GCM-RCM annual projected precipitation changes for German regions, in mm and in percentage.*

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **DE2 RCP85** | 16.4 | 37.7 | 69.1 | +, 0.1* | 109.8▲ | 133▲ | 166.6▲ | +, 0.11* | 25.4▲ | 53.3▲ | 89.7▲ | +, 0.15* |
| | 1.60 | 3.80 | 6.90 | | 11.9 | 14.4 | 18 | | 2.8 | 5.9 | 9.9 | |
| **DE2 RCP26** | -128.5 | -126.3 | -132.9 | +, 0.01 | 12▲ | 15.8▲ | 7.8▲ | +, 0.01 | -9.8▲ | -7▲ | -8.8▲ | +, 0.01 |
| | -12.90 | -12.70 | -13.40 | | 1.3 | 1.7 | 0.8 | | -1.1 | -0.8 | -1 | |
| **DEA RCP85** | 12.2 | 36.4 | 80 | +, 0.16* | 78.9▲ | 98.8▲ | 137▲ | +, 0.13* | 29.2▲ | 58.4▲ | 107.6▲ | +, 0.22* |
| | 1.20 | 3.70 | 8.10 | | 8.9 | 11.2 | 15.5 | | 3.4 | 6.9 | 12.7 | |
| **DEA RCP26** | -37.7 | -33.3 | -54.7 | +, -0.01 | 11.3▲ | 15.9▲ | -5▲ | +, -0.01 | 11▲ | 11.9▲ | 7.7▲ | +, 0 |
| | -3.70 | -3.30 | -5.40 | | 1.3 | 1.8 | -0.6 | | 1.3 | 1.4 | 0.9 | |
| **DED RCP85** | 19.8 | 40.8 | 69.2 | +, 0.11* | 77.9▲ | 94.3▲ | 117.2▼ | +, 0.1* | 39.9▲ | 62.8▲ | 93.5▲ | +, 0.17* |
| | 2.20 | 4.60 | 7.80 | | 11.8 | 14.3 | 17.8 | | 6.2 | 9.8 | 14.5 | |
| **DED RCP26** | -131.1 | -125.6 | -140.9 | -, -0.01 | 21▲ | 25.9▲ | 13▲ | -, -0.01 | 11.3▲ | 12.2▲ | 12.9▲ | +, -0.01 |
| | -12.90 | -12.30 | -13.90 | | 3.2 | 3.9 | 2 | | 1.8 | 1.9 | 2 | |
| **DEF RCP85** | 38.2 | 64.8 | 120.7 | +, 0.25* | 100.8▲ | 124.4▲ | 175.8▲ | +, 0.23* | 46.8▲ | 68▲ | 127.8▲ | +, 0.24* |
| | 4.10 | 7 | 13.10 | | 12.4 | 15.3 | 21.6 | | 6 | 8.7 | 16.3 | |
| **DEF RCP26** | -21.4 | -4.8 | -20.2 | +, -0.01 | 33.8▲ | 48.4▲ | 34.7▲ | +, -0.01 | 16.9▲ | 7.9▲ | 11.9▲ | +, -0.01 |
| | -2.20 | -0.50 | -2 | | 4.2 | 5.9 | 4.3 | | 2.2 | 1 | 1.5 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective RR. ($*$) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

However, in contrast, the positive bias between historical GCM-RCM simulations and observations in in DED (East Germany) is significantly larger ($RMSE$=270mm). Therefore, after BC, it can be observed that the shift from uncorrected to BC projections in DE2, DEA, and DEF is relatively small, and uncorrected projections, BC-Eval and BC-Hist remain close together. In contrast, after BC-Eval and BC-Hist in DED, BC projections (both BC-Hist and BC-Eval) of total annual precipitation make a significant downward shift compared to other regions (Fig. 6.10 III).

After BC, projected changes in precipitation increase under BC-Eval and BC-Hist in all regions for RCP8.5, apart from DEF under RCP2.6 (Fig. 6.10 I-IV (B)). Changes in BC-Eval are all above 100mm under RCP8.5: 166mm (DE2), 137mm (DEA), 117mm (DED), and 176mm (DEF) by the end of the century. Under RCP2.6, these changes are significantly smaller: 8mm (DE2), -5 mm (DEA), 13mm (DED), 35mm (DEF) for BC-Eval by the end of the century.

Ensemble projected changes in precipitation with the BC-Hist calibration are smaller relative to BC-Eval under both emission scenarios, but are all comparatively greater than uncorrected projected changes (Table 6.14 (B)). Significant increasing trends in precipitation are observed in DE2, DEA, DED, and DEF but only under RCP8.5.

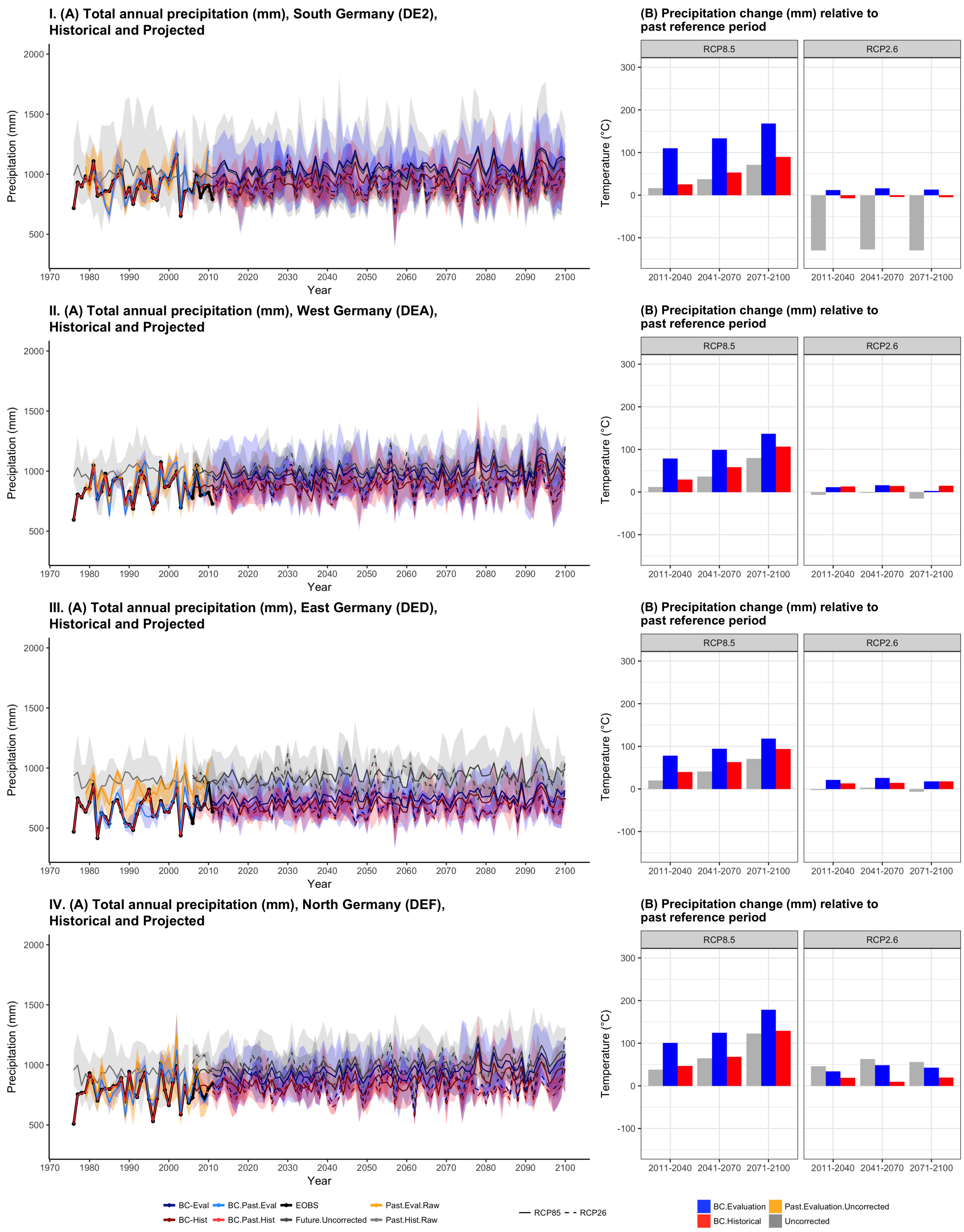



*Figure 6.10: Regional ensemble BC GCM-RCM historical and projected changes to total annual precipitation for I. South (DE2), II. West (DEA), III. East (DED), and IV. North (DEF) Germany until the end of the century using different calibration for the bias correction (BC-Eval and BC-Hist). (A) shows the range of simulations under RCP8.5 and RCP2.6. (B) shows changes in these climate variables relative to the respective past RR.*

### 6.3.3 National-level changes to summer (JJA) climate indices

The number of days above 31°C ('hot days') and total precipitation during the summer months of June-August are important climate indices which aim to represent the climate component of crop heat stress, as discussed in previous chapters (See Chapter 1 and the SCCM evaluation in Chapter 3). These summer (JJA) climate indices are calculated from daily future Tmax and precipitation simulations for the UK (Fig. 6.11), Germany (Fig. 6.12), and four German regions (Figs. 6.13) and 6.14).

#### 6.3.3.1 Changes in the number of days above 31°C

Because the indices are averaged over the entire country, daily average temperatures of above 31°C over the entire UK were not typically observed in the climate analysis (Chapter 3), and this is also observed in the climate projections. The changes to hot days are low in uncorrected projections, BC-Eval and BC-Hist for the UK: projected increases are typically less than one day per summer by the end of the century for uncorrected and BC-Eval projections. BC-Hist GCM-RCMs project comparatively more hot days over the UK by the end of the century, but these remain under 2 days per summer.

For example, RCA-HadGEM typically projects the highest increases in the number of hot days under RCP8.5: 1 (uncorrected), 1.1 (BC-Eval) and 1.7 (BC-Hist). Some trends from GCM-RCMs are unable to be computed because of zero values in the past and in the future (Table 6.15). Significant ($p$<0.05) trends are observed for the UK, but they are generally have small $R^2$ values, for example uncorrected RCP8.5 and RCP2.6 ensemble means ($R^2$=0.05 and $R^2$=0.03), BC-Eval RCA-HadGEM ($R^2$=0.05) and BC-Hist RCA-HadGEM ($R^2$=0.09).

In contrast to the projections of little to no changes over the UK, more hot days are projected for Germany: the ensemble mean of uncorrected projections under RCP8.5 projects 6.6 more days per summer while RCP2.6 projects 4.2 more days by the end of the century (Table 6.17). Generally, RCP2.6 projects fewer hot days compared to RCP8.5 for the GCM-RCMs RCA-HadGEM and RCA-MPI. RCP2.6 and RCP8.5 projections are observed to follow the same trajectories for regional Tmax, where projections from RCP2.6 are higher in the early century but projected changes level off while those under RCP8.5 continue to increase for Germany.

After BC, the number of projected hot days increases for Germany. For example, the BC-Eval ensemble mean projects 9.3 (6.2) more hot days under RCP8.5 (RCP2.6). For BC-Hist, this is 9.3 (10.6) more hot days under RCP8.5 (RCP2.6). The number of projected hot days is highest using BC-Hist, followed by BC-Eval and then uncorrected projections (Fig. 6.12 I (B)) by the end of the century under both scenarios for Germany.

#### 6.3.3.2 Summer (JJA) precipitation

For JJA precipitation, climate projections generally show negative trends, in addition to projected decreases. For example, the ensemble RCP8.5 mean of uncorrected JJA precipitation projects decreases under 30mm, along with a significant negative trend for the UK ($R^2$=0.16). Trends in uncorrected projections of precipitation are negative for the UK based on simulations from CCLM-MPI, RACMO-ECEARTH, RCA-IPSL, and RCA-MPI; however, while significant, the $R^2$ values are relatively small ($R^2$< 0.16 for uncorrected RCP8.5 projections (Table 6.12). Over Germany, reductions of 20 (85)mm are projected under RCP8.5 (RCP2.6), and with significant negative trends for the RCP8.5 ensemble mean ($R^2$=0.06). Projections from CCLM-MPI, RCA-IPSL, RCA-MPI and the RCP8.5 ensemble mean have negative trends ($R^2$<0.14), but RCP8.5 RCA-CC has an increasing trend ($R^2$=0.08).

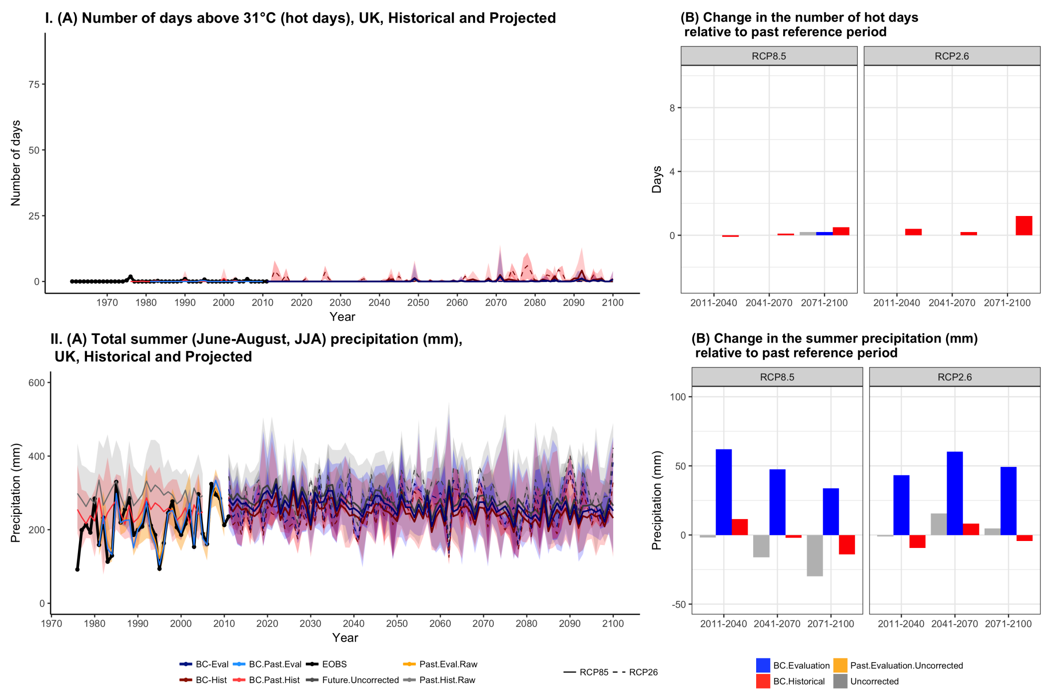



*Figure 6.11: Ensemble BC GCM-RCM historical and projected changes to I. the number of days above 31°C in JJA, and II. Total JJA precipitation (mm) for the UK until the end of the century using BC-Eval and BC-Hist. I-II (A) shows the range of simulations under RCP8.5 and RCP2.6. I-II (B) shows changes in these climate indices relative to the respective RR.*

*Table 6.15: GCM-RCM projected changes to the number of hot days for the UK, in number of days and percentage.*

| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CCLM-MPI** | 0 | 0 | 0.07 | +, 0.01 | 0 | 0 | 0▼ | - | 0 | 0.27▲ | 0.37▲ | +, 0.05* |
| | - | - | - | | - | - | - | | - | - | - | |
| **RACMO-ECEARTH** | 0 | 0 | 0 | - | 0 | 0 | 0.23▲ | +,0.04* | -0.27▼ | -0.27▼ | -0.03▼ | +,0.05* |
| | - | - | - | | - | - | - | | -101.25 | -101.25 | -11.25 | |
| **RCA-CC** | 0 | 0 | 0 | - | -0.07▼ | -0.07▼ | -0.07▼ | - | 0 | 0 | 0 | - |
| | - | - | - | | -105 | -105 | -105 | | - | - | - | |
| **RCA-HADGEM** | 0 | 0.27 | 0.97 | +, 0.05* | -0.07▼ | 0.37▲ | 1.07▲ | +, 0.05* | -0.3▼ | 0.57▲ | 1.74▲ | +, 0.09* |
| | - | - | - | | -105 | 555 | 1605 | | -90 | 171 | 522 | |
| **RCA-IPSL** | 0 | 0 | 0.03 | +, 0.02 | -0.07▼ | -0.07▼ | -0.03▼ | +, 0.02 | 0 | 0.03▲ | 0.3▲ | +, 0.04* |
| | - | - | - | | -105 | -105 | -45 | | - | - | - | |
| **RCA-MPI** | 0 | 0 | 0.13 | +, 0.01 | -0.07▼ | -0.07▼ | 0.1▼ | +, 0.01 | -0.03▼ | 0.07▲ | 0.97▲ | +, 0.13* |
| | - | - | - | | -105 | -105 | 150 | | -90 | 210 | 2910 | |
| ***RCP85_Mean*** | 0 | 0.04 | 0.2 | +, 0.08* | -0.02▼ | 0.05▲ | 0.22▲ | +, 0.08* | -0.1▼ | 0.11▲ | 0.54▲ | +, 0.16* |
| | - | - | - | | -90 | 225 | 990 | | -94.74 | 104.21 | 511.58 | |
| **RCA-HADGEM_RCP26** | 0 | 0 | 0.07 | +, 0.03* | -0.07▼ | -0.07▼ | 0▼ | +, 0.03* | 0.57▲ | 0.33▲ | 1.86▲ | +, 0.01 |
| | - | - | - | | -105 | -105 | 0 | | - | - | - | |
| **RCA-MPI_RCP26** | 0 | 0 | 0 | - | -0.07▼ | -0.07▼ | -0.03▼ | +, 0 | 0.27▲ | 0.03▲ | 0.6▲ | +, 0.01 |
| | - | - | - | | -105 | -105 | -45 | | 101.25 | 11.25 | 225 | |
| ***RCP26_Mean*** | 0 | 0 | 0.03 | +, 0.03* | -0.02▼ | -0.02▼ | 0.03 | +, 0.04* | 0.55▲ | 0.32▲ | 1.38▲ | +, -0.01 |
| | - | - | - | | -90 | -90 | 135 | | - | - | - | |
| | **Uncorrected** | | | | **BC-Eval** | | | | **BC-Hist** | | | |

Table 6.16: GCM-RCM projected JJA precipitation changes for the UK, in mm and in percentage.

| | Uncorrected 2011-2040 | Uncorrected 2041-2070 | Uncorrected 2071-2100 | Uncorrected Trend ($R^2$) | BC-Eval 2011-2040 | BC-Eval 2041-2070 | BC-Eval 2071-2100 | BC-Eval Trend ($R^2$) | BC-Hist 2011-2040 | BC-Hist 2041-2070 | BC-Hist 2071-2100 | BC-Hist Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CCLM-MPI** | -11.5 | -33.8 | -48.1 | -, 0.1* | 90.1▲ | 64.3▲ | 49▲ | -, 0.13* | -10▲ | -30.8▲ | -43.6▲ | -, 0.1* |
| | -4.4 | -12.8 | -18.2 | | 42.2 | 30.1 | 22.9 | | -4.1 | -12.6 | -17.9 | |
| **RACMO-ECEARTH** | -1.2 | -16.5 | -30.3 | -, 0.07* | 12.7▲ | -4.7▲ | -20.6▲ | -, 0.08* | 12.1▲ | -4.2▲ | -18.8▲ | -, 0.07* |
| | -0.5 | -7.2 | -13.2 | | 5.9 | -2.2 | -9.5 | | 5.5 | -1.9 | -8.5 | |
| **RCA-CC** | 18.7 | 27.3 | 40.6 | +, 0.01 | 132.7▲ | 142.3▲ | 155.4▲ | +, 0.02 | 31.8▲ | 38.6▲ | 54.1▲ | +, 0.02 |
| | 5.1 | 7.5 | 11.1 | | 61.8 | 66.3 | 72.4 | | 10.5 | 12.7 | 17.8 | |
| **RCA-HADGEM** | 3.7 | -8.9 | -28 | -, 0.03 | 12.3▲ | 0.3▲ | -17.6▲ | -, 0.03 | 19.3▲ | 5.1▲ | -11▲ | -, 0.03 |
| | 1.4 | -3.5 | -11 | | 5.7 | 0.1 | -8.2 | | 8.5 | 2.3 | -4.9 | |
| **RCA-IPSL** | -4.2 | -18.3 | -43.2 | -, 0.09* | 43.3▲ | 30.4▲ | 6.2▲ | -, 0.09* | 13.8▲ | 1.3▲ | -21.5▲ | -, 0.09* |
| | -1.4 | -6.1 | -14.5 | | 20.2 | 14.2 | 2.9 | | 6.1 | 0.6 | -9.5 | |
| **RCA-MPI** | -17.2 | -46.9 | -71.4 | -, 0.11* | 80.8▲ | 52.5▲ | 28.6▲ | -, 0.11* | 2.5▲ | -22.2▲ | -43.7▲ | -, 0.11* |
| | -4.9 | -13.5 | -20.5 | | 37.6 | 24.4 | 13.3 | | 0.9 | -8.4 | -16.6 | |
| ***RCP85_Mean*** | -1.9 | -16.2 | -29.9 | -, 0.16* | 62▲ | 47.5▲ | 33.7▲ | -, 0.17* | 11.6▲ | -2.1▲ | -13.9▲ | -, 0.16* |
| | -0.6 | -5.5 | -10.2 | | 28.9 | 22.1 | 15.7 | | 4.7 | -0.8 | -5.6 | |
| **RCA-HADGEM_RCP26** | -2.9 | 20 | -1.3 | +, 0 | 12.4▲ | 35.8▲ | 14.4▲ | +, 0 | -12.2▼ | 12▼ | -10.1▼ | +, 0 |
| | -1.1 | 7.6 | -0.5 | | 5.8 | 16.7 | 6.7 | | -5 | 4.9 | -4.1 | |
| **RCA-MPI_RCP26** | 94.5 | 104.6 | 99 | +, -0.01 | 73.9▼ | 84.8▼ | 78.7▲ | +, -0.01 | 24.1▼ | 35.2▼ | 29▼ | +, -0.01 |
| | 41.2 | 45.6 | 43.2 | | 34.4 | 39.5 | 36.6 | | 11 | 16 | 13.2 | |
| ***RCP26_Mean*** | -73.2 | -56.7 | -71.7 | +, 0 | 43.2▲ | 60.3▲ | 45▲ | +, 0 | -65.3▲ | -47.6▲ | -63.3▲ | +, 0 |
| | -20 | -15.5 | -19.6 | | 20.1 | 28.1 | 21 | | -21.5 | -15.7 | -20.9 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective RR. ($*$) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

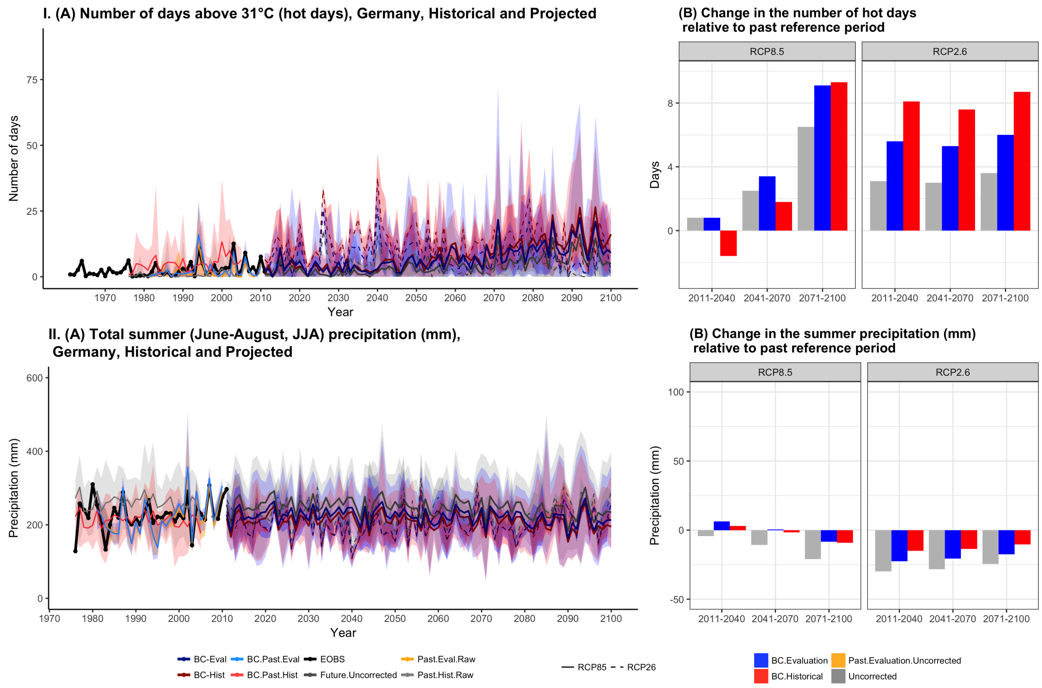



*Figure 6.12: Ensemble BC GCM-RCM historical and projected changes to I. the number of days above 31°C in JJA, and II. Total JJA precipitation (mm) for the UK until the end of the century using BC-Eval and BC-Hist. I-II (A) shows the range of simulations under RCP8.5 and RCP2.6. I-II (B) shows changes in these climate indices relative to the respective RR.*

*Table 6.17: GCM-RCM projected changes to the number of hot days for Germany, in number of days and percentage.*

| | Uncorrected 2011-2040 | Uncorrected 2041-2070 | Uncorrected 2071-2100 | Uncorrected Trend ($R^2$) | BC-Eval 2011-2040 | BC-Eval 2041-2070 | BC-Eval 2071-2100 | BC-Eval Trend ($R^2$) | BC-Hist 2011-2040 | BC-Hist 2041-2070 | BC-Hist 2071-2100 | BC-Hist Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CCLM-MPI** | 0 | 0.4 | 1.9 | +, 0.16* | -2.2▼ | -1.8▼ | -0.3▼ | +, 0.17* | 1.5▲ | 4.1▲ | 9.1▲ | +, 0.31* |
| | 0 | 400 | 1900 | | -97.1 | -79.4 | -13.2 | | 125 | 341.7 | 758.3 | |
| **RACMO-ECEARTH** | -0.1 | 0.7 | 2 | +, 0.11* | -1.7▲ | -0.1▼ | 3.1▲ | +, 0.22* | -1.3▲ | 1.6▲ | 7.2▲ | +, 0.28* |
| | -18.8 | 131.2 | 375 | | -66.2 | -3.9 | 120.8 | | -32.8 | 40.3 | 181.5 | |
| **RCA-CC** | -0.2 | 0 | 0.8 | +, 0.05* | -2.1▲ | -1.8▼ | 0▼ | +, 0.08* | -1.4▲ | -1▼ | 3.3▲ | +, 0.18* |
| | -54.5 | 0 | 218.2 | | -72.4 | -62.1 | 0 | | -43.7 | -31.2 | 103.1 | |
| **RCA-HADGEM** | 2.8 | 8 | 18.2 | +, 0.25* | 8.4▲ | 14.5▼ | 28▲ | +, 0.25* | -1.3▼ | 4.2▼ | 15.5▼ | +, 0.29* |
| | 89.4 | 255.3 | 580.9 | | 289.7 | 500 | 965.5 | | -21.1 | 68.1 | 251.4 | |
| **RCA-IPSL** | 1.4 | 2.8 | 7.1 | +, 0.09* | 0.7▼ | 4.2▲ | 10.8▼ | +, 0.17* | -5.5▼ | -0.3▼ | 8.9▲ | +, 0.26* |
| | 280 | 560 | 1420 | | 24.1 | 144.8 | 372.4 | | -58.1 | -3.2 | 94 | |
| **RCA-MPI** | 0.8 | 3 | 9.6 | +, 0.25* | 0.6▼ | 4.5▲ | 13.3▼ | +, 0.29* | -1.7▼ | 2.3▼ | 12.3▼ | +, 0.32* |
| | 400 | 1500 | 4800 | | 20.7 | 155.2 | 458.6 | | -31.1 | 42.1 | 225 | |
| ***RCP85_Mean*** | 0.8 | 2.5 | 6.6 | +, 0.41* | 0.8 | 3.4▲ | 9.3▲ | +, 0.48* | -1.6▼ | 1.8▼ | 9.3▲ | +, 0.59* |
| | 99.3 | 310.3 | 819.3 | | 31 | 131.9 | 360.8 | | -32.6 | 36.7 | 189.4 | |
| **RCA-HADGEM_RCP26** | 6 | 5.6 | 6.1 | +, -0.01 | 9.1▲ | 8.3▲ | 8.8▲ | +, -0.01 | 13.3▼ | 12.7▼ | 13.1▼ | +, 0.01 |
| | 6000 | 5600 | 6100 | | 313.8 | 286.2 | 303.4 | | 1108.3 | 1058.3 | 1091.7 | |
| **RCA-MPI_RCP26** | 1.2 | 1.5 | 2.2 | +, -0.01 | 1.4▲ | 1.7▲ | 2.8▲ | +, -0.01 | 7.4▲ | 7.3▲ | 9.2▲ | +, -0.01 |
| | 225 | 281.2 | 412.5 | | 48.3 | 58.6 | 96.6 | | 186.6 | 184 | 231.9 | |
| ***RCP26_Mean*** | 3.5 | 3.5 | 4.2 | +, -0.01 | 5.6▲ | 5.3▲ | 6.2▲ | +, -0.01 | 9.8▲ | 9.4▲ | 10.6▼ | +, -0.01 |
| | 954.5 | 954.5 | 1145.5 | | 217.2 | 205.6 | 240.5 | | 306.2 | 293.8 | 331.2 | |

*Table 6.18: GCM-RCM projected JJA precipitation changes for Germany, in mm and in percentage.*

| | Uncorrected 2011-2040 | Uncorrected 2041-2070 | Uncorrected 2071-2100 | Uncorrected Trend ($R^2$) | BC-Eval 2011-2040 | BC-Eval 2041-2070 | BC-Eval 2071-2100 | BC-Eval Trend ($R^2$) | BC-Hist 2011-2040 | BC-Hist 2041-2070 | BC-Hist 2071-2100 | BC-Hist Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CCLM-MPI** | -9.4 | -30 | -53.5 | -, 0.14* | 55.5▲ | 36.7▲ | 14.2▲ | -, 0.18* | -7.4▲ | -23.3▲ | -40.7▲ | -, 0.13* |
| | -3.1 | -10 | -17.8 | | 24.8 | 16.4 | 6.3 | | -3.2 | -10.2 | -17.9 | |
| **RACMO-ECEARTH** | -1.8 | -3 | -20.2 | -, 0.01 | -9▲ | -10▼ | -26.2▼ | -, 0.01 | 5.3▲ | 4.8▲ | -12.6▲ | -, 0 |
| | -0.8 | -1.3 | -8.5 | | -4 | -4.5 | -11.7 | | 2.4 | 2.2 | -5.7 | |
| **RCA-CC** | 9.2 | 8.3 | 44.4 | +, 0.08* | 69.4▼ | 68.6▼ | 103.9▼ | +, 0.08* | 14.9▼ | 12.6▼ | 50.4▲ | +, 0.1* |
| | 2.9 | 2.6 | 13.9 | | 31 | 30.6 | 46.4 | | 5.9 | 5 | 19.9 | |
| **RCA-HADGEM** | -8.4 | 1.6 | -11.5 | -, -0.01 | -67.8▼ | -57.9▼ | -69.1▲ | -, -0.01 | -0.8▼ | 9.3▲ | -2.7▲ | -, -0.01 |
| | -4.4 | 0.8 | -6 | | -30.3 | -25.9 | -30.9 | | -0.5 | 5.2 | -1.5 | |
| **RCA-IPSL** | -19.1 | -21 | -42.8 | -, 0.04* | -35.4▲ | -36.6▲ | -56.7▲ | -, 0.04* | -5.6▲ | -5.5▲ | -24.2▼ | -, 0.03 |
| | -8.1 | -8.9 | -18.1 | | -15.8 | -16.3 | -25.3 | | -3.1 | -3.1 | -13.6 | |
| **RCA-MPI** | 3.7 | -20.6 | -43.9 | -, 0.08* | 24.6▼ | 2.1▲ | -19.2▼ | -, 0.08* | 11.6▼ | -7.2▲ | -25.8▼ | -, 0.06* |
| | 1.3 | -7.5 | -15.9 | | 11 | 0.9 | -8.6 | | 5.3 | -3.3 | -11.8 | |
| ***RCP85_Mean*** | -4.3 | -10.8 | -20.8 | -, 0.06* | 6.3▲ | 0.5▲ | -8.4▲ | -, 0.05* | 3▲ | -1.6▲ | -8.9▲ | -, 0.03 |
| | -1.7 | -4.2 | -8 | | 2.8 | 0.2 | -3.8 | | 1.4 | -0.8 | -4.2 | |
| **RCA-HADGEM_RCP26** | -102.5 | -108.4 | -87.8 | +, 0.01 | -53.9▲ | -58.9▲ | -39.7▲ | +, 0.01 | -41.7▲ | -47.2▲ | -26.4▲ | +, 0.01 |
| | -34.2 | -36.1 | -29.3 | | -24.1 | -26.3 | -17.7 | | -18.3 | -20.7 | -11.6 | |
| **RCA-MPI_RCP26** | 25.7 | 34.7 | 19 | -, -0.01 | 8.7▼ | 17.4▼ | 1.9▼ | -, -0.01 | -12.7▼ | -4.2▼ | -19.4▼ | -, -0.01 |
| | 10.8 | 14.6 | 8 | | 3.9 | 7.8 | 0.8 | | -5.7 | -1.9 | -8.7 | |
| ***RCP26_Mean*** | -88.9 | -87.4 | -85.5 | +, -0.01 | -22.5▲ | -20.6▲ | -19.3▲ | +, -0.01 | -55.5▲ | -54▲ | -51.7▲ | +, -0.01 |
| | -27.9 | -27.4 | -26.8 | | -10.1 | -9.2 | -8.6 | | -21.9 | -21.3 | -20.4 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective RR. ($*$) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

After BC, negative JJA precipitation trends remain significant for CCLM-MPI, RACMO-ECEARTH, RCA-IPSL, RCA-MPI, and the RCP8.5 ensemble for both BC-Eval and BC-Hist. RCP2.6 JJA precipitation simulations do not have any significant linear trends for the UK. For Germany, CCLM-MPI, RCA-CC, RCA-IPSL, RCA-MPI and the RCP8.5 ensemble retain their significant linear trends for BC-Eval and BC-Hist, apart from BC-Hist RCA-IPSL (Table 6.13). After BC, the relative projected changes over Germany for JJA precipitation are smaller than uncorrected projections under both scenarios (Fig. 6.12 II (B)). In contrast, BC projections for UK JJA precipitation show large projected increases using BC-Eval for both scenarios (Fig. 6.11 II (B)).

Overall, therefore, there are only small projected increases in the number of hot days over the entire UK, but there are more over Germany under both emission scenarios. In terms of JJA precipitation, the ensemble means of GCM-RCM simulations over the UK and Germany show negative trends.

### 6.3.4 Regional changes to summer (JJA) climate indices

#### 6.3.4.1 Changes in the number of days above 31°C

At the regional level, the magnitude of projected changes to the number of hot days and total JJA precipitation is observed to depend greatly on the region: for example, based on the climate analysis in Chapter 3, South Germany (DE2) has a comparatively warmer average climate than the more northern state of DEF. As may be expected, DE2 therefore has the largest projected changes in the number of hot days compared to other states, 11.3 (7.1) days based on the uncorrected RCP8.5 (RCP2.6) ensemble mean (Table 6.19). This is followed by eastern Germany (DED) and western Germany (DEA) with 7.5 (5.4) and 6.4 (4.7) more hot days under the RCP8.5 (RCP2.6) scenario. DEF is the coolest region and has projected increases in the number of hot days by only 2 (1.4) days by the end of the century under RCP8.5 (RCP2.6) emissions scenario.

The difference in the scenario forcing can be observed in the projected changes to hot days: again, the projections from RCP8.5 start small but continuously increase until the end of the century; in contrast, the number of projected hot days under RCP2.6 are relatively the same through the three future 30-year periods. Because of these different trajectories, significant positive trends are found only in the projections forced by RCP8.5 in all regions ($R^2$=0.58, 0.38, 0.43 and 0.17 for DE2, DEA, DED and DEF respectively). These increasing trends remain significant after BC using both BC-Eval and BC-Hist, with $R^2$ values increasing after correction.

After BC, the number of projected hot days generally increases for all regions using BC-Eval; results using BC-Hist are more mixed, with decreases in the RCP2.6 ensemble mean for DE2, and some early- and mid-century decreases relative to the uncorrected projected changes are also observed (Fig. 6.13 (B)). The increase in the number of hot days under BC-Eval (RCP8.5) is: 14.1, 7.4, 8.4, 4.8 for each region respectively (DE2, DEA, DED, DEF). For BC-Hist (RCP8.5), this is 8.7, 8.7, 9.8 and 7.2 respectively.

#### 6.3.4.2 Summer (JJA) precipitation

Overall, trends in projected JJA precipitation changes are unclear (Fig. 6.14 A and B). There are no significant trends found for the future, apart from DE2 (South Germany), which is projected to have decreases in precipitation, although the $R^2$ value is small ($R^2$=0.05) and the ensemble projected changes are also small (20mm) (Table 6.20). Larger decreases are found for DE2 under the RCP2.6 scenario, up to 100mm less over the June-August period – however, no significant trends are found for any of the projections under RCP2.6 for any region.

DEA, DED, and DEF are also projected to have small (<21 mm) changes by the end of the century under RCP8.5, and larger projected decreases are found under RCP2.6, up to 113mm less for DED, for example. After BC, these projected changes reduced relative to the uncorrected projections. Significant trends are found for the ensemble mean of JJA rainfall for DE2, DEA, and DEF (BC-Eval), but none are observed for BC-Hist.

This concludes the results section; in the following discussion, these climate projections are summarized and used to revisit and answer the chapter research questions.

*Table 6.19: GCM-RCM projected changes to the number of hot days in German regions, in days and in percentage.*

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **DE2 RCP85 Mean** | 1.6 | 5 | 11.3 | +, 0.58* | 2.4▲ | 6.6▲ | 14.1▲ | +, 0.61* | -5.7▼ | -1▼ | 8.7▲ | +, 0.68* |
| | 108.7 | 339.6 | 767.5 | | 71.3 | 196 | 418.8 | | -58.5 | -10.3 | 89.3 | |
| **DE2 RCP26 Mean** | 6 | 6.4 | 7.1 | +, 0 | 9.6▲ | 9.6▲ | 10.1▲ | -, -0.01 | -0.6▼ | 1.9▼ | 3.4▼ | +, 0.04* |
| | 600 | 640 | 710 | | 285.1 | 285.1 | 300 | | -5.5 | 17.5 | 31.4 | |
| **DEA RCP85 Mean** | 0.7 | 2.4 | 6.4 | +, 0.38* | 0.6▼ | 2.7▲ | 7.4▲ | +, 0.41* | -1.1▼ | 2.1▼ | 8.7▲ | +, 0.53* |
| | 52.7 | 180.8 | 482 | | 22.2 | 100 | 274.1 | | -22.8 | 43.4 | 180 | |
| **DEA RCP26 Mean** | 3.7 | 3.7 | 4.7 | +, -0.01 | 4.6▲ | 4▲ | 5.4▲ | +, -0.01 | 4.3▲ | 4▲ | 6.2▲ | +, 0 |
| | 555 | 555 | 705 | | 170.4 | 148.1 | 200 | | 104 | 96.8 | 150 | |
| **DED RCP85 Mean** | 0.9 | 3 | 7.5 | +, 0.43* | 0.4▼ | 3 | 8.4▲ | +, 0.47* | -1.4▼ | 2.2▼ | 9.8▲ | +, 0.58* |
| | 63.5 | 211.8 | 529.4 | | 12.9 | 97.1 | 271.9 | | -24.9 | 39.1 | 174.1 | |
| **DED RCP26_Mean** | 4.6 | 4.8 | 5.4 | +, -0.01 | 5.3▲ | 5▲ | 5.7▲ | +, -0.01 | 3.2▼ | 4.5▼ | 6.1▲ | +, 0.02 |
| | 511.1 | 533.3 | 600 | | 171.6 | 161.9 | 184.5 | | 68.1 | 95.7 | 129.8 | |
| **DEF RCP85 Mean** | 0.1 | 0.6 | 2 | +, 0.17* | 0.6▲ | 1.8▲ | 4.8▲ | +, 0.31* | 0.6▲ | 2.6▲ | 7.2▲ | +, 0.41* |
| | 28.6 | 171.4 | 571.4 | | 74 | 221.9 | 591.8 | | 31.2 | 135.3 | 374.6 | |
| **DEF RCP26 Mean** | 1.1 | 0.7 | 1.4 | +, -0.01 | 2.5▲ | 2.3▲ | 3.6▲ | +, 0 | 3.2▲ | 2.1▲ | 4.2▲ | +, 0 |
| | 1100 | 700 | 1400 | | 308.2 | 283.6 | 443.8 | | 400 | 262.5 | 525 | |

*Table 6.20: GCM-RCM projected changes to total JJA precipitation in German regions, im mm and and in percentage.*

| | Uncorrected 2011-2040 | Uncorrected 2041-2070 | Uncorrected 2071-2100 | Uncorrected Trend ($R^2$) | BC-Eval 2011-2040 | BC-Eval 2041-2070 | BC-Eval 2071-2100 | BC-Eval Trend ($R^2$) | BC-Hist 2011-2040 | BC-Hist 2041-2070 | BC-Hist 2071-2100 | BC-Hist Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **DE2 RCP85 Mean** | -7.6 | -15.6 | -23.5 | -, 0.05* | 11▲ | 3.5▲ | -3.5▲ | -, 0.04* | 3.8▲ | -2.1▲ | -7.5▲ | -, 0.02 |
| | -2.7 | -5.6 | -8.4 | | 4.2 | 1.3 | -1.3 | | 1.6 | -0.9 | -3.1 | |
| **DE2 RCP26 Mean** | -102.1 | -103.1 | -97.7 | +, 0 | -37▲ | -37.6▲ | -31.7▲ | +, 0 | -81▲ | -83.5▲ | -75.7▲ | +, 0 |
| | -33 | -33.3 | -31.6 | | -14.2 | -14.4 | -12.1 | | -27.5 | -28.3 | -25.7 | |
| **DEA RCP85 Mean** | -5.5 | -9.9 | -21.7 | -, 0.03 | 30.6▲ | 25.4▲ | 12.8▲ | -, 0.05* | 9.1▲ | 6▲ | -3.3▲ | -, 0.01 |
| | -2.2 | -4 | -8.7 | | 14.9 | 12.4 | 6.2 | | 4.3 | 2.8 | -1.6 | |
| **DEA RCP26 Mean** | -62 | -59.2 | -64.3 | +, -0.01 | 9.7▲ | 12.3▲ | 7.3▲ | +, -0.01 | -48.6▲ | -47▲ | -47.9▲ | +, -0.01 |
| | -21 | -20 | -21.8 | | 4.7 | 6 | 3.6 | | -19 | -18.4 | -18.7 | |
| **DED RCP85 Mean** | -7.2 | -14.8 | -14.5 | -, 0 | -3.6▲ | -9.7▲ | -9.7▲ | -, 0 | 16▲ | 9.9▲ | 12.7▲ | -, -0.01 |
| | -2.7 | -5.5 | -5.4 | | -1.6 | -4.4 | -4.4 | | 8.3 | 5.2 | 6.6 | |
| **DED RCP26 Mean** | -111.3 | -118.1 | -113.2 | +, -0.01 | -24.8▲ | -30.2▲ | -26.5▲ | +, -0.01 | -58.6▲ | -65.4▲ | -57.5▲ | +, -0.01 |
| | -31.1 | -33 | -31.6 | | -11.4 | -13.8 | -12.1 | | -24 | -26.8 | -23.5 | |
| **DEF RCP85 Mean** | 6.4 | 3 | -8.2 | -, 0.02 | 38.7▲ | 34.3▲ | 23.6▲ | -, 0.03* | 28.4▲ | 21.2▲ | 13.4▲ | -, 0.01 |
| | 2.6 | 1.2 | -3.3 | | 18.8 | 16.7 | 11.5 | | 13.4 | 10 | 6.3 | |
| **DEF RCP26 Mean** | -58.2 | -51.6 | -48.3 | +, 0 | 14▲ | 20.1▲ | 23.3▲ | +, 0 | -49.6▲ | -49.7▲ | -42.1▲ | +, 0 |
| | -18.9 | -16.8 | -15.7 | | 6.8 | 9.8 | 11.3 | | -19.3 | -19.3 | -16.4 | |

*Raw changes are in white rows and percentage in gray. All changes are relative to the respective RR. (∗) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

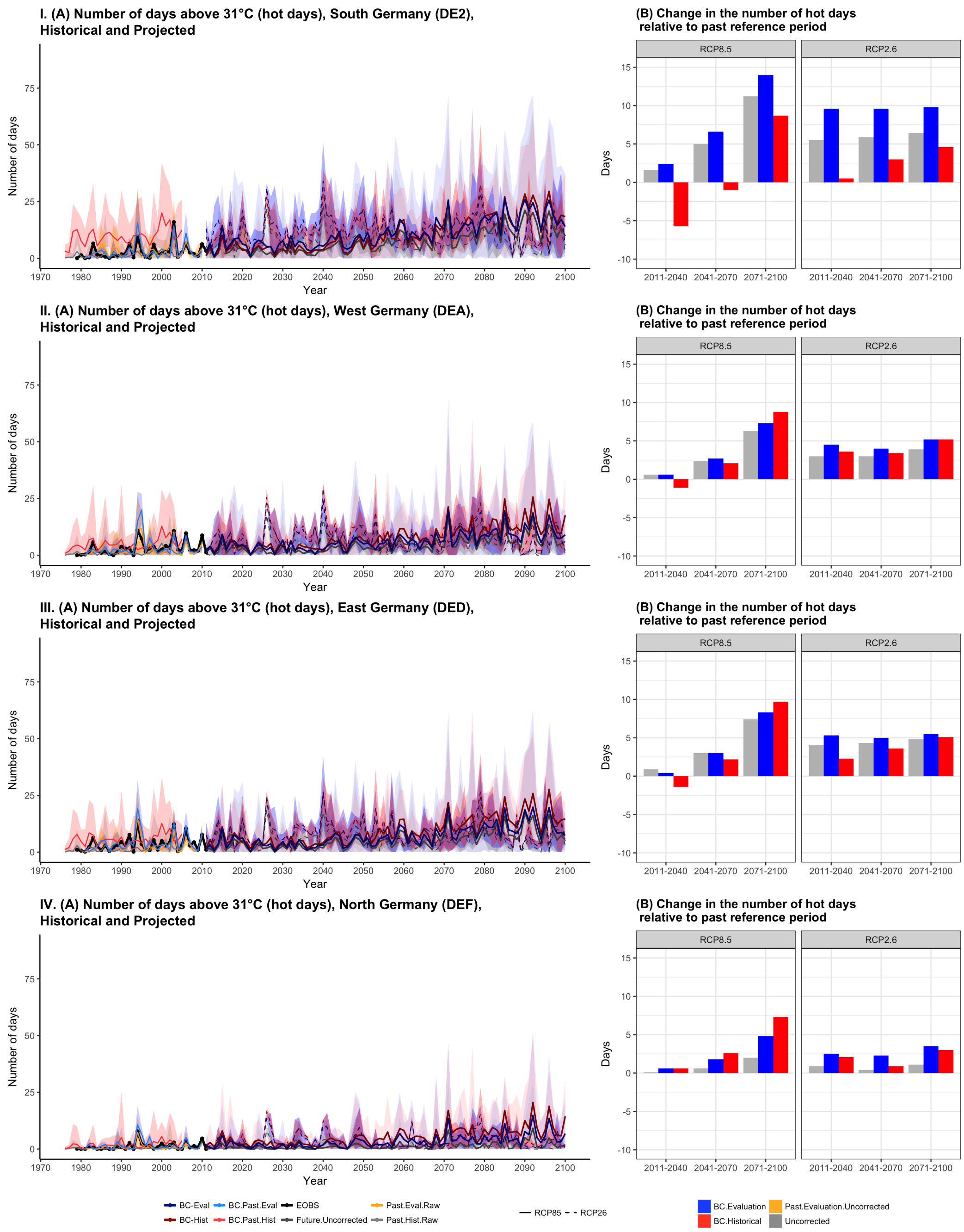



*Figure 6.13: Regional ensemble BC GCM-RCM historical and projected changes to the number of days above 31°C in JJA, for I. South (DE2), II. West (DEA), III. East (DED), and IV. North (DEF) Germany until the end of the century using different calibration for the bias correction (BC-Eval and BC-Hist). (A) shows the range of simulations under RCP8.5 and RCP2.6. (B) shows changes in these climate variables relative to the respective past RR.*

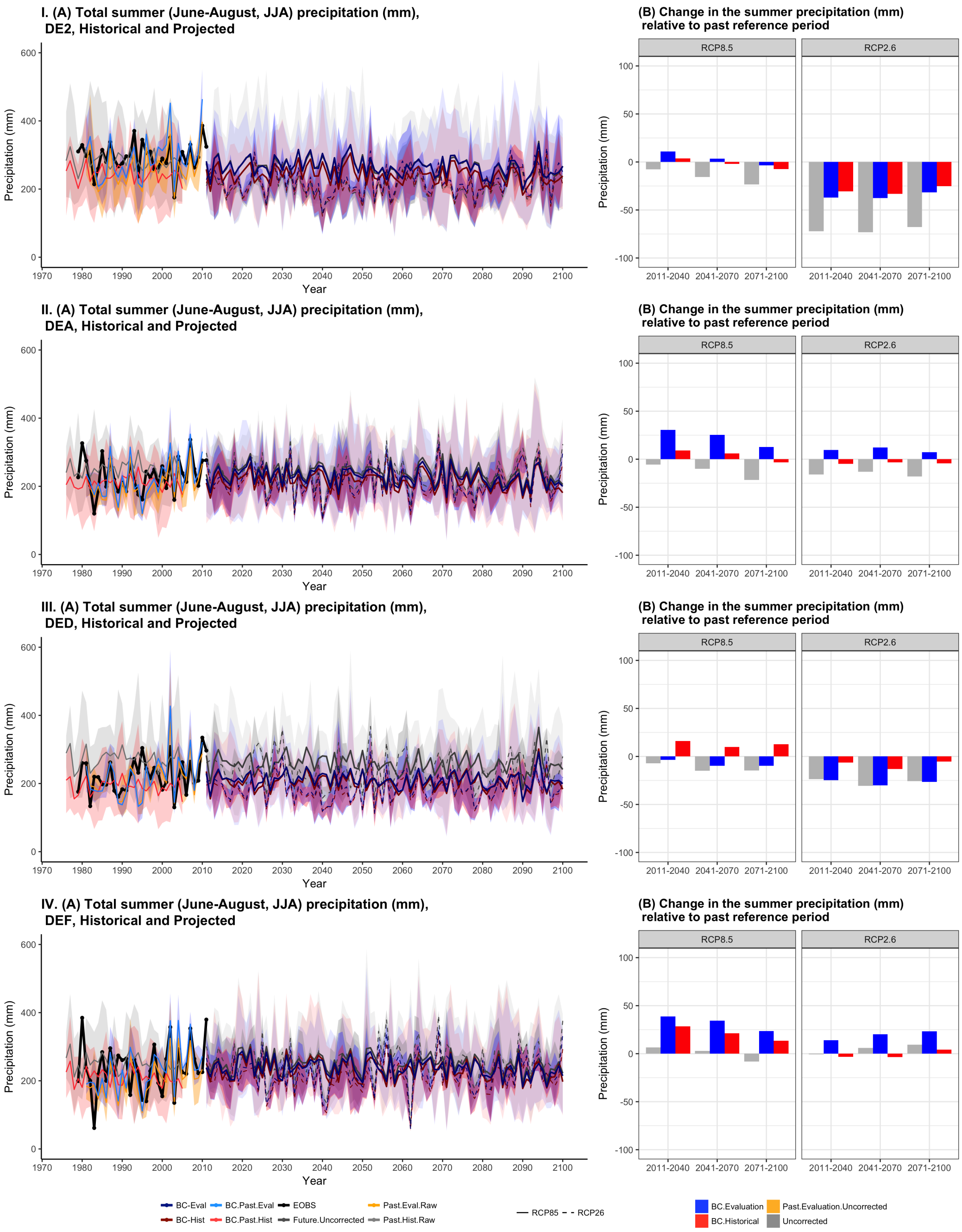



*Figure 6.14: Regional ensemble BC GCM-RCM historical and projected changes to total JJA precipitation (mm), for I. South (DE2), II. West (DEA), III. East (DED), and IV. North (DEF) Germany until the end of the century using different calibration for the bias correction (BC-Eval and BC-Hist). (A) shows the range of simulations under RCP8.5 and RCP2.6. (B) shows changes in these climate variables relative to the respective past RR.*

#### 6.3.4.3 Summary of significant differences between ensemble projections

A summary of when ensemble mean BC projections are significantly different to uncorrected projections is shown in Table 6.21. This shows the results of a Student's t-test, where the difference between the means of two samples is tested. After BC, most projections are significantly different to their uncorrected projection counterparts, indicative of the shifts made after adjustment.

*Table 6.21: Summary of significant differences, with respect to uncorrected projections (RCP8.5 and 2.6).*

| | | BC-Eval | | | BC-Eval RCP2.6 | | | BC-Hist | | | BC-Hist RCP2.6 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Region** | **Variable** | **2011-2040** | **2041-2070** | **2070-2100** | **2011-2040** | **2041-2070** | **2071-2100** | **2011-2040** | **2041-2070** | **2071-2100** | **2011-2040** | **2041-2070** | **2071-2100** |
| **UK** | **Tmax** | | | | | | | | | | | | |
| | **Tmin** | | | | | | | | | | | | |
| | **Precip** | | | | | | | | | | | | |
| | **Hot.Days** | | | | | | | | | | | | |
| | **JJA.Precip** | | | | | | | | | | | | |
| **Germany** | **Tmax** | | | | | | | | | | | | |
| | **Tmin** | | | | | | | | | | | | |
| | **Precip** | | | | | | | | | | | | |
| | **Hot.Days** | | | | | | | | | | | | |
| | **JJA.Precip** | | | | | | | | | | | | |
| **DE2** | **Tmax** | | | | | | | | | | | | |
| | **Tmin** | | | | | | | | | | | | |
| | **Precip** | | | | | | | | | | | | |
| | **Hot.Days** | | | | | | | | | | | | |
| | **JJA.Precip** | | | | | | | | | | | | |
| **DEA** | **Tmax** | | | | | | | | | | | | |
| | **Tmin** | | | | | | | | | | | | |
| | **Precip** | | | | | | | | | | | | |
| | **Hot.Days** | | | | | | | | | | | | |
| | **JJA.Precip** | | | | | | | | | | | | |
| **DED** | **Tmax** | | | | | | | | | | | | |
| | **Tmin** | | | | | | | | | | | | |
| | **Precip** | | | | | | | | | | | | |
| | **Hot.Days** | | | | | | | | | | | | |
| | **JJA.Precip** | | | | | | | | | | | | |
| **DEF** | **Tmax** | | | | | | | | | | | | |
| | **Tmin** | | | | | | | | | | | | |
| | **Precip** | | | | | | | | | | | | |
| | **Hot.Days** | | | | | | | | | | | | |
| | **JJA.Precip** | | | | | | | | | | | | |

*A shaded box indicates a significant difference between uncorrected and BC projections based on a t-test ($p$<0.05).*

In the UK, where very few hot days are found before and after BC, this index is mostly zero hence its similarity to the respective RR. For Germany, as can be observed in Fig. 6.3, ensemble means of BC-Eval, BC-Hist and uncorrected projections are very close to each other, and hence they are not found to be significantly different across all 30-year intervals for BC-Hist, and for the last 30-year period for BC-Eval.

At the regional level, a handful of temperature and precipitation simulations are found to be not significantly different, for instance BC-Eval JJA precipitation for both scenarios in DE2, DEA, and DEF (only RCP8.5). The closeness of ensemble mean JJA precipitation across raw and BC projections can also be observed in Fig. 6.14. For DE2, BC-Eval DE2 annual precipitation (both scenarios) and most of BC-Hist RCP2.6 simulations for other variables, apart from the hot days are also ot significantly different to their uncorrected projections. The mean of BC-Eval and BC-Hist Tmin in DED and DEF, respectively, is also found to be not significant to raw projections across early- and mid-century intervals.

What this indicates is that a large majority of simulations do change and shift after BC. Based on these results, it is argued that when uncorrected projections capture past E-OBS observations of climate fairly well, BC-Eval and BC-Hist simulations are quite close to each other and to the raw projections, resulting in only small modifications after BC. What does this indicate, in terms of GCM and RCM error? In the discussion, these results are reviewed and presented in several cases.

## 6.4 Discussion

In this section, the results from the chapter are summarized and put into context with the research questions at the beginning of the chapter, which focus on the analyzing the projected changes to temperature and precipitation from the chosen GCM-RCMs, and how BC affects these projected changes. The following discussion also aims to connect the results to the debate on the appropriate use of BC for the future. How useful are the different calibration approaches in processing climate model output for future impact projections?

### 6.4.1 Summary of projected changes

In the results of the analysis of downscaled climate projections from EURO-CORDEX, the following results are found, in the context of the first two research questions posed at the beginning of the chapter:

- Projections of both Tmax and Tmin from EURO-CORDEX GCM-RCMs show annual warming trends over the UK, Germany, and the four German regions examined in the study. **(RQ 1)**
- Depending on the future emission scenario, individual GCM-RCM pairs project different magnitudes of warming, with uncorrected future temperature projections from RCA-HadGEM typically projecting the largest changes in temperature.
- In terms of total annual precipitation, uncorrected projections, approximately half of the GCM-RCMs show increasing trend over the UK and Germany, albeit with generally small $R^2$ values. The ensemble mean over both countries and all four German regions also show significant increases under RCP8.5. The ensemble mean projections of total annual precipitation under RCP2.6 did not show any significant trends for the UK, Germany, or any of the German regions.
- The number of days above 31°C are shown to have significant increasing trends. Warm days are projected to increase over Germany by around 6.6 (4.2) days under the RCP8.5 (RCP2.6) scenario in the uncorrected projections. While the UK projections also show

significant increases in the number of hot days, the increase in the number of days over the entire UK is low, under 1-2 days per summer. In German regions, South Germany (DE2) is projected to have 11 more hot days per summer by the end of the century based on uncorrected RCP8.5 projections. RCP2.6 also projects more hot days, but these are typically lower than RCP8.5 projections, and there are no significant increasing linear trends observed.

- Summer precipitation totals are projected to decrease in the UK and Germany, but ensemble projected changes are small (30mm, RCP8.5 UK) and in Germany (20mm, RCP8.5 Germany). Projected decreases under RCP2.6 are greater than the RCP8.5 projections (70mm UK, 85mm Germany), but did not show any significant trends. Similar trends are observed for regional Germany, where projected decreases in JJA precipitation are greater when forced by RCP2.6.
- The 'evolution' of temperature projections forced by the high-emission scenario RCP8.5 are distinct from the lower RCP2.6 scenario: typically projected temperature changes under RCP8.5 start out low but continuously climb until the end of the century. In contrast, projected temperatures under RCP2.6 start out high do not increase further by the mid- and late-century. **(RQ2)**
- For precipitation, the divergence of precipitation projections between the two different scenarios is less distinct compared to temperature, but projected changes under RCP2.6 with RCA-HadGEM and RCA-MPI, the GCM-RCMs available for the study, are less than their projections forced by RCP8.5 in all study regions. **(RQ2)**

The effect of BC on these projected changes in temperature, precipitation, and seasonal summer climate – the foci of research questions 3 and 4 – are discussed in the following section, along with how they can be used in impact assessment **(RQ5)**.

## 6.4.2 Analyzing the effects of BC on projections

### 6.4.2.1 'Practical' BC for use in impact assessment

The results of the chapter show that, as expected, the use of BC shifts the beginning of the early century projections closer to the mean of the past BC evaluation simulations (for BC-Eval) and past BC historical simulations (for BC-Hist), which are more similar to observations in their mean and distribution than uncorrected past simulations. The results shown in this chapter are evidence as to why the use of BC is seen as a needed practice in impact assessment, particularly in impact studies – because of the large biases in uncorrected historical projections that are carried over to future climate projections.

A particular example is for Tmax, which is a critical climate variable for wheat, because of its sensitivity to heat stress. In uncorrected projections, while changes of over 2°C are projected, the mean of uncorrected future projections is shown to be below the mean of observations in the UK and Germany (Figs. 6.2-I and 6.3-I). It is shown in the results of this chaper how the use of BC methods applied to future projections, particularly for temperatures, shifts simulations to a plausible range. This shift is important for crop yield projections, as realistic climate change projections are needed for developing appropriate and effective adaptation strategies and better targeting global emissions reduction goals (Ramirez-Villegas et al., 2013).

Therefore, this large temperature shift/jump after BC is particularly significant for the crop models chosen for the study. The PCM CERES-Wheat/DSSAT used in the study requires daily temperature input, and the SCCM depends on the hot-day index which is calculated from daily Tmax. Because yield simulations are sensitive to precipitation and temperature biases, impact assessments need climate input data at high spatial and temporal resolutions with minimal biases. Without BC, it has been shown that large biases in yields are due to biases in rainfall during the growing season inherited from GCMs (Macadam et al., 2016). When BC is applied, the largest positive yield biases are reduced because the underlying biases in growing season rainfall are reduced (Macadam et al., 2016). Even when downscaled, the biases from climate models hamper the direct application of RCM output in impact studies (Casanueva et al., 2016).

There are thus several arguments in favor of BC: the design of BC is to bridge the gap between the the information provided by climate models and data needs to make quantitative climate impact projections (Hempel et al., 2013). In addition to being a common method in climate change impact studies (Cannon et al., 2015), it has been reported in numerous studies that some form of correction must be applied to climate projections before their use in climate change impact assessment (Iizumi et al., 2017, Hawkins et al., 2013b, Piani et al., 2010). For these reasons, and as shown by the results, it is argued that BC, despite criticisms, assumptions and issues, as previously discussed (See Section 2.4.2.1 and Chapter 5 discussion), serves a practical purpose in impact assessment; how these biases affect yield projections is examined in the subsequent chapter.

#### 6.4.2.2 Bias correction and the climate change signal

The results of this chapter also address the third research question for the chapter, which is how BC affects climate projections. Several changes to the projected climate changes in temperature and precipitation can be observed after BC. Most projections are significantly different to their uncorrected projections (See Table 6.21), and this affects their projected changes. For example, the ensemble projected temperature change relative to the respective RR (past BC evaluation simulations for BC-Eval and past BC historical simulations for BC-Hist) is smaller than the uncorrected projected temperature change in the UK, Germany, and all four German regions. There are some exceptions, for example the RCP8.5 BC-Hist ensemble Tmax mean for DEF. For precipitation, projected annual changes to precipitation typically increase after correction for the RCP8.5 scenario.

In terms of the average annual Tmax and Tmin and precipitation totals, most RCP8.5 trends remain significant, and therefore robust, even after BC in all the study regions. Significant warming is projected for all regions until the end of the century before and after BC. Projections forced by RCP2.6 remain non-significant for all GCM-RCM simulations of Tmax and Tmin in the UK and Germany. However, there are a small number of changes where after BC, trends become insignificant, although most changes are to trends with relatively low $R^2$ values to begin with:

- BC-Hist Tmin, RCP2.6 RCA-MPI, Germany
- BC-Eval and BC-Hist Tmin, RCP2.6 ensemble mean, DE2
- BC-Hist Tmin, RCP2.6 ensemble mean, DEA
- BC-Eval and BC-Hist Tmin, RCP2.6 ensemble mean, DED
- BC-Hist Tmin, RCP2.6 ensemble mean, DEF
- BC-Eval and BC-Hist total annual precipitation, RCP8.5 RACMO-ECEARTH, Germany
- BC-Hist hot days, RCP2.6 RCA-HadGEM, UK
- BC-Hist hot days, RCP2.6 ensemble mean, UK
- BC-Hist JJA rain, RCP8.5 RCA-IPSL, Germany
- BC-Hist JJA rain, RCP8.5 ensemble mean, Germany
- BC-Hist JJA rain, RCP8.5 ensemble mean, DE2

There are also a number of simulations with linear trends that become significant after correction, also with small $R^2$ values:

- BC-Eval and BC-Hist RCP8.5 RCA-HadGEM, Germany
- BC-Hist hot days, RCP2.6 ensemble mean, DE2
- BC-Hist JJA rain, RCP8.5 ensemble mean, DEA
- BC-Hist JJA rain, RCP8.5 ensemble mean, DEF

These changes in trends are also evidence of why there remains significant scientific criticism and controversy in the use of BC. As previously discussed in Chapter 5, the BC issue is still much debated by the scientific community (Dosio, 2016). BC is heavily criticized because of its potential to exaggerate high extremes while over-correcting low extremes, the potential that BC methods may change the climate signal, and disrupt physical consistency (Maraun et al., 2017, Maraun, 2016, Sippel et al., 2016, Hempel et al., 2013, Maraun, 2013). BC approaches hinge on the fundamental assumption that the GCM produces skillful input in the first place, in order for the correction to be effective, as well as the assumption that the GCM can produce a plausible representation of climate change (Maraun, 2016). It has been argued that the preservation of the climate change signal is crucial (Hempel et al., 2013); projected changes and trends by climate models should be preserved in a way that the sensitivity of the climate model is preserved and not affected by BC.

To contrast the criticism of BC, there are also studies that show that BC can actually improve the climate change signal: in a study, Gobiet et al. (2015) show that while quantile mapping does modify the climate change signal, it also removes the intensity-dependent errors in the original GCM output, potentially leading to an improved signal (Gobiet et al., 2015). Other studies have also shown that quantile mapping improves the correspondence with observed changes in some locations and degrades it in others (Maurer and Pierce, 2014), but overall they find that the influence of BC does not systematically degrade projected differences.

These findings contrast with studies that report that the transfer function via quantile mapping can be considered as a ‘leap of faith’ that may lead to a false certainty about the robustness of the adjusted projection (Grillakis et al., 2017, p.890). The fact that bias adjustment affects the change of mean climate is known: BC can reduce or increase temperature-related climate change over Europe (Dosio, 2016). In the results of this chapter, it is shown that most increasing (warming) trends in temperature remain robust after correction, and that precipitation trends are also mostly robust. However, it is also shown that there are some changes in the significance (but no significant changes in the direction) of the linear trend. However, the projected change depends on the calibration setup (BC-Eval or BC-Hist) used for the correction, which is discussed in the next section to address the fourth research question of the chapter.

#### 6.4.2.3 Comparing BC-Eval and BC-Hist: case examples

Two calibration methods are used in this chapter. This dual approach was adopted because scientific opinions within different communities of practice differ on how to calibrate the correction for future projections, given that both GCMs and RCMs contribute to biases. It has been argued that the use of evaluation simulations to assess RCM skill is useful as the choice of downscaling RCM can be a source of uncertainty in climate projections (e.g. Kotlarski et al., 2014, Hwang et al., 2014, Menut et al., 2013, Oettli et al., 2011). In contrast, the more commonly used BC-Hist approach is argued to be a useful and practical way of correcting the future GCM output with the purpose of using them in impact assessments as it minimizes error from both the GCM and RCM simultaneously (Pasten Zapata, 2017).

Despite these differences in design and purpose between BC-Eval and BC-Hist, a comparison of their results is useful to explore how BC affects future projections, and this approach is a relatively new way of understanding climate model error that is introduced by the GCM-RCM pairing, which is research question 5 in this chapter. Through the comparison of temperature and precipitation simulations calibrated with BC-Eval and BC-Hist, four different cases have been identified: (1) a poorly-performing GCM (high bias in the past historical simulations) is paired with a well-performing RCM (low bias in evaluation simulations); (2) a well-performing GCM (with low bias in the historical simulations) is paired with a poorly-performing RCM; (3) both the GCM and RCM are skillful in past historical and past evaluation simulations, or (4) both the GCM and RCM perform poorly and have large biases. Similar cases have also been identified in a hydrological context (Pasten Zapata, 2017).

**Case 1: A biased GCM and with a well-performing RCM.** In this example of total annual precipitation over Germany from CCLM-MPI simulations, past historical GCM-RCM simulations have a large positive bias relative to observations (Fig. 6.15), while uncorrected evaluation simulations are well-correlated to observations. The future GCM output thus retains this bias, and BC-Eval projections calibrated to the BC evaluation simulations are only shifted slightly compared to uncorrected projections. BC-Hist simulations are further negatively shifted since the contributing error is mostly from the driving GCM.

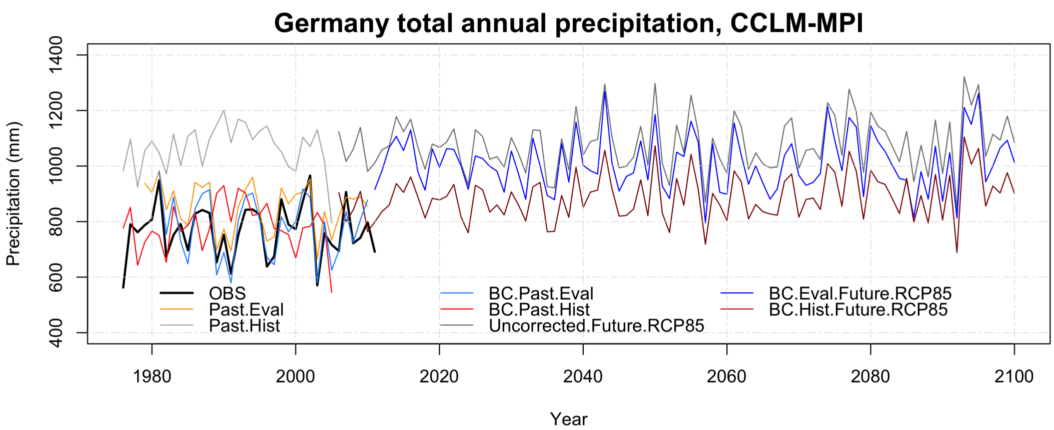



*Figure 6.15: Total annual precipitation, past and projected, Germany, CCLM-MPI.*

**Case 2: A well-performing GCM and with a biased RCM.** In this example, the opposite to Case 1 is observed: the error from the contributing RCM (RCA) is larger than from the GCM. Past historical simulations from RCA-IPSL have smaller biases relative to observations compared to the uncorrected past evaluation simulations, so after BC the projections shift downward in order to compensate for this positive bias from the RCM.

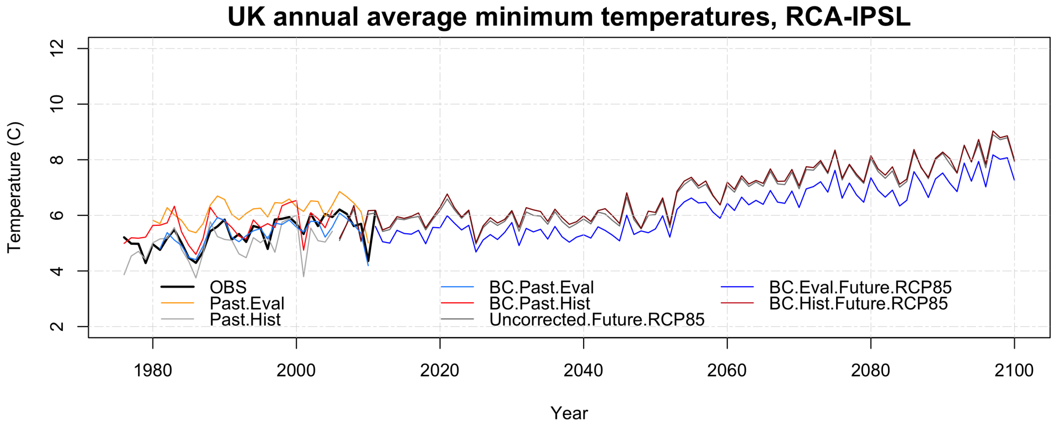



*Figure 6.16: Annual average Tmin, UK, RCA-IPSL*

**Case 3: Both the GCM and RCM are skillful.** Both the evaluation and historical simulations perform relatively well compared to observations and have minimal biases (Fig. 6.17). Therefore, after BC, BC-Eval and BC-Hist total annual precipitation values change very little relative to uncorrected projections in the time-series of projections of total annual precipitation.

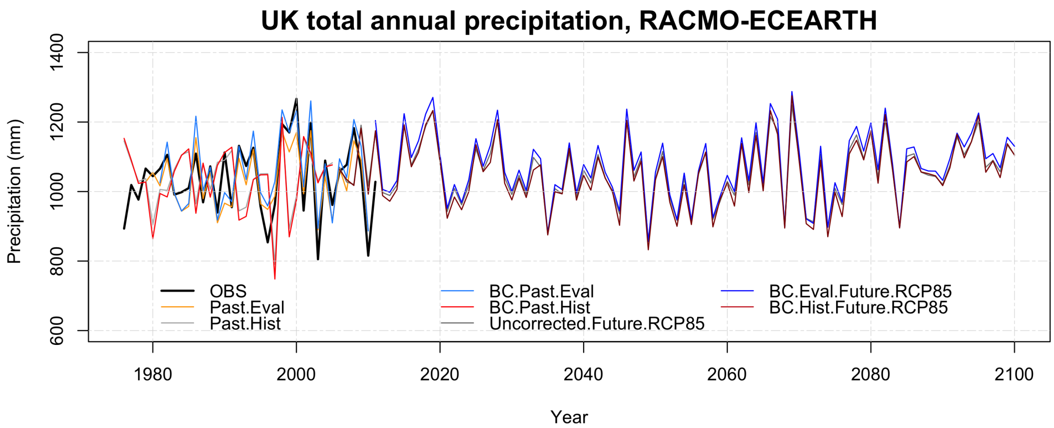



*Figure 6.17: Total annual precipitation, past and projected, UK, RACMO-ECEARTH*

**Case 4: Both the GCM and RCM are unskillful (have large biases).** In this case of annual Tmax, both the RACMO evaluation simulation and the RACMO-ECEARTH past historical simulation have similar large negative biases relative to observations. After BC, it can be observed that BC-Eval and BC-Hist are both shifted positively relative to uncorrected projections.

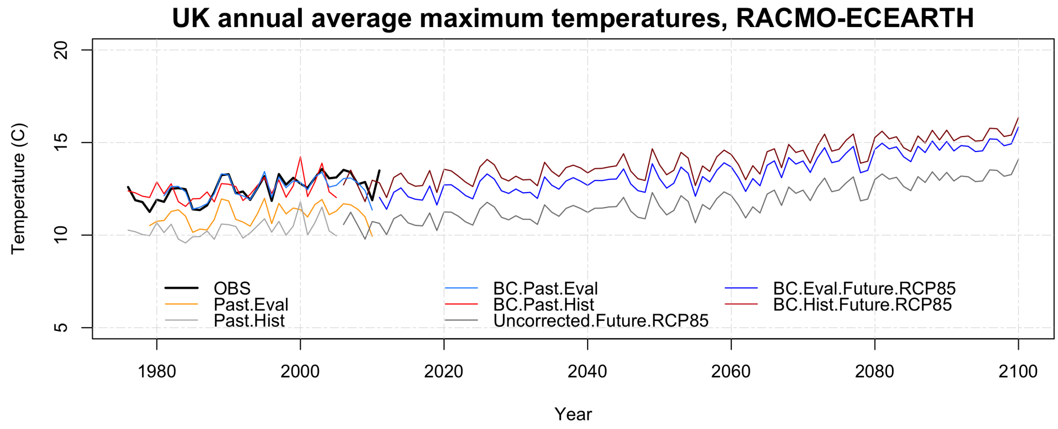



*Figure 6.18: Annual average Tmax, UK, RACMO-ECEARTH*

These four cases demonstrate that the choice of the GCM-RCM combination, as well as the application of BC, are important contributors to the uncertainty (range) of future climate projections. What these cases show, rather than that one calibration approach (BC-Eval or BC-Hist) is more suitable than the other, is that the choice of both the GCM and RCM – and how they perform when used together over a particular domain – is very significant to produce a plausible set of climate projections.

#### 6.4.2.4 Using BC calibration as a method to understand combined error of GCM-RCM choice

The approach used in this chapter of using BC-Eval and BC-Hist to compare how they affect future climate projections is thus presented as an option for evaluating how particular GCM-RCM pairs can influence the projected changes in temperature and precipitation. This relatively novel way of understanding the combined error of GCM-RCM error, along with the error of the RCM alone (See Chapter 5 for RCM evaluation results) can become a supplementary method in the process of selection of climate models for a study. In addition, other methods are recommended for this purpose and they are outlined in the following discussion as reviewing and ranking based on previous performances, and using ensembles.

#### 6.4.2.5 Reviewing previous climate model performances

GCM-RCMs are often reviewed and selected carefully through a review of their performances in previous studies. For instance, this study reviews the chosen GCM-RCMs in Section 4.3.2.2, Chapter 4. The chosen RCMs (CCLM, RACMO, RCA) generally had acceptable performances, but noted cool biases for Tmax and wet biases over Europe (Kotlarski et al., 2014). Among the chosen GCMs, MPI, CC, and HadGEM were typically noted for their better performances compared to other GCMs in CMIP5 used for EURO-CORDEX based on a Model Performance Index score (Jury et al., 2015). In contrast, the lower resolution IPSL model has been reviewed poorly for its simulation of precipitation and was classified as containing “biases" in a key GCM evaluation study (McSweeney et al., 2015).

EC-ECEARTH was found to have “significant biases" in the same study, while CC, MPI, and HadGEM had satisfactory performances for their simulation of temperature and precipitation cycles, storm tracks, and circulation over Europe (McSweeney et al., 2015). EC-EARTH was found to perform well for dynamic variables but less favorably for surface temperature (Hazeleger et al., 2010), and this is reflected in the Case 4 example (Fig. 6.18) in this study where past RACMO-ECEARTH simulations are poorly related to past E-OBS over the UK: the cool biases of RACMO for Tmax are combined with EC-EARTH’s challenges in simulating surface temperatures and thus uncorrected projections also contain a significant cool bias.

Therefore, the results shown in this chapter which use two different calibration reference simulations (past BC-Eval or past BC-Hist) means that it is possible to investigate the effect of RCM-only and joint GCM-RCM error, and by using the case examples as indicative demonstrations of the error introduced by the pairing, this information can be used to minimize biases in the future. This further emphasizes that impact assessments should also perform careful selection of the climate models and considering which variables are to be simulated. Given the large number of simulations available, the added complexity and uncertainty of BC (choice of method, and considering its criticism), the complexity of impact models themselves, impact modelers are given great challenges. It has thus been reported that the [climate change] practitioner's dilemma is no longer the lack of down-scaled projections; it is how to choose an appropriate dataset, assess its credibility, and use it wisely (Ekström et al., 2015, Barsugli et al., 2013). These challenges are contextualized with the demands of using climate model output in impact studies in the next section.

#### 6.4.2.6 Implications for impact studies: ensembles

In addition to a review of previous performances, some studies use a ranking approach to assess and select a subset climate models from of the large number of available simulations (e.g. Jury et al., 2015, McSweeney et al., 2015) to create a climate ensemble; it is also possible to start with an ensemble and further subset the best-performing members from the group. However, by reducing the model ensemble one also reduces the information about the uncertainty in the projections and the ensembles (Wilcke and Bärring, 2016).

It is argued that the reduction of ensembles to only those that have minimal biases or other measures of ideal performance comes into conflict with the motivation of using multiple models in climate change research to cover and characterize different sources of uncertainties (e.g. Yip et al., 2011, Wilby and Dessai, 2010, Hawkins and Sutton, 2009). To further support this argument, while the objective of multi-model ensembles is often to narrow the range of uncertainty, it has been reported that this consensus-seeking approach is too limited, and that exploring the differences between models (e.g. climate, or impact models) is in itself a valuable approach (Knutti and Sedláček, 2012, Challinor et al., 2014). The criteria for selection of a GCM or RCM is also very context-dependent: the 'one-model-one vote' approach is also flawed as all projections should be considered to have a non-negligible likelihood of occurring (McSweeney et al., 2015), while in the real world the choice of climate model may be dependent solely on the project partners (Wilcke and Bärring, 2016).

However, to create and utilize a very large ensemble of climate models for projecting future climate change and its impacts (here, in the context of crop modeling) is also problematic. The use of very large ensembles is severely limited by the real-world constraints of time, limited computing resources, and incomplete GCM-RCM combination simulations (Wilcke and Bärring, 2016). For the latter reason, for example, in this study only two chosen GCM-RCMs from the r1i1p1 ensemble were available for RCP2.6 at the time of analysis. Many impact modelers may have problems in handling GCM-RCM ensembles that are 'too big' (Wilcke and Bärring, 2016, p.191), making the research project not feasible.

In addition, the time and resource constraints of using large ensembles will be faced with other real-world issues of providing projections in a timely manner. There is great motivation on the part of decision makers, in both the public and private sectors, to acquire and understand information on climate change that can inform their decisions (Lemos and Rood, 2010). Thus, there is also great demand on climate modelers to synthesize climate and crop modeling information quickly and efficiently: a recent example of this is the anticipated report on the impacts of half a degree of warming following the 2015 Paris Agreement: the main proponents of the report due in 2018 ask: "Can such research be carried out in time with a high enough level of reliability to properly inform such a momentous policy decision?" (Mitchell et al., 2016, p.736).

The complexity of the discourse surrounding BC makes the use of climate projections for impact projections far from straightforward. While BC has obvious benefits in making climate projections more directly usable, it is argued that its limitations also hinder and limit the direct application and use of climate projections, even after their correction. Therefore, while recent developments show promise in making BC overcome its main challenges, and as BC methods become more sophisticated and consider trend-preserving techniques (e.g. Grillakis et al., 2017, Sippel et al., 2016, Hempel et al., 2013), it remains that BC methods should be applied with caution and consideration of the context of their use, and that they are a stop-gap measure to the need to improve underlying climate model performances.

## 6.5 Conclusion

In this chapter, it was investigated how temperatures, precipitation, and summer climate indices relevant for crop growth are projected to change under different future emission scenarios. It has been shown that future projections of temperature show significant warming in both Tmax and Tmin over the UK and Germany under both RCP8.5 and 2.6. Projections also show some increases in total annual precipitation, unclear trends in JJA precipitation, and increases in the number of days where heat stress could occur for heat-sensitive crops like wheat.

It is also shown that errors from both GCMs and RCMs propagate into future climate projections, and that the two calibration approaches to BC (quantile-quantile mapping) are able to demonstrate how the combination of GCM and RCM can affect the shift in projections and projected climate changes after correction. The comparative results in this chapter present a relatively new way of understanding how the calibration of the bias correction (on evaluation or historical simulations) can be used to understand how GCM-RCM pairings can also affect future climate change projections.

While BC continues to be used as a standard step in impact assessment, there is ongoing criticism of the scientific basis for BC that come into conflict with the demands of rapid impact assessments, alongside real-world constraints. Based on the results of this chapter and the scientific work around BC, it is concluded that it is crucial that climate change projections and the impact assessments that rely on them must account for biases from both the climate models themselves, as well as the methods used to make projections more credible.

# Chapter 7

# Multi-method comparison of the projected impacts of climate change on yield

## 7.1 Introduction

In this chapter, projections of future climate from global and regional climate models (GCM-RCMs) are used in two distinct methods of simulating crop yield – a process-based crop model (PCM) and statistical crop-climate models (SCCMs) – in order to address two key research objectives. The first objective is to quantify how increasing temperatures and changing rainfall due to climate change will affect wheat production in the UK and Germany. The intended outcome of the first objective is to add to the existing body of knowledge on the impacts of climate change on wheat yields. The second objective is to provide a characterization of how yield projections are affected by the choices of GCMs and downscaling RCMs, bias correction, emission scenarios, and impact model methods. This characterization of uncertainty – a key concept discussed throughout the thesis – is achieved using an uncertainty partitioning approach.

### 7.1.1 Comparing the methods of crop models

Numerous scientific studies report that wheat yields at a global and European scale are generally expected to be negatively affected by climate change and variability, due to sensitivity to high temperatures, and potentially increased occurrence of heat stress and drought (Trnka et al., 2015, Asseng et al., 2014, Porter et al., 2014, Olesen et al., 2011, Trnka et al., 2011, Porter and Semenov, 2005). The additional adverse effects of excess water are also projected to potentially reduce wheat yields in the future (Zampieri et al., 2017). It has been argued that some of these adverse effects can be mediated by adaptation and genetic breeding (Moore and Lobell, 2014, Semenov et al., 2014, Reidsma et al., 2009).

Many of these findings are dependent on the crop models, which are useful tools for adaptation (Chenu et al., 2017). Crop models have decades of development history and have been used widely around the world to advance knowledge on crop interactions with genetics, the environment, and management (the $G \times E \times M$ pillars of agronomy) for a variety of purposes (Chenu et al., 2017). However, many crop modeling studies rely on the results of a single chosen crop model and/or method, which can be either process-based or statistical in its approach (See Chapter 2 for an extensive discussion on the differences between approaches).

In light of the need to better characterize crop modeling uncertainty, there are large coordinated scientific efforts to generate impact projections from multi-model ensembles. Multi-model ensembles (MMEs) are able to give a significant contribution to the characterization of uncertainty and are increasingly used in climate impact assessments (Challinor et al., 2013). Many comparisons are under the umbrella of the Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP, Warszawski et al., 2014) of which the Agricultural Model Intercomparison and Improvement Project (AgMIP, Rosenzweig et al., 2013) is part. Results from a 30-member crop model comparison from AgMIP estimate that global wheat production will fall by 6% for each degree of further temperature increase, in addition to becoming more variable (Asseng et al., 2014), with more recent studies investigating the differences in agricultural impacts from 1.5° and 2°C warming, in light of the Paris Agreement (Schleussner et al., 2018, Ruane et al., 2018, 2017). MMEs are also useful in exploring the differences between individual climate models, for instance the Coupled Model Intercomparison Project (CMIP6, Eyring et al., 2016).

Despite the usefulness of MMEs to characterize uncertainty (in this case, the range of possible future yield outcomes), reducing uncertainty is still dependent on the continued improvement of crop models in climate change impact assessments. Improving the crop models can improve the accuracy of simulations and reduce the number of models needed in a MME to a more practical number for impact studies (Maiorano et al., 2017). However, it has been argued in this study (e.g. Chapter 2, 3) that MMEs and communities of practice like AgMIP are largely focused on PCMs, and that PCM-only MMEs and the community of practice surrounding MIPs miss the role and function of statistical approaches in understanding and comparing projected climate change impacts.

Reasons for this exclusion are that statistical approaches have been long criticized as they are often thought to be lacking in complexity (e.g. Semenov et al., 2012). Evidence exists, however, that SCCMs could be useful for understanding impacts and adaptation where sufficient data is present (e.g. Ray et al., 2015, Iizumi et al., 2013, Hawkins et al., 2013a, Lobell and Burke, 2010). Multi-method comparisons between PCMs and SCCMs, also discussed in Chapter 2, are seen as a way to better elucidate the differences between crop modeling methods. In spite of this valuable comparison, multi-method crop model studies are at a relatively early stage (Moore et al., 2017, Watson et al., 2015), so this study aims to contribute to that field.

A review of the literature shows that crop model studies fall into one of two ‘camps’, PCMs or SCCMs (Lobell and Asseng, 2017, p.1) – in this work they are referred to as ‘communities of practice’ (See Chapter 1). Based on the review of climate and impact modeling studies that have informed this thesis, it is argued that differences between the practices of these crop modeling communities have led to relatively few comparisons. It is also argued that there is a disciplinary gap between the climate modeling and the impact modeling communities of practice, leading to opposing views on what may seem like straightforward and useful processes. For example, downscaling and bias correction – both common methods in the impact assessment process – are highly debated and criticized within climate modeling studies (e.g. the added value debate, and the practice of bias correction (Maraun et al., 2017, Maraun, 2016), see Chapters 4 and 6).

This gap is a reason why the results of key multi-method comparative studies are interesting: despite major methodological differences, initial comparative studies have found no major systematic differences in the projected impacts of major crops to warming from either PCMs or SCCMs, apart from the issue of how the $CO_2$ effect is modeled (Lobell and Asseng, 2017, Liu et al., 2016). Other factors such as uncertainties in climate data and differences in calibration also contribute to the observed differences between the respective yield projections of different methods (Watson et al., 2015). In light of this scientific discourse, this chapter aims to contribute to this emerging area of study of crop model method comparison, and to support robust crop impact assessment by using methods to decompose and quantify uncertainty.

#### 7.1.1.1 Uncertainty decomposition in impact assessments

The concept and practice of partitioning uncertainty is used often in climate projections, and has also been used in crop impact studies. In climate studies, uncertainty partitioning (also termed uncertainty decomposition) is the method of determining the fraction of uncertainty caused by three main identified sources of uncertainty in climate projections: climate models, future emission scenario, and natural or internal variability (Yip et al., 2011, Hawkins and Sutton, 2011, 2009), with other statistical analysis of uncertainty completed with Bayesian methods (e.g. Northrop and Chandler, 2014). In a recent climate projections study, the use of different bias correction methods and forcing data sets was found to be a significant contributor total uncertainty, potentially more than different GCMs and RCPs (Iizumi et al., 2017).

Uncertainty partitioning using analysis of variance (ANOVA) statistical tests is also a well-practiced method in hydrological impact studies, which, similar to agricultural impact studies, also follow the impact chain from RCP-driven GCM to impact model (e.g. Hattermann et al., 2018, Vetter et al., 2017, Bosshard et al., 2013). In an agricultural impact assessment study for wheat in China (Vermeulen et al., 2013), uncertainty was partitioned into climate models, crop models, and natural variability. Their contribution to uncertainty changed over time, but climate model uncertainty was larger than that of the crop model (Vermeulen et al., 2013). In another study that investigated wheat projections in India, uncertainty decomposition methods found that crop model uncertainty dominates the fractional uncertainty in yield projections (Koehler et al., 2013). This was largely due to how temperature-driven process were represented, with uncertainty from these crop developmental processes larger than climate model uncertainty. Bias correction was not found to be a major contributing factor to uncertainty in their study (Koehler et al., 2013).

Given the relatively recent focus on multi-methods and re-evaluating the purpose and application of bias correction, there are opportunities to contribute knowledge to this emerging field to study the 'intermediate' steps in the actual impact assessment cascade (See Chapter 1, Fig. 1.4) and one of the identified research niches is to characterize how these steps between climate and crop models affect yield projections. In particular, focusing on a multi-method context was an identified research gap in the literature review.

### 7.1.2 Chapter approach and objectives

The approach of this chapter is to build upon the methods and results of previous chapters, firstly: it is informed by the literature review to determine research opportunities in the crop-climate discipline and a review of the development and performance of GCMs, RCMs, and crop modeling methods. Secondly, the yield projections of this chapter use two different approaches to crop modeling that were comparatively evaluated and used for a yield hindcast (Chapters 3 and 5). Lastly, the climate model output for the PCM and SCCM are taken from the state-of-the-art European Coordinated Regional Downscaling Experiment (EURO-CORDEX, Jacob et al., 2014), and these were evaluated for the past (Chapters 4, 5) an bias-corrected for the future with two calibration methods (See Chapter 6).

#### 7.1.2.1 Chapter research questions

The objectives of the chapter are: (1) to generate wheat yield projections until the end of the century through two different crop modeling methods, and (2) to characterize their uncertainty as components of downscaled global climate models (GCM-RCMs), bias correction (BC), and the crop modeling method. Although emission scenario uncertainty is also taken into consideration, the approach is limited by availability of data (See the following Data and Methods section). The research questions are:

**(1)** How are wheat yields in the UK and Germany projected to be affected by changes in temperature and precipitation?

**(2)** How much do the projections of PCMs and SCCMs differ, including between different emission scenarios?

**(3)** How do yield projections change after bias correction of climate input?

**(4)** How much variance in yield projections is caused by the choice of climate model, crop model method, and whether climate model output is bias corrected?

**(5)** Based on the results of research question 4, are there opportunities to reduce or better characterize the most significant source of uncertainty identified for the study?

The aim of the chapter is to provide evidence to fill the gap in the need for more comparative method studies while considering current issues and debates in climate and crop modeling.

## 7.2 Data and methods

The data and methods of this chapter are largely dependent on the previous' chapters results. This section contains a a brief summary of the climate and crop models used in previous chapters, followed by the methods needed to address the chapter research questions. Figure 7.1 gives an overview of the chapter design.

### 7.2.1 Climate model output

Daily projections of future precipitation, and maximum and minimum temperature from six GCM-RCM combinations were processed and bias-corrected in Chapter 6, and the summer climate indices developed for the statistical crop-climate model (SCCM) were also derived from these projections. The six GCM-RCM combinations, detailed in Section 4.3.2 and Table 4.1 in Chapter 4 are: CCLM-MPI, RACMO-ECEARTH, RCA-CC, RCA-HadGEM, RCA-IPSL, and RCA-MPI. Simulations forced by the RCP8.5 and RCP2.6 emission scenarios (Moss et al., 2010, van Vuuren et al., 2011) are taken from EURO-CORDEX (Jacob et al., 2014).

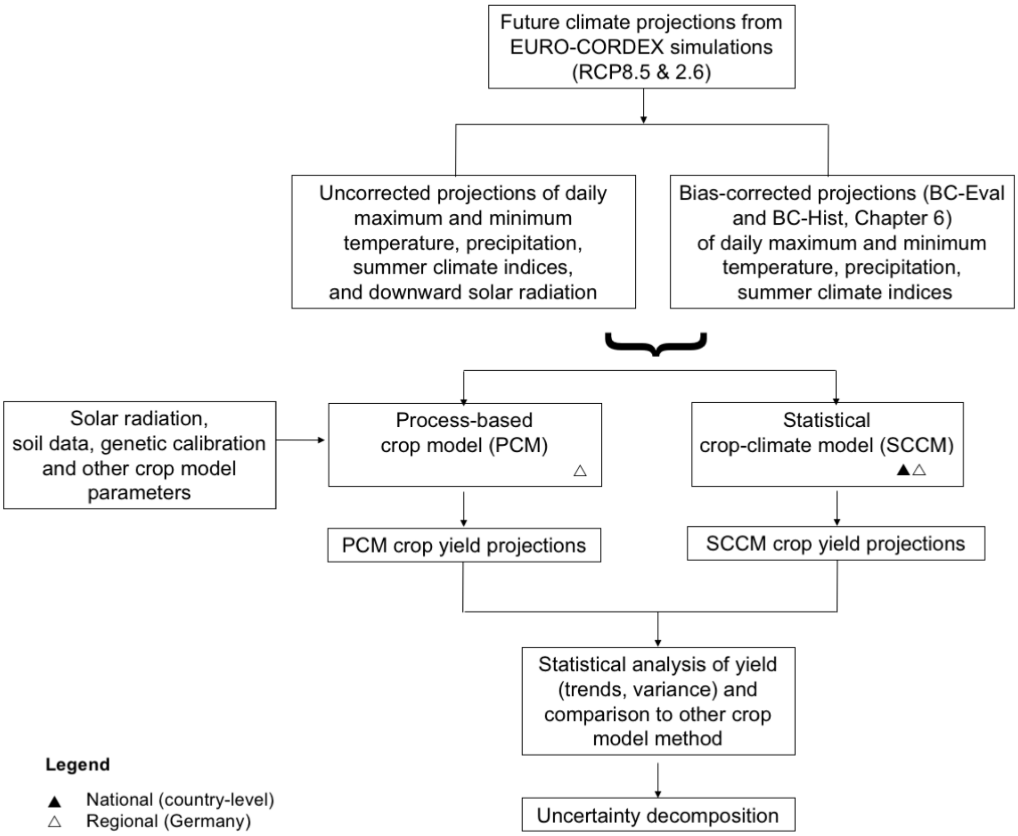



*Figure 7.1: Overview of Chapter 7 research design.*

Climate model output is processed to a regular lat-lon grid over the UK, Germany, and four German states representing South, West, East, and North Germany (NUTS codes DE2, DEA, DED, and DEF). Additional solar radiation data for the PCM is also obtained from EURO-CORDEX under RCP8.5 and RCP2.6. It is not corrected in order to focus on the effect of BC on temperature and precipitation, and because BC of solar radiation can be challenging given the difficulty in obtaining high-quality observational data due to different equipment, their calibration and maintenance, and sparse ground stations (Urraca et al., 2017b). However, emerging datasets and satellite observations show promise in improving the evaluation of solar radiation (Frank et al., 2018, Urraca et al., 2017b,a).

#### 7.2.1.1 Bias correction method

GCM-RCM outputs were bias-corrected in the Chapter 6 using quantile-quantile mapping (QQ mapping), which included a reduction of drizzle days and five-fold cross-validation step, based on the methods from Maraun and Widmann (2015), Teutschbein and Seibert (2012) and Piani et al. (2010). In Chapters 5 and 6, BC historical GCM-RCM and RCM-only evaluation simulations (reanalysis-driven) from a past reference period (1976-2005 and 1981-2010, respectively) were used to calibrate the correction for future climate projections, namely called BC-Hist and BC-Eval. By doing so, the effect on the range of simulations based on the BC calibration was examined more carefully as a method of exploring the error from a GCM-RCM pair.

In this chapter, the climate model output corrected by these two different calibration approaches (BC-Eval and BC-Hist), along with uncorrected climate model projections, all forced by two different emission scenarios (RCP8.5 and 2.6) are used as input to the crop models.

### 7.2.2 Review of selected crop models and their calibration

#### 7.2.2.1 Process-based and statistical crop model

The PCM selected for the study is CERES-Wheat/DSSAT (Jones et al., 2003), which was evaluated in Chapter 3. Crop growth in CERES-Wheat is based on radiation use efficiency (RUE) and crop thermal time/ growing degree-days (GDD), which are computed based on the daily maximum and minimum temperatures (Jones et al., 2003) (See Chapter 3). Similar to previous chapters, crop parameters from Nain and Kersebaum (2007) are used, and yield is only simulated with the PCM at the regional German level. For other PCM parameters, fine-scale soil data (1 km gridded resolution) is taken from a soil database (IRI et al., 2015). Based on a long-term field experiment in Germany (Merbach and Schulz, 2013), optimal levels of mineral-based fertilizers are provided. These crop calibration parameters are maintained for all regions to simplify the comparison of yield responses to temperature and precipitation, and default parameters are maintained whenever possible to facilitate comparison. The $CO_2$ fertilization effect is also not included in simulations, and this is acknowledged to be a limitation of the study. It is not included here for the simplicity of comparing results between the PCM (which can include the $CO_2$ effect) and SCCMs. Challenges in including the $CO_2$ effect remains a significant limitation of statistical approaches to crop modeling (Lobell and Asseng, 2017).

National and regional SCCMs have been developed and evaluated in Chapter 3, based generally on the work of Hawkins et al. (2013a) and Lobell and Burke (2010). The SCCMs for the chosen geographical study areas (both country- and regional-level) are based on heat stress indices for wheat, where $T_H$ is the number of days above 31°C between June and August (JJA), $\bar{P_S}$ is total JJA precipitation. The coefficient of the time trend is maintained as year 2010 from where future yield projections begin. The significance of these climate indices varies per region. For Germany, DE2 and DED (South and East Germany) these are: $T_H$, $\bar{P_S}$ and $T_H \times \bar{P_S}$. The UK and DEA only have $T_H$ and $\bar{P_S}$, while DEF only has $\bar{P_S}$ as a predictor. For more information on both the SCCMs and PCM, see Chapter 3.

#### 7.2.2.2 Revisiting PCM and SCCM evaluation results

In this chapter, the robustness and credibility of the future yield projections are dependent on the skillful performance of the crop model, for example in their evaluation (See Chapter 3 and Challinor et al., 2017). However, it was shown that the while the SCCM showed highly satisfactory performance relative to observed regional yields, the PCM only showed significant correlation for one region (DED). While the PCM simulated median yields for other regions within the range of the observed regional yields in DEA and DEF, it was argued that the use of an input-intensive field-based PCM at the regional scale brings errors due to input and aggregation error (See Chapter 3).

There was some improvement to the PCM simulated yields when the climate input used for CERES-Wheat was corrected with BC (Chapter 5), but its limitations at the regional level still remain. However, the multi-method comparison in this chapter is contigent on the feasibility of the projections from the PCM, which has been previously evaluated and validated for the regional level in Germany (e.g. Nain and Kersebaum, 2007), for where it is used in this chapter. Despite the unimpressive PCM evaluation results in Chapter 3, the PCM is still used here in the chapter – while recognizing its limitations – in order to provide a perspective on how this common approach (e.g. field-scale models are used at larger scales half the time (Challinor et al., 2017)) may add to the uncertainty of yield projections.

### 7.2.3 Yield comparison: statistical methods

The projected yields from both crop modeling approaches are compared to each other in 30-year intervals: 2011-2040, 2041-2070, and 2071-2100 (early, mid and late century, respectively) and to hindcasted yields (with the respective method and past bias-corrected climate model output) as baseline/reference yields from 1976-2005. In essence, this means that BC-Hist projected yields are compared to past BC-Hist yield simulations, BC-Eval projected yields to past BC-Eval yield simulations, respective to each crop model method. Although observations are commonly used as a baseline in impact studies (e.g. Hattermann et al., 2018, Soltani et al., 2016), this comparison allows for an better account of climate model biases.

How yields compare using uncorrected and BC climate projections (BC-Eval and BC-Hist) is also explored. The mean and variance (coefficient of variation or CV, the standard deviation over the mean) of yields are used to assess change relative to the hindcast reference period. Analysis also includes using linear regression and the Student's t-test to identify significant trends and differences in mean relative to the baseline yield, respectively.

### 7.2.4 Uncertainty decomposition

#### 7.2.4.1 Defining the sources of uncertainty

An uncertainty decomposition method is used to characterize and partition uncertainty from different sources in the impact modeling cascade. Following the framing of Ruiz-Ramos and Mínguez (2010), uncertainty in yield projections is generally defined as coming from (1) climate models, (2) crop models, and (3) the methods to link them, and this framework is adopted for the study. This partitioning is used to define the cascade of uncertainty in the impact projections (Fig. 7.2).

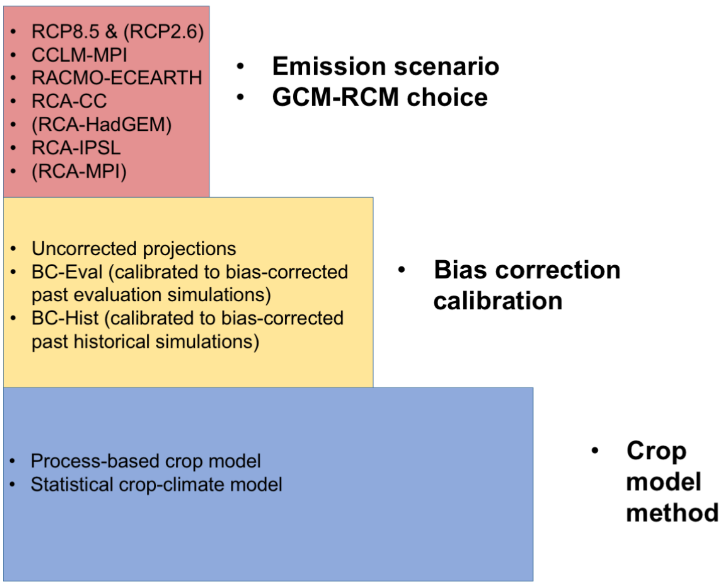



*Figure 7.2: Research model of the cascade of uncertainty in yield projections.*

Following this framing, the uncertainty in yield projections in this chapter is partitioned into three main effects, namely: (1) the climate model (GCM-RCM) choice, (2) the crop modeling method (either PCM or SCCM), and (3) the BC calibration, defined as using either (1) uncorrected, (2) BC-Eval, or (3) BC-Hist climate projections. While considering scenario uncertainty would also provide valuable insight into the analysis, because of the availability of only two GCM-RCMs for the RCP2.6 emission scenario, uncertainty is calculated separately for RCP8.5 and RCP2.6 to avoid imbalanced weighting of the GCMs. In future work, an additional analysis of crop model parameters would also provide a more thorough analysis of what components of the crop modeling method contribute to uncertainty.

For the uncertainty decomposition, the research design has resulted in 48 projections of yield for each German region: 36 for RCP8.5 and 12 for RCP2.6. This is broken down as follows: for RCP8.5, 6 GCM-RCMs $\times$ 3 types of correction calibration $\times$ 2 crop modeling methods = 36 future yield projections; For RCP2.6, 2 GCM-RCMs $\times$ 3 types of correction calibration $\times$ 2 crop modeling methods = 12 future yield projections.

#### 7.2.4.2 ANOVA as uncertainty partitioning method

The uncertainty partitioning method is performed using an analysis of variance (ANOVA) approach to quantify the identified sources of uncertainty in the impact modeling cascade. ANOVA is useful to characterize and split uncertainty into contributing sources and it additionally allows the determination of significant variations in the impact chain (Hattermann et al., 2018, Vetter et al., 2017). ANOVA is a form of statistical hypothesis testing for more than two groups, where the variation between and among groups is tested, where the null hypothesis is that all groups are simply random samples of the same population (Vetter et al., 2017). A similar ANOVA approach has been adopted in a climate model context (e.g. Yip et al., 2011), and in several hydrological impact studies (e.g. Hattermann et al., 2018, Vetter et al., 2017, Bosshard et al., 2013).

Adapting the approach of the Hattermann et al. (2018) study, the total sum of squares (SST) is used to express the total variation that can be attributed to the various factors, identified in Fig. 7.2. SST is calculated as deviations of single yield projections ($Y$) from the grand mean of yield projections ($\bar{Y}$) (Equation 7.1), which is defined per 30-year future period. In this work, due to the scenario availability limitation, the three factors used for variance decomposition are defined as the GCM-RCMs ($N$=6), crop model method (CM) ($N$=2) and BC calibration (BC) ($N$=3) for each RCP.

$$SST = \sum_{i=1}^{N_{GCM}} \sum_{i=1}^{N_{CM}} \sum_{i=1}^{N_{BC}} (Y_{ijk} - \bar{Y})^2 \tag{7.1}$$

Further first- and second-order interactions (e.g. between the GCM-RCM and crop modeling method, GCM and BC) are also included in the ANOVA, with the SST therefore being broken down into smaller sums of squares (SS) (Equation 7.2). In essence, interaction effects in ANOVA indicate that there may be a variable or effect that can influence the relationship between an independent and dependent variable: in this case, yield and the chosen effects. There is some debate on how to interpret main effects when interactions are also significant, and these limitations of using ANOVA are discussed in the next subsection.

$$SST = SS_{GCM} + SS_{CM} + SS_{BC} + SS_{GCM \times CM} + SS_{GCM \times BC} + SS_{CM \times BC} + SS_{GCM \times CM \times BC} \tag{7.2}$$

To interpret ANOVA results, $R$ software ($aov$) is used to calculate the F-test value, which is used for determining the significance of any variation in the levels (GCMs, crop model methods and BC calibration). In this study variance is analyzed over the three future 30-year periods to analyze if uncertainty from these sources also changes over time. Ternary plots ($R$ $ggtern$ (Hamilton, 2018)) are used to represent the fractional uncertainty from each source, per scenario, similar to other studies (e.g. Hattermann et al., 2018, Vetter et al., 2017, Bosshard et al., 2013) from which the uncertainty approach is based.

#### 7.2.4.3 Limitations to uncertainty analysis

There are limitations to the use of ANOVA as an approach to analyze uncertainty. For example, due to the design of the study with a small combination of GCM-RCMs, ANOVA is applied here in fixed factor mode. A factor is fixed when the levels under study are the only ones of interest, and the conclusions drawn from the analysis apply only to this specific setting (Hattermann et al., 2018). The use of ANOVA in a climate model context has also been previously criticized when it was used with an unequal numbers of runs at each GCM-scenario combination (e.g. Yip et al., 2011 criticized by Northrop and Chandler, 2014). Additionally, the study's limited information at the country-level scale means that this uncertainty analysis is only applied to the regional scale in Germany where PCM and SCCM crop projections can be compared. In addition, while ANOVA has been a valuable tool in agronomical studies, there has been some debate how it should be reported (including interaction terms). The use (and misuse) of ANOVA has led to a debate on statistical rigor in science journals (e.g. McIntosh, 2015).

In light of these criticisms of ANOVA approaches, the yield projections under the different RCPs are analyzed separately in this study to prevent imbalanced weighting, although this prevents the analysis of scenario uncertainty jointly with GCM-RCM, BC, and method interactions. To handle interaction terms for the ANOVA in this study, although the focus is given to the main effects (GCM-RCM, BC, and Method), when the interaction term is significant and has a large F-value, these results are also reported as non-negligible. This is because when interaction terms are not considered, the importance of individual uncertainty sources is potentially overestimated (Bosshard et al., 2013). In order to perform a more thorough quantification of interaction terms, multiple realizations of each impact modeling chain combination with different realizations of GCMs, BC, and crop modeling methods (Bosshard et al., 2013, Yip et al., 2011) are needed, which can be done in future work, and therein also investigate the effect of natural variability on uncertainty.

It is clear that in future studies, more available simulations with different GCM-RCMs, and RCPs could make the analysis of uncertainty more robust. However, despite these limitations, it is argued that the analysis of uncertainty in this comparative and interdisciplinary context is relatively novel in the field. Evidence from this analysis can be used to further characterize uncertainty in the impact chain.

#### 7.2.4.4 Objectives and limitations of the comparative approach

The research design of the study is to use a PCM (DSSAT/CERES-Wheat, Jones et al., 2003, see Chapter 3 for a detailed description of its modeled processes) and a climate index-based SCCM (See Chapter 3 for its development and Hawkins et al., 2013a, Lobell and Burke, 2010), existing regional crop calibration parameters for the PCM (Nain and Kersebaum, 2007), and not including the effect of $CO_2$, to be able to feasibly compare SCCM and PCM results. While this design limits the analysis of phenology, $CO_2$ effect, among other non-climate factors, the focus on temperature and precipitation gives insight as to how wheat yields could potentially respond to these important influences.

Additionally, focus is given to the effect of BC, which has been well-discussed in previous chapters as a contentious step in the climate modeling community, but generally accepted in impact analyses. It is crucial here to re-iterate the focus of the research, which is to fin 'value exploring the differences' (e.g. Challinor et al., 2014, p.78) not only between GCMs or crop models, but between the crop modeling methods.

## 7.3 Results

In this section, the yield projections generated by different crop modeling methods are reported and compared, giving focus to before and after using BC climate model output and relative to the respective baseline yield hindcast. The results of the partitioning of uncertainty are also reported. Similar to the previous chapters, the ensemble results are shown here for clarity, but additional information on the individual GCM-RCM-driven yield projections is also provided in the results tables. Again, projected changes are relative to the respective past raw, BC-Eval, BC-Hist simulations (SCCM or PCM).

### 7.3.1 Crop yield projections: comparing the effect of crop model and BC methods

#### 7.3.1.1 National statistical yield projections

At the country level, it can be observed that future yields in the UK are projected to increase relative to baseline yields, and that yields are fairly constant until the end of the century (Fig. 7.4a). Ensemble projections show that yields may increase for the UK under both RCP8.5 and RCP2.6 ($R^2$=0.16 and 0.17, respectively). These gains are fairly small, approximately 1.4 ton increases relative to the baseline yield (raw historical-driven yields) for the

RCP8.5 ensemble mean for all three 30-year future periods (Table 7.1), with larger gains projected under RCP2.6 (1.7 t/ha).

UK yield projections from individual GCM-RCMs also show increases relative to the baseline, with these projected changes generally decreasing after using BC-Hist and BC-Eval climate model output for the SCCM, although projected ensemble mean yield increases are also approximately 1 t/ha: for example, gains under the RCP8.5 ensemble mean are 0.9 t/ha (BC-Eval) and 1.4 t/ha (BC-Hist).

In contrast, large decreases in yield are projected for Germany under the RCP8.5 scenario ($R^2$=0.32) (Fig. 7.4b, Table 7.2). Raw RCA-MPI under RCP2.6 also shows a significant negative trend, although the $R^2$ is small ($R^2$=0.05) and the ensemble RCP2.6 scenario does not show significant trends with raw climate model output. While the projected changes in yield are all positive relative to their respective baseline yield, it can be observed that these projected yield changes decrease across the three 30-year future periods. For example, the projected changes in yield from raw simulations from RCA-MPI show 1.9 t/ha increases relative to the RCA-MPI raw historical yield baseline between 2011-2040.

By the mid-century (2041-2070), this decreases to 1.8 and then 1.3 t/ha by the end of the century. This pattern of yield increases, followed by decreasing yields in the mid- and late- century can generally be observed across all yield projections from individual GCM-RCMs, under both scenarios, and with different BC calibration methods for Germany (Table 7.2). Some significant yield decreases lower than the reference period are projected for the last 30 years of the century by RCA-HadGEM.

In terms of variation, it can be observed that the CV – the extent of variability with regard to the mean – is low for all ensembles of UK yield projections for all BC calibration approaches, which means that yield simulations are fairly close to each other for the UK (Fig. 7.3A). CV is observed to increase for Germany – which means that yields deviate more from the mean – for both future emission scenarios by the late century (2071-2100) (Fig. 7.3B).

Although the results for the country-level in the UK are limited to the statistical approach (due to lack of UK region yield data, including reported calibration parameters), in the next section, yield projections from both the SCCM and the PCM for the four German regions are reported.

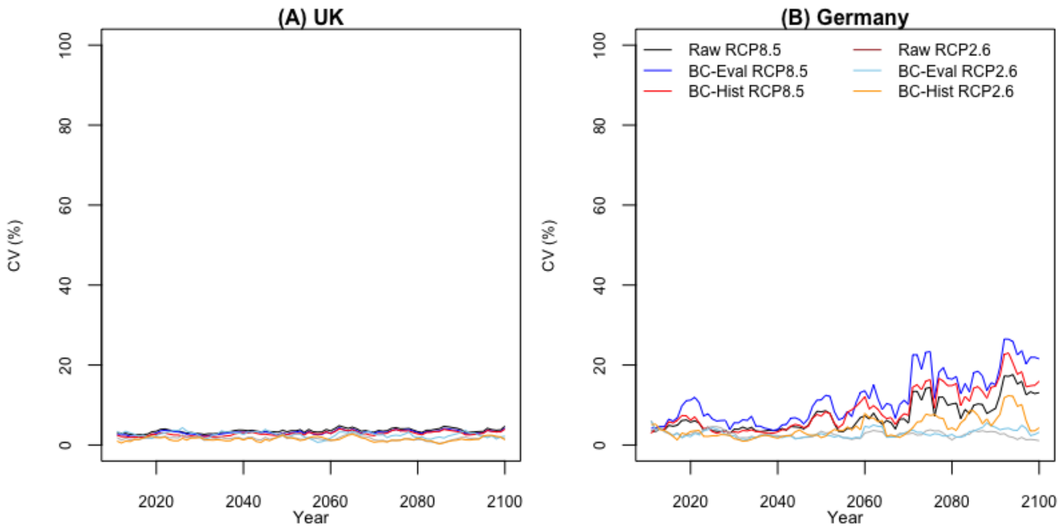



*Figure 7.3: Coefficient of variation for SCCM yield projections, (A) UK and (B) Germany. Lines represent the 5-year simple moving average for the ensemble mean of yield projections, under BC calibration approach (Raw, BC-Eval and BC-Hist), for both the RCP8.5 and RCP2.6 scenario.*

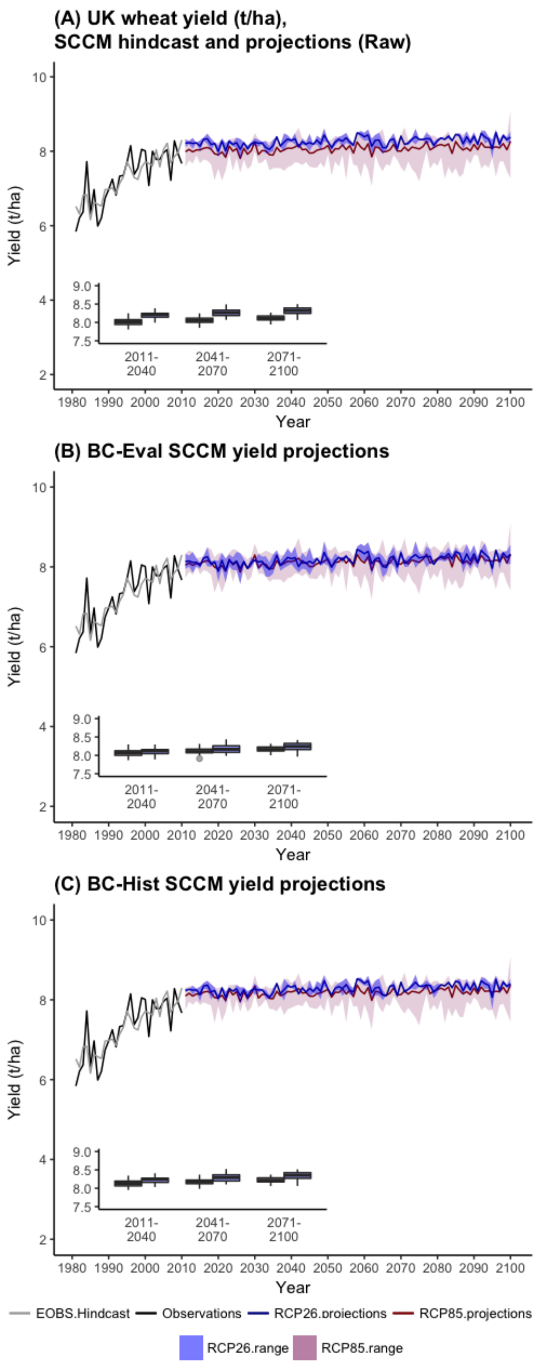



Fig. 7.4a. UK.

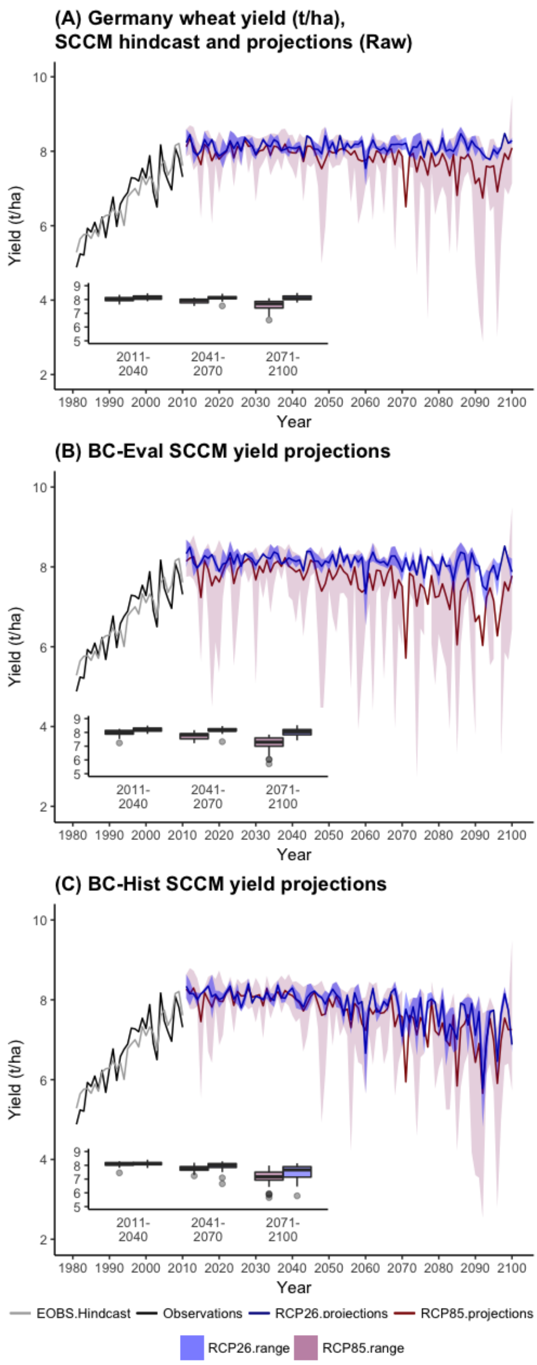



Fig. 7.4b. Germany.

*Figure 7.4: Yield projections with the SCCM: Fig. 7.4a is for the UK and Fig. 7.4b is for Germany. (A) uses uncorrected climate model output, (B) uses BC-Eval and (C) uses BC-Hist using scenarios RCP8.5 and RCP2.6. Inset plot shows the spread of simulations for early, mid, and late century intervals (30-year intervals). The SCCM E-OBS hindcast is also shown as reference of crop model performance.*

*Table 7.1: UK SCCM crop yield projections.*

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **CCLM-MPI** | 1.4 | 1.5 | 1.5 | +, 0.11* | 0.7▼ | 0.8▼ | 0.8▼ | +, 0.13* | 1.4 | 1.4▼ | 1.5 | +, 0.11* |
| | 20.3 | 21.1 | 23.6 | | 9 | 10.3 | 11.1 | | 20 | 20.8 | 22.5 | |
| **RACMO-ECEARTH** | 1.3 | 1.4 | 1.4 | +, 0.06* | 0.9▼ | 1▼ | 1.1▼ | +, 0.07* | 1.3 | 1.3▼ | 1.4 | +, 0.06* |
| | 19.6 | 20.9 | 22.2 | | 12.9 | 13.8 | 14.5 | | 18.6 | 19.5 | 20.6 | |
| **RCA-CC** | 1.3 | 1.2 | 1.2 | -, 0 | 0.5▼ | 0.5▼ | 0.4▼ | -, 0.01 | 1.2▼ | 1.2 | 1.1▼ | -, 0.01 |
| | 18.9 | 18.2 | 17.3 | | 6.9 | 6.4 | 5.9 | | 17.8 | 17.4 | 16.5 | |
| **RCA-HadGEM** | 1.3 | 1.4 | 1.4 | +, 0.03 | 0.9▼ | 1▼ | 1.1▼ | +, 0.03 | 1.3 | 1.3▼ | 1.4 | +, 0.03 |
| | 20.6 | 20 | 21.6 | | 12.9 | 13.5 | 14.4 | | 19 | 19 | 19.8 | |
| **RCA-IPSL** | 1.3 | 1.4 | 1.5 | +, 0.07* | 0.8▼ | 0.9▼ | 1▼ | +, 0.07* | 1.3 | 1.3▼ | 1.4▼ | +, 0.07* |
| | 20.8 | 21 | 21.8 | | 11.4 | 12 | 13.1 | | 18.9 | 19.4 | 20.4 | |
| **RCA-MPI** | 1.4 | 1.5 | 1.6 | +, 0.11* | 0.7▼ | 0.8▼ | 0.9▼ | +, 0.11* | 1.3▼ | 1.4▼ | 1.5▼ | +, 0.11* |
| | 20.2 | 23.5 | 23.4 | | 9.5 | 10.9 | 12.1 | | 19.1 | 21.3 | 21.6 | |
| ***RCP85_Mean*** | 1.3 | 1.4 | 1.4 | +, 0.16* | 0.8▼ | 0.8▼ | 0.9▼ | +, 0.17* | 1.3 | 1.3▼ | 1.4 | +, 0.16* |
| | 20.1 | 21.5 | 21.6 | | 10.4 | 11.2 | 11.9 | | 18.6 | 19.8 | 20.2 | |
| **RCA-HadGEM_RCP26** | 1.3 | 1.4 | 1.5 | +, 0.11* | 0.7▼ | 0.8▼ | 0.8▼ | +, 0.13* | 1.3 | 1.4 | 1.4▼ | +, 0.11* |
| | 19.7 | 22 | 22.3 | | 9.1 | 10.4 | 11.2 | | 18.8 | 20.3 | 20.9 | |
| **RCA-MPI_RCP26** | 1.8 | 1.8 | 1.9 | +, 0.06* | 0.9▼ | 1▼ | 1.1▼ | +, 0.07* | 1.4▼ | 1.5▼ | 1.6▼ | +, 0.06* |
| | 26 | 28.3 | 28.2 | | 12.9 | 13.7 | 14.5 | | 20.9 | 22.2 | 22.7 | |
| ***RCP26_Mean*** | 1.6 | 1.6 | 1.7 | +, 0.17* | 0.8▼ | 0.9▼ | 0.9▼ | +, 0.19* | 1.4▼ | 1.4▼ | 1.5▼ | +, 0.16* |
| | 22.9 | 25.1 | 25.2 | | 11 | 12.1 | 12.8 | | 19.9 | 21.3 | 21.8 | |

*Table 7.2: Germany SCCM crop yield projections.*

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **CCLM-MPI** | 2 | 2.1 | 2.1 | +, 0.01 | 1.4▼ | 1.4▼ | 1.4▼ | +, 0.01 | 1.9▼ | 1.7▼ | 1.3▼ | -, 0.21* |
| | 33.5 | 33.1 | 35.1 | | 20.3 | 21.4 | 21.4 | | 29.9 | 28.5 | 21.4 | |
| **RACMO-ECEARTH** | 2 | 1.9 | 1.9 | -, 0.05* | 1.7▼ | 1.5▼ | 1.3▼ | -, 0.18* | 2.1▲ | 1.8▼ | 1.4▼ | -, 0.25* |
| | 31.8 | 30.2 | 30.4 | | 25.2 | 22.5 | 19.3 | | 34.5 | 31.9 | 23.3 | |
| **RCA-CC** | 1.9 | 1.9 | 1.7 | -, 0.1* | 1.4▼ | 1.3▼ | 1.1▼ | -, 0.18* | 2▲ | 2▲ | 1.6▼ | -, 0.31* |
| | 31.4 | 31.9 | 27.6 | | 20.5 | 20 | 16.9 | | 33.5 | 31.5 | 27.2 | |
| **RCA-HadGEM** | 1.7 | 1.1 | 0.1 | -, 0.22* | 0.8▼ | 0.1▼ | -1.4▼ | -, 0.22* | 2.1▲ | 1.5▲ | 0.3▲ | -, 0.27* |
| | 28.8 | 18 | 1.1 | | 12.4 | 1.8 | -20.5 | | 34.8 | 24.6 | 4.9 | |
| **RCA-IPSL** | 1.9 | 1.8 | 1.5 | -, 0.04* | 1.6▼ | 1.2▼ | 0.6▼ | -, 0.12* | 2.6▲ | 2▲ | 1.1▼ | -, 0.21* |
| | 31.5 | 29.6 | 24.8 | | 23.8 | 18.6 | 9 | | 43.4 | 34.1 | 17.5 | |
| **RCA-MPI** | 1.9 | 1.8 | 1.3 | -, 0.16* | 1.3▼ | 1.1▼ | 0.3▼ | -, 0.23* | 2▲ | 1.7▼ | 0.8▼ | -, 0.27* |
| | 29.9 | 31.1 | 20.9 | | 20.2 | 17 | 4.4 | | 33.5 | 29 | 13.1 | |
| ***RCP85_Mean*** | 1.9 | 1.8 | 1.4 | -, 0.32* | 1.3▼ | 1.1▼ | 0.5▼ | -, 0.41* | 2.1▲ | 1.8 | 1.1▼ | -, 0.53* |
| | 30.3 | 29.1 | 23.2 | | 20 | 16.5 | 7.9 | | 37.6 | 30 | 18 | |
| **RCA-HadGEM_RCP26** | 1.8 | 1.8 | 1.8 | +, 0.01 | 1.5▼ | 1.5▼ | 1.5▼ | +, 0.01 | 2.2▲ | 2▲ | 1.6▼ | -, 0.21* |
| | 28.2 | 30 | 29.4 | | 22.4 | 23.4 | 22.9 | | 36.1 | 33.7 | 26.2 | |
| **RCA-MPI_RCP26** | 2.2 | 2.1 | 2 | -, 0.05* | 1.7▼ | 1.6▼ | 1.4▼ | -, 0.18* | 2.2 | 1.9▼ | 1.5▼ | -, 0.25* |
| | 34.5 | 33.8 | 32.8 | | 26.4 | 23.8 | 20.4 | | 36.2 | 31.7 | 25 | |
| ***RCP26_Mean*** | 2 | 1.9 | 1.9 | -, -0.01 | 1.5▼ | 1.5▼ | 1.4▼ | -, 0.09* | 2.2▲ | 1.9 | 1.5▼ | -, 0.34* |
| | 31.3 | 31.9 | 31.1 | | 23.4 | 22.6 | 20.6 | | 36.2 | 32.7 | 25.6 | |

*Projected changes (t/ha) are in white rows and percentage in gray. All changes are relative to the respective baseline yield hindcast. ($*$) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

### 7.3.2 Regional comparison of crop yield projections

In this section, key changes and results are highlighted per region to answer the research questions: how do different methods project yield changes? Does the BC method also affect the sign and/or magnitude of projected changes?

#### 7.3.2.1 Projected changes and trends in yield and its variability

**(1) DE2, South Germany**

Yield projections for this region show significant negative trends across the three 30-year future periods when using the SCCM with GCM-RCM output (Uncorrected, BC-Eval and BC-Hist) for RCP8.5: $R^2$=0.45, 0.35 and 0.5, respectively (Fig. 7.6a, Table 7.3). Uncorrected and BC-Eval ensemble yield projections under RCP2.6 show small but significant positive trends ($R^2$<0.06) while RCP2.6 BC-Hist yields show a negative trend ($R^2$=0.19). Among the GCM-RCMs, uncorrected RCA-HadGEM used for the SCCM typically projects large negative yield changes, with a change of -2.9 t/ha (relative to baseline yield) projected by the end of the century. RCA-MPI, RCA-IPSL, and the RCP8.5 ensemble mean also project decreases in yield that would bring mean yields lower than the reference period for all three BC calibrations. Generally, it can be observed that very poor yields are anticipated in the latter part of the century with RCP8.5.

For the PCM projections (Fig. 7.6b), although the trends in yields are significant, they have small $R^2$ values: for example, the RCP8.5 ensemble mean using raw GCM-RCM output ($R^2$=0.07). RCP8.5 BC-Eval or BC-Hist do not show any significant trends. Ensemble RCP2.6 projections also show small but significant positive linear trends ($R^2$<0.06) for all BC approaches (Table 7.4). The projected changes in yield from individual GCM-RCMs are largely negative when raw GCM-RCM output is used, although these losses are quite small (under 1 t/ha). However, after BC, most of these projected changes increase relative to the yield projections made with raw GCM-RCM output. For example, after BC-Eval, yield changes are mostly positive (although under 1 t/ha), apart from RCA-HadGEM, which still projects yield losses of about 0.8 (0.1) t/ha RCP8.5 (RCP2.6) from 2071-2100. In contrast, after BC-Hist, RCA-HadGEM projects positive yield changes under both scenarios.

The t-test results show that most SCCM mean yield projections are significantly different to the baseline, apart from RCP8.5 for the middle of the century (2041-2070) for all BC (See Table 7.11 at the end of the Results). Based on the projected changes and t-test results, unlike the SCCM projections where yields steadily decrease, RCP8.5 ensemble PCM yields tend to show small increases, with more variation in the mid-century. This is reflected in the analysis of CV (Fig. 7.5 I), where PCM yields show more even deviation relative to the mean. The CV for the SCCM yields, in contrast, steadily grows towards more variation for RCP8.5.

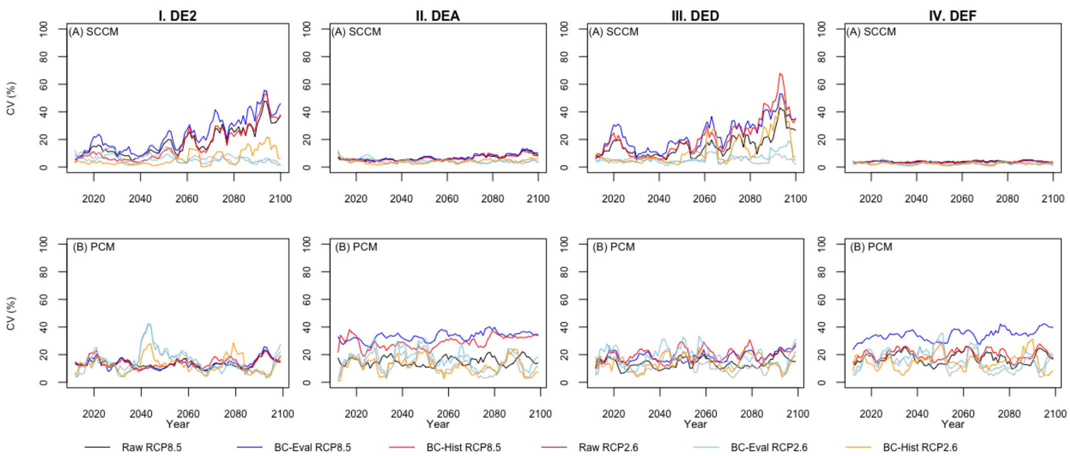



*Figure 7.5: Coefficient of variation for SCCM and PCM yield projections, German regions (I-IV). Lines represent the 5-year simple moving average for the ensemble mean of yield projections with a different BC calibration approach (Raw, BC-Eval and BC-Hist), for both the RCP8.5 and RCP2.6 scenario.*

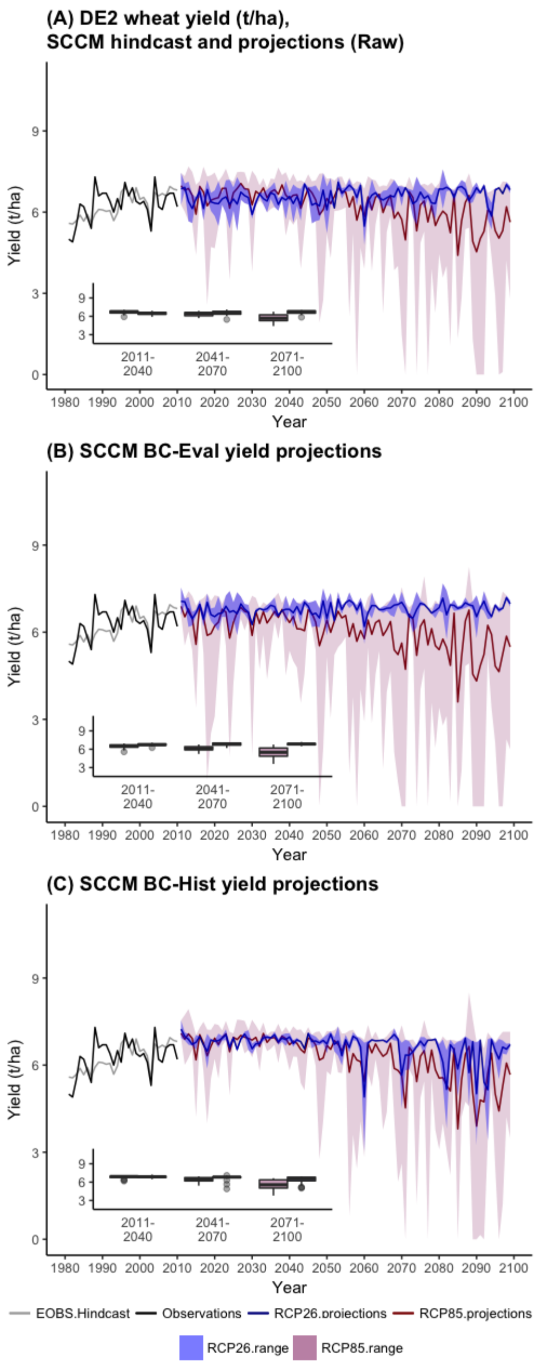



*Fig. 7.6a. DE2 SCCM.*

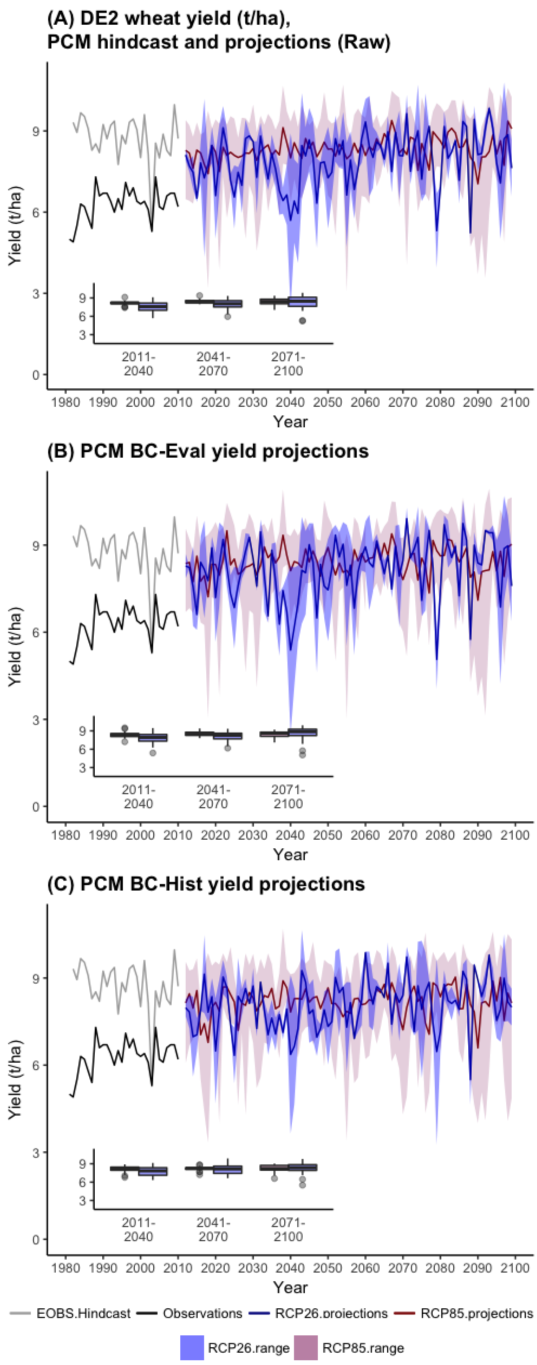



*Fig. 7.6b. DE2 PCM.*

*Figure 7.6: Yield projections for DE2 (South Germany): Fig. 7.6a shows the SCCM projections and Fig. 7.6b shows the PCM projections. (A) uses uncorrected climate model output, (B) uses BC-Eval and (C) uses BC-Hist. Inset plot shows spread of simulations for early, mid, and late century (30-year intervals). The SCCM and PCM E-OBS hindcasts are also shown as reference of crop model performance.*

Table 7.3: South Germany (DE2) SCCM crop yield projections.

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **CCLM-MPI** | 0.9 | 1.1 | 1.4 | +, 0.15* | 0.4▼ | 0.5▼ | 0.6▼ | +, 0.08* | 1▲ | 0.8▼ | 0.3▼ | -, 0.14* |
| | 18.4 | 18.6 | 23.3 | | 6.6 | 8.4 | 9.8 | | 16.3 | 14.4 | 5.9 | |
| **RACMO-ECEARTH** | 0.9 | 0.9 | 0.8 | -, 0 | 0.8▼ | 0.8▼ | 0.8 | -, -0.01 | 1▲ | 0.9 | 0.7▼ | -, 0.05* |
| | 14.2 | 13.8 | 13.3 | | 12.9 | 13.4 | 13 | | 19.1 | 17.4 | 11.7 | |
| **RCA-CC** | 1 | 1 | 0.8 | -, 0.09* | 0.7▼ | 0.7▼ | 0.8 | -, -0.01 | 1.3▲ | 1.2▲ | 1.2▲ | -, 0 |
| | 17 | 18.4 | 13.2 | | 12.3 | 11.6 | 12.3 | | 22.8 | 20.3 | 20.7 | |
| **RCA-HadGEM** | 0 | -1 | -2.9 | -, 0.29* | -0.5▼ | -1.5▲ | -3.2▲ | -, 0.22* | 1.5▲ | 0.7▲ | -1.3▼ | -, 0.32* |
| | 0.8 | -15.7 | -46.3 | | -9.2 | -24.5 | -54.2 | | 26.2 | 12.6 | -26 | |
| **RCA-IPSL** | 0.7 | 0 | -1 | -, 0.21* | 0.6▼ | -0.1▼ | -1.3▲ | -, 0.22* | 2▲ | 1.1▲ | -0.4▼ | -, 0.3* |
| | 10.9 | 0 | -18.4 | | 10.7 | -2.1 | -21.3 | | 35.3 | 19.3 | -6.4 | |
| **RCA-MPI** | 0.7 | 0.4 | -0.6 | -, 0.22* | 0.7 | 0.3▼ | -0.6 | -, 0.14* | 1.1▲ | 0.7▲ | -0.4▼ | -, 0.21* |
| | 12.3 | 6.6 | -9.3 | | 11.7 | 4.7 | -9.3 | | 19.6 | 12.1 | -7.4 | |
| ***RCP85_Mean*** | 0.7 | 0.4 | -0.2 | -, 0.45* | 0.4▼ | 0▼ | -0.6▲ | -, 0.35* | 1.3▲ | 0.9▲ | 0▲ | -, 0.5* |
| | 11.3 | 6.3 | -4.1 | | 6.4 | 0.8 | -9.3 | | 26.8 | 15.6 | 0.5 | |
| **RCA-HadGEM_RCP26** | -0.1 | 0.1 | 0.3 | +, 0.15* | 0.6▲ | 0.7▲ | 0.8▲ | +, 0.08* | 1.6▲ | 1.4▲ | 0.9▲ | -, 0.14* |
| | -1.3 | 1.4 | 5.5 | | 9.9 | 11.8 | 12.5 | | 29.4 | 25.4 | 17 | |
| **RCA-MPI_RCP26** | 0.6 | 0.6 | 0.6 | -, 0 | 1▲ | 1.1▲ | 1▲ | -, -0.01 | 1.2▲ | 1▲ | 0.8▲ | -, 0.05* |
| | 10.4 | 10.3 | 9.5 | | 17.2 | 17.7 | 16.9 | | 22.2 | 17.9 | 15.1 | |
| ***RCP26_Mean*** | 0.3 | 0.4 | 0.5 | +, 0.06* | 0.7▲ | 0.7▲ | 0.8▲ | +, 0.04* | 1.4▲ | 1.2▲ | 0.9▲ | -, 0.19* |
| | 4.6 | 5.8 | 7.5 | | 11.1 | 12.3 | 12.3 | | 25.8 | 21.6 | 16 | |

Table 7.4: South Germany (DE2) PCM crop yield projections

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **CCLM-MPI** | -0.9 | -0.7 | -0.8 | +, 0.02 | 0.1▲ | 0.3▲ | 0.1▲ | +, -0.01 | -0.8▲ | -0.6▲ | -0.5▲ | +, 0 |
| | -10.6 | -8 | -9.1 | | 0.8 | 3 | 1.2 | | -9.2 | -8.1 | -6.3 | |
| **RACMO-ECEARTH** | -0.5 | -0.2 | -0.1 | +, 0.01 | 0.2▲ | 0.1▲ | 0.4▲ | +, -0.01 | 0.8▲ | 0.4▲ | 0.6▲ | -, 0 |
| | -6.1 | -2.3 | -1.6 | | 2 | 1.6 | 4.7 | | 10.4 | 5.5 | 6.7 | |
| **RCA-CC** | 0 | 0.2 | 0.2 | +, 0.03 | 0.5▲ | 0.7▲ | 0.7▲ | +, 0.01 | 0.2▲ | 0.4▲ | 0.7▲ | +, 0.03* |
| | -0.3 | 2.6 | 2.6 | | 6.3 | 8.5 | 8.1 | | 2 | 4.6 | 9.3 | |
| **RCA-HadGEM** | -0.2 | 0.3 | -0.2 | +, -0.01 | -0.8▲ | -0.3▼ | -0.8▲ | +, -0.01 | 0▲ | 0.6▲ | 0.3▲ | +, -0.01 |
| | -2.7 | 3.5 | -2.4 | | -9.7 | -3.3 | -10 | | 0.1 | 8 | 3.5 | |
| **RCA-IPSL** | -0.1 | 0.1 | 0.5 | +, 0.04* | 0.2▲ | 0.2▲ | 0.5 | +, 0.01 | 0.7▲ | 0.6▲ | 0.7▲ | +, -0.01 |
| | -1.6 | 1.4 | 5.3 | | 2.8 | 2.4 | 5.8 | | 8.6 | 8.1 | 8.6 | |
| **RCA-MPI** | -0.1 | 0 | 0 | +, -0.01 | 0.6▲ | 0.7▲ | 0.5▲ | -, -0.01 | 0.1▲ | 0.1▲ | 0 | +, -0.01 |
| | -0.7 | -0.3 | -0.3 | | 7 | 8.6 | 6.3 | | 0.8 | 1.2 | 0.3 | |
| ***RCP85_Mean*** | -0.3 | -0.1 | -0.1 | +, 0.07* | 0.1▲ | 0.2▲ | 0.2▲ | +, 0.01 | 0.2▲ | 0.3▲ | 0.3▲ | +, 0 |
| | -3.8 | -0.6 | -0.8 | | 0.7 | 2.6 | 1.8 | | 2.1 | 3 | 3.7 | |
| **RCA-HadGEM_RCP26** | -0.3 | -0.4 | 0.2 | +, 0.02 | -0.6▼ | -0.7▼ | -0.1▼ | +, 0 | 0.2▲ | 0.2▲ | 0.7▲ | +, 0.01 |
| | -4.6 | -4.1 | 3 | | -7.7 | -8.6 | -1.7 | | 3 | 1.9 | 8.7 | |
| **RCA-MPI_RCP26** | -0.6 | 0 | 0 | +, 0.04* | 0▲ | 0.7▲ | 0.8▲ | +, 0.05* | -0.6 | 0.1▲ | -0.1▼ | +, 0.03 |
| | -8.6 | 0.6 | 0.2 | | 0.3 | 8.4 | 9.2 | | -8.1 | 1.3 | -1.3 | |
| ***RCP26_Mean*** | -0.5 | -0.2 | 0.1 | +, 0.06* | -0.4▼ | -0.2 | 0.2▲ | +, 0.04* | -0.2▼ | 0.1▲ | 0.3▲ | +, 0.05* |
| | -6.6 | -1.8 | 1.6 | | -5.5 | -1.9 | 2 | | -2.6 | 1.6 | 3.7 | |

*Projected changes (t/ha) are in white rows and percentage in gray. All changes are relative to the respective baseline yield hindcast. ($*$) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

**(2) DEA, West Germany**

A comparison of yield projections shows that the spread of PCM yield simulations after BC is larger than both the SCCM range and the raw PCM yield simulation range (Figs 7.7a and 7.7b). This large contrast in variation around the mean can also be observed in the CV of yields (Fig. 7.5 II).

In terms of linear trends, the RCP8.5 ensemble mean of SCCM projected yields shows significant negative trends under all BC (Table 7.5). This is in contrast to the SCCM yield projections with uncorrected GCM-RCM output: only RCA-CC, RCA-HadGEM, RCA-MPI and the ensemble mean have significant trends. Individual SCCM GCM-RCM yield projections also all show significant negative trends, apart from BC-Eval CCLM-MPI. PCM yield projections only show positive significant trends for RCP2.6 yield projections, but BC-Eval PCM projections show negative trends for RCA-MPI and the RCP8.5 mean ($R^2$=0.17 and 0.04), and positive trends for all BC-Eval RCP2.6 yield projections (Table 7.6). Under BC-Hist, only RCA-HadGEM shows a significant negative (positive) trend with the PCM under RCP8.5 (2.6) (both with $R^2$=0.09).

In terms of changes, most GCM-RCMs used as input to the SCCM and PCM project yield losses under RCP8.5. For the SCCM, this is shown as 'decreasing increases' in yield while the PCM generally projects yield changes lower than the baseline. For example, with the SCCM, CCLM-MPI projects initial increases of 1.8 t/ha (0.9 and 1.6 for BC-Eval and BC-Hist) which fall in the next two 30-year future periods. CCLM-MPI used with the PCM also projects growing losses, but generally lower than the baseline yields: -0.6, 0.2, and -0.8 t/ha (Uncorrected, BC-Eval and BC-Hist) by the end of the century. Some large projected decreases relative to their respective baseline are from BC-Eval RCA-MPI (between 4-5 t/ha losses), and BC-Hist RCA-HadGEM (between 3-4 t/ha losses).

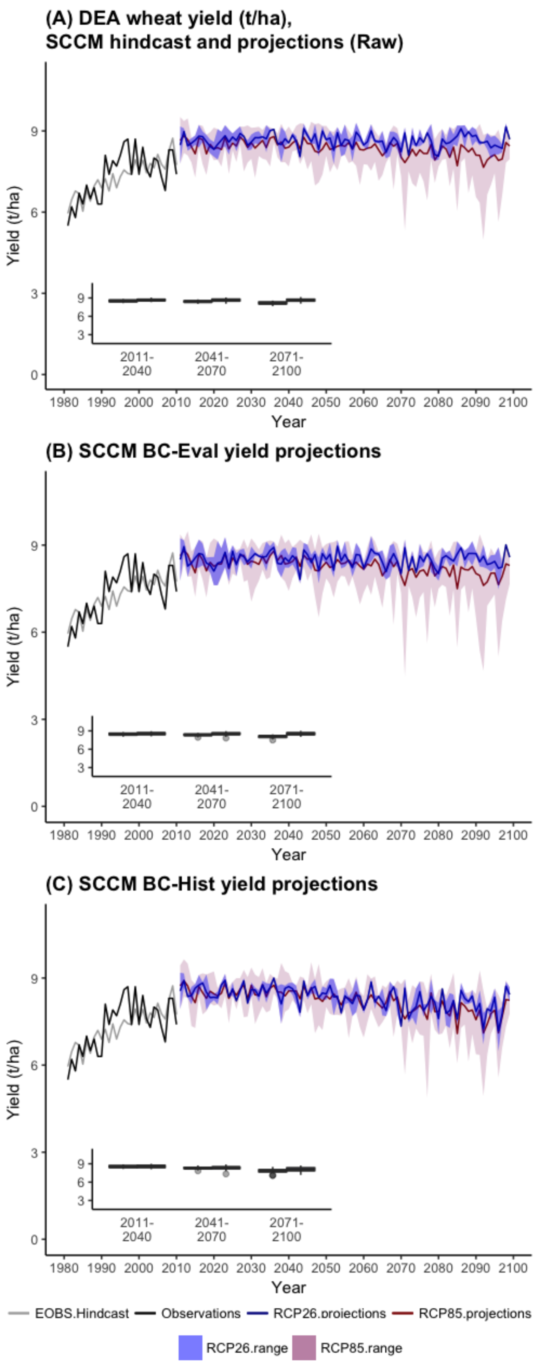



Fig. 7.7a. DEA SCCM.

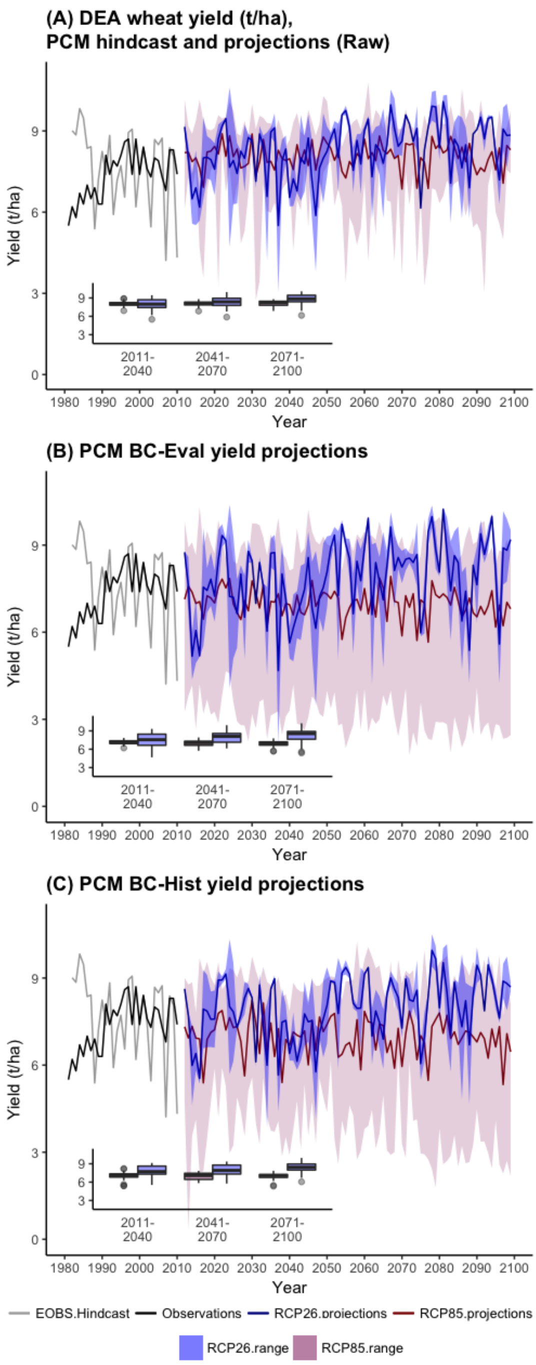



Fig. 7.7b. DEA PCM.

*Figure 7.7: Yield projections for DEA (West Germany): Fig. 7.7a shows the SCCM projections and Fig. 7.7b shows the PCM projections. (A) uses uncorrected climate model output, (B) uses BC-Eval and (C) uses BC-Hist. Inset plot shows spread of simulations for early, mid, and late century (30-year intervals). The SCCM and PCM E-OBS hindcasts are also shown as reference of crop model performance.*

Table 7.5: West Germany (DEA) SCCM crop yield projections.

| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CCLM-MPI** | 1.8 | 1.9 | 1.8 | +, 0 | 0.9▼ | 1.1▼ | 1.2▼ | +, 0.1* | 1.6▼ | 1.5▼ | 1.2▼ | -, 0.12* |
| | 26.2 | 26.2 | 27.9 | | 12.5 | 14.7 | 16.1 | | 23.3 | 21.8 | 18.3 | |
| **RACMO-ECEARTH** | 1.7 | 1.6 | 1.6 | -, 0.02 | 1.5▼ | 1.3▼ | 1.2▼ | -, 0.1* | 1.9▲ | 1.6 | 1.3▼ | -, 0.23* |
| | 24.7 | 22.4 | 23.5 | | 21 | 18 | 16.9 | | 28 | 21.3 | 18.6 | |
| **RCA-CC** | 1.5 | 1.5 | 1.2 | -, 0.06* | 1▼ | 0.9▼ | 0.6▼ | -, 0.1* | 1.5 | 1.5 | 0.9▼ | -, 0.22* |
| | 21.8 | 21.9 | 16.8 | | 13.1 | 12.7 | 8.1 | | 22.6 | 21.7 | 13.3 | |
| **RCA-HadGEM** | 1.6 | 1.1 | 0.5 | -, 0.32* | 1.1▼ | 0.6▼ | -0.2▼ | -, 0.29* | 2▲ | 1.6▲ | 0.9▲ | -, 0.36* |
| | 24.4 | 16.4 | 6.9 | | 15.4 | 8.7 | -2.1 | | 30.5 | 23.3 | 11.7 | |
| **RCA-IPSL** | 1.8 | 1.7 | 1.5 | -, 0.02 | 1.5▼ | 1.4▼ | 1.1▼ | -, 0.06* | 1.6▼ | 1.3▼ | 0.8▼ | -, 0.16* |
| | 26 | 24.4 | 21.9 | | 20.8 | 18.9 | 14.5 | | 22.7 | 18.7 | 11.2 | |
| **RCA-MPI** | 1.6 | 1.6 | 1.2 | -, 0.1* | 1.1▼ | 1▼ | 0.6▼ | -, 0.14* | 1.5▼ | 1.4▼ | 0.8▼ | -, 0.22* |
| | 22.2 | 23.5 | 17.2 | | 14.6 | 13.4 | 8.3 | | 23 | 20.7 | 12.6 | |
| ***RCP85_Mean*** | 1.7 | 1.5 | 1.3 | -, 0.33* | 1.2▼ | 1▼ | 0.7▼ | -, 0.4* | 1.7 | 1.5 | 1▼ | -, 0.56* |
| | 23.6 | 22.9 | 18.9 | | 16.1 | 14.3 | 10.1 | | 23.2 | 21.2 | 14.2 | |
| **RCA-HadGEM_RCP26** | 1.6 | 1.6 | 1.6 | +, 0 | 1▼ | 1.1▼ | 1.3▼ | +, 0.1* | 1.9▲ | 1.7▲ | 1.5▼ | -, 0.12* |
| | 22.4 | 24.3 | 23.7 | | 13.8 | 15.8 | 17.2 | | 27.9 | 24.9 | 21.4 | |
| **RCA-MPI_RCP26** | 2 | 1.9 | 1.9 | -, 0.02 | 1.5▼ | 1.3▼ | 1.2▼ | -, 0.1* | 1.7▼ | 1.4▼ | 1.1▼ | -, 0.23* |
| | 29.3 | 28 | 27.8 | | 20.8 | 17.8 | 16.7 | | 25.6 | 20.3 | 16.5 | |
| ***RCP26_Mean*** | 1.8 | 1.8 | 1.8 | -, -0.01 | 1.2▼ | 1.2▼ | 1.2▼ | -, -0.01 | 1.8 | 1.6▼ | 1.3▼ | -, 0.29* |
| | 25.8 | 26.2 | 25.7 | | 17 | 16.5 | 16.6 | | 26.7 | 22.6 | 19 | |
| | **Uncorrected** | | | | **BC-Eval** | | | | **BC-Hist** | | | |

Table 7.6: West Germany (DEA) PCM crop yield projections.

| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CCLM-MPI** | -0.7 | -0.8 | -0.6 | -, -0.01 | 0.3▲ | 0.3▲ | 0.2▲ | -, 0 | -1.2▼ | -1.1▼ | -0.8▼ | +, -0.01 |
| | -8.1 | -9.8 | -7.3 | | 4.4 | 4.1 | 2.7 | | -15.6 | -15.7 | -10.4 | |
| **RACMO-ECEARTH** | -0.5 | -0.5 | -0.3 | +, -0.01 | -0.4▼ | -0.7▲ | -0.6▲ | -, -0.01 | 0▲ | -0.2▼ | 0.1▲ | +, -0.01 |
| | -6.5 | -6.1 | -3.9 | | -5.5 | -9.1 | -7.1 | | -0.2 | -2.5 | 1.5 | |
| **RCA-CC** | 0.1 | 0 | -0.1 | -, 0.01 | 0.6▲ | 0.7▲ | 0.6▲ | +, -0.01 | 0.2▲ | 0.4▲ | 0.5▲ | +, 0 |
| | 1.2 | 0.3 | -1 | | 7.5 | 9.6 | 7.5 | | 2.3 | 4.9 | 6.3 | |
| **RCA-HadGEM** | -0.3 | 0 | -0.4 | -, -0.01 | -1▲ | -0.8▼ | -1.2▲ | -, -0.01 | -3.3▲ | -3.1▼ | -4.1▲ | -, 0.09* |
| | -3.6 | -0.3 | -4.9 | | -13.1 | -10.3 | -15.4 | | -41.2 | -42.8 | -62.3 | |
| **RCA-IPSL** | 0.2 | 0 | 0.2 | -, -0.01 | 0.5▲ | 0 | 0.4▲ | -, -0.01 | 1.5▲ | 0.6▲ | 1▲ | -, 0.01 |
| | 2.2 | -0.3 | 1.9 | | 6.4 | 0.1 | 4.8 | | 17.8 | 8.6 | 12.1 | |
| **RCA-MPI** | -0.4 | -0.1 | -0.2 | +, -0.01 | -4.4▲ | -4.7▲ | -5.2▲ | -, 0.17* | -0.5▲ | -0.4▲ | -0.4▲ | -, -0.01 |
| | -4.5 | -1.2 | -2.9 | | -55.6 | -59.3 | -65.7 | | -6.4 | -5.3 | -5.6 | |
| ***RCP85_Mean*** | -0.3 | -0.2 | -0.3 | -, -0.01 | -0.7▲ | -0.8▲ | -0.9▲ | -, 0.04* | -0.6▲ | -0.6▲ | -0.6▲ | -, 0 |
| | -3.2 | -2.7 | -3 | | -8.5 | -10 | -11.5 | | -8.5 | -7.8 | -8.4 | |
| **RCA-HadGEM_RCP26** | -0.1 | 0.2 | 0.8 | +, 0.07* | -1.1▼ | -0.5▼ | 0▼ | +, 0.06* | 0.2▲ | 0.3▲ | 1.2▲ | +, 0.09* |
| | -1.5 | 1.9 | 10.2 | | -13.4 | -6.5 | -0.1 | | 2.2 | 3.6 | 14.8 | |
| **RCA-MPI_RCP26** | -0.4 | 0 | 0.2 | +, 0.04* | 0▲ | 0.6▲ | 0.7▲ | +, 0.04* | -0.5▲ | 0 | 0▼ | +, 0.02 |
| | -5.7 | 0.3 | 2.1 | | -0.1 | 7.1 | 8.7 | | -6.8 | -0.4 | -0.1 | |
| **RCP26_Mean** | -0.3 | 0.1 | 0.5 | +, 0.11* | -0.4▲ | 0.2▲ | 0.5 | +, 0.09* | -0.2▼ | 0.1 | 0.6▲ | +, 0.1* |
| | -3.6 | 1.1 | 6.2 | | -4.9 | 2.1 | 6.2 | | -2.3 | 1.6 | 7.4 | |
| | **Uncorrected** | | | | **BC-Eval** | | | | **BC-Hist** | | | |

*Projected changes (t/ha) are in white rows and percentage in gray. All changes are relative to the respective baseline yield hindcast. (*) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

**(3) DED, East Germany**

DED ensemble mean yield projections from the SCCM under RCP8.5 show significant negative trends with all raw and BC GCM-RCM output (Fig. 7.8b, Table 7.7). Before BC, only the RCP8.5 SCCM yield projections from RACMO-ECEARTH, RCA-HadGEM, RCA-MPI, and RCP2.6 RCA-MPI had significant negative trends. After BC-Eval, RCP8.5 RCA-IPSL also gains a significant negative trend, and after BC-Hist all individual GCM-RCMs yield projections apart from RCA-CC have significant negative trends.

Projections with the SCCM for RCP2.6 also show significant negative trends, with the $R^2$ value growing after BC, e.g. the RCP26 mean has $R^2$=0.03, 0.11, and 0.24 for raw, BC-Eval and BC-Hist yield projections. Similar to other regions, DED SCCM yield projections generally show 'decreasing increases', meaning that after an initial increase in the early future period, these yield gains fall by the end of the century. In the case of RCP8.5, Raw, BC-Eval and BC-Hist RCA-HadGEM, BC-Eval and BC-Hist RCA-MPI, these projected changes are negative for 2071-2100.

In contrast, few significant trends are observed for PCM yields, but raw projected changes are mostly negative, although these are typically less than 1 t/ha (Table 7.8). After BC, negative yield losses respective to the baseline yield are still projected for BC-Hist CCLM-MPI, BC-Eval and BC-Hist RACMO-ECEARTH, BC-Eval and BC-Hist RCA-HadGEM, and BC-Hist RCA-MPI, with the largest yield losses at around 1 t/ha for BC-Eval RCA-HadGEM. Similarly, RCA-HadGEM SCCM yield projections also projected the largest losses. SCCM projections also show large variability compared to PCM CV (Fig. 7.5 III).

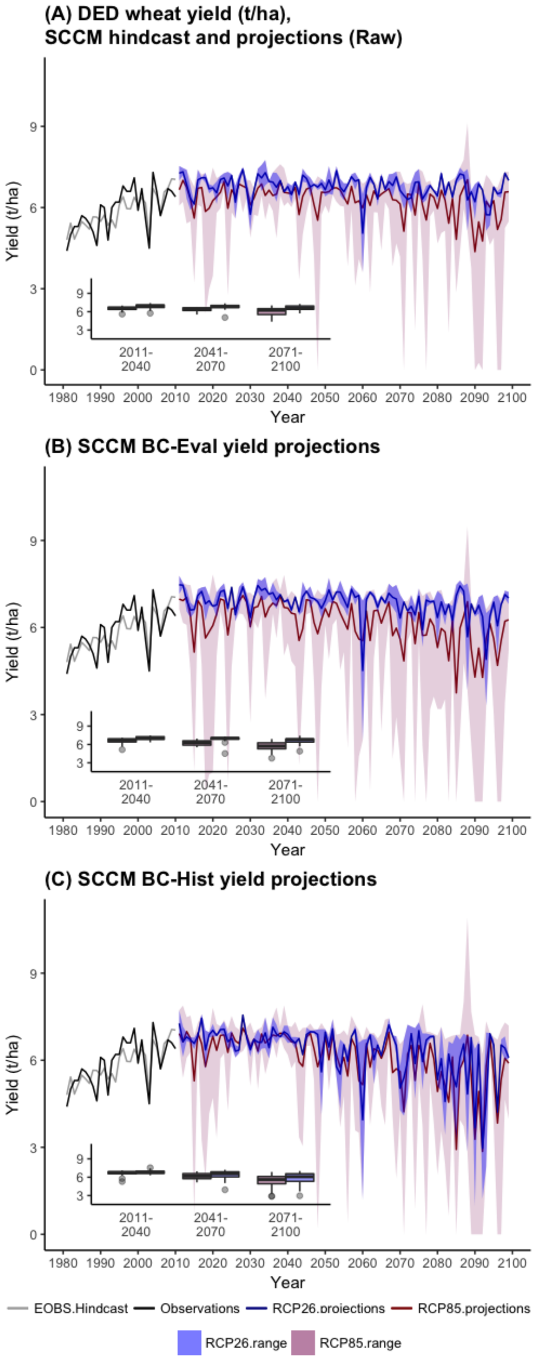



*Fig. 7.8a. DED SCCM.*

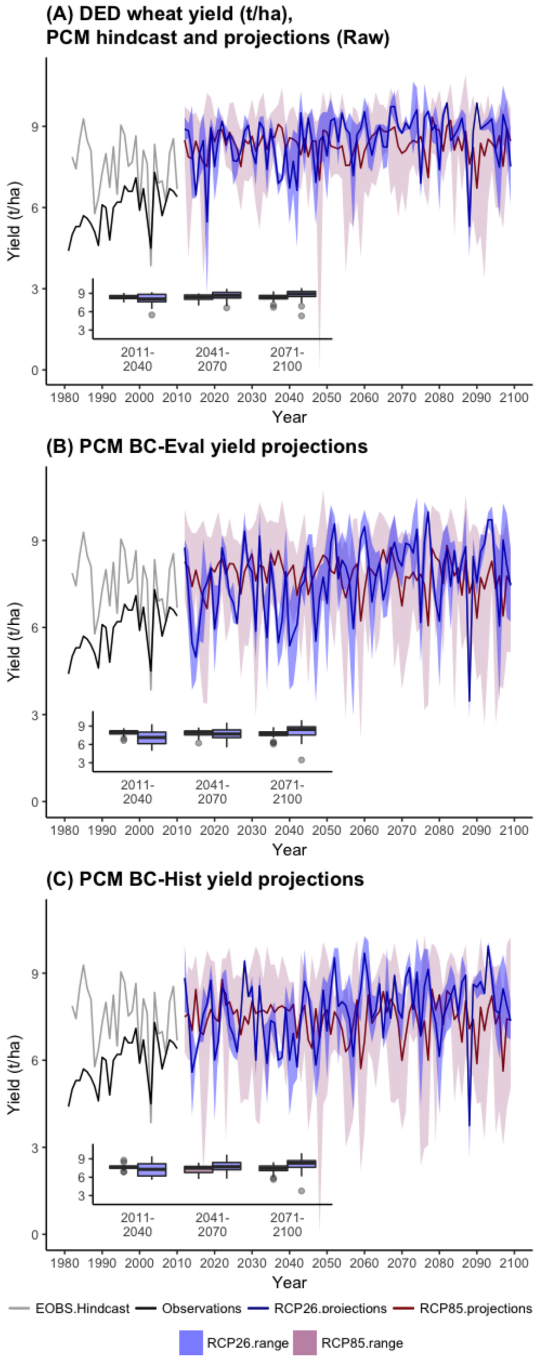



*Fig. 7.8b. DED PCM.*

*Figure 7.8: Yield projections for DED (East Germany): Fig. 7.8a shows the SCCM projections and Fig. 7.8b shows the PCM projections. (A) uses uncorrected climate model output, (B) uses BC-Eval and (C) uses BC-Hist. Inset plot shows spread of simulations for early, mid, and late century (30-year intervals). The SCCM and PCM E-OBS hindcasts are also shown as reference of crop model performance.*

Table 7.7: East Germany (DED) SCCM crop yield projections.

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **CCLM-MPI** | 1.4 | 1.5 | 1.3 | -, -0.01 | 1.2▼ | 1.3▼ | 1.2▼ | +, -0.01 | 1.2▼ | 0.9▼ | 0.1▼ | -, 0.13* |
| | 26.9 | 28.1 | 28.7 | | 20.9 | 23.2 | 22.5 | | 21.2 | 16.2 | 2.7 | |
| **RACMO-ECEARTH** | 1.4 | 1.2 | 1.1 | -, 0.07* | 1.4 | 1▼ | 0.6▼ | -, 0.15* | 1.5▲ | 1▼ | 0.6▼ | -, 0.12* |
| | 27.1 | 22.5 | 20.7 | | 26 | 18.8 | 11 | | 32.1 | 21.8 | 11 | |
| **RCA-CC** | 1.5 | 1.5 | 1.4 | -, -0.01 | 1.3▼ | 1.3▼ | 1.3▼ | -, 0 | 1.6▲ | 1.5 | 1.7▲ | +, -0.01 |
| | 29.3 | 27.7 | 26.3 | | 24.1 | 21.9 | 22.1 | | 32.3 | 27.6 | 31.9 | |
| **RCA-HadGEM** | 0.7 | 0.2 | -0.9 | -, 0.09* | 0▼ | -0.8▼ | -2.1▲ | -, 0.1* | 1.7▲ | 1▲ | -0.5▼ | -, 0.16* |
| | 15.1 | 3.8 | -17.2 | | -0.8 | -15.4 | -40 | | 33.4 | 22.5 | -11.3 | |
| **RCA-IPSL** | 1.2 | 1.2 | 0.8 | -, 0 | 1.3▲ | 0.9▼ | 0.3▼ | -, 0.06* | 2▲ | 1.2 | 0.4▼ | -, 0.11* |
| | 22.9 | 22 | 15 | | 25.1 | 16.6 | 4.6 | | 38 | 24.3 | 7.4 | |
| **RCA-MPI** | 1.3 | 1.3 | 0.7 | -, 0.06* | 1.4▲ | 1▼ | -0.1▼ | -, 0.16* | 1.3 | 0.8▼ | -0.4▼ | -, 0.15* |
| | 22.7 | 27.9 | 12.2 | | 24 | 18.5 | -2.2 | | 25.2 | 16.4 | -9.3 | |
| ***RCP85_Mean*** | 1.3 | 1.1 | 0.7 | -, 0.15* | 1▼ | 0.7▼ | 0.1▼ | -, 0.27* | 1.6▲ | 1.1 | 0.3▼ | -, 0.32* |
| | 23.4 | 21.9 | 14 | | 18.3 | 12.3 | 1 | | 34.5 | 20.7 | 6.2 | |
| **RCA-HadGEM_RCP26** | 1.4 | 1.5 | 1.3 | -, -0.01 | 1.6▲ | 1.7▲ | 1.6▲ | +, -0.01 | 2.1▲ | 1.8▲ | 1.1▼ | -, 0.13* |
| | 26.1 | 28.9 | 24.8 | | 29.8 | 31.9 | 28.5 | | 45.4 | 34.2 | 21.6 | |
| **RCA-MPI_RCP26** | 1.7 | 1.5 | 1.4 | -, 0.07* | 1.7 | 1.3▼ | 0.9▼ | -, 0.15* | 1.6▼ | 1.1▼ | 0.7▼ | -, 0.12* |
| | 32 | 28.1 | 25.6 | | 32.2 | 25.1 | 16.5 | | 33.9 | 20.4 | 13.4 | |
| ***RCP26_Mean*** | 1.6 | 1.5 | 1.3 | -, 0.03* | 1.4▼ | 1.3▼ | 1▼ | -, 0.11* | 1.8▲ | 1.4▼ | 0.9▼ | -, 0.24* |
| | 29.1 | 28.5 | 25.2 | | 26.5 | 24 | 18.2 | | 39.7 | 27.3 | 17.5 | |

Table 7.8: East Germany (DED) PCM crop yield projections.

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **CCLM-MPI** | -0.9 | -0.7 | -0.7 | +, -0.01 | 0.8▲ | 0.9▲ | 1▲ | +, -0.01 | -0.7▲ | -1▼ | -0.6▲ | -, -0.01 |
| | -10.4 | -8.4 | -7.9 | | 11.5 | 13 | 13.4 | | -9.8 | -14.4 | -7.9 | |
| **RACMO-ECEARTH** | -0.1 | -0.2 | 0 | +, -0.01 | 0.5▲ | 0.1▲ | -0.1▼ | -, 0.02 | 1.1▲ | 0.1▲ | -0.1▼ | -, 0.06* |
| | -0.8 | -2.2 | -0.2 | | 6.8 | 1.3 | -1.4 | | 15.5 | 2 | -1.3 | |
| **RCA-CC** | 0.4 | 0.1 | 0 | -, 0.02 | 1.3▲ | 1.4▲ | 1.4▲ | +, -0.01 | -0.5▼ | -0.3▼ | 0 | +, 0 |
| | 4.8 | 1.4 | 0.5 | | 17.9 | 20.2 | 19 | | -7.4 | -4.5 | -0.5 | |
| **RCA-HadGEM** | -0.2 | 0.1 | -0.5 | -, -0.01 | -0.5▲ | 0▼ | -1▲ | -, -0.01 | -0.2 | 0.2▲ | -0.4▼ | -, -0.01 |
| | -2.4 | 1.7 | -6.2 | | -7.3 | -0.4 | -13.3 | | -2.1 | 3 | -6.7 | |
| **RCA-IPSL** | 0.3 | -0.2 | 0.4 | +, 0 | 1.1▲ | 0.7▲ | 1▲ | +, -0.01 | 1.9▲ | 1.1▲ | 1.8▲ | +, -0.01 |
| | 3 | -2.2 | 5.1 | | 15.1 | 9.3 | 14.8 | | 24 | 14.8 | 23.7 | |
| **RCA-MPI** | 0 | 0 | 0 | +, -0.01 | 1.3▲ | 0.8▲ | 0.7▲ | -, 0.02 | 0.2▲ | -0.3▼ | -0.2▼ | -, 0 |
| | -0.3 | -0.2 | -0.1 | | 18 | 11.2 | 9.3 | | 3.4 | -4.2 | -2.1 | |
| ***RCP85_Mean*** | -0.1 | -0.1 | -0.1 | +, -0.01 | 0.8▲ | 0.7▲ | 0.5▲ | -, 0.01 | 0.3▲ | 0▲ | 0.1▲ | -, 0.01 |
| | -1 | -1.6 | -1.5 | | 10.6 | 9.4 | 7.3 | | 5.1 | -0.5 | 1 | |
| **RCA-HadGEM_RCP26** | 0 | 0 | 0.5 | +, 0 | -0.4▼ | -0.3▼ | 0.4▼ | +, 0.03 | -0.1▼ | -0.1▼ | 0.5 | +, 0.01 |
| | 0.5 | 0.1 | 5.8 | | -5.7 | -3.6 | 6.1 | | -1 | -0.8 | 7.1 | |
| **RCA-MPI_RCP26** | -0.7 | 0.3 | 0.1 | +, 0.06* | 0.2▲ | 1.3▲ | 1.3▲ | +, 0.07* | -0.8▲ | 0.1▼ | 0.1 | +, 0.06* |
| | -9.2 | 3.8 | 1.3 | | 3.1 | 18.5 | 18.1 | | -11.4 | 1.4 | 1.5 | |
| ***RCP26_Mean*** | -0.3 | 0.2 | 0.3 | +, 0.06* | 0▲ | 0.6▲ | 0.9▲ | +, 0.09* | -0.5▲ | 0▼ | 0.3 | +, 0.07* |
| | -4.4 | 2 | 3.5 | | -0.6 | 8.2 | 12.8 | | -6.2 | 0.3 | 4.3 | |

*Projected changes (t/ha) are in white rows and percentage in gray. All changes are relative to the respective baseline yield hindcast. (∗) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

**(4) DEF, North Germany**

Although individual GCM-RCMs show some small significant positive trends, no significant trends are observed for the RCP8.5 ensemble means for either crop modeling method (Tables 7.9, 7.10), and the significant positive trends for the ensemble SCCM RCP2.6 projections (raw and BC-Eval) have small $R^2$ values ($R^2$<0.07). This means that in terms of linear trends, the effects of future climate on yields in this region are unclear based on the SCCM and PCM. The spread and CV of SCCM projections in DEF is also very small (Fig. 7.9a) compared to the PCM projections (Fig. 7.9b) and CV (Fig. 7.5 IV).

In terms of projected changes, small yield gains (approximately 1 t/ha) are projected for the raw, BC-Eval and BC-Hist RCP8.5 ensemble mean of SCCM projections. Also with the SCCM, individual GCM-RCMs project increases relative to the baseline period; however, unlike other regions these do not show the pattern of 'decreasing increases' observed in other regions, and projected changes are fairly even through the three future periods. SCCM BC-Eval and BC-Hist projected changes are generally lower than the SCCM raw projected changes (Table 7.9).

For the PCM, the raw ensemble mean of yield projections under RCP8.5 shows no projected yield changes (Table 7.10). However, after BC, small negative yield losses (lower than the baseline yields) of under 0.5 t/ha are projected with BC-Eval and BC-Hist. Negative projected changes are also observed for many individual GCM-RCM yield projections over different future periods, and the effect of BC on these changes relative to the raw projected changes varies. The largest yield losses are projected with RCP8.5 BC-Eval CCLM-MPI: -4.6 t/ha by the end of the century. Overall, the projected changes to wheat yields in this region are less clear than in other regions based on ensemble SCCM and PCM yields.

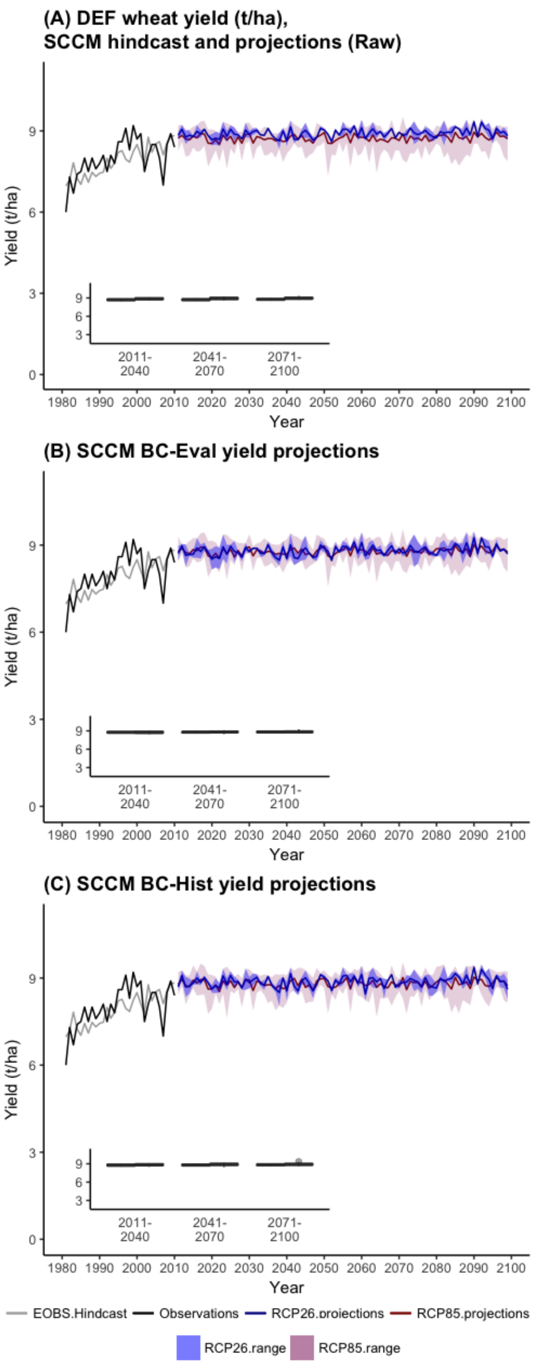



*Fig. 7.9a. DEF SCCM.*

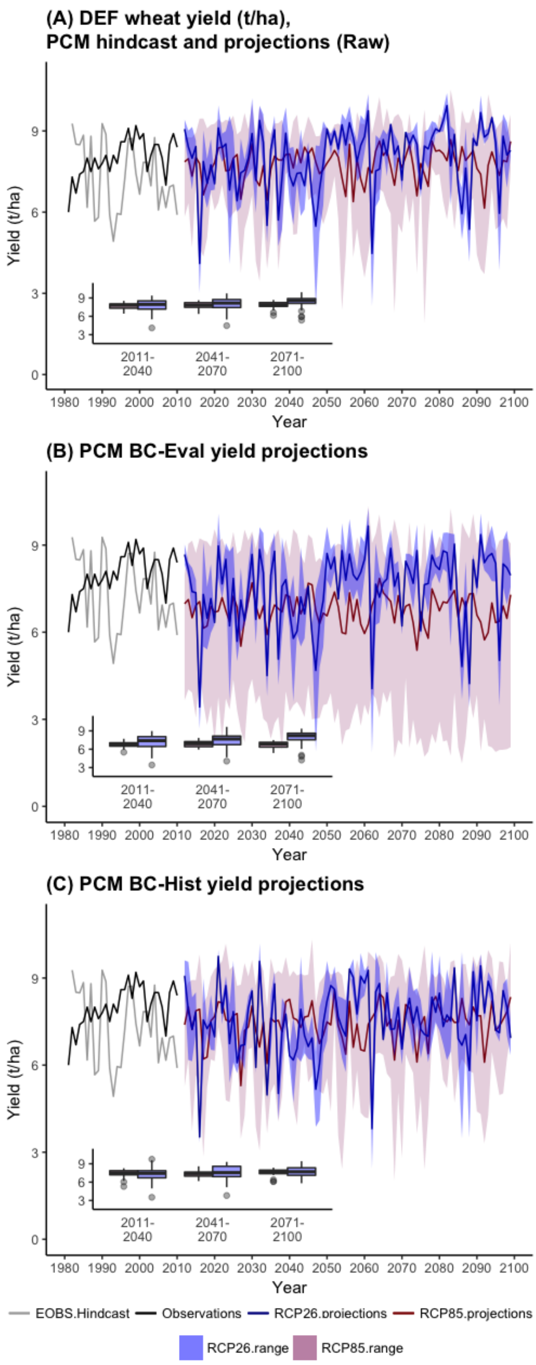



*Fig. 7.9b. DEF PCM.*

*Figure 7.9: Yield projections for DEF (North Germany): Fig. 7.9a shows the SCCM projections and Fig. 7.9b shows the PCM projections. (A) uses uncorrected climate model output, (B) uses BC-Eval and (C) uses BC-Hist. Inset plot shows spread of simulations for early, mid, and late century (30-year intervals). The SCCM and PCM E-OBS hindcasts are also shown as reference of crop model performance.*

Table 7.9: North Germany (DEF) SCCM crop yield projections.

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **CCLM-MPI** | 1.3 | 1.3 | 1.4 | +, 0.07* | 0.7▼ | 0.8▼ | 0.8▼ | +, 0.09* | 1.3 | 1.4▲ | 1.4 | +, 0.08* |
| | 16.6 | 17.1 | 18.8 | | 8.8 | 9.7 | 10.6 | | 17.1 | 17.7 | 18.9 | |
| **RACMO-ECEARTH** | 1.2 | 1.2 | 1.3 | +, -0.01 | 1▼ | 1▼ | 1.1▼ | +, -0.01 | 1.1▼ | 1.2 | 1.2▼ | -, -0.01 |
| | 15.5 | 16.6 | 17.4 | | 12.5 | 13.1 | 13.5 | | 14.7 | 14.7 | 15.7 | |
| **RCA-CC** | 1.1 | 1 | 0.9 | -, 0.02 | 0.6▼ | 0.5▼ | 0.5▼ | -, 0.02 | 0.9▼ | 1 | 0.9 | -, 0 |
| | 14 | 13.2 | 12.3 | | 7.3 | 6.7 | 6 | | 12.2 | 12.9 | 11.5 | |
| **RCA-HadGEM** | 1.2 | 1.1 | 1.2 | -, -0.01 | 1.2 | 1.1 | 1.1▼ | -, -0.01 | 1.2 | 1.1 | 1.1▼ | -, -0.01 |
| | 17 | 14.7 | 15.7 | | 14.7 | 13.5 | 13.9 | | 15.6 | 13.8 | 14 | |
| **RCA-IPSL** | 1.3 | 1.3 | 1.3 | +, -0.01 | 1▼ | 1▼ | 1▼ | +, -0.01 | 1▼ | 1▼ | 1▼ | +, -0.01 |
| | 17.3 | 16.7 | 17.4 | | 12.2 | 12.1 | 12.9 | | 13.2 | 12.8 | 13.7 | |
| **RCA-MPI** | 1.2 | 1.3 | 1.4 | +, 0.09* | 0.7▼ | 0.8▼ | 1▼ | +, 0.09* | 1.1▼ | 1.2▼ | 1.3▼ | +, 0.07* |
| | 15 | 17.9 | 18.5 | | 9 | 10.7 | 12.1 | | 14.2 | 16.6 | 17.1 | |
| ***RCP85_Mean*** | 1.2 | 1.2 | 1.3 | +, 0.02 | 0.8▼ | 0.9▼ | 0.9▼ | +, 0.03 | 1.1▼ | 1.1▼ | 1.2▼ | +, 0.01 |
| | 16 | 16.5 | 16.7 | | 10.7 | 11 | 11.5 | | 13.8 | 15 | 15.1 | |
| **RCA-HadGEM_RCP26** | 1.1 | 1.2 | 1.2 | +, 0.07* | 0.7▼ | 0.8▼ | 0.8▼ | +, 0.09* | 1.1 | 1.1▼ | 1.2 | +, 0.08* |
| | 14.4 | 16 | 16.3 | | 9 | 9.9 | 10.7 | | 13.8 | 15.1 | 15.6 | |
| **RCA-MPI_RCP26** | 1.6 | 1.6 | 1.7 | +, -0.01 | 1▼ | 1▼ | 1.1▼ | +, -0.01 | 1.2▼ | 1.3▼ | 1.3▼ | -, -0.01 |
| | 20.4 | 22.1 | 22 | | 12.4 | 13 | 13.4 | | 15.8 | 16.7 | 16.7 | |
| ***RCP26_Mean*** | 1.3 | 1.4 | 1.4 | +, 0.05* | 0.8▼ | 0.9▼ | 1▼ | +, 0.07* | 1.2▼ | 1.2▼ | 1.2▼ | +, 0.02 |
| | 17.4 | 19.1 | 19.2 | | 10.7 | 11.4 | 12 | | 14.8 | 15.9 | 16.2 | |

Table 7.10: North Germany (DEF) PCM crop yield projections.

| | Uncorrected | | | | BC-Eval | | | | BC-Hist | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) | 2011-2040 | 2041-2070 | 2071-2100 | Trend ($R^2$) |
| **CCLM-MPI** | -0.5 | -0.6 | -0.2 | +, -0.01 | -3.6▼ | -3.7▼ | -4.6▼ | -, 0.23* | -1.1▼ | -1.2▼ | -0.7▼ | +, -0.01 |
| | -7.1 | -8.2 | -1.9 | | -51.4 | -52.1 | -61.8 | | -13.9 | -15.8 | -8.9 | |
| **RACMO-ECEARTH** | -0.3 | -0.5 | -0.3 | +, -0.01 | -0.2▼ | -0.5 | -0.2▼ | +, -0.01 | -0.3 | -0.8▲ | -0.2▼ | +, -0.01 |
| | -3.9 | -6.6 | -3.9 | | -2.8 | -6.8 | -2.6 | | -4.7 | -13.1 | -2.9 | |
| **RCA-CC** | 0 | -0.1 | -0.3 | -, 0 | 0.7▲ | 0.8▲ | 0.5▲ | -, -0.01 | 0.4▲ | 0.8▲ | 0.2▲ | -, -0.01 |
| | -0.5 | -1.2 | -4.1 | | 9.8 | 11.5 | 6.7 | | 5.1 | 9.7 | 3.3 | |
| **RCA-HadGEM** | -0.5 | 0.1 | 0.5 | +, 0.04* | -1.2▲ | -0.5▼ | -0.3▼ | +, 0.03 | -0.2▼ | 0.4▲ | 0.6▲ | +, 0.02 |
| | -5.4 | 1.9 | 6.6 | | -16.7 | -7 | -4.7 | | -2.4 | 6.3 | 9.4 | |
| **RCA-IPSL** | -0.1 | 0 | 0 | +, -0.01 | 0.6▲ | 0.5▲ | 0.5▲ | +, -0.01 | 1.5▲ | 0.8▲ | 1.1▲ | -, 0 |
| | -1 | -0.2 | -0.3 | | 8 | 6.3 | 7.3 | | 18.1 | 11.4 | 13.7 | |
| **RCA-MPI** | -0.5 | -0.2 | -0.5 | -, -0.01 | 0.4▲ | 0.6▲ | 0.5▲ | -, -0.01 | -0.7▲ | -0.6▲ | -0.4▼ | +, 0 |
| | -7 | -2.7 | -7.1 | | 5.7 | 7.9 | 6.2 | | -10.1 | -7.2 | -5.8 | |
| ***RCP85_Mean*** | -0.3 | -0.2 | -0.1 | +, 0 | -0.5▲ | -0.4▲ | -0.5▲ | -, -0.01 | -0.1▼ | -0.1▼ | 0.1▲ | +, 0 |
| | -4 | -2.6 | -1.7 | | -6.1 | -5 | -7 | | -1.2 | -1 | 1.3 | |
| **RCA-HadGEM_RCP26** | -0.2 | 0.2 | 0.7 | +, 0.03 | -0.9▼ | -0.4▼ | 0.1▼ | +, 0.03* | 0▲ | 0.1▼ | 1▲ | +, 0.06* |
| | -3 | 2.5 | 8.6 | | -11.5 | -5.9 | 0.9 | | -0.1 | 1.4 | 13.6 | |
| **RCA-MPI_RCP26** | -0.5 | -0.5 | -0.2 | +, 0 | 0.4▲ | 0.4▲ | 0.6▲ | +, 0 | -0.7▲ | -0.4▼ | -0.9▲ | -, -0.01 |
| | -6.4 | -5.3 | -3 | | 5.2 | 5.5 | 8.6 | | -10.5 | -4.8 | -11.7 | |
| ***RCP26_Mean*** | -0.3 | -0.1 | 0.2 | +, 0.03 | 0▲ | 0.2▲ | 0.6▲ | +, 0.03 | -0.4▲ | -0.1 | 0.1▼ | +, 0.02 |
| | -4.7 | -1.4 | 2.8 | | -0.2 | 2.7 | 7.7 | | -5.3 | -1.7 | 1 | |

*Projected changes (t/ha) are in white rows and percentage in gray. All changes are relative to the respective baseline yield hindcast. ($*$) indicates a significant ($p$ <0.05) trend (+/-). In addition, a ▲(▼) indicates a relative increase (decrease) to the uncorrected projected change.*

### 7.3.3 Summary of projected yield changes

#### 7.3.3.1 Significant differences of yield projections

In this section, the significance of the projections relative to the respective past yield hindcast are summarized for country-level SCCM and a comparison of regional SCCM and PCM results. At the national level, yields simulated with the SCCM all show significant differences for all three future periods respective to their baseline yields. At the regional level, the mean of SCCM yield projections for DEA (West Germany) and DEF (North Germany) are significantly different to their respective baseline yields (e.g. raw projections to raw historical hindcast), based on a t-test (Table 7.11).

For yields simulated for German regions with a PCM, it can be observed that in the mid-century (2041-2070), yield projections are generally not significantly different to their baseline yields. Based on the lack of significant linear trends as reported in the previous section, this indicates that yield changes are small in the mid-century across all BC projections. Changes during this mid-century future period are typically less than half a ton (both gains and losses) particularly in DE2 (Table 7.4) and DEF (Table 7.10) for the mean ensemble yields.

#### 7.3.3.2 Comparing projected percentage changes to yield between methods

The mean projected percentage yield changes (relative to the respective baseline yield) per BC method (Raw, BC-Eval and BC-Hist) and averages across all these simulations are shown for each region in Figures 7.10 (RCP8.5) and 7.11 (RCP2.6). This analysis shows that all SCCM and PCM projected percentage changes in yield are significantly different to each other under RCP2.6.

For DEA and DEF, across all future periods, SCCM and PCM projected percentage changes are significantly different from one another under RCP8.5. However, there are some instances where the differences are not significant. For instance, projected yield changes in DE2 (2041-2070 and 2071-2100) and DED (2071-2100) are of similar magnitudes. These similarities and differences in projected yield changes are further explored in the discussion section. In the following section, the yield projections are analyzed to characterize how their ranges (uncertainty) are influenced by the choices in the different GCM-RCMs, crop modeling methods and the BC approach.

*Table 7.11: Significant differences between ensemble projected and baseline yields (SCCM and PCM), German regions.*

| | | SCCM | | | PCM | | |
|---|---|---|---|---|---|---|---|
| **Region** | **Correction** | **2011-2040** | **2041-2070** | **2071-2100** | **2011-2040** | **2041-2070** | **2071-2100** |
| DE2 | Raw RCP8.5 | | | | | | |
| | Raw RCP2.6 | | | | | | |
| | BC-Eval RCP8.5 | | | | | | |
| | BC-Eval RCP2.6 | | | | | | |
| | BC-Hist RCP8.5 | | | | | | |
| | BC-Hist RCP2.6 | | | | | | |
| DEA | Raw RCP8.5 | | | | | | |
| | Raw RCP2.6 | | | | | | |
| | BC-Eval RCP8.5 | | | | | | |
| | BC-Eval RCP2.6 | | | | | | |
| | BC-Hist RCP8.5 | | | | | | |
| | BC-Hist RCP2.6 | | | | | | |
| DED | Raw RCP8.5 | | | | | | |
| | Raw RCP2.6 | | | | | | |
| | BC-Eval RCP8.5 | | | | | | |
| | BC-Eval RCP2.6 | | | | | | |
| | BC-Hist RCP8.5 | | | | | | |
| | BC-Hist RCP2.6 | | | | | | |
| DEF | Raw RCP8.5 | | | | | | |
| | Raw RCP2.6 | | | | | | |
| | BC-Eval RCP8.5 | | | | | | |
| | BC-Eval RCP2.6 | | | | | | |
| | BC-Hist RCP8.5 | | | | | | |
| | BC-Hist RCP2.6 | | | | | | |

*A shaded box indicates a significant difference between projected yields and their baseline/reference, based on a t-test ($p$<0.05).*

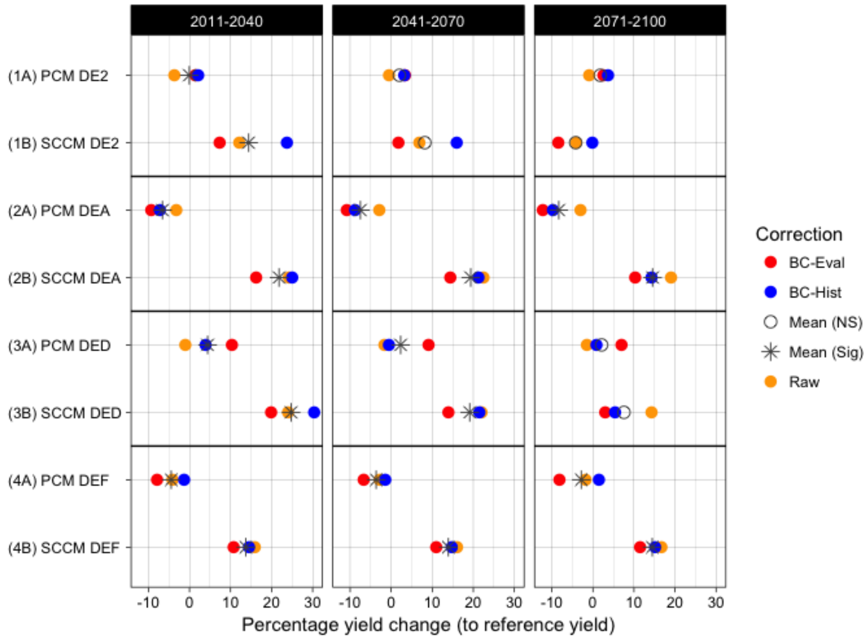



*Figure 7.10: Summary of regional projected yield changes: crop modeling method and BC comparison for RCP8.5. A * indicates that the mean of the SCCM and PCM change is significantly different (based on a t-test).*

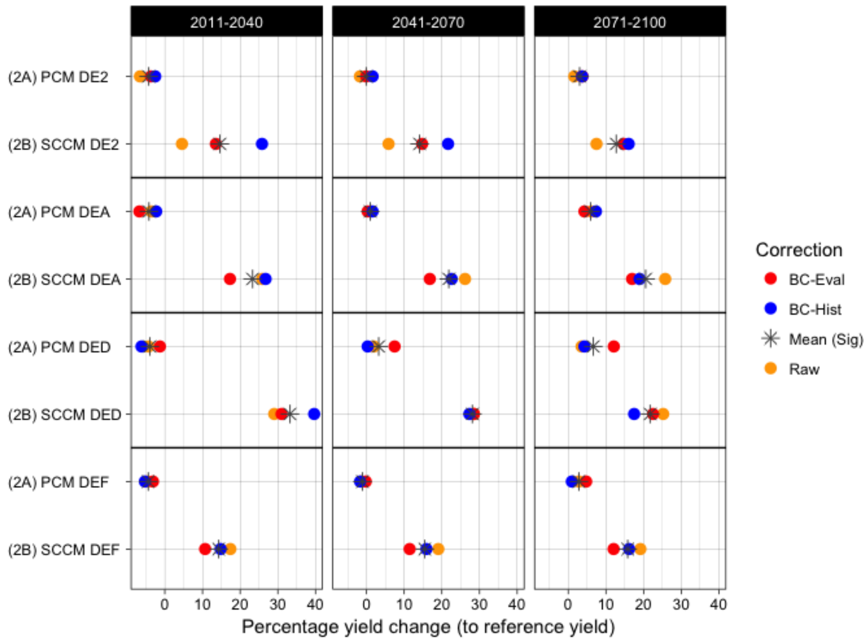



*Figure 7.11: Summary of regional projected yield changes: method and BC comparison for RCP2.6. A * indicates that the mean of the SCCM and PCM change is significantly different (based on a t-test).*

### 7.3.4 Uncertainty analysis

ANOVA results show that, in this study, the uncertainty in yield projections is overwhelmingly influenced by the choice in crop modeling method across all four German regions when considering GCM-RCM, BC approach, and the respective interaction terms (Figs. 7.12, 7.13). For example, in the ternary plots for RCP8.5 and RCP2.6, across all future 30-year intervals, most values are located in the lower right corner, indicating that the choice of crop modeling method is the dominant share of uncertainty in the study, when considering only the main effects included in the ANOVA. F-test values are used to determine the significance of the fractional partitions, with the percentages shown in Tables 7.12 and 7.13.

For DE2, the crop model method contribution to uncertainty is largest in the middle of the century (2041-2070) with up to 80% of the variance attributed to the crop model method (Fig. 7.13). Approximately 50% of the uncertainty in DED and DEF are also explained by the crop modeling method across all 30-year periods, but this variance contribution also decreases between 2071-2100. However, in DEA, while the choice in crop modeling method is still a significant contributor to uncertainty, it has a smaller influence of approximately 20% in the early- and mid-century, and this falls to only 8% by the last 30-year period. This can also be observed in the ternary plots, where DEA is often farther from the rightmost corner (Method) and closer to the GCM corner. For DEA, the uncertainty from the choice in GCMs becomes more dominant (24%) than the crop model method for 2071-2100.

The fall in the influence of crop modeling method is also observed for DEA under RCP2.6, where other interactions between the GCM and the method (49.2% in 2041-2070), GCM and the BC approach (25.9% in 2071-2100) become more dominant compared to the crop modeling method (Table 7.13), which can confound the results of the ANOVA.

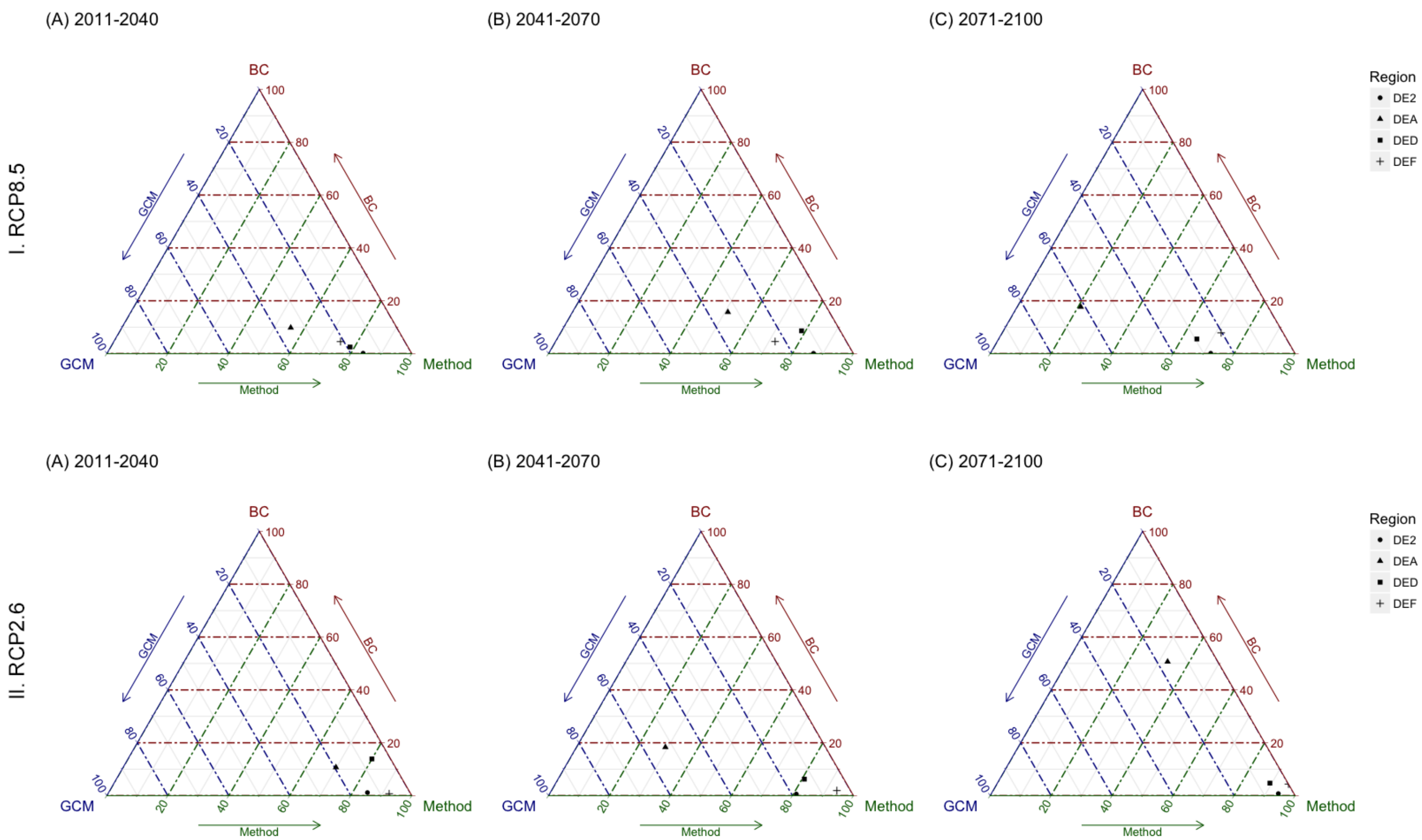



*Figure 7.12: Ternary plots for German regional yield simulations to show uncertainty partitions, in percentage of the sum of squares with ANOVA: under I. RCP8.5 and II. RCP2.6 for (A) early century, (B) mid-century and (C) late century. Uncertainty is decomposed into: GCM-RCM combination (GCM), crop modeling method (Method) and Bias correction calibration (BC).*

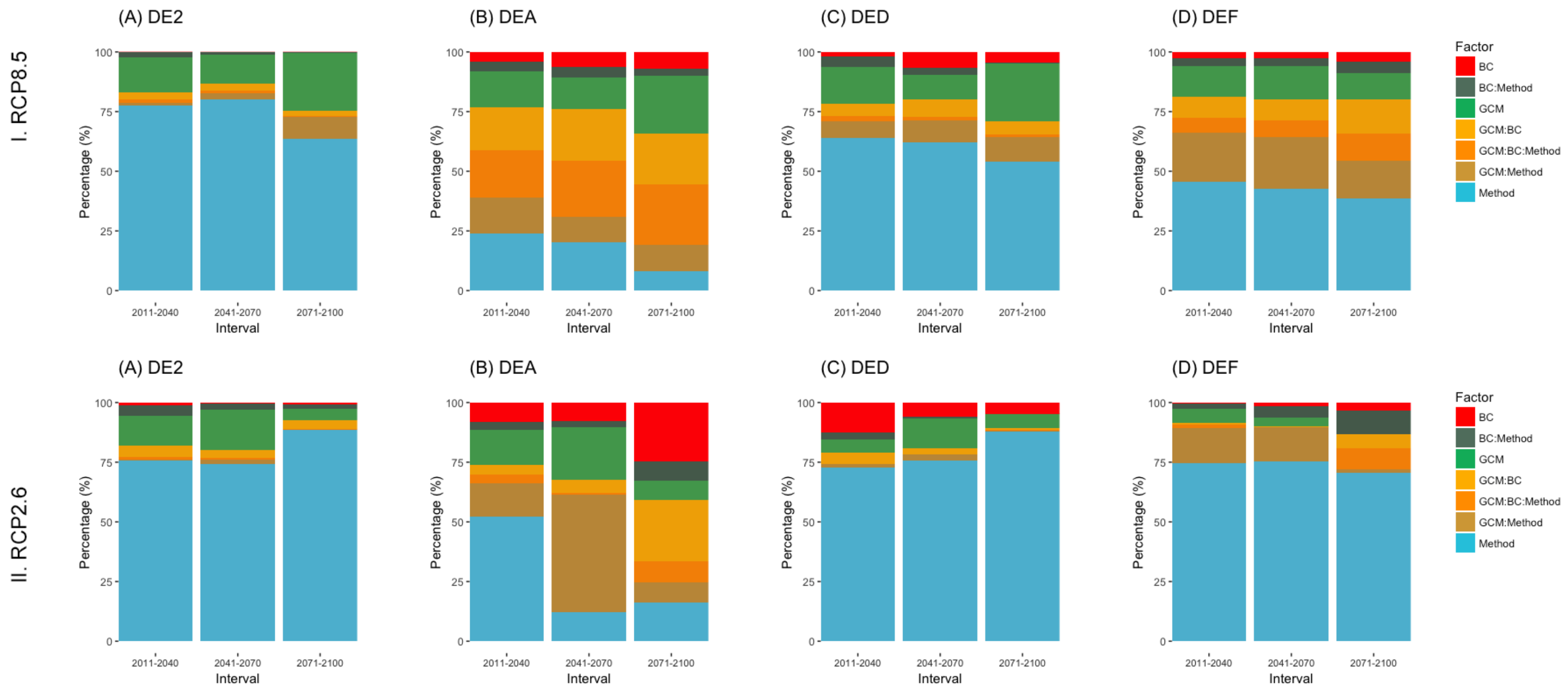



*Figure 7.13: Uncertainty partitions per region, interval and scenario, considering interactions. Areas show the percentage of the sum of squares with ANOVA: using the under I. RCP8.5 and II. RCP2.6 for (A) early century, (B) mid-century and (C) late century. Uncertainty is decomposed into: GCM-RCM combination (GCM), BC calibration, crop modeling method (Method), and their second-order interactions.*

*Table 7.12: Fractional uncertainty, RCP8.5.*

| | DE2 | | | DEA | | | DED | | | DEF | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | 2011-2040 | 2041-2070 | 2071-2100 | 2011-2040 | 2041-2070 | 2071-2100 | 2011-2040 | 2041-2070 | 2071-2100 |
| **GCM** | 14.7%* | 12%* | 24.3%* | 15.0% | 13.3% | 24.2% | 15.5%* | 10%* | 24.4%* | 12.9%* | 13.8%* | 10.9% |
| **BC** | 0.1% | 0.1% | 0.1% | 4.2% | 6.3% | 7.0% | 2%* | 6.8%* | 4.6%* | 2.8% | 2.7% | 4.2% |
| **Method** | 77.4%* | 80%* | 63.7%* | 23.8%* | 20.4%* | 8.2% | 64.1%* | 62%* | 53.9%* | 45.5%* | 42.5%* | 38.5%* |
| **GCM:BC** | 3.1% | 3.1%* | 2.2%* | 18.1% | 21.5% | 21.2% | 5.2% | 7.5%* | 5.3%* | 8.6% | 9.0% | 14.3% |
| **GCM:Method** | 1.2% | 2.7%* | 8.9%* | 15.1% | 10.4% | 11.0% | 6.6%* | 9.4%* | 10.4%* | 20.7%* | 21.9%* | 15.9% |
| **BC:Method** | 2.1%* | 1.1%* | 0.2% | 4.0% | 4.4% | 3.1% | 4.3%* | 2.9%* | 0.1% | 3.3% | 3.3% | 4.8% |
| **GCM:BC:Method** | 1.4% | 0.9% | 0.5% | 19.8% | 23.7% | 25.4% | 2.2% | 1.3% | 1.3% | 6.3% | 6.8% | 11.5% |

*Table 7.13: Fractional uncertainty, RCP2.6.*

| | DE2 | | | DEA | | | DED | | | DEF | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2011-2040 | 2041-2070 | 2071-2100 | 2011-2040 | 2041-2070 | 2071-2100 | 2011-2040 | 2041-2070 | 2071-2100 | 2011-2040 | 2041-2070 | 2071-2100 |
| **GCM** | 12.5%* | 16.9%* | 4.8%* | 14.6% | 22%* | 8.1% | 5.6%* | 12.2%* | 5.8%* | 5.7% | 3.6%* | 0.0% |
| **BC** | 1.0% | 0.5% | 0.7% | 8.0% | 7.7% | 24.8% | 12.6%* | 5.9%* | 4.7%* | 0.6% | 1.5% | 3.2% |
| **Method** | 75.6%* | 74.3%* | 88.6%* | 52.2%* | 12.2%* | 16.1% | 72.7%* | 75.9%* | 88%* | 74.7%* | 75.4%* | 70.40%* |
| **GCM:BC** | 4.7% | 3.4% | 3.7%* | 4.1% | 5.5% | 25.9% | 4.6%* | 2.5%* | 0.9% | 0.7% | 0.1% | 5.9% |
| **GCM:Method** | 0.5% | 1.9% | 0.1% | 14.0% | 49.2%* | 8.4% | 1.5%* | 2.5%* | 0.2% | 14.5%* | 14.3%* | 1.6% |
| **BC:Method** | 4.6% | 2.3% | 2.0% | 3.5% | 2.7% | 7.8% | 2.9% | 1.0% | 0.2% | 2.2% | 4.9%* | 10.1% |
| **Residuals** | 1.1% | 0.6% | 0.1% | 3.6% | 0.6% | 8.9% | 0.2% | 0.1% | 0.2% | 1.5% | 0.1% | 8.7% |

*($*$) indicates a significant F-test value ($p$ <0.05).*

Based on this analysis, generally, after crop modeling method, it can be reported that that the next largest source of variance is the choice of GCM-RCM in the study. BC is the least important influence on the fractional uncertainty contribution in the study for both RCP8.5 and RCP2.6 scenarios in the study design, typically under 10%.

These results – crop projections (UK, Germany), regional PCM and SCCM yield comparisons, and the uncertainty decomposition – are discussed in the next section.

## 7.4 Discussion

In this chapter, the research questions are addressed through the comparison of SCCM and PCM yield projections, and the uncertainty decomposition analysis. For the first part of the discussion, the results are framed to help answer the first research question, which is how are wheat yields in the UK and Germany projected to be affected by climate change – in addition, how do the results in this chapter compare to previous findings?

### 7.4.1 Outlook for future wheat yields

#### 7.4.1.1 National-level yields for the UK (SCCM)

UK yield projections in this chapter were produced with a SCCM based on the number of hot days and total JJA precipitation (See Chapter 3 for more information on the crop model approaches). Because the projections for the increase in the number of hot days over the entire UK are quite low, and the projected decreases in summer rainfall, while significant, are also small (~30mm decreases based on the RCP8.5 ensemble mean from EURO-CORDEX simulations, see Chapter 6), the resulting UK yield projections are observed to behave fairly constantly into the future – meaning without large changes in mean or variability – relative to the baseline yields.

The results of the chapter with the UK SCCM show that there is actually potential for yields to increase under climate change scenarios, approximately up to 1.4 t/ha based on the ensemble uncorrected SCCM projections. Ensemble yield projections under RCP8.5 and RCP2.6 showed positive trends, and projected changes were significantly different to the baseline yields. These potential yield gains are approximately equal under RCP8.5 and RCP2.6 (~1-1.5 t/ha), with yields using BC input showing similar trends (Fig. 7.14).

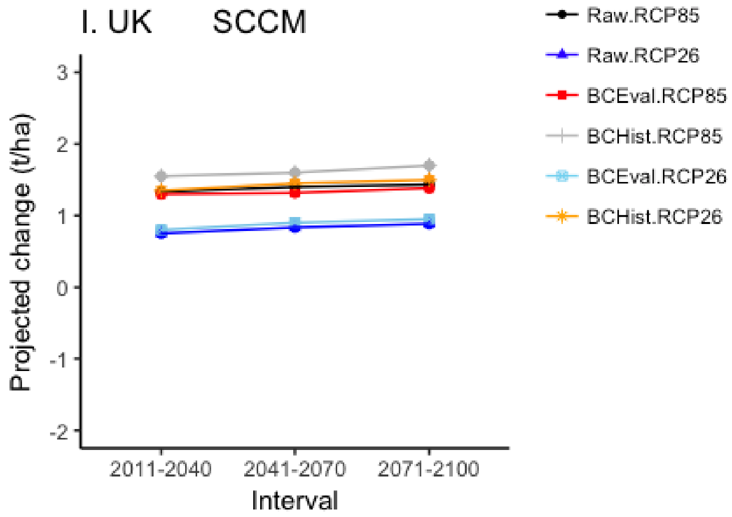



*Figure 7.14: Average projected yield changes, UK SCCM.*

#### 7.4.1.2 Comparison of UK yield projections to previous studies

The SCCM results of potential gains in wheat yield for the UK (~10-20% relative to respective baselines) are in general agreement with other impact projections, e.g. the UK Climate Change Risk Assessment 2017 (CCRA, Brown et al., 2016), modeling and review studies (e.g. Rial-Lovera et al., 2017, Röder et al., 2014, Cho et al., 2012, Ferrara et al., 2010) which also project small potential increases in yield for the UK. It should be noted that the previous UK CCRA in 2012, developed with statistical approaches, was heavily criticized (e.g. Semenov et al., 2012) for overestimating the potential percentage increase in yield, as the relationship between yields and temperature was overestimated, and the influence of factors such as heat stress around flowering were underestimated (Brown et al., 2016, Semenov et al., 2012).

Therefore, while the SCCM results for the UK shown here prove to be robust enough to share similar results to previous studies, and the heat-based indices are specific to heat stress around flowering (anthesis), the lack of inclusion of other important genetic-environment-management ($G \times E \times M$) interactions, and crucially, the $CO_2$ fertilization effect – which may even potentially make these yield gains larger – make these crop projections a relatively simple analysis of the effects of changes to temperature and precipitation on UK wheat yield. Additionally, known issues with statistical models such as the issue of response stationarity mean that these results are also affected by these limitations.

Known important crop growth influences such irrigation, radiation, changing crop varieties and management also strongly influence crop yields. Provided with sufficient data on these numerous influencing factors, a regional analysis of the UK wheat production sector's risk to climate change could provide more robust results, and allow for a comparison with a PCM (similar to the following discussion for Germany) to better characterize uncertainty in the yield projections for the UK. However, the issues discussed surrounding the application of a field-based PCM to a regional scale (e.g. scale and aggregation error, see Chapter 3) should also be recalled in the regional yield context.

Despite these limitations, a key finding of this analysis, shared with other UK wheat studies, is that based on empirical relationships, wheat yields can actually show positive responses to changes in temperature and precipitation; it is argued that there may be unique opportunities for adaptation that can take advantage of these potential yield gains, however small. Beneficial changes to wheat varieties, technology and management practices, existing adaptation measures, and improved technology – which have already been shown to influence yield more than climate (e.g. Moore and Lobell, 2015, 2014, Semenov et al., 2012, Semenov and Shewry, 2011, Semenov, 2009) – are also argued to improve the outlook for wheat yields, if adaptation measures are planned, timed, and managed well.

#### 7.4.1.3 National and regional-level yields for Germany (SCCM)

In contrast to the yield projections in the UK, the SCCM projections of yield for Germany, and German regions, show evidence of negative yield trends and increased yield variability into the future. At the country-level, SCCM projections under the RCP8.5 scenario (across raw and all BC approaches) and RCP2.6 (across BC-Eval and BC-Hist), show that the projected increases in the number of hot days and reduction in summer precipitation (See Chapter 6), could reduce the yield gains over time, as compared to the respective baseline yields.

Trends of 'decreasing increases' (from ~20% to as low as 2% relative to the respective baseline) were observed for Germany, and the ensemble RCP8.5 SCCM DE2, DEA and DED yield projections (Figs. 7.15, 7.16). SCCM yield projections for the northernmost German state (DEF) did not show significant trends for either emission scenario. Projected changes for Germany and its regions, apart from DEF, showed diminishing positive yield changes relative to baseline yields (1976-2005). The positive sign of change relative to the SCCM hindcast means that wheat yields are not projected to go as low as the baseline, but could show reductions to yield over time, as shown in Fig. 7.16. The positive projected yield changes are also related to the rapid rise in observed yields until 2010, which means that the baseline yield level is lower than where yield projections begin in 2011.

#### 7.4.1.4 Comparison of German yield projections to previous studies

These projected decreases in yield are in agreement with previous studies which have focused on the impact of climate change on wheat production in Germany, which is one of the largest producers of wheat in Europe, second only to France (Lüttger and Feike, 2018), where warming climatic conditions have been found to pose threats to wheat production (Strer et al., 2018, Trnka et al., 2015). For example, a crop modeling study

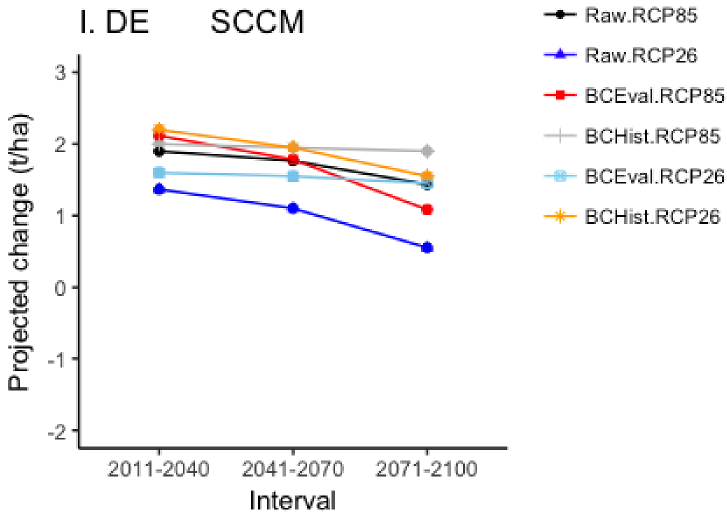



*Figure 7.15: Average projected yield changes, Germany SCCM).*

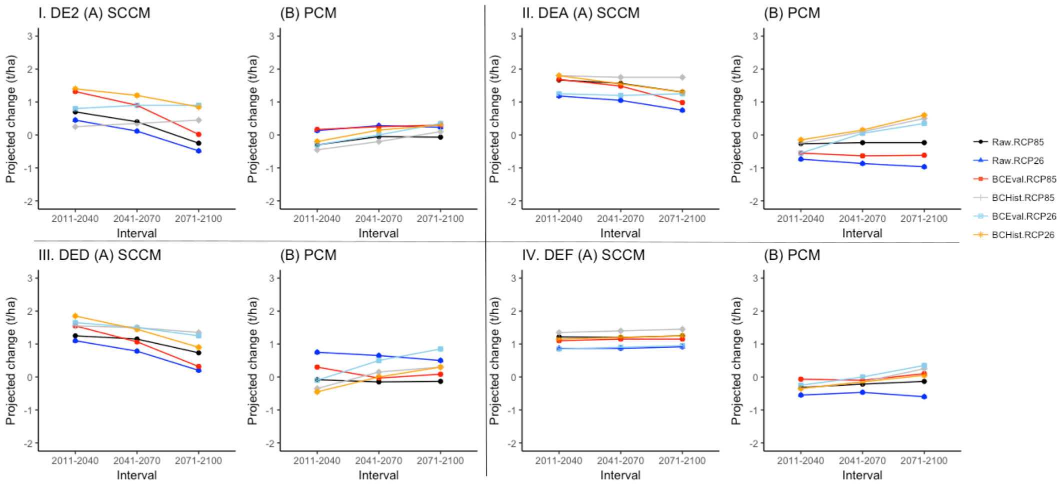



*Figure 7.16: Projected yield changes, German regions (I-IV). SCCM and PCM yield projections are shown in subfigures (A) and (B), respectively.*

projected negative changes in wheat yield when considered without the $CO_2$ effect (Kersebaum and Nendel, 2014), as was also projected in the statistical approach of this study. Another modeling study divided Europe into 'environmental zones' considering climate, where Germany was considered to be composed of the Atlantic North (roughly the DEF region in this study), and Continental climates (other regions). Their study found that the impacts of climate change on winter wheat are negative across most of these zones (Olesen et al., 2011).

While the general results of SCCM projections for Germany are shown to match previous studies (i.e. decreasing yields due to increases in heat stress), how do the PCM projections compare?

### 7.4.2 Comparing crop modeling methods: projected yield changes and trends

In order to compare the regional Germany SCCM and PCM yield projections, which have different relative baselines (e.g. PCM projections to past PCM hindcasts, and similarly for SCCMs), percentage yield changes are shown in the comparison of yield projections in Figs. 7.10 and 7.11 in the Results section. These results show that yield projections from different crop modeling methods have varying levels of agreement, depending on the region. For example, mean changes in yield were generally dissimilar across crop modeling methods in German regions in DEA and DEF, while yield projections showed some agreement in DE2 and DED.

Differences were also observed in trends between yield projected by different crop modeling methods: ensemble RCP8.5 SCCM yield projections typically showed significant negative trends in DE2, DEA, and DED, but RCP8.5 PCM yield projections typically showed no significant trends for any regions. RCP2.6 yield projections from either crop modeling method typically did not show significant trends, or had very small $R^2$ values.

SCCM projections generally showed patterns of 'decreasing increases', meaning that projected changes in yield were positive respective to the past yield baseline, but these yields decreased over time towards the end of the century (Fig. 7.16). In contrast, apart from DEF where both the SCCM and PCM showed relatively 'flat' yield trends and changes, PCMs used with raw climate model output tended to project negative yield losses below baseline yields but did not have clear patterns for these decreases.

While some studies have found projections of yields to generally be in agreement between crop modeling methods – for example, in multi-method studies that have found generally similar outcomes between methods when investigating yield responses to temperature (Liu et al., 2016), and to both temperature and precipitation (Lobell and Asseng, 2017) – the results of this chapter show that there are observable differences between yield projections to be explored when using different crop modeling methods. How and why do these differences arise, and what can they be attributed to in the modeling/simulation process? These are explained in the following subsections.

#### 7.4.2.1 Role of temperature and precipitation in wheat yield projections

Firstly, one potential reason for the differences between crop model method yield projections is how temperature and precipitation are utilized and modified in the PCM and in the SCCMs. It is argued that the way temperature and precipitation are related to yield within the mechanisms of the crop model has an influence on the differences between yield projections. For instance, SCCMs, including the ones used in this study, generally have much fewer predictors for yield, particularly when compared to PCMs that use daily weather, and also contain various input parameters related to management and genetics. Thus, more weight is given to the predictive power of climate in the SCCMs of the study.

Given the relatively small changes to the number of hot days for the UK (See Chapter 6), the results show that SCCM UK wheat yield projections show very low variation around the mean. In addition, the smaller coefficient of JJA precipitation in SCCMs makes the temperature-based index more critical for yields in SCCMs. The more flexible structure of SCCMs also means that heat stress can be more easily and directly represented in

SCCMs. This is useful as it has been found that uncertainties in crop yield predictions in most regions are dominated by uncertainties in future temperature, rather than precipitation (Watson et al., 2015, Lobell and Burke, 2008).

In contrast, the way temperature and precipitation can be specifically represented as heat stress indices in SCCMs is mismatched with the way climate is represented as daily values in PCMs such as CERES-Wheat/DSSAT. CERES/DSSAT uses daily temperature to calculate other parameters, including, leaf area index, photosynthesis, nitrogen fixation, and soil temperature (Jones et al., 2003), which follow less linear mechanisms than those used by the SCCMs in the study. In addition, the rate of wheat development in the CERES module is governed by thermal time in the form of growing degree-days (GDD) based on the daily maximum and minimum temperatures (Jones et al., 2003). There are certain GDD required to progress from one growth stage to another, including the crucial anthesis and grain-filling stages, which in turn also interact with other input parameters.

The different ways the climate data are modified and used within the SCCM and PCM, alongside the large differences in their structure and input parameters, manifests clearly in the variation between yield projections. Hence, it is argued that while SCCMs can be devised to fit a variety of empirical data and scales, as well as focus on relevant climate indices, or other predictive factors (e.g. soil, Kristensen et al., 2011), PCMs are much less flexible and are limited by their structure and represented processes in order to make yield projections. It was also previously shown in Chapter 3 how yield simulations are highly sensitive to the genetic input parameters used in the PCM, so the importance of aggregation and input parameter errors cannot be discounted; this is revisited in the next section.

### 7.4.3 Revisiting evaluation results and their impacts on future yield projections

In Chapter 3, it was shown how the evaluation performance of the PCM was very different to the well-correlated yield simulations from the SCCM approach to both a yield baseline hindcast (with climate observations) and actual yield observations. Although BC of RCM evaluation climate output was able to improve the PCM yield simulations in Chapter 5 (reducing error relative to the yield hindcast), it is important to consider how the performance of the crop model method in the past will influence future yield simulations.

The issue of 'reasonable performance' was discussed in Chapter 3, where it was reported that, as with this study, many field-based PCMs are applied to scales beyond their original design (e.g. Challinor et al., 2017, Ewert et al., 2011, Hansen and Jones, 2000), which leads to issues of spatial and aggregation error of input parameters. Despite this scale mismatch (See also Chapter 2 discussion on the limitations of the PCM approach), the included plant growth processes, proven simulation ability of PCMs, and the importance of comparing the outcomes of two vastly different crop modeling approaches has been argued to at least justify the approach used in this chapter to explore their differences.

However, as seen in the results, the difference in crop modeling method can influence the outcomes for yield projections, even though they were driven by the same climate data. In this case, the decomposition of uncertainty with ANOVA reveals how important the choice of crop modeling method is.

#### 7.4.3.1 Dominance of uncertainty by crop modeling method

Although care must be taken when extrapolating results beyond this study, in the results of the uncertainty analysis (Section 7.3.4, Figs. 7.12, 7.13), it was shown that crop modeling method dominates the uncertainty based on ANOVA results. Given the large methodological differences between crop modeling methods, this result is not unexpected. The large differences in methods – and their associated limitations – have led to a large range of plausible yield projections (See Figs. 7.4a-7.9b), one that is greater than other factors such as the choice of GCM-RCM, whether these GCM-RCMs were bias-corrected or not, and also inclusive of potential interactions between the main chosen effects.

This is an opportunity for future research, and related to reearch question 5: knowing that crop modeling method contributes most to large uncertainties in future crop yield projections, how can this be addressed? Some recommendations can be made here, also following the recommendations for 'good practices in crop modeling' and benchmarks for evaluating skill (e.g. Challinor et al., 2017, Müller et al., 2017): for example, reducing the uncertainties associated with the crop modeling method itself. This means reducing input error and aggregation by using crop models at appropriate scales; when impossible (e.g. for lack of an appropriate model for a crop, or lacking regional data), the deficiencies of the modeling process should be transparently outlined and documented.

Another opportunity to address RQ5 would be to reduce the number of extraneous variables that can influence yield simulation outcomes: for example, by using and carefully selecting experimental parameters to provide better consistency in data and improve the comparability to outcomes from other methods, in this case the SCCM projections. Given the differences in outcome and the influence of the crop modeling method, should one crop modeling approach be used over the other?

### 7.4.4 Advantages and disadvantages of different crop modeling approaches

In this study, SCCMs and PCMs have been shown to produce both diverging and similar yield projections. There are numerous reasons why statistical approaches can be considered adequate for modeling the impacts of future climate change. For instance, general recommendations from multi-method studies are that statistical approaches are useful for testing relatively simple crop-climate relationships and extrapolate based on observed relationships (Watson et al., 2015). Additionally, empirical modeling is considered suitable for analyzing past and current crop-yield patterns (Soltani et al., 2016). On the other hand, statistical approaches are critiqued for their lack of complexity and stationarity, and that they are missing the representation of numerous factors apart from climate that influence plant growth and development (See Chapter 2 for a further discussion on the limitations of statistical approaches).

There are also numerous reasons why PCMs remain advantageous over SCCMs in several contexts. For example, the differences between PCMs and SCCMs in how they handle $CO_2$ is well known. In particular, SCCMs are criticized because they do not include $CO_2$ effects. Although the study does not focus on $CO_2$ effects, the impact that its inclusion into crop models is still important to discuss. For instance, $CO_2$ in modeling studies may change yield projections: in the same Kersebaum and Nendel (2014) study which found negative yield effects due to climate change in Germany, the addition of the $CO_2$ effect and resulting reduced transpiration made yield projections shift towards increasing yields. $CO_2$ is an important consideration for climate change studies, as increased atmospheric $CO_2$ concentration directly increases photosynthesis in C3 plants like wheat, and also decreases stomatal conductance, which thereby increases crop water use efficiency (Kersebaum and Nendel, 2014). Therefore, while climate change has the potential to reduce crop yield, the fertilization effect of $CO_2$ tends to increase yield (Erda et al., 2005) – although it should be noted that C3 crop photosynthesis increases beyond a $CO_2$ concentration of 1000 ppm (Kersebaum and Nendel, 2014).

Apart from the importance of the $CO_2$ effect, the response of wheat crop phenology to climate change also needs to be considered. Although it is well-acknowledged that extreme temperatures and heat stress are likely to reduce wheat yields (Rezaei et al., 2015, Asseng et al., 2014), including phenology responses to climate change and $CO_2$ also can also affect potential yield responses. For example, in their study, Rezaei et al. (2015) found that because of warming causing an acceleration of crop phenology, a cooling effect was observed: earlier wheat crop heading (the stage prior to anthesis/flowering) compensated for the enhanced warming and heat stress due to climate change. Despite this beneficial avoidance of anthesis occurring around the hottest days of the year (and when more hot days are possible), heat stress could still then adversely affect the crop during the grain filling stage, which occurs after anthesis (Rezaei et al., 2015). Since different cultivars (varieties) of wheat also respond differently to temperature and precipitation changes, the choice of cultivar is also an important factor to be considered in climate change impact assessments (Rezaei et al., 2018), something that can be considered by PCMs.

Other important factors that influence yield include the inputs and management used for wheat growing. For instance, in a recent analysis of the relationship between wheat yield volatility, inputs (e.g. capital, labor, energy, fertilizer, *inter alia*) and weather in Germany, both inputs and weather affected yield projections, with a slight majority of the projected changes attributed to input variability (Albers et al., 2017). When input choices were left out, it was found that weather impacts and common shocks would be overestimated (Albers et al., 2017). Given that PCMs are able to incorporate $CO_2$ effects, phenology and inputs can be considered justification that the systems-based understanding of cropping systems, including wheat physiology, can be considered advantageous for many reasons compared to statistical approaches.

However, PCMs run at significant cost (Lobell and Asseng, 2017), and have been criticized for their input-intensive nature (Lobell and Burke, 2010), and infrequent reporting of crucial calibration parameters, among other issues related to upscaling and input parameters (See Chapter 2 and 3 discussion). In addition, a crop modeling study that considers all these above-mentioned factors that influence yield would be very difficult to implement fairly in a comparative approach with statistical approaches. It is argued that the comparison of an extremely complex or intensively calibrated PCM to a structurally simple SCCM would lead to unbalanced results on how different crop modeling methods compare to each other, and if one projection is more plausible than the other, hence the research design was to use regional crop parameters and maintain simulation defaults whenever possible.

It is thus important to ask, how can climate change impact assessments move forward with multi-method comparisons, while considering the challenges of comparing fundamentally different approaches? While more recommendations are discussed in the subsequent Conclusions chapter, it is argued that while it makes sense to consider the complexity of crop growth and development in modeling studies to project a well-informed crop response to climate change, finding the data and calibrating it to run PCMs to simultaneously factor in these numerous parameters (cultivar, inputs, phenology, climate, $CO_2$) is a challenging task. Furthermore, it could make yield projections highly specific and difficult to extrapolate beyond the selected geographical region for which the PCM was run.

Therefore, the comparison of SCCMs and PCMs remains challenging. Any expectations that they will have the same projected changes, are likewise challenging, given the numerous differences between crop modeling methods. It is argued that rather than this be detrimental to the emerging field of crop model method comparison, this should encourage more comparisons and investigation. It is additionally argued that the depth of agronomical knowledge from developing and using PCMs and their ability to model complex $G \times E \times M$ interactions, alongside the powerful, rapid and transparent methods of statistical approaches, are therefore worth the larger effort, cost, and time in using both methods comparatively.

### 7.4.5 Other influences on uncertainty

After crop modeling method, GCMs were typically the next largest influence on the uncertainty. In climate model uncertainty decomposition studies, climate model uncertainty has been shown to be a major source of uncertainty particularly in the early and mid-century for temperature (Hawkins and Sutton, 2009) and precipitation (Hawkins and Sutton, 2011). In the same vein, hydrological impact studies which also use a single impact model have found that climate models dominate fractional uncertainty in most seasons and compared to other factors like RCP scenario and hydrological model uncertainty (Hattermann et al., 2018, Vetter et al., 2017, Bosshard et al., 2013). In agricultural impact studies, climate model uncertainty was greater than a single regional crop model (GLAM), inclusive of the effect of adaptation and natural variability (Vermeulen et al., 2013). In a study that specifically investigated crop model parameterizations, (Koehler et al., 2013) found that the representation of temperature-driven processes in the crop model (also GLAM) was on average larger than climate model uncertainty, indicating the relative importance of crop development.

If the large influence of the crop modeling method was excluded from the analysis, it has been shown that the choice of GCM-RCM is very influential over the fractional uncertainty in impact assessments, although this is also temporally variable (e.g. Hawkins and Sutton, 2009, Northrop and Chandler, 2014, Vetter et al., 2017, Hattermann et al., 2018). This brings the results from Chapter 6 back into focus, where the importance of selecting GCM-RCM combinations with small biases was emphasized.

Based on the results of the uncertainty decomposition, DEA was the only region where the crop modeling method did not unequivocally dominate over other sources of uncertainty, other interactions became more important in the ANOVA. For example, in DEA (RCP8.5) in the mid-century, the contribution of uncertainty from the interaction between GCM and BC method, as well as GCM-BC-Method, was larger than crop model method uncertainty, although these contributions are not significant based on the results of the F-test (Table 7.12). By the end of the century, GCMs, although still not significant based on the F-test, overtake crop modeling method in fractional uncertainty.

These results show that the consideration of first- and second- order terms in the ANOVA can reveal relevant interactions between the choice of method and the other variables in the impact assessment cascade. First- and second-order interactions in the ANOVA also show interesting results which characterize the contribution of BC to uncertainty in yield projections, and these are discussed in the next section.

#### 7.4.5.1 Effect of bias correction on projected changes to yield

In the uncertainty decomposition analysis, the choice of whether BC was applied or not, including the calibration approach (BC-Eval or BC-Hist) in the end did not show a large influence over yield variability. The influence of BC was dwarfed by crop model method and GCM-RCM uncertainty (Figs. 7.12-7.13). In the study of Koehler et al. (2013) which analyzed uncertainty in yield projections with a regional PCM and a 17-member projection ensemble and corrected by linear and change-factor methods, BC was also not found to be a significant source of uncertainty.

In the yield projections of the study, the use of BC-Eval or BC-Hist occasionally modified the significance of the linear trend (particularly for the SCCM results), for example in DEA and DED SCCM simulations where $R^2$ values for the negative tended to increase after BC (Tables 7.5 and 7.7). The uncertainty analysis reveals that in lieu of BC itself, the fractional uncertainty of the interaction between GCM and BC method was found to be more important across all regions and both emission scenarios (Tables 7.12 and 7.13), meaning the way that BC modified the selected GCM-RCMs was more influential on yield uncertainty than the BC approach on its own.

This links back to the importance of the choice of GCM-RCM combination for impact studies, which was discussed at the end of Chapter 6. Poorly-performing GCM-RCMs which contain large biases change significantly after BC, as shown in Chapter 6 where projections 'jump' from the historical range and magnitude of temperature and precipitation to a corrected range. This shift and jump in climate model output in turn results in changes to yield simulations. In contrast, when GCM-RCMs have small biases, BC does not have a large impact on climate model output, and resulting yield simulations are then fairly similar as well. Whether BC is in itself important is argued to be thus more dependent on the choice of GCMs and RCMs.

### 7.4.6 Novel results and implications

The clear influence of crop modeling method in the study, including its interactions with GCM-RCM and BC method, is a key finding. The choice of PCM or SCCM in impact analyses can lead to a large range of plausible outcomes for wheat yields in the face of climate change. Previous studies have made clear progress in better characterizing uncertainty in agricultural impact projections in the cascade of uncertainty. However, they often do not take multi-method approaches to impact assessment.

To the author's knowledge, although there are emerging studies which have analyzed multi-method approaches, there have not been many studies which have taken an fractional uncertainty approach to analyzing yield projections between methods, including a focus on bias correction, so the results presented here are novel. Although it is important to note that these findings are contextual to the research design and data (recalling that ANOVA is used here in fixed-factor mode), there are many more research opportunities to explore from this point to analyze how much crop modeling method affects uncertainty in other geographies, including the $CO_2$ effect, and considering scenario uncertainty, crop model parameterization, among other future research pathways. These opportunities for future work, as well as other recommendations for multi-method comparisons in the context of uncertainty, are discussed in the final Conclusions chapter.

## 7.5 Conclusion

The impacts of climate change, specifically changes to temperature and precipitation, are shown in the research to present both opportunities and threats to wheat yields in the UK and Germany. While small increases are projected for the UK, yield projections are generally toward yield decreases across Germany. While computationally costly process-based crop models have been often placed in contention with the structurally simple statistical approaches to modeling yield, the results here show that their projected yield changes can also sometimes be in agreement. However, there remain differences that are valuable to be explored further, and by following good practices in crop modeling, extraneous variables and input errors that influence uncertainty and yield projections can also be reduced.

Continuing multi-method comparisons for yield projections can be costly in terms of time and effort, but the uncertainty analysis in this chapter reveals how important the influence that the choice of crop modeling method has on yield projections under climate change.

# Chapter 8

# Conclusions

In this chapter, the results of the analysis are reviewed to derive conclusions and recommendations about the outcomes and novel findings of the study. Opportunities to deepen the research are highlighted, alongside critical discussion of key limitations in the study. The research spans the disciplines of climate and crop sciences in an attempt to provide knowledge with a focus on uncertainty by comparing different methods – with the context of the communities of practice that drive these different choices – to understand how intermediate steps and decisions can have impacts on the range of yield projections in a study. To do so, the study has used a number of methods in climate modeling, crop modeling, and the necessary steps to link the two. By doing so, an uncertainty analysis focusing on contributions from climate and crop models, as well as bias correction, downscaling and crop modeling method, was completed.

The research was conducted focusing on the 'what' and 'how' in attempt to better understand the methods behind impact assessment. Firstly, what are the impacts of climate change on wheat production in Europe? Secondly, how is this typically or normally performed? By addressing these questions, a number of key areas that need better focus on in the impact simulation cascade were identified, and are discussed here.

## 8.1 Revisiting research gaps

The review of related literature identified the following gaps in climate-crop impact assessments. Some of these are the following:

- Going beyond the assumption that RCMs automatically provide more skill or information.

  ◇ This was addressed in Chapter 4 where RCMs showed some advantages over their driving GCMs, but results showed that these gains depended on the climate variable. For example, maximum and minimum temperatures from RCMs showed better agreement, but RCM precipitation was not necessarily better than GCM precipitation.

- Investigating the effect of the BC method on climate and yield projections.

  ◇ This was addressed in Chapters 5 and 7. It was shown that all BC methods improved climate simulations relative to observations, but quantile-quantile mapping also improved other features of the climate model output. The improvement of climate model output has downstream improvements in yield simulations.

  ◇ In Chapter 7, it was shown that the calibration of BC (BC-Eval or BC-Hist, calibrated on the RCM evaluation period or the historical period, respectively) affected the magnitude or yield changes, and in some instances changed the direction or significance of the future yield trend.

- Differentiating the error contributed by the choice of the GCM and RCM in a way to understand how this affects future projected climate changes.

  ◇ This was addressed in Chapter 6. BC-Eval and BC-Hist were shown to be able to distinguish several 'cases' of pairings of GCMs and RCMs, which emphasized the importance of selection of well-perfoming GCM-

RCM pairs as this may have impacts on future climate projections and further downstream in impact assessments.

- Investigating the differences between crop modeling approaches.
- Better characterization of the 'intermediate' steps in the linkages between climate and crop models (See Figs. 1.2-1.4); and
- Remembering the larger context of uncertainty in impact assessments.

◇ The last three points are further discussed in the following section.

## 8.2 General discussion of results

**Analysis of the impacts of climate change on wheat yields**

A main objective of the work was to analyze how climate change may impact wheat yields in the future periods until the end of this century. The projections of yield for the chosen sites, the UK and Germany, show that changes in temperature and precipitation will affect yields, providing both threats to wheat production and opportunities for adaptation (Chapter 7). The sensitivity of wheat to rising temperatures particularly around flowering is well-known, but warming climates may actually be beneficial for some wheat-growing areas in Europe, and this was shown in the results with the statistical crop-climate model for the UK, where projections ranged from 10-20% increases relative to a historical baseline.

While yield projections were also of similar magnitudes for Germany, the increases in the current and next decade were followed by yield declines. The trends and magnitude of these changes varied between crop modeling method, and also if bias correction (with different calibrations, BC-Eval or BC-Hist) were applied.

The analysis of wheat yields from the past (1981-2010, Chapter 3) also showed that periods of heat stress with many hot days and periods of low summer rain, have already had impacts on wheat yields in both the UK and Germany. In this analysis, it was also shown that yields in the UK and Germany have been stagnating, particularly after the year 1999 (Chapter 3), and this has previously been linked to both climate and external factors. With the climate change projections from multiple climate models (both directly from simulations and bias-corrected), showing potential increases in the number of hot days, more so for Germany and its regions than the UK, it is clear that heat stress is a present and future risk for wheat yields.

**Uncertainty decomposition and analysis in a multi-method context**

A key focus and result of this study is the decomposition of uncertainty that revealed the choice of crop model method (process-based or statistical) as a major source of uncertainty in the yield projections of the study, followed by climate model and bias correction (Chapter 7). This is a key finding for the study, but one that cannot have been entirely unexpected: it was thoroughly discussed in Chapter 1, the literature review in Chapter 2 and the evaluation of crop model performance in Chapter 3 that there are fundamental differences between the two main methods of simulating crop yields (PCMs and SCCMs). The result of the uncertainty analysis is contributed to in part by the differences in the theoretical purpose and practices of the compared crop model method.

These fundamental differences are manifested in the type of input and processes modeled in process-based models (PCMs) and statistical crop-climate models (SCCMs). Statistical approaches are fed with relatively simple empirical data (large-scale climate indices and yield patterns), but PCMs rely on a large set of input parameters to drive fine-scale processes behind the crop model, for example (in DSSAT/CERES-Wheat) calculating the number of growing degree days to determine the wheat crop growth stage, how much of the available fertilizer has been assimilated, to how this affects the number of grains per head of wheat, and finally their weight to calculate yield. These differences, alongside the original intended scale of application (field-based for the PCM, regional and larger for the SCCM) are also responsible for the input and aggregation errors that led to diverging crop model performances in their evaluation in Chapter 3.

It was discussed in Chapter 7 that the input- and detail-oriented approaches of PCMs contain valuable information (e.g. the $CO_2$ effect) fundamental to an accurate representation of crop growth and development. However, it was also discussed that the opportunities and benefits provided by statistical approaches (e.g. transparency, simplicity, and greater applicability in perhaps data-poor regions) are valuable; furthermore, many process-based approaches are themselves built on empirical data and statistical models.

Each approach also has their associated limitations: stationarity and (over)simplicity for SCCMs, and challenging implementation for PCMs. Some of the challenges and issues of comparing crop modeling methods were summarized, based on the literature review (Chapter 2), as calibration differences, scale mismatches, upscaling and aggregation errors, and stationarity. The research also revealed the complexity and importance of calibrating parameters such as the genetic cultivar coefficients. Because of the large number of processes these parameters control, better reporting of their use in studies should be promoted for reproducibility – but regional and long-term field experiments that provide valuable calibration data for PCMs should also be supported as they provide invaluable knowledge from the field that can be used for validation and evaluation.

These differences, while valuable, have resulted in different communities of practice, which are argued to underpin the development and collaboration between these different ‘camps’ of crop model methods. These differences (between models, and between communities) are also perhaps are why crop model comparison studies are only recently coming into focus despite both camps having decades of scientific research and development. For example, the prominent multi-model ensembles that compare crop model differences are largely limited to PCMs.

Apart from the need for more multi-method studies, it was argued that multi-method comparisons should be contextualized within the cascade of uncertainty from climate models to impact models, which this study has achieved. It has also been reported that multi-method studies for agriculture and crops are still in the beginning stages, so the results of the study contribute to growing body of knowledge on the differences between crop modeling methods while using uncertainty decomposition methods.

Rather than the challenges of comparison being a basis for continuing to work in disciplinary (or crop model method) silos, it is argued that there are valuable similarities, as well as differences, that should be the focus of initiatives and efforts in understanding crop yields, and hence future food security.

### Downscaling, bias correction and calibration

The work also focused on extensively on climate models, particularly around methods which are typically taken as standard in impact assessment: downscaling and bias correction (Chapters 4 and 5). Dynamical downscaling has been shown as a way to develop spatially and temporally higher-resolution climate simulations, which are typically required by crop models. However there is an ongoing scientific debate around added value, as regional climate models (RCMs) still rely on global climate models (GCMs) that have many limitations and parameterized processes (as reviewed in Chapter 2) that can lead to bias in climate simulations relative to observations, even when downscaled (Chapter 4).

While some added value from RCMs was found in the study, the gains were not unequivocal, and the tests to determine this added value were challenging. Thus, it is a key recommendation of the study that while added value can be found in downscaling, it is important to justify the use of downscaling methods in impact assessment.

This recommendation is also important as RCMs were shown in the work to also introduce their own error into climate simulations (Chapter 5). Some of these errors were post-processed using a number of bias correction methods that ranged from simple scaling to more complex distribution-based transformation. Also echoing the different communities of practice within crop modeling, bias correction is also a debated topic within the climate modeling discipline, where it is sometimes seen as solely a post-processing step that does not address underlying error. While this is true – that bias correction cannot improve the actual representation of climate processes within models – both simple and complex bias correction methods (e.g. scaling and quantile mapping) were shown here in the work to reduce errors in RCM simulations. Bias correction was able to improve yield hindcasts generated by two different crop modeling methods (Chapter 5).

Two different calibration approaches to bias correction of future climate change projections were also presented as a new way of thinking about GCM-RCM combinations and how different pairs of climate models can affect projected changes in temperature and precipitation (Chapter 6), and these were also analyzed in the context of uncertainty and multi-method approaches (Chapter 7). This approach has value for selecting GCM-RCM pairs in impact assessment, as it has been shown in previous studies (see Chapter 7 discussion, and papers from e.g. Hawkins and Sutton (2009)) that the choice of climate models has a large impact on uncertainty. An important outcome of the research is understanding how poorly (or well-) paired GCMs and RCMs are, considering the focus and discussion on the cascade of uncertainty.

## 8.3 Limitations of the study

Some of the limitations of the study have already been discussed in each relevant chapter. These previously mentioned limitations of the study include the lack of analysis at the regional level for the UK, which would have made the multi-method comparison of UK wheat yields possible. However, due to data limitations with UK regional data at the time of writing, this was not a feasible step as climate-crop analyses are more robust with data from more than 30 years. Not enough climate model runs for RCP2.6 simulations also hindered the inclusion of scenario uncertainty in the decomposition analysis (Chapter 7). The analysis would have been enhanced by including more simulations, and considering other experimental ensembles of GCMs, and RCMs. The relatively small number of simulations also limits the ANOVA in the uncertainty decomposition to a fixed-factor mode, so using the results of the study outside of the research design should be done with caution.

Other challenges included the lack of regional data for use with the PCM at the regional level in Germany. This includes both the genetic parameters and regional data for evaluation and calibration. While the use of generic or even default parameters has been reported as a standard practice, it is apparent that carefully calibrating each region would have provided more local-specific results which could have improved the correlation of yield hindcasts to observations, for example in DE2 (South Germany). In addition, the work with PCM considered temperature and precipitation changes, but not changes in $CO_2$. However, these choices are rationalized as necessary steps in the analysis since the focus of the study was to create yield projections from the PCM that were feasibly comparable to the SCCM. Other limitations which merit further investigation are how to manage the ‘technology’ trend in the SCCM that remained the same over time (i.e. dealing with stationarity).

More extensive tests and statistical analyses – such as a possible focus on the duration of extreme events, a focus on winter climate, and a more complete characterization of added value, for example – are also not carried out in the study, but are presented as ways the work can be taken forward into the future.

## 8.4 Recommendations for future research and impact analyses

These identified limitations are also ways to outline several opportunities to continue the work of this thesis: for example, including more climate models, scenarios, and simulations can provide a better characterization and quantification of uncertainty using ANOVA, and also provide a fuller picture of both known and unknown unknowns. Additionally including other crop models (both PCMs and SCCMs), and a focus on their input parameters and yield responses, can also enhance the uncertainty analysis carried out in the research. As mentioned in Chapter 3, accessible databases of crop yields, climate, crop parameters should be supported for better dissemination of information, but also reproducibility.

The work carried out in this thesis was originally designed to offer steps in better characterizing and understanding uncertainty in yield projections. Some of the recommendations from the work include, as previously discussed, going beyond implicit claims: be it for the value of downscaling, the choice of one bias correction method (or calibration over the other), or that one crop modeling method is superior to the other. Other recommendations are following examples and guidelines for good crop modeling practice: this means thorough and rigorous evaluation and calibration with sufficient input information, an understanding of the underlying processes in models, and critical analysis of the results.

Uncertainty analyses, such as the ANOVA method used in this study, should also become a more integrated component of impact assessment studies, considering how important the cascade of uncertainty is. An additional recommendation is to consider how the selection of GCMs and RCMs (although these are mostly decided upon for reasons of data availability or convenience, as discussed in previous chapters) can affect future climate and yield projections, so developing a systematic way for selection is recommended (e.g. ranking, reviewing previous studies, or the BC-Eval/BC-Hist approach presented here).

Another big-picture recommendation, in the vein of promoting more multi-method studies, is that the different communities of practice (climate, crop, and the in-between) collaborate and communicate more effectively to better resolve and develop join approaches to understanding food security.

**Resolving the crop model method 'conflict'**

A large focus of the work was comparing crop modeling methods, so some recommendations are also discussed here: the results of the work agree with previous recommendations that SCCMs/statistical approaches are a rapid and useful way of understanding future changes to yield based on a number of chosen indices. They are also extremely useful in understanding past relationships between yields and climate.

However, given the complexity of how crops are actually cultivated and produced – including the choice of what variety of wheat, how intensively it is managed, and how it can also be affected by many other factors apart from climate, PCMs provide decades of valuable agronomical knowledge and a systems-based understanding of crop development. At present, their ability to handle changes in $CO_2$ and incorporate changes to cropping patterns, also make them very suitable for future yield projections.

Both approaches have serious limitations and issues. But rather than these two methods be pitted against each other – as they are frequently done in previous studies – these approaches can be complementary. For example, the relatively inflexible structure of PCMs to focus on important periods of crop development, such as heat stress, can be aided by statistical approaches. While SCCMs are argued to be too basic to include complex crop responses, they are extremely useful in places with limited data, and can thus guide where future research and support are needed. Both approaches provide ways to understand future climate change impacts, and both also help in understanding uncertainty.

### 8.4.1 Concluding remarks

The impacts of climate change on important food crops like wheat are not just a distant threat: climate change and variability are already affecting agriculture. Gaining knowledge on potential impacts and ways human communities around the world can adapt are invaluable for adaptation and food security. While research that uses climate and crop models has made tremendous progress in representing the complexity of the interactions between temperature, precipitation, and yield, the cascade of uncertainty remains due to different methods, the inherent shortcomings of models, and our own limitations in knowledge.

A key recommendation that is important to this work is that both the strengths and limitations of the impact assessment studies (and the methods and data they use) are communicated transparently to better understand where more work is needed, in order to support more collaborative initiatives to prepare, adapt, and transform to our changing climate.

# Climate analysis and RCM evaluation

In this appendix, supplementary information is provided on the evaluation of reanalysis-driven RCM simulations from 1981-2010, particularly on the seasonal and daily timestep, as well as individual regional RCM evaluation-driven crop simulations (SCCM and PCM). It is structured in the following way:

**A. Country, regional, and site climate analysis: Observations of climate**

1. Temperature trends
2. Precipitation trends
3. Climate index trends

**B. Seasonal analysis (Hot day index and JJA total precipitation)**

1. Statistical analysis (correlation, RMSE and mean bias) for the UK and Germany climate simulations
2. Statistical analysis (correlation, RMSE and mean bias) for regional German climate simulations

**C. Daily analysis of Tmax, Tmin and precipitation**

1. Statistical analysis (correlation, RMSE, mean bias, and Kolmogorov-Smirnov (KS) test statistics) for the UK and Germany climate simulations
2. Statistical analysis (correlation, RMSE and mean bias, and Kolmogorov-Smirnov (KS) test statistics) for regional German climate simulations
3. Empirical cumulative distribution function and probability density function graphs for the UK and Germany climate simulations
4. Empirical cumulative distribution function and probability density function graphs for regional German climate simulations
5. Taylor diagrams for climate model simulations of the UK and Germany climate
6. Taylor diagrams for climate model simulations of regional German climate

7. Statistical analysis (correlation, RMSE and mean bias) for SCCM yield simulations for regional German climate simulations using uncorrected and bias-corrected RCM output
8. Statistical analysis (correlation, RMSE and mean bias) for PCM yield simulations for regional German climate simulations using uncorrected and bias-corrected RCM output

## A. Country, regional, and site climate analysis

In this section of results, a brief analysis of climate trends is discussed for temperature, precipitation, and the summer climate indices for the UK, Germany and the four German states.

## A1. Temperature trends

### National level

E-OBS data between 1961-2013 shows that the highest average temperatures (Tavg) in the UK and Germany are, as expected, in the summer months of July and August (Fig. A1A, A2A). For example, E-OBS data shows Tavg above 20°C in July and August in Germany. The December-February (DJF) months typically had the coolest Tavg, followed by March-May (MAM), September-November (SON) for both countries (Figs. A1B, A2B). Seasonal and annual Tavg show significant positive trends for the UK and Germany (Fig. A1C, A2C), Table A1).

### German regional level and site level

Climate analysis of the four chosen German regions (1979-2014, to match regional yield data) show similar monthly, seasonal and annual temperature trends compared to national-level climate averages (Figs. A3-A6). The coolest Tmax, on average, are observed in DEF, the northernmost German state used in the analysis. Trend analysis showed that DE2 and DEF all show significant increasing trends for Tavg, Tmax and Tmin while DEA and DED only show increasing trends for annual Tmax. All regions show significant increasing Tavg trends for MAM and JJA seasons, with DEF also showing a significant increasing trend for SON (Table A1).

The monthly, seasonal and annual cycle of temperature and precipitation are also analyzed for BL for the period 1978-2014 (Fig. A7). The hottest months are between JJA. Tmin is, on average, below freezing during DJF. Seasonal Tavg shows no significant trends but annual Tavg has increased since 1978 in BL.

*Table A1: Temperature trend analysis.*

| **Region** | **Annual** | | | **DJF** | | | **MAM** | | |
|---|---|---|---|---|---|---|---|---|---|
| | Tavg | Tmax | Tmin | Tavg | Tmax | Tmin | Tavg | Tmax | Tmin |
| UK | 0.18* | NS | 0.42* | 0.07* | NS | 0.08* | 0.09* | NS | 0.21* |
| Germany | 0.28* | 0.31* | 0.2* | 0.06* | 0.07* | NS | 0.21* | 0.26* | 0.11* |
| DE2 (S Germany) | 0.23* | 0.28* | 0.13* | NS | NS | NS | 0.15* | 0.21* | NS |
| DEA (W Germany) | 0.16* | 0.26* | NS | NS | NS | NS | 0.1* | 0.2* | NS |
| DED (E Germany) | NS | 0.15* | NS | NS | NS | NS | NS | 0.11* | NS |
| DEF (N Germany) | 0.23* | 0.27* | 0.17* | NS | NS | NS | 0.11* | 0.17* | NS |
| BL (site) | 0.2* | 0.15* | 0.11* | NS | NS | NS | 0.13* | NS | NS |

*Table A1 continued.*

| **Region** | **JJA** | | | **SON** | | |
|---|---|---|---|---|---|---|
| | Tavg | Tmax | Tmin | Tavg | Tmax | Tmin |
| UK | 0.08* | NS | 0.4* | 0.09* | NS | 0.17* |
| Germany | 0.27* | 0.22* | 0.32* | NS | NS | NS |
| DE2 (S Germany) | 0.16* | 0.2* | NS | NS | NS | NS |
| DEA (W Germany) | 0.1* | 0.17* | NS | NS | NS | NS |
| DED (E Germany) | 0.09* | 0.16* | NS | NS | NS | NS |
| DEF (N Germany) | 0.15* | 0.15* | 0.1* | 0.12* | 0.11* | 0.11* |
| BL (site) | 0.22* | 0.18* | 0.12* | NS | NS | NS |

*($*$) indicates statistical significance ($p$ <0.05), and a positive trend.*

## A2. Precipitation trends

E-OBS precipitation data for the UK and Germany shows that distinct rainy periods: in the UK, most precipitation is received between October to January (Fig. A1D) while Germany has generally lower precipitation totals and receives relatively even rainfall throughout the year, with a small peak in the JJA months (Fig. A2D); this is reflective of their respective geography and corresponding climate. German states show similar climate patterns to Germany overall, with DE2 and DED also receiving more rainfall in JJA. DEA and DEF show more even patterns of rainfall distribution throughout the year. The precipitation records for BL are similar to German regions and Germany overall: the most precipitation occurs during the summer and the driest month is typically February.

Both countries do not show any significant trends in seasonal precipitation, but there is a small positive trend in annual UK precipitation, although the $R^2$ value is small ($R^2$=0.09). Apart from these, the only significant linear trends are a negative trend in precipitation between March-May in DEA ($R^2$=0.15), and increasing precipitation in JJA for the BL site ($R^2$=0.1). Apart from DEA, no region showed significant trends in annual or seasonal rainfall (Figs. A3-A6).

## A3. Climate index trend analysis

Analysis of E-OBS data shows that there are significantly more days above 31°C ($T_H$) averaged over Germany compared to the UK between 1961-2013 (Fig. A8A-B). In Germany, hot days are observed particularly in the years that also showed low summer precipitation: for example, 1964, 1976, 1983, 1994, 2003, 2006 and 2010. While the UK had notable hot years in the summers of 1976, 1990, 2003, and 2006, the $T_H$ index is low and there are no significant trends. $T_H$ in Germany increased significantly between 1961-2013 ($R^2$=0.13).

At the regional level, analysis of summer climate predictors shows that North Germany (DEF) experienced relatively fewer hot days than other German regions: for example during the heat wave of 2003, there were seventeen observed days above 31°C in JJA averaged over the whole DE2 region; in DEF this was only three days. High numbers of hot days also coincide with relatively low summer precipitation, again for years 1994, 2003, and 2006, giving evidence for some interaction between the two climate predictors (Figs. A9). The number of hot days is observed to be increasing in DE2 and DEA ($R^2$=0.11 and 0.09, respectively).

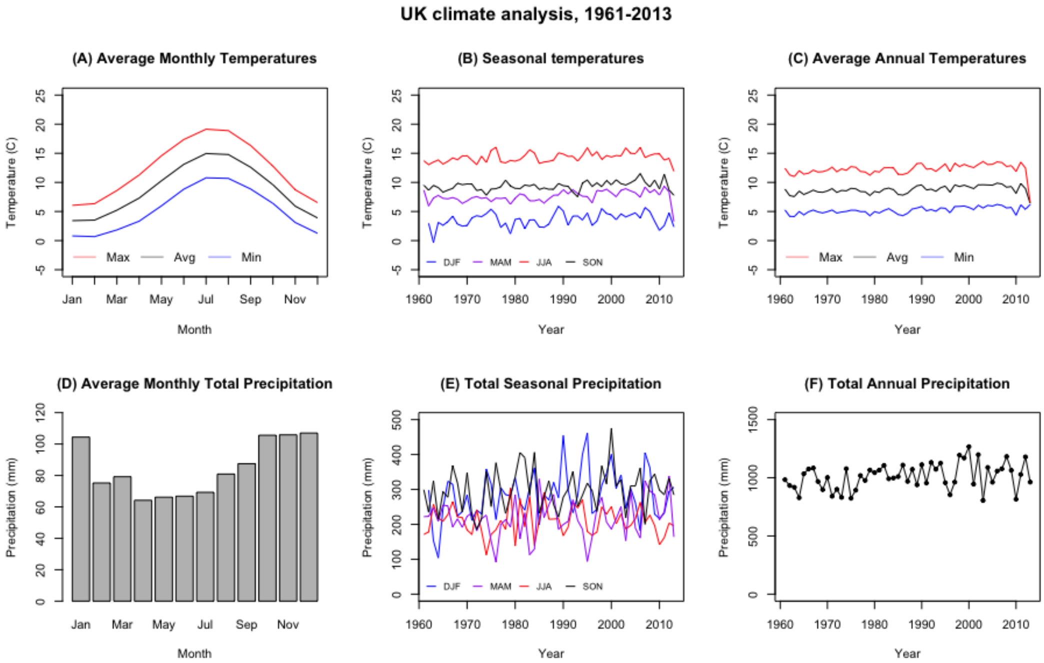



*Figure A1: UK climate averages, 1961-2013.*

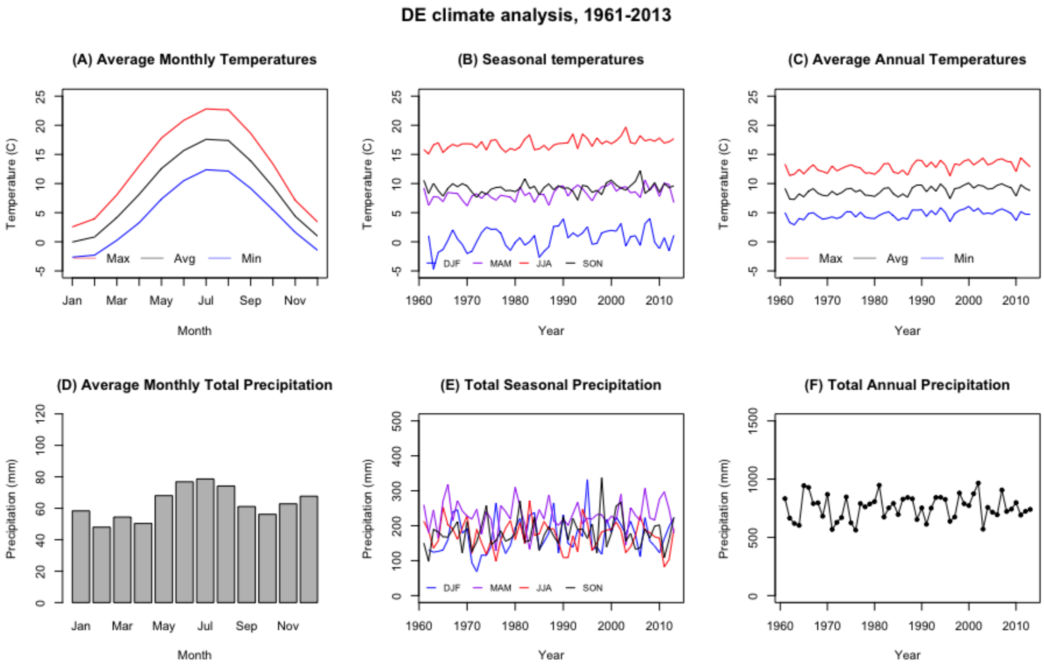



*Figure A2: Germany climate averages, 1961-2013.*

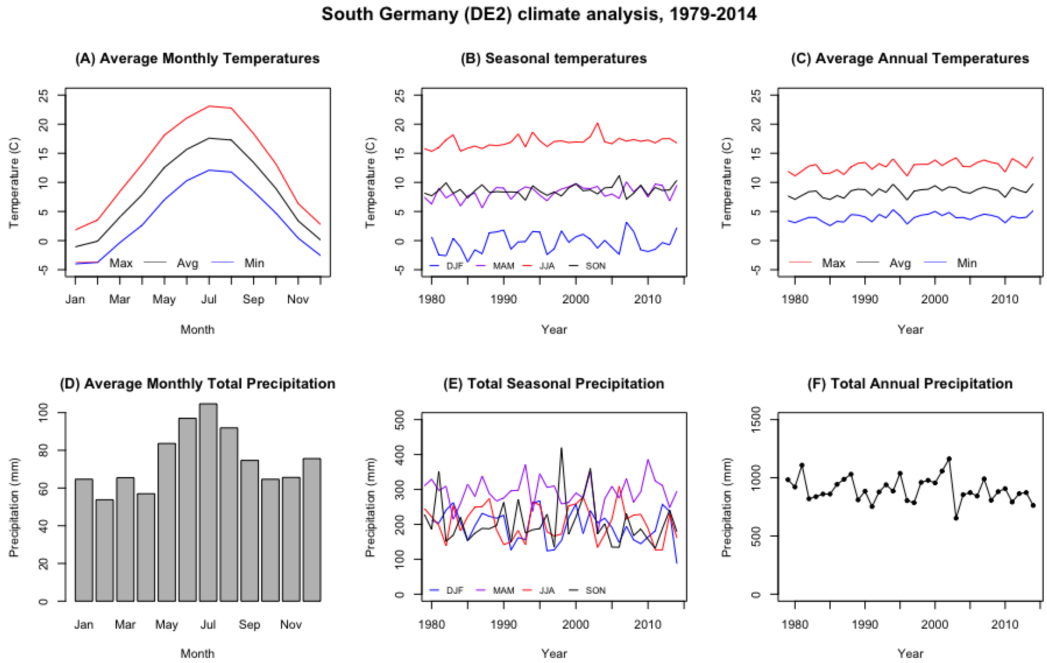



*Figure A3: South Germany (DE2) climate averages, 1979-2014.*

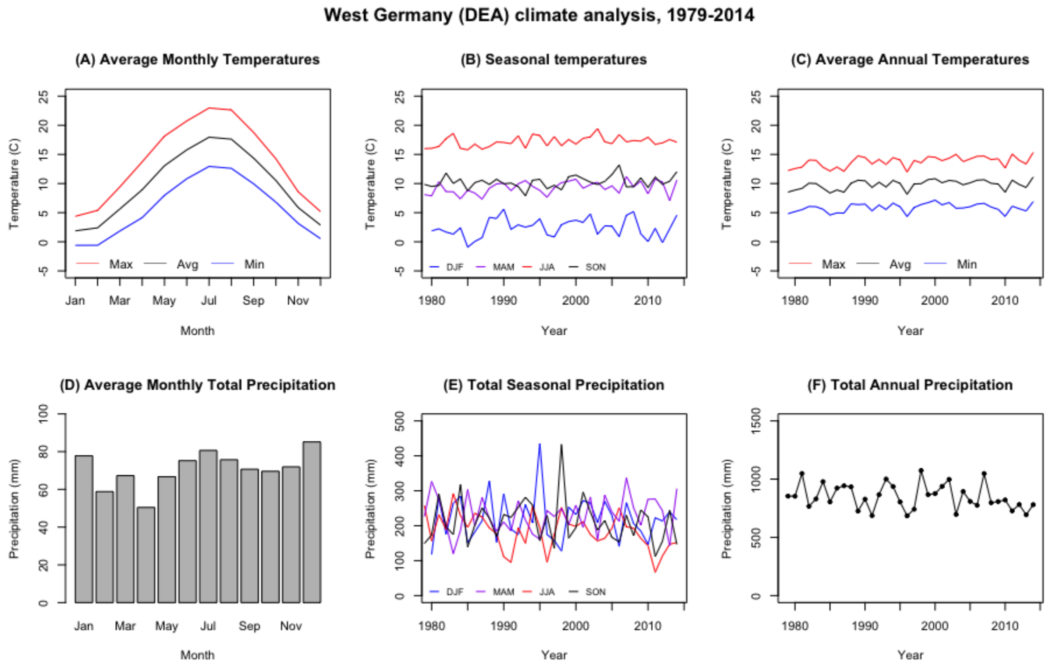



*Figure A4: West Germany (DEA) climate averages, 1979-2014.*

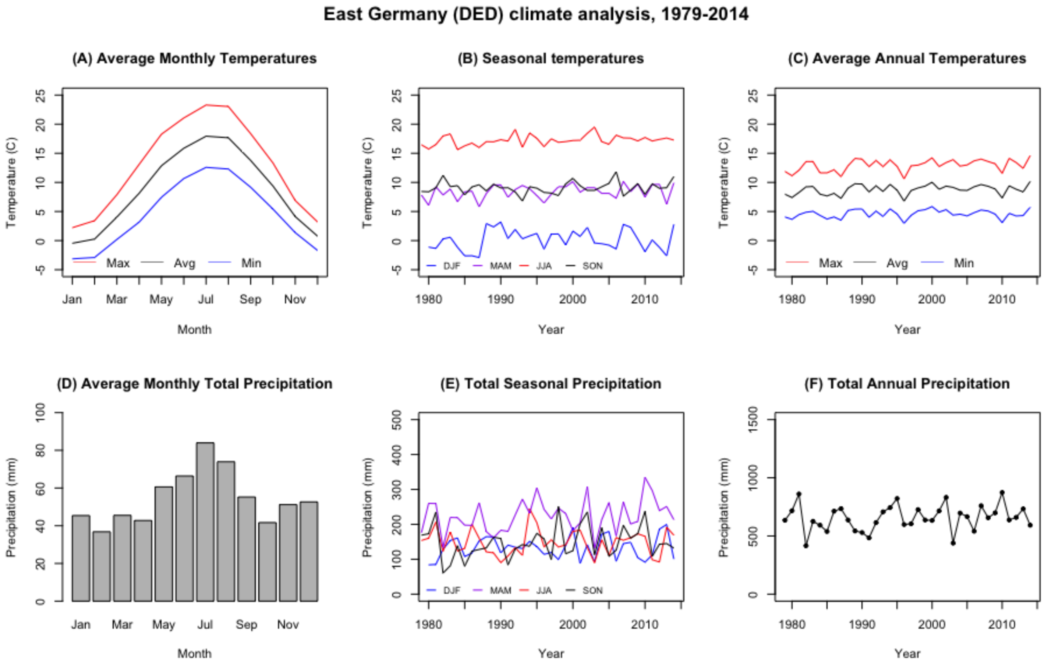



*Figure A5: East Germany (DED) climate averages, 1979-2014.*

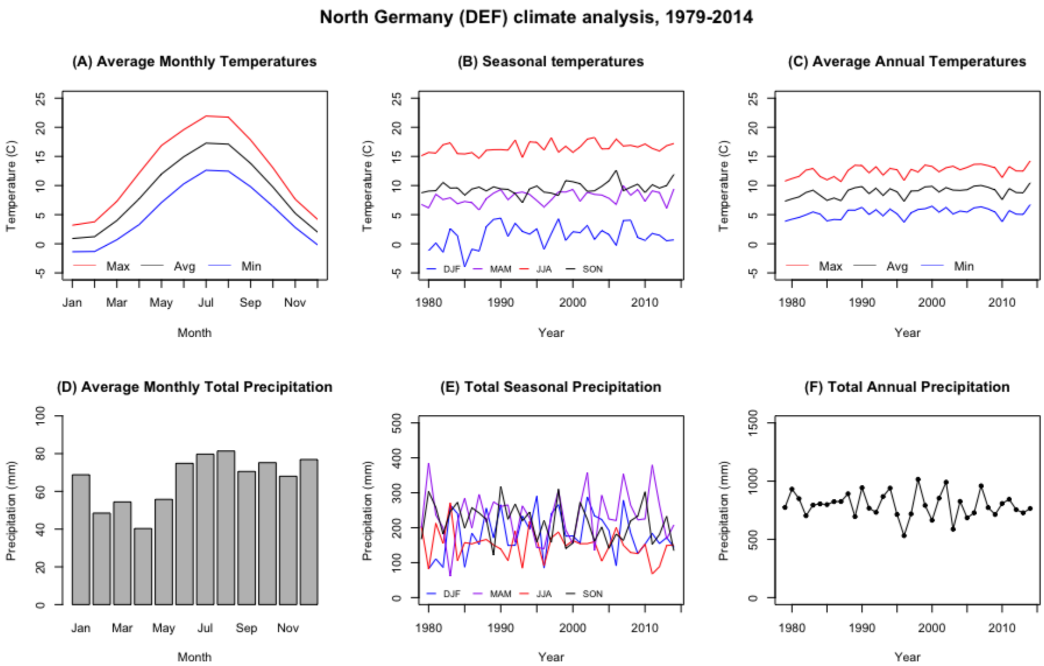



*Figure A6: North Germany (DEF) climate averages, 1979-2014.*

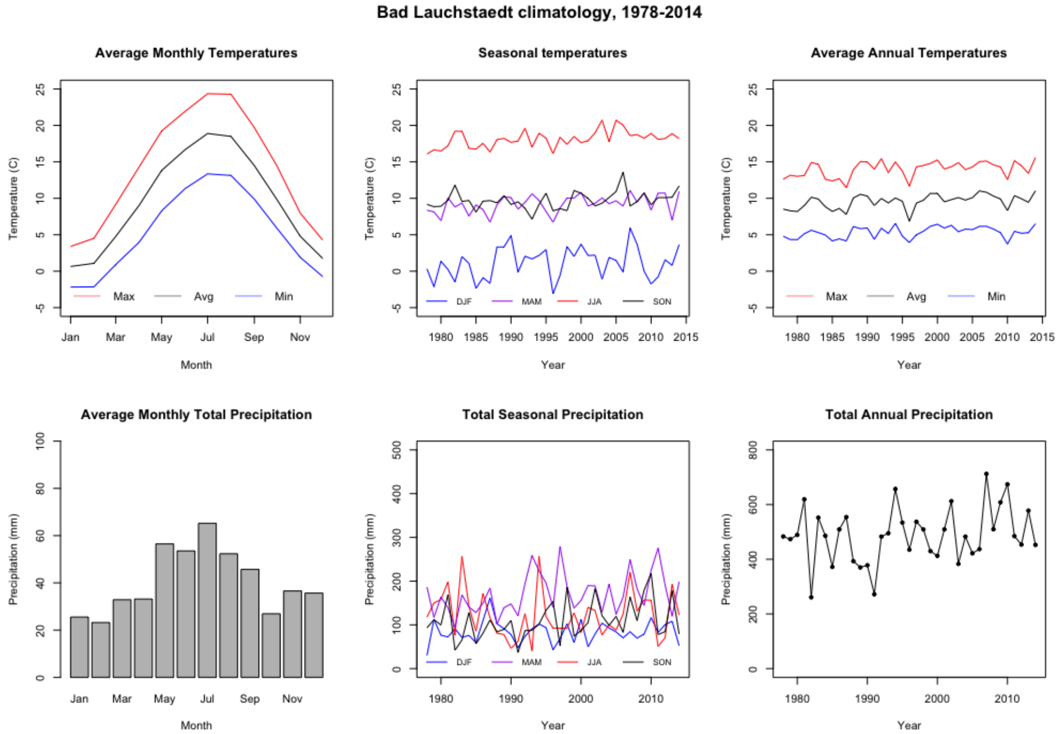



*Figure A7: Bad Lauchstädt climate averages, 1978-2014.*

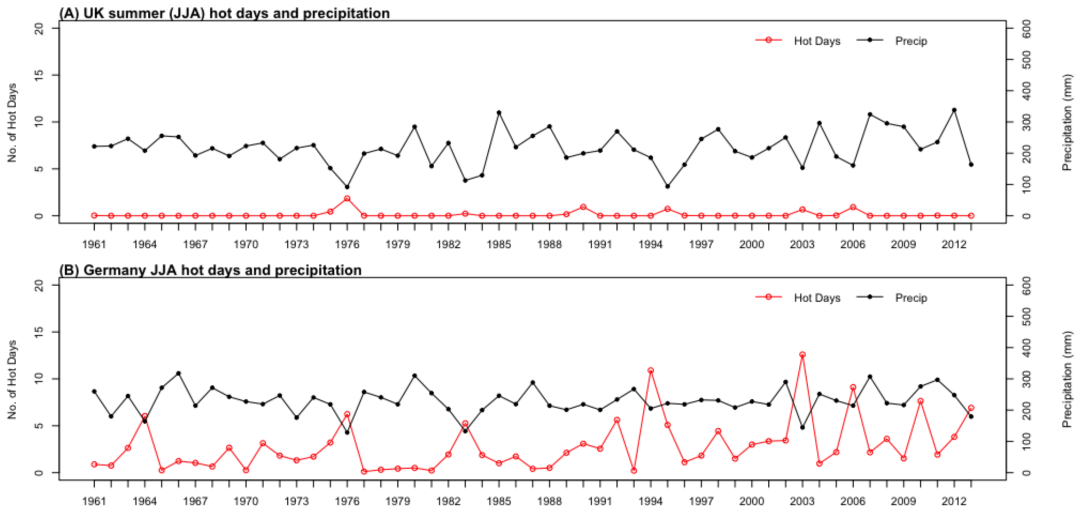



*Figure A8: Hot day count ($T_H$) and mean summer precipitation ($\bar{P}_S$) in (A) the UK and (B) Germany.*

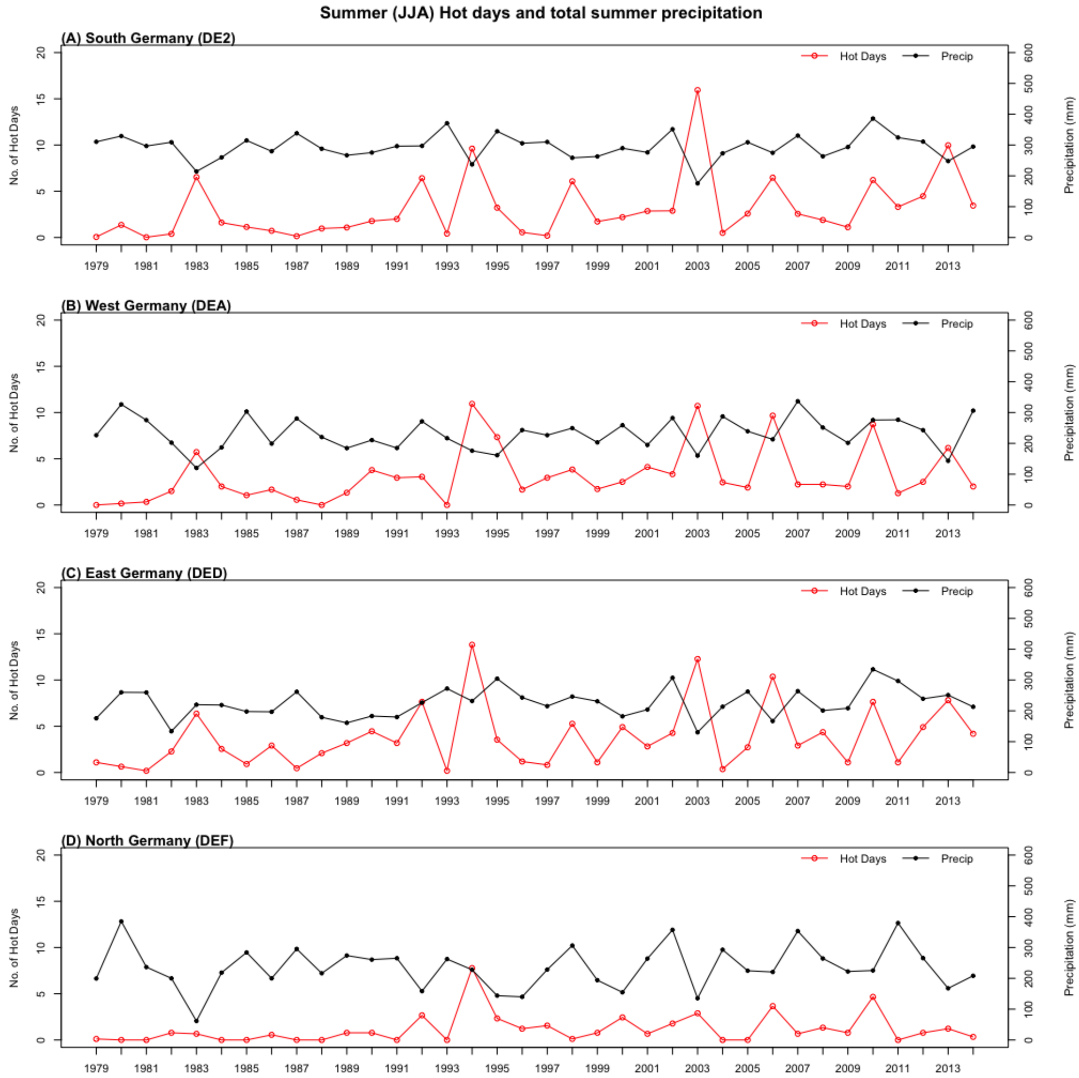



*Figure A9: Hot day count ($T_H$) and mean summer precipitation ($\bar{P}_S$) in German regions (A)-(D).*

# B. Seasonal analysis (Hot day index and JJA total precipitation)

## B1. National level

*Table A2: Statistical comparison between seasonal (summer, June-August or JJA) climate indices, UK and Germany, 1981-2010. (* is $p$ <0.05).*

*Table A2a. Hot day index (days above 31° C).*

| **Model and BC method** | **UK** | | | **Germany** | | |
|---|---|---|---|---|---|---|
| | **Correl.** | **RMSE** | **Mean bias** | **Correl.** | **RMSE** | **Mean bias** |
| **Raw CCLM** | NA | 0.31 | -0.13 | 0.72, * | 3.49 | -2.73 |
| **Raw RACMO** | NA | 0.31 | -0.13 | 0.68, * | 2.97 | -2 |
| **Raw RCA** | NA | 0.31 | -0.13 | 0.61, * | 4.72 | 0.63 |
| **Linear BC CCLM** | NA | 0.31 | -0.13 | 0.85, * | 2.24 | -1.13 |
| **Linear BC RACMO** | NA | 0.31 | -0.13 | 0.64, * | 2.73 | 0.13 |
| **Linear BC RCA** | NA | 0.27 | 0.01 | 0.66, * | 5.94 | 2 |
| **Variance BC CCLM** | NA | 0.31 | -0.13 | 0.85, * | 2.2 | -1.03 |
| **Variance BC RACMO** | NA | 0.31 | -0.13 | 0.64, * | 2.54 | -0.8 |
| **Variance BC RCA** | NA | 0.31 | -0.13 | 0.64, * | 4.03 | -0.57 |
| **QQ BC CCLM** | NA | 0.31 | -0.13 | 0.8, * | 2.28 | -1.1 |
| **QQ BC RACMO** | NA | 0.31 | -0.13 | 0.54, * | 2.91 | -0.8 |
| **QQ BC RCA** | NA | 0.33 | -0.06 | 0.58, * | 4.47 | -0.47 |

*Table A2b. Total JJA precipitation.*

| **Model and BC method** | **UK** | | | **Germany** | | |
|---|---|---|---|---|---|---|
| | **Correl.** | **RMSE** | **Mean bias** | **Correl.** | **RMSE** | **Mean bias** |
| Raw CCLM | 0.96, * | 47.32 | -42.98 | 0.7, * | 35.7 | -16.32 |
| Raw RACMO | 0.92, * | 24.85 | -1.83 | 0.73, * | 35.08 | 0.78 |
| Raw RCA | 0.8, * | 55.41 | 40.24 | 0.53, * | 57.03 | -4.55 |
| Linear BC CCLM | 0.96, * | 17.16 | -0.89 | 0.71, * | 33.5 | 0.25 |
| Linear BC RACMO | 0.92, * | 24.29 | 0.7 | 0.73, * | 35.1 | 1.29 |
| Linear BC RCA | 0.8, * | 35.9 | 1.21 | 0.53, * | 59.61 | 0.91 |
| Variance BC CCLM | 1, * | 0 | 0 | 1, * | 0 | 0 |
| Variance BC RACMO | 1, * | 0 | 0 | 1, * | 0 | 0 |
| Variance BC RCA | 1, * | 0 | 0 | 1, * | 0 | 0 |
| QQ BC CCLM | 0.96, * | 18.25 | -4.46 | 0.71, * | 32.19 | -2.79 |
| QQ BC RACMO | 0.91, * | 25.75 | -2.18 | 0.71, * | 41.31 | -2.69 |
| QQ BC RCA | 0.8, * | 37.84 | -3.25 | 0.52, * | 66.16 | -2.54 |

## B2. Regional level

*Table A3: Statistical comparison between seasonal (summer, June-August or JJA) hot day index (days above 31°C), German regions, 1981-2010. (* is $p$ <0.05).*

| Model and BC method | DE2 (South Germany) | | | DEA (West Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.75, * | 3.03 | -1.93 | 0.67, * | 3.38 | -2.54 |
| **Raw RACMO** | 0.63, * | 2.7 | -0.83 | 0.62, * | 2.7 | -1.17 |
| **Raw RCA** | 0.62, * | 7 | 3.87 | 0.64, * | 5.42 | 1.59 |
| **Linear BC CCLM** | 0.81, * | 2.75 | -0.09 | 0.77, * | 2.21 | -1.01 |
| **Linear BC RACMO** | 0.58, * | 3.21 | 0.71 | 0.59, * | 4.16 | 1.09 |
| **Linear BC RCA** | 0.64, * | 7.47 | 4.01 | 0.68, * | 6.83 | 3.03 |
| **Variance BC CCLM** | 0.81, * | 2.95 | 0.14 | 0.77, * | 2.2 | -0.74 |
| **Variance BC RACMO** | 0.6, * | 2.72 | -0.49 | 0.66, * | 3.16 | 0.23 |
| **Variance BC RCA** | 0.66, * | 4.41 | 0.07 | 0.67, * | 4.22 | 0.26 |
| **QQ BC CCLM** | 0.81, * | 2.99 | 0.41 | 0.71, * | 2.54 | -0.47 |
| **QQ BC RACMO** | 0.59, * | 2.71 | -0.39 | 0.51, * | 4.16 | 0.39 |
| **QQ BC RCA** | 0.64, * | 4.71 | 0.17 | 0.62, * | 4.97 | 0.33 |

*Table A3 continued.*

| Model and BC method | DED (East Germany) | | | DEF (North Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.84, * | 2.98 | -2.23 | 0.42, * | 1.81 | -0.83 |
| **Raw RACMO** | 0.61, * | 3.27 | -1.33 | 0.81, * | 1.36 | -0.8 |
| **Raw RCA** | 0.61, * | 5.08 | 1.24 | 0.73, * | 2.18 | 0.17 |
| **Linear BC CCLM** | 0.87, * | 2.07 | -0.69 | 0.57, * | 1.57 | -0.46 |
| **Linear BC RACMO** | 0.64, * | 3.34 | 0.77 | 0.82, * | 1.01 | -0.1 |
| **Linear BC RCA** | 0.68, * | 7.02 | 3.37 | 0.69, * | 3.26 | 1 |
| **Variance BC CCLM** | 0.87, * | 2.14 | -0.36 | 0.56, * | 1.72 | -0.26 |
| **Variance BC RACMO** | 0.62, * | 3.5 | -0.03 | 0.78, * | 1.14 | 0.07 |
| **Variance BC RCA** | 0.71, * | 4.9 | 0.94 | 0.7, * | 3.1 | 0.84 |
| **QQ BC CCLM** | 0.82, * | 2.76 | -0.03 | 0.43, * | 2.18 | 0 |
| **QQ BC RACMO** | 0.55, * | 3.63 | 0.07 | 0.68, * | 1.48 | 0.2 |
| **QQ BC RCA** | 0.7, * | 4.99 | 0.87 | 0.7, * | 3.31 | 0.8 |

*Table A4: Statistical comparison between seasonal (summer, June-August or JJA) total precipitation, German regions, 1981-2010. (* is $p$ <0.05).*

| Model and BC method | DE2 (South Germany) | | | DEA (West Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.48, * | 58.89 | 18 | 0.55, * | 66.97 | -49.81 |
| **Raw RACMO** | 0.58, * | 54.89 | -21.46 | 0.6, * | 48.96 | -6.95 |
| **Raw RCA** | 0.44, * | 106.9 | -87.41 | 0.57, * | 58.57 | -8.95 |
| **Linear BC CCLM** | 0.49, * | 53.13 | -1.87 | 0.58, * | 50.87 | -0.6 |
| **Linear BC RACMO** | 0.59, * | 54.14 | -0.54 | 0.59, * | 50.01 | 0.69 |
| **Linear BC RCA** | 0.42, * | 89.93 | 0.49 | 0.57, * | 62.07 | 1.97 |
| **Variance BC CCLM** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **Variance BC RACMO** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **Variance BC RCA** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **QQ BC CCLM** | 0.5, * | 52.87 | -0.57 | 0.62, * | 45.11 | -5.3 |
| **QQ BC RACMO** | 0.57, * | 64.3 | 0.62 | 0.58, * | 56.99 | -1.8 |
| **QQ BC RCA** | 0.42, * | 97.46 | 0.49 | 0.57, * | 64.9 | -2.92 |

*Table A4 continued.*

| Model and BC method | DED (East Germany) | | | DEF (North Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.46, * | 59.6 | -26.92 | 0.58, * | 82.3 | -61.23 |
| **Raw RACMO** | 0.43, * | 60.46 | -2.49 | 0.6, * | 59.22 | -21.08 |
| **Raw RCA** | 0.5, * | 70.44 | 20.51 | 0.3 | 84.12 | 9.58 |
| **Linear BC CCLM** | 0.47, * | 58.83 | 1.46 | 0.59, * | 63.01 | -1.08 |
| **Linear BC RACMO** | 0.45, * | 59.77 | 0.36 | 0.58, * | 58.64 | 3 |
| **Linear BC RCA** | 0.49, * | 66.63 | -1.01 | 0.31 | 82.9 | 3.11 |
| **Variance BC CCLM** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **Variance BC RACMO** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **Variance BC RCA** | 1, * | 0 | 0 | 1, * | 0 | 0 |
| **QQ BC CCLM** | 0.47, * | 56.74 | -2.19 | 0.59, * | 60.78 | -2.33 |
| **QQ BC RACMO** | 0.47, * | 70.38 | 4.02 | 0.54, * | 67.72 | 1.78 |
| **QQ BC RCA** | 0.49, * | 78.3 | 3.07 | 0.32 | 87.82 | 0.45 |

# C. Daily analysis of Tmax, Tmin and precipitation

## C1. Statistical analysis (correlation, RMSE and mean bias) for the UK and Germany climate simulations

*Table A5: Statistical comparison between daily values of RCM evaluation simulations and observations, UK and Germany, 1981-2010. (* is $p < 0.05$).*

*Table A5a. Maximum temperature.*

| Model and BC method | UK | | | Germany | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.94, * | 2.31 | -1.27 | 0.97, * | 2.47 | -1.49 |
| **Raw RACMO** | 0.94, * | 2.4 | -1.48 | 0.96, * | 2.63 | -0.92 |
| **Raw RCA** | 0.92, * | 2.43 | -1.13 | 0.94, * | 3.1 | -0.9 |
| **Linear BC CCLM** | 0.94, * | 1.92 | 0 | 0.97, * | 1.93 | 0 |
| **Linear BC RACMO** | 0.95, * | 1.76 | -0.01 | 0.96, * | 2.38 | -0.02 |
| **Linear BC RCA** | 0.93, * | 2.05 | -0.01 | 0.94, * | 2.83 | -0.02 |
| **Variance BC CCLM** | 0.95, * | 1.72 | 0 | 0.97, * | 1.95 | 0 |
| **Variance BC RACMO** | 0.95, * | 1.74 | -0.01 | 0.96, * | 2.33 | -0.02 |
| **Variance BC RCA** | 0.94, * | 1.97 | -0.01 | 0.95, * | 2.76 | -0.02 |
| **QQ BC CCLM** | 0.95, * | 1.73 | -0.08 | 0.97, * | 1.98 | -0.09 |
| **QQ BC RACMO** | 0.95, * | 1.77 | -0.08 | 0.96, * | 2.36 | -0.09 |
| **QQ BC RCA** | 0.93, * | 1.98 | -0.08 | 0.94, * | 2.79 | -0.09 |

*Table A5b. Minimum temperature.*

| Model and BC method | UK | | | Germany | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.94, * | 1.8 | 0.53 | 0.96, * | 2.03 | 0.53 |
| **Raw RACMO** | 0.93, * | 1.74 | 0.36 | 0.95, * | 2.3 | -0.86 |
| **Raw RCA** | 0.92, * | 1.95 | 0.74 | 0.93, * | 2.38 | 0.27 |
| **Linear BC CCLM** | 0.93, * | 1.66 | 0 | 0.96, * | 1.86 | 0.01 |
| **Linear BC RACMO** | 0.93, * | 1.67 | -0.01 | 0.95, * | 2.02 | -0.01 |
| **Linear BC RCA** | 0.92, * | 1.79 | 0 | 0.93, * | 2.35 | 0 |
| **Variance BC CCLM** | 0.93, * | 1.69 | 0 | 0.96, * | 1.84 | 0.01 |
| **Variance BC RACMO** | 0.93, * | 1.7 | 0 | 0.95, * | 1.98 | -0.01 |
| **Variance BC RCA** | 0.92, * | 1.82 | 0 | 0.93, * | 2.32 | 0 |
| **QQ BC CCLM** | 0.93, * | 1.69 | -0.06 | 0.96, * | 1.87 | 0 |
| **QQ BC RACMO** | 0.93, * | 1.72 | -0.07 | 0.95, * | 2.02 | -0.01 |
| **QQ BC RCA** | 0.92, * | 1.83 | -0.06 | 0.93, * | 2.34 | 0 |

*Table A5c. Precipitation.*

| Model and BC method | UK | | | Germany | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.72, * | 2.36 | -0.36 | 0.7, * | 2.11 | 0.11 |
| **Raw RACMO** | 0.7, * | 2.4 | -0.04 | 0.68, * | 2.15 | 0.22 |
| **Raw RCA** | 0.64, * | 2.75 | 0.36 | 0.58, * | 2.43 | 0.31 |
| **Linear BC CCLM** | 0.72, * | 2.47 | -0.01 | 0.7, * | 2.11 | 0 |
| **Linear BC RACMO** | 0.71, * | 2.4 | 0 | 0.68, * | 2.08 | 0.01 |
| **Linear BC RCA** | 0.66, * | 2.56 | -0.01 | 0.58, * | 2.33 | 0.01 |
| **Variance BC CCLM** | 0.75, * | 2.32 | 0 | 0.75, * | 1.9 | 0 |
| **Variance BC RACMO** | 0.73, * | 2.37 | 0 | 0.72, * | 2.01 | 0 |
| **Variance BC RCA** | 0.69, * | 2.56 | 0 | 0.64, * | 2.29 | 0 |
| **QQ BC CCLM** | 0.72, * | 2.48 | 0.03 | 0.71, * | 2.1 | 0.02 |
| **QQ BC RACMO** | 0.69, * | 2.87 | 0.04 | 0.68, * | 2.3 | 0.03 |
| **QQ BC RCA** | 0.65, * | 2.91 | 0.03 | 0.58, * | 2.58 | 0.03 |

*Table A6: Kolmogorov-Smirnov (KS) test statistics on the distribution of daily maximum and minimum temperature, and precipitation from RCM evaluation simulations and observations, UK and Germany, 1981-2010.*

| Model and BC method | UK | | | Germany | | |
|---|---|---|---|---|---|---|
| | Tmax | Tmin | Precip | Tmax | Tmin | Precip |
| **Raw CCLM** | 0.09, * | 0.06, * | 0.06, * | 0.07, * | 0.06, * | 0.11, * |
| **Raw RACMO** | 0.14, * | 0.04, * | 0.12, * | 0.06, * | 0.07, * | 0.12, * |
| **Raw RCA** | 0.11, * | 0.06, * | 0.1, * | 0.07, * | 0.02, * | 0.11, * |
| **Linear BC CCLM** | 0.02, * | 0.02 | 0.06, * | 0.01 | 0.02, * | 0.11, * |
| **Linear BC RACMO** | 0.01 | 0.02 | 0.11, * | 0.02 | 0.02 | 0.12, * |
| **Linear BC RCA** | 0.02 | 0.01 | 0.09, * | 0.02, * | 0.02 | 0.11, * |
| **Variance BC CCLM** | 0.01 | 0.01 | 0.06, * | 0.01 | 0.01 | 0.11, * |
| **Variance BC RACMO** | 0.01 | 0.01 | 0.08, * | 0.01 | 0.01 | 0.12, * |
| **Variance BC RCA** | 0.01 | 0.01 | 0.08, * | 0.01 | 0.01 | 0.11, * |
| **QQ BC CCLM** | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02, * |
| **QQ BC RACMO** | 0.01 | 0.02 | 0.05, * | 0.01 | 0.01 | 0.03, * |
| **QQ BC RCA** | 0.01 | 0.01 | 0.04, * | 0.01 | 0.01 | 0.03, * |

*($*$) indicates statistical significance ($p$ <0.05).*

## C2. Statistical analysis (correlation, RMSE and mean bias) for the German regional climate simulations

*Table A7: Statistical comparison of daily values of maximum temperature from RCM evaluation simulations and observations, German regions, 1981-2010. (* is $p$ <0.05).*

| Model and BC method | DE2 (South Germany) | | | DEA (West Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.96, * | 2.93 | -1.7 | 0.96, * | 2.87 | -1.61 |
| **Raw RACMO** | 0.94, * | 3.16 | -0.87 | 0.94, * | 2.93 | -1.16 |
| **Raw RCA** | 0.92, * | 3.68 | -0.72 | 0.92, * | 3.32 | -1.01 |
| **Linear BC CCLM** | 0.97, * | 2.35 | 0 | 0.96, * | 2.33 | 0.01 |
| **Linear BC RACMO** | 0.95, * | 2.96 | -0.02 | 0.95, * | 2.63 | -0.02 |
| **Linear BC RCA** | 0.93, * | 3.44 | -0.01 | 0.93, * | 3.08 | -0.02 |
| **Variance BC CCLM** | 0.96, * | 2.39 | 0 | 0.96, * | 2.33 | 0.01 |
| **Variance BC RACMO** | 0.95, * | 2.88 | -0.02 | 0.95, * | 2.59 | -0.02 |
| **Variance BC RCA** | 0.93, * | 3.3 | -0.01 | 0.93, * | 3.01 | -0.02 |
| **QQ BC CCLM** | 0.96, * | 2.41 | -0.05 | 0.96, * | 2.35 | -0.08 |
| **QQ BC RACMO** | 0.95, * | 2.9 | -0.06 | 0.95, * | 2.62 | -0.08 |
| **QQ BC RCA** | 0.93, * | 3.32 | -0.06 | 0.93, * | 3.05 | -0.08 |

*Table A7 continued.*

| Model and BC method | DED (East Germany) | | | DEF (North Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.96, * | 3 | -1.53 | 0.96, * | 2.49 | -1.09 |
| **Raw RACMO** | 0.94, * | 3.15 | -0.96 | 0.95, * | 2.76 | -1.09 |
| **Raw RCA** | 0.92, * | 3.66 | -1 | 0.93, * | 3.05 | -0.92 |
| **Linear BC CCLM** | 0.96, * | 2.55 | 0 | 0.96, * | 2.22 | 0 |
| **Linear BC RACMO** | 0.95, * | 2.95 | -0.02 | 0.95, * | 2.4 | -0.02 |
| **Linear BC RCA** | 0.93, * | 3.42 | -0.01 | 0.94, * | 2.8 | -0.02 |
| **Variance BC CCLM** | 0.96, * | 2.58 | 0 | 0.96, * | 2.23 | 0 |
| **Variance BC RACMO** | 0.95, * | 2.9 | -0.02 | 0.95, * | 2.44 | -0.02 |
| **Variance BC RCA** | 0.93, * | 3.39 | -0.01 | 0.93, * | 2.82 | -0.02 |
| **QQ BC CCLM** | 0.96, * | 2.6 | -0.05 | 0.96, * | 2.25 | -0.12 |
| **QQ BC RACMO** | 0.95, * | 2.93 | -0.06 | 0.95, * | 2.47 | -0.12 |
| **QQ BC RCA** | 0.93, * | 3.42 | -0.05 | 0.93, * | 2.85 | -0.12 |

*Table A8: Statistical comparison of daily values of minimum temperature from RCM evaluation simulations and observations, German regions, 1981-2010. (* is $p < 0.05$).*

| Model and BC method | DE2 (South Germany) | | | DEA (West Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.95, * | 2.43 | 0.79 | 0.93, * | 2.35 | 0.06 |
| **Raw RACMO** | 0.94, * | 3.08 | -1.44 | 0.92, * | 2.67 | -0.93 |
| **Raw RCA** | 0.91, * | 2.91 | 0.28 | 0.9, * | 2.77 | -0.05 |
| **Linear BC CCLM** | 0.95, * | 2.25 | 0.01 | 0.93, * | 2.24 | 0.01 |
| **Linear BC RACMO** | 0.93, * | 2.55 | 0 | 0.92, * | 2.44 | -0.01 |
| **Linear BC RCA** | 0.91, * | 2.87 | 0.01 | 0.9, * | 2.76 | -0.01 |
| **Variance BC CCLM** | 0.95, * | 2.23 | 0.01 | 0.93, * | 2.23 | 0.01 |
| **Variance BC RACMO** | 0.94, * | 2.46 | 0 | 0.92, * | 2.36 | -0.01 |
| **Variance BC RCA** | 0.92, * | 2.78 | 0 | 0.9, * | 2.7 | -0.01 |
| **QQ BC CCLM** | 0.95, * | 2.25 | 0 | 0.93, * | 2.25 | 0.01 |
| **QQ BC RACMO** | 0.93, * | 2.48 | 0 | 0.92, * | 2.39 | 0.01 |
| **QQ BC RCA** | 0.92, * | 2.8 | 0 | 0.9, * | 2.71 | 0.02 |

*Table A8 continued.*

| Model and BC method | DED (East Germany) | | | DEF (North Germany) | | |
|---|---|---|---|---|---|---|
| | Correl. | RMSE | Mean bias | Correl. | RMSE | Mean bias |
| **Raw CCLM** | 0.94, * | 2.58 | 0.42 | 0.93, * | 2.4 | 0.59 |
| **Raw RACMO** | 0.92, * | 3.13 | -1.13 | 0.92, * | 2.44 | 0.14 |
| **Raw RCA** | 0.9, * | 3.1 | 0.32 | 0.9, * | 2.73 | 0.34 |
| **Linear BC CCLM** | 0.94, * | 2.43 | 0.01 | 0.93, * | 2.24 | 0 |
| **Linear BC RACMO** | 0.92, * | 2.81 | 0 | 0.92, * | 2.37 | -0.02 |
| **Linear BC RCA** | 0.9, * | 3.07 | 0 | 0.9, * | 2.7 | -0.02 |
| **Variance BC CCLM** | 0.94, * | 2.41 | 0.01 | 0.93, * | 2.29 | 0 |
| **Variance BC RACMO** | 0.92, * | 2.71 | 0 | 0.92, * | 2.42 | -0.02 |
| **Variance BC RCA** | 0.9, * | 3.02 | 0 | 0.91, * | 2.67 | -0.02 |
| **QQ BC CCLM** | 0.94, * | 2.45 | 0.03 | 0.93, * | 2.32 | -0.06 |
| **QQ BC RACMO** | 0.92, * | 2.74 | 0.02 | 0.92, * | 2.46 | -0.06 |
| **QQ BC RCA** | 0.9, * | 3.05 | 0.03 | 0.9, * | 2.69 | -0.06 |

*Table A9: Statistical comparison of daily values of precipitation from RCM evaluation simulations and observations, German regions, 1981-2010. (* is $p$ <0.05).*

| Model and BC method | DE2 (South Germany) | | | DEA (West Germany) | | |
|---|---|---|---|---|---|---|
| | **Correl.** | **RMSE** | **Mean bias** | **Correl.** | **RMSE** | **Mean bias** |
| **Raw CCLM** | 0.6, * | 3.7 | 0.46 | 0.55, * | 3.59 | -0.2 |
| **Raw RACMO** | 0.58, * | 3.48 | 0.18 | 0.56, * | 3.54 | 0.16 |
| **Raw RCA** | 0.44, * | 3.7 | -0.18 | 0.46, * | 3.89 | 0.24 |
| **Linear BC CCLM** | 0.6, * | 3.4 | 0 | 0.53, * | 3.93 | 0 |
| **Linear BC RACMO** | 0.57, * | 3.41 | 0 | 0.55, * | 3.49 | 0 |
| **Linear BC RCA** | 0.43, * | 3.94 | 0.01 | 0.45, * | 3.79 | 0.01 |
| **Variance BC CCLM** | 0.66, * | 3.15 | 0 | 0.63, * | 3.28 | 0 |
| **Variance BC RACMO** | 0.62, * | 3.3 | 0 | 0.59, * | 3.46 | 0 |
| **Variance BC RCA** | 0.5, * | 3.82 | 0 | 0.52, * | 3.75 | 0 |
| **QQ BC CCLM** | 0.61, * | 3.53 | 0.04 | 0.57, * | 3.76 | 0.04 |
| **QQ BC RACMO** | 0.57, * | 3.77 | 0.05 | 0.54, * | 4 | 0.07 |
| **QQ BC RCA** | 0.42, * | 4.48 | 0.06 | 0.45, * | 4.27 | 0.07 |

*Table A9 continued.*

| Model and BC method | DED (East Germany) | | | DEF (North Germany) | | |
|---|---|---|---|---|---|---|
| | **Correl.** | **RMSE** | **Mean bias** | **Correl.** | **RMSE** | **Mean bias** |
| **Raw CCLM** | 0.43, * | 3.83 | 0.19 | 0.47, * | 3.72 | -0.33 |
| **Raw RACMO** | 0.44, * | 3.78 | 0.4 | 0.47, * | 3.65 | 0.05 |
| **Raw RCA** | 0.34, * | 4.16 | 0.58 | 0.38, * | 4.22 | 0.36 |
| **Linear BC CCLM** | 0.42, * | 3.86 | 0.01 | 0.46, * | 4.12 | -0.01 |
| **Linear BC RACMO** | 0.45, * | 3.47 | 0.01 | 0.47, * | 3.66 | 0.01 |
| **Linear BC RCA** | 0.34, * | 3.7 | 0 | 0.38, * | 3.95 | 0.01 |
| **Variance BC CCLM** | 0.52, * | 3.36 | 0 | 0.56, * | 3.5 | 0 |
| **Variance BC RACMO** | 0.5, * | 3.43 | 0 | 0.51, * | 3.68 | 0 |
| **Variance BC RCA** | 0.4, * | 3.78 | 0 | 0.44, * | 3.95 | 0 |
| **QQ BC CCLM** | 0.44, * | 3.8 | -0.01 | 0.49, * | 3.99 | 0.03 |
| **QQ BC RACMO** | 0.44, * | 3.97 | 0.03 | 0.46, * | 4.23 | 0.07 |
| **QQ BC RCA** | 0.33, * | 4.22 | 0.02 | 0.36, * | 4.53 | 0.06 |

*Table A10: Kolmogorov-Smirnov (KS) test statistics on the distribution of daily maximum and minimum temperature, and precipitation from RCM evaluation simulations and observations, German regions, 1981-2010.*

| **Model and BC method** | **DE2 (South Germany)** | | | **DEA (West Germany)** | | |
|---|---|---|---|---|---|---|
| | **Tmax** | **Tmin** | **Precip** | **Tmax** | **Tmin** | **Precip** |
| **Raw CCLM** | 0.53, * | 0.35, * | 0.55, * | 0.57, * | 0.22, * | 0.57, * |
| **Raw RACMO** | 0.13, * | 0.52, * | 0.7, * | 0.15, * | 0.49, * | 0.71, * |
| **Raw RCA** | 0.07, * | 0.06, * | 0.23, * | 0.08, * | 0.05, * | 0.2, * |
| **Linear BC CCLM** | 0.05, * | 0.1, * | 0.23, * | 0.07, * | 0.06, * | 0.26, * |
| **Linear BC RACMO** | 0.06, * | 0.02, * | 0.18, * | 0.08, * | 0.03, * | 0.23, * |
| **Linear BC RCA** | 0.02, * | 0.01 | 0.23, * | 0.01 | 0.03, * | 0.2, * |
| **Variance BC CCLM** | 0.02, * | 0.02, * | 0.23, * | 0.01 | 0.02, * | 0.26, * |
| **Variance BC RACMO** | 0.02, * | 0.01 | 0.18, * | 0.02, * | 0.03, * | 0.23, * |
| **Variance BC RCA** | 0.01 | 0.01 | 0.23, * | 0.01 | 0.02, * | 0.2, * |
| **QQ BC CCLM** | 0.01 | 0.01 | 0.23, * | 0.01 | 0.02 | 0.26, * |
| **QQ BC RACMO** | 0.01 | 0.01 | 0.18, * | 0.01 | 0.02, * | 0.23, * |
| **QQ BC RCA** | 0.01 | 0.01 | 0.02, * | 0.01 | 0.02, * | 0.02, * |

*Table A10 continued.*

| **Model and BC method** | **DED (East Germany)** | | | **DEF (North Germany)** | | |
|---|---|---|---|---|---|---|
| | **Tmax** | **Tmin** | **Precip** | **Tmax** | **Tmin** | **Precip** |
| **Raw CCLM** | 0.53, * | 0.31, * | 0.65, * | 0.52, * | 0.25, * | 0.6, * |
| **Raw RACMO** | 0.13, * | 0.5, * | 0.76, * | 0.2, * | 0.51, * | 0.71, * |
| **Raw RCA** | 0.07, * | 0.06, * | 0.27, * | 0.06, * | 0.07, * | 0.23, * |
| **Linear BC CCLM** | 0.05, * | 0.08, * | 0.3, * | 0.09, * | 0.04, * | 0.32, * |
| **Linear BC RACMO** | 0.07, * | 0.02, * | 0.27, * | 0.09, * | 0.03, * | 0.26, * |
| **Linear BC RCA** | 0.02, * | 0.02, * | 0.27, * | 0.01 | 0.03, * | 0.23, * |
| **Variance BC CCLM** | 0.01 | 0.02, * | 0.3, * | 0.01 | 0.02 | 0.32, * |
| **Variance BC RACMO** | 0.02, * | 0.02, * | 0.27, * | 0.02 | 0.03, * | 0.26, * |
| **Variance BC RCA** | 0.01 | 0.01 | 0.27, * | 0.01 | 0.02 | 0.23, * |
| **QQ BC CCLM** | 0.01 | 0.02, * | 0.3, * | 0.01 | 0.01 | 0.32, * |
| **QQ BC RACMO** | 0.02 | 0.02 | 0.27, * | 0.01 | 0.02, * | 0.26, * |
| **QQ BC RCA** | 0.02 | 0.02 | 0.03, * | 0.01 | 0.02, * | 0.02, * |

*($*$) indicates statistical significance ($p$ <0.05).*

# C3. Cumulative distribution and probability density functions: National level

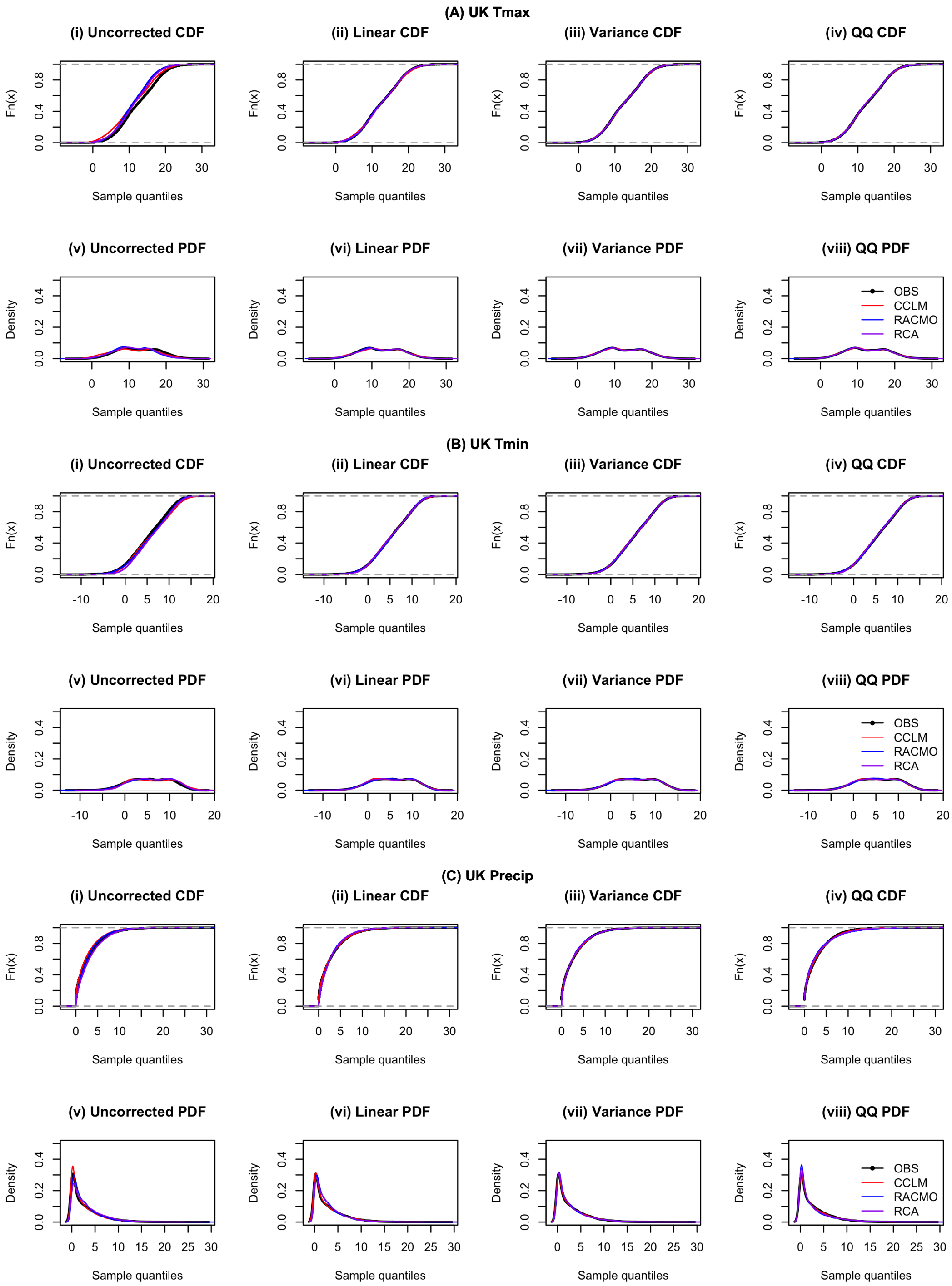



*Figure A10: Empirical cumulative and probability distribution function (i-iv: CDF and v-viii: PDF) plots for daily observed and evaluation RCM-simulated values of (A) maximum temperature, (B) minimum temperature, and (C) precipitation.*

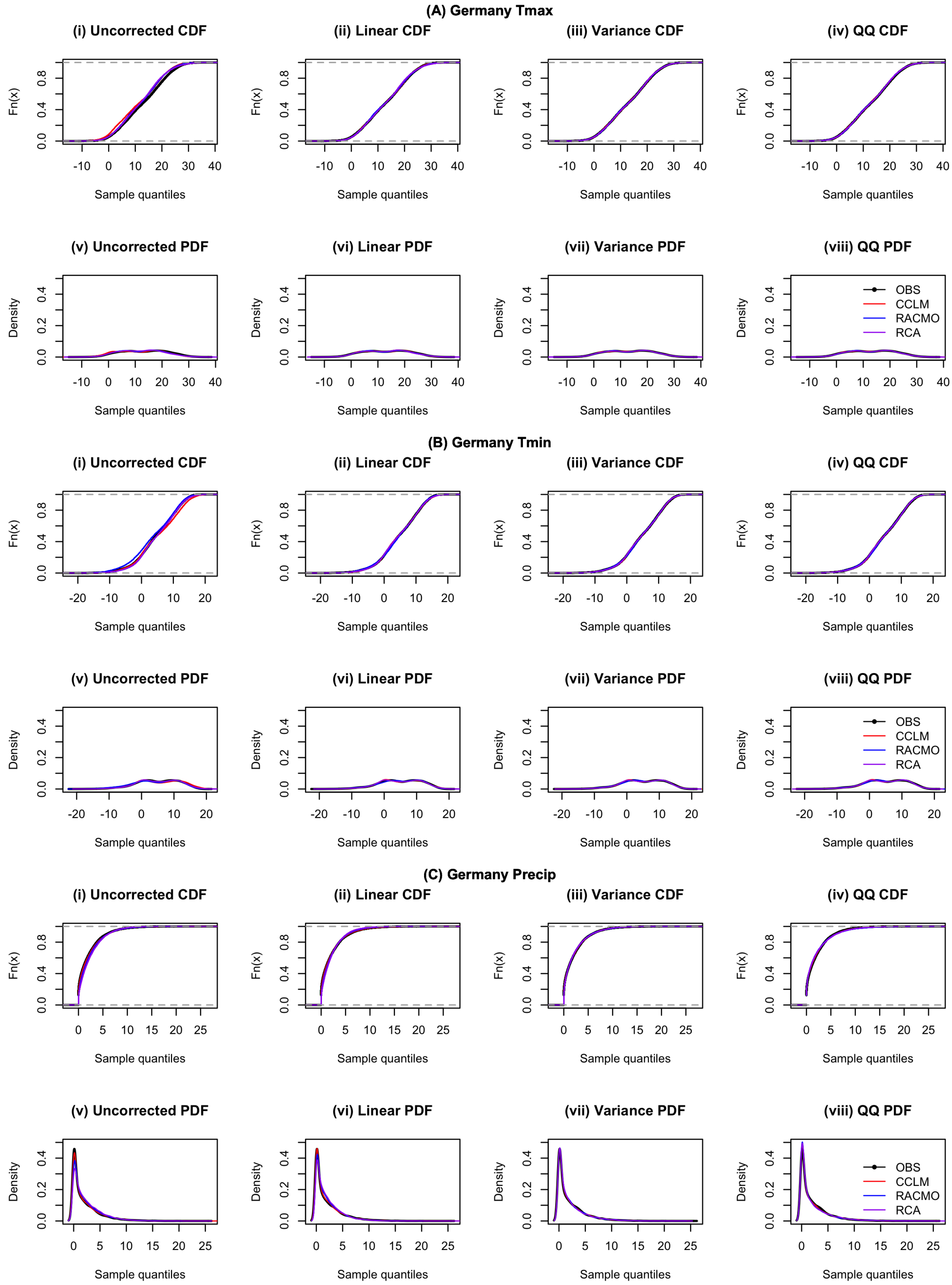



*Figure A11: Empirical cumulative and probability distribution function (i-iv: CDF and v-viii: PDF) plots for daily observed and evaluation RCM-simulated values of (A) maximum temperature, (B) minimum temperature, and (C) precipitation.*

## C4. Cumulative distribution and probability density functions: regional level

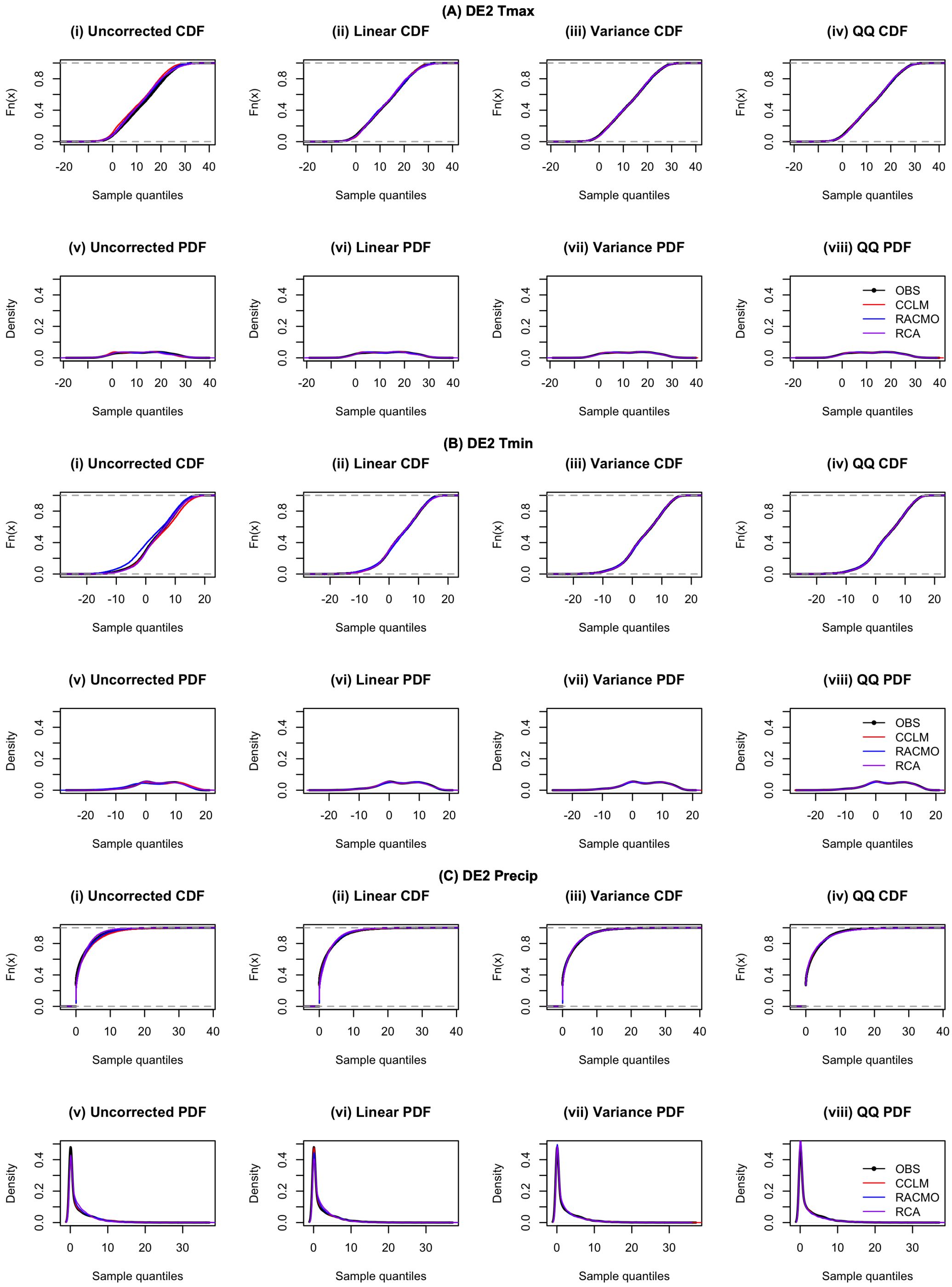



*Figure A12: Empirical cumulative and probability distribution function (i-iv: CDF and v-viii: PDF) plots for daily observed and evaluation RCM-simulated values of (A) maximum temperature, (B) minimum temperature, and (C) precipitation.*

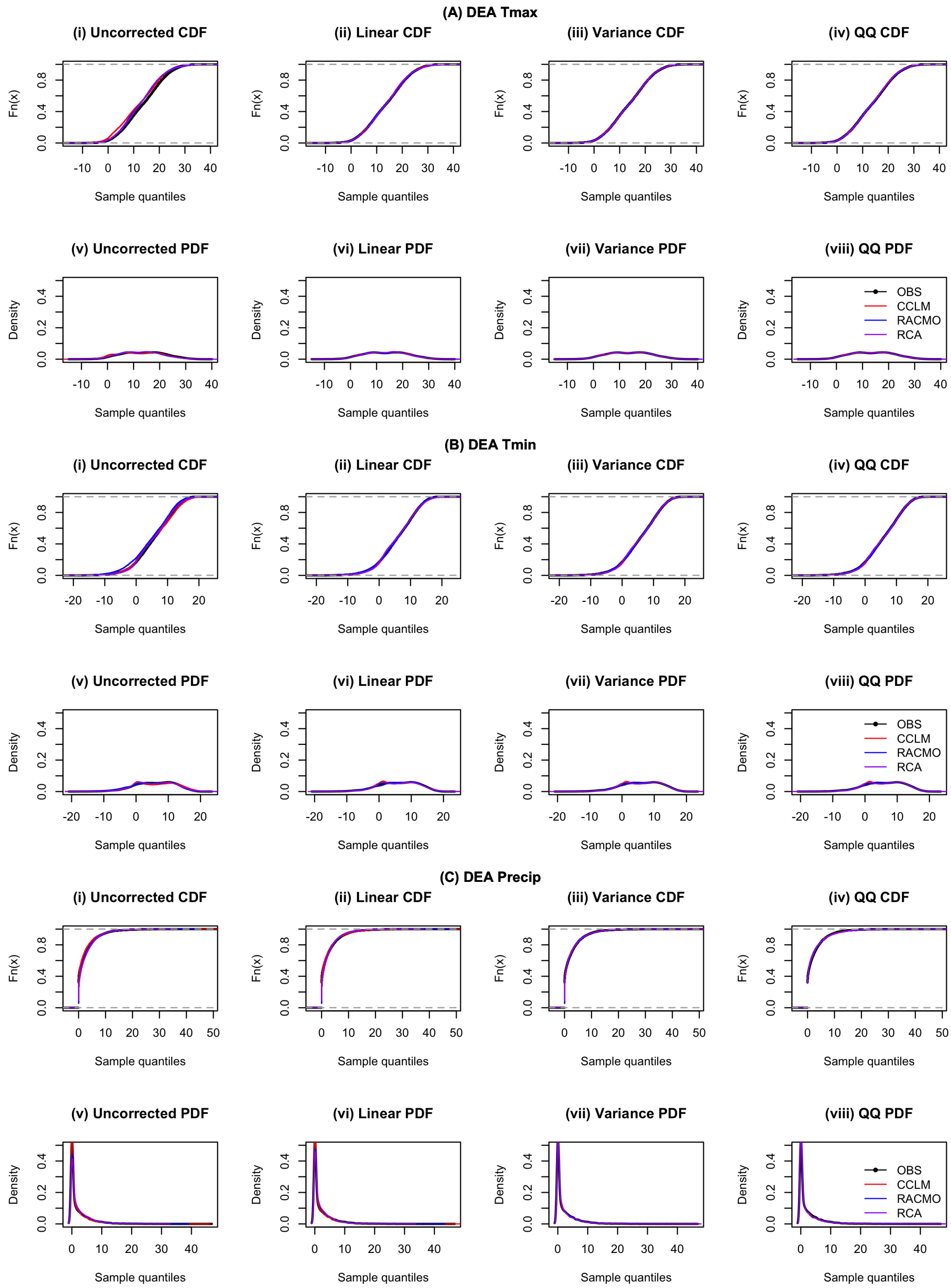



*Figure A13: Empirical cumulative and probability distribution function (i-iv: CDF and v-viii: PDF) plots for daily observed and evaluation RCM-simulated values of (A) maximum temperature, (B) minimum temperature, and (C) precipitation.*

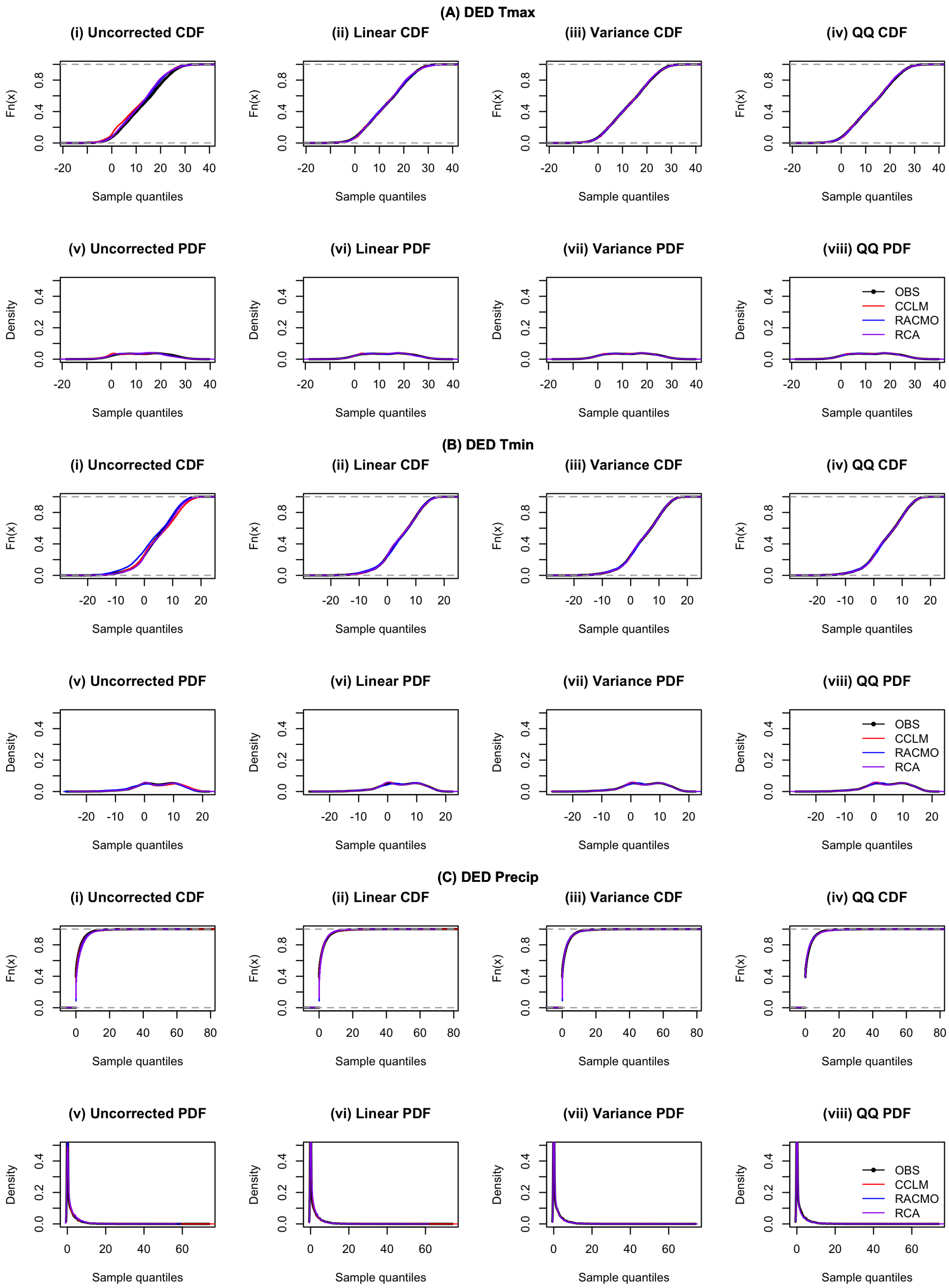



*Figure A14: Empirical cumulative and probability distribution function (i-iv: CDF and v-viii: PDF) plots for daily observed and evaluation RCM-simulated values of (A) maximum temperature, (B) minimum temperature, and (C) precipitation.*

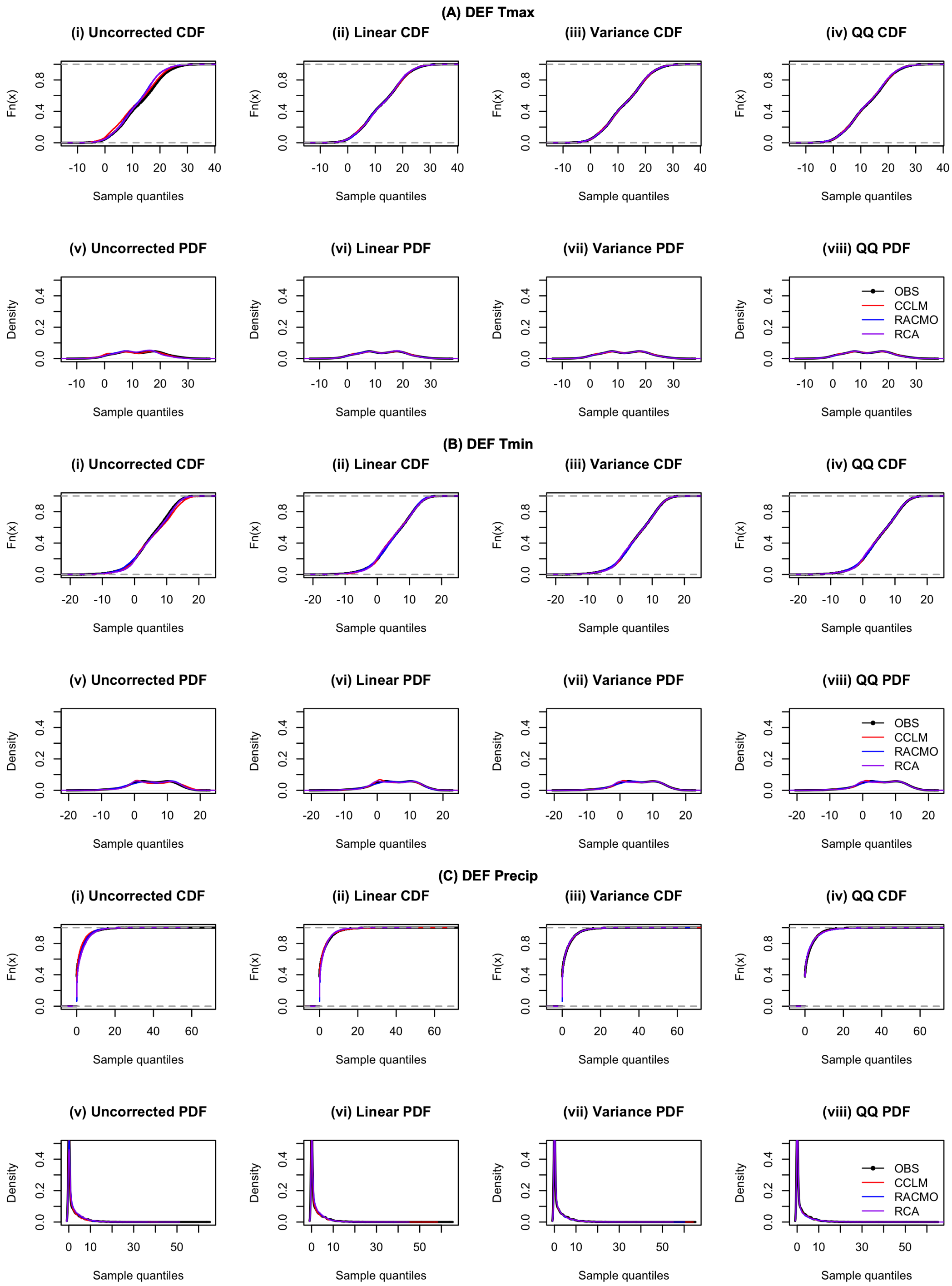



*Figure A15: Empirical cumulative and probability distribution function (i-iv: CDF and v-viii: PDF) plots for daily observed and evaluation RCM-simulated values of (A) maximum temperature, (B) minimum temperature, and (C) precipitation.*

# C5. Taylor diagrams for RCM evaluation simulations of temperature and precipitation relative to observations: National level

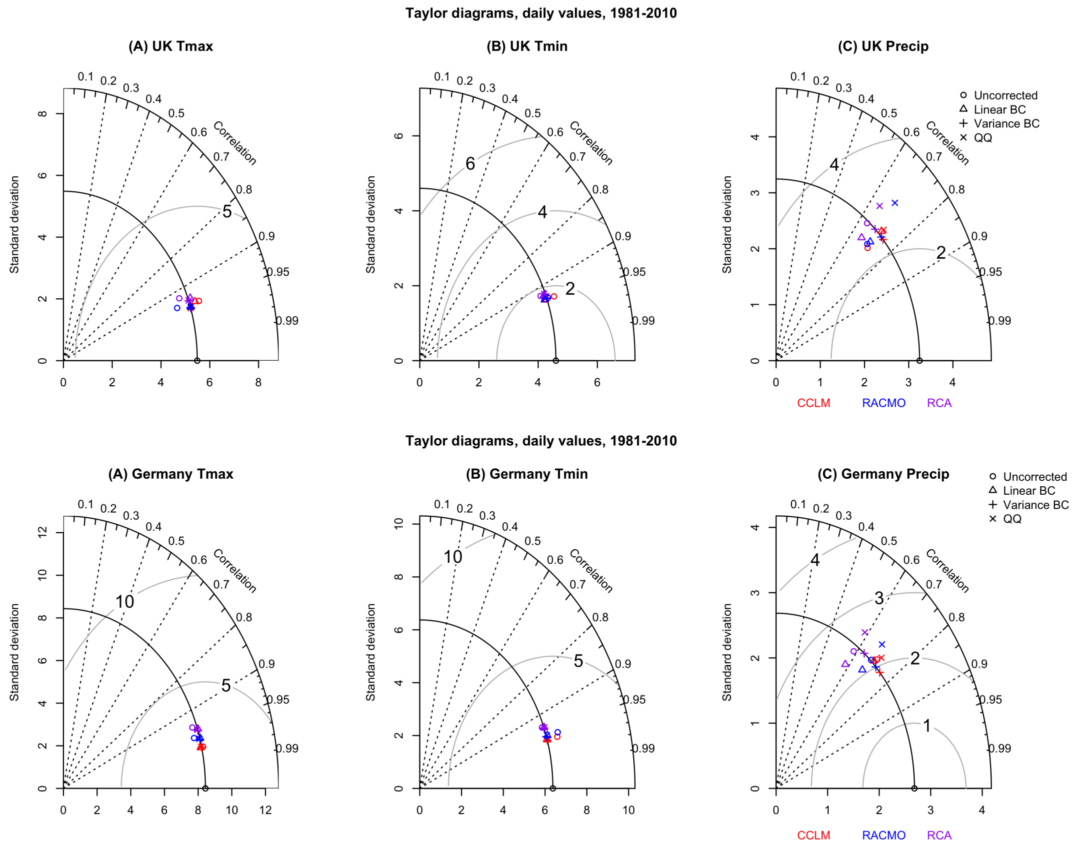



*Figure A16: Taylor diagrams, daily (A) maximum temperature, (B) minimum temperature, and (C) precipitation for the UK and Germany, 1981-2010.*

## C6. Taylor diagrams, regional level

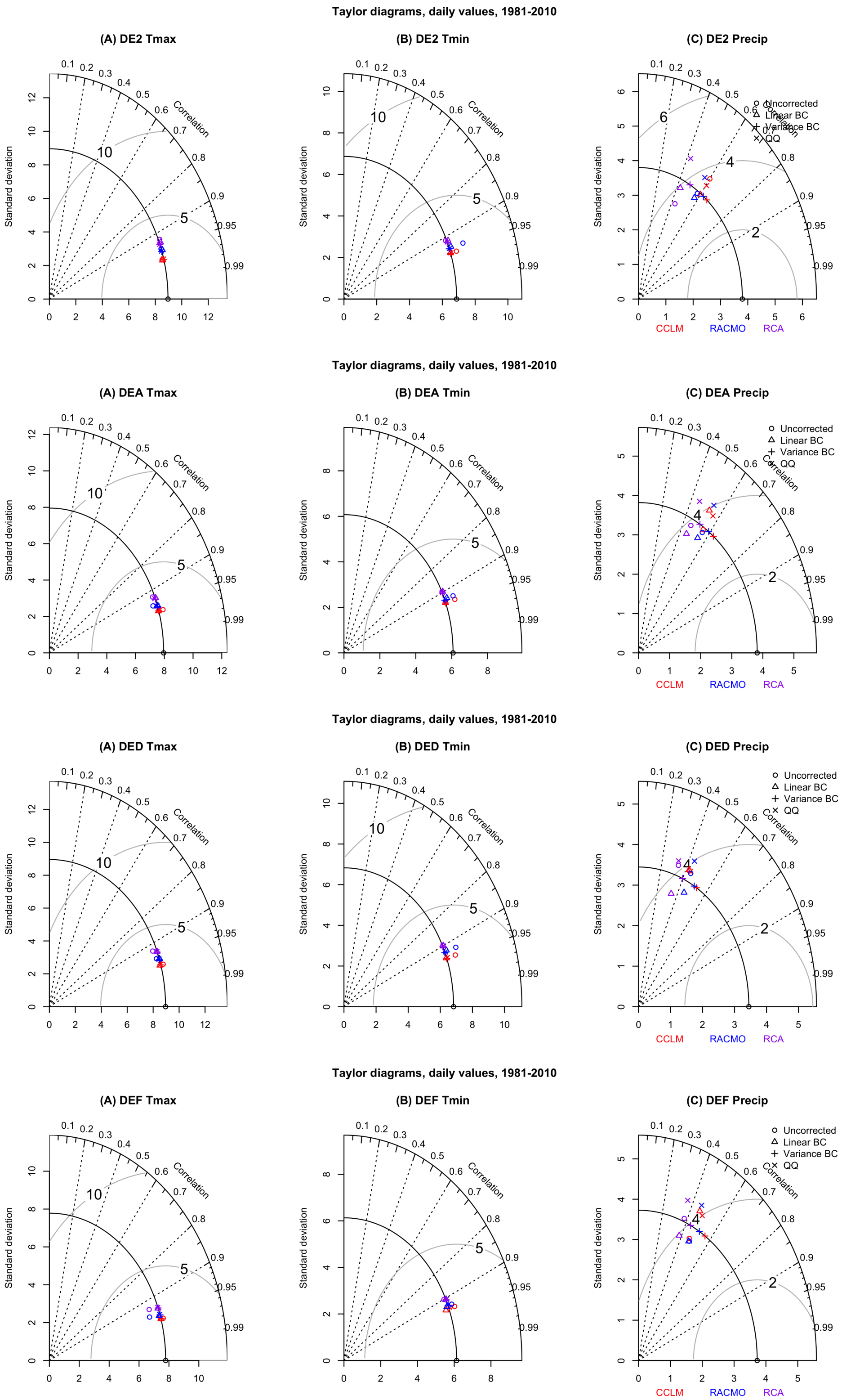



*Figure A17: Taylor diagrams, daily (A) maximum temperature, (B) minimum temperature, and (C) precipitation for German regions, 1981-2010.*

## C7. Statistical analysis for SCCM yield simulations for regional German climate simulations using uncorrected and bias-corrected RCM output

*Table A11: SCCM yield simulations, DE2 (South Germany).*

| Model and BC method | Correlation | | RMSE | | Mean bias | |
|---|---|---|---|---|---|---|
| | EOBS | Obs | EOBS | Obs | EOBS | Obs |
| **Raw CCLM** | 0.62, * | 0.59, * | 0.45 | 0.53 | -0.15 | -0.18 |
| **Raw RACMO** | 0.69, * | 0.4, * | 0.35 | 0.56 | 0.06 | 0.03 |
| **Raw RCA** | 0.53, * | 0.17, | 1.02 | 1.24 | -0.32 | -0.35 |
| **Linear BC CCLM** | 0.72, * | 0.68, * | 0.37 | 0.46 | -0.12 | -0.14 |
| **Linear BC RACMO** | 0.62, * | 0.28, | 0.41 | 0.64 | -0.03 | -0.06 |
| **Linear BC RCA** | 0.55, * | 0.23, | 0.96 | 1.15 | -0.47 | -0.49 |
| **Variance BC CCLM** | 0.88, * | 0.62, * | 0.24 | 0.49 | -0.09 | -0.12 |
| **Variance BC RACMO** | 0.77, * | 0.43, * | 0.31 | 0.56 | -0.02 | -0.05 |
| **Variance BC RCA** | 0.79, * | 0.49, * | 0.32 | 0.56 | -0.1 | -0.13 |
| **QQ BC CCLM** | 0.73, * | 0.7, * | 0.38 | 0.46 | -0.13 | -0.16 |
| **QQ BC RACMO** | 0.58, * | 0.37, * | 0.4 | 0.58 | -0.05 | -0.08 |
| **QQ BC RCA** | 0.45, * | 0.26 | 0.85 | 0.99 | -0.31 | -0.34 |

*Table A12: SCCM yield simulations, DEA (West Germany).*

| Model and BC method | Correlation | | RMSE | | Mean bias | |
|---|---|---|---|---|---|---|
| | EOBS | Obs | EOBS | Obs | EOBS | Obs |
| **Raw CCLM** | 0.92, * | 0.69, * | 0.61 | 0.8 | 0.55 | 0.48 |
| **Raw RACMO** | 0.88, * | 0.64, * | 0.35 | 0.68 | 0.14 | 0.07 |
| **Raw RCA** | 0.88, * | 0.63, * | 0.37 | 0.72 | -0.08 | -0.15 |
| **Linear BC CCLM** | 0.89, * | 0.63, * | 0.33 | 0.69 | 0.08 | 0.01 |
| **Linear BC RACMO** | 0.84, * | 0.57, * | 0.42 | 0.78 | -0.1 | -0.17 |
| **Linear BC RCA** | 0.83, * | 0.63, * | 0.51 | 0.79 | -0.27 | -0.34 |
| **Variance BC CCLM** | 0.97, * | 0.68, * | 0.18 | 0.64 | 0.05 | -0.02 |
| **Variance BC RACMO** | 0.94, * | 0.66, * | 0.25 | 0.68 | -0.03 | -0.1 |
| **Variance BC RCA** | 0.9, * | 0.59, * | 0.34 | 0.76 | -0.03 | -0.1 |
| **QQ BC CCLM** | 0.9, * | 0.61, * | 0.3 | 0.7 | 0.07 | 0 |
| **QQ BC RACMO** | 0.8, * | 0.53, * | 0.43 | 0.79 | -0.03 | -0.1 |
| **QQ BC RCA** | 0.84, * | 0.62, * | 0.43 | 0.73 | -0.02 | -0.09 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $r$, smaller $RMSE$ or bias) relative to the yield hindcast or observations.*

*Table A13: SCCM yield simulations, DED (East Germany).*

| Model and BC method | Correlation | | RMSE | | Mean bias | |
|---|---|---|---|---|---|---|
| | EOBS | Obs | EOBS | Obs | EOBS | Obs |
| **Raw CCLM** | 0.77, * | 0.71, * | 0.51 | 0.6 | 0.27 | 0.22 |
| **Raw RACMO** | 0.63, * | 0.42, * | 0.54 | 0.76 | 0.04 | -0.01 |
| **Raw RCA** | 0.62, * | 0.44, * | 1.11 | 1.27 | -0.51 | -0.56 |
| **Linear BC CCLM** | 0.82, * | 0.71, * | 0.39 | 0.56 | -0.02 | -0.07 |
| **Linear BC RACMO** | 0.64, * | 0.37, * | 0.66 | 0.91 | -0.18 | -0.23 |
| **Linear BC RCA** | 0.55, * | 0.38, * | 1.67 | 1.79 | -0.74 | -0.79 |
| **Variance BC CCLM** | 0.92, * | 0.69, * | 0.27 | 0.58 | -0.02 | -0.07 |
| **Variance BC RACMO** | 0.83, * | 0.56, * | 0.4 | 0.7 | 0.1 | 0.05 |
| **Variance BC RCA** | 0.86, * | 0.73, * | 0.42 | 0.6 | -0.07 | -0.12 |
| **QQ BC CCLM** | 0.85, * | 0.71, * | 0.39 | 0.58 | -0.1 | -0.15 |
| **QQ BC RACMO** | 0.49, * | 0.26 | 0.79 | 1.02 | -0.15 | -0.2 |
| **QQ BC RCA** | 0.48, * | 0.28 | 1.44 | 1.61 | -0.49 | -0.54 |

*Table A14: SCCM yield simulations, DEF (North Germany).*

| Model and BC method | Correlation | | RMSE | | Mean bias | |
|---|---|---|---|---|---|---|
| | EOBS | Obs | EOBS | Obs | EOBS | Obs |
| **Raw CCLM** | 0.91, * | 0.65, * | 0.34 | 0.57 | 0.25 | 0.16 |
| **Raw RACMO** | 0.91, * | 0.6, * | 0.24 | 0.57 | 0.09 | -0.01 |
| **Raw RCA** | 0.79, * | 0.7, * | 0.35 | 0.53 | -0.04 | -0.13 |
| **Linear BC CCLM** | 0.89, * | 0.66, * | 0.26 | 0.55 | 0 | -0.09 |
| **Linear BC RACMO** | 0.9, * | 0.59, * | 0.24 | 0.59 | -0.01 | -0.11 |
| **Linear BC RCA** | 0.8, * | 0.7, * | 0.34 | 0.52 | -0.01 | -0.11 |
| **Variance BC CCLM** | 1, * | 0.62, * | 0.03 | 0.58 | 0 | -0.09 |
| **Variance BC RACMO** | 1, * | 0.62, * | 0.03 | 0.58 | 0 | -0.09 |
| **Variance BC RCA** | 1, * | 0.62, * | 0.03 | 0.58 | 0 | -0.09 |
| **QQ BC CCLM** | 0.89, * | 0.65, * | 0.25 | 0.56 | 0.01 | -0.08 |
| **QQ BC RACMO** | 0.86, * | 0.56, * | 0.28 | 0.61 | -0.01 | -0.1 |
| **QQ BC RCA** | 0.77, * | 0.7, * | 0.36 | 0.52 | 0 | -0.1 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $r$, smaller $RMSE$ or bias) relative to the yield hindcast or observations.*

## C8. Statistical analysis for PCM yield simulations for regional German climate simulations using uncorrected and bias-corrected RCM output

*Table A15: PCM yield simulations, DE2 (South Germany).*

| Model and BC method | Correlation | | RMSE | | Mean bias | |
|---|---|---|---|---|---|---|
| | EOBS | Obs | EOBS | Obs | EOBS | Obs |
| **Raw CCLM** | 0.48, * | 0.2 | 0.88 | 2.39 | -0.17 | 2.21 |
| **Raw RACMO** | 0.29 | -0.06 | 1.49 | 2.56 | -0.37 | 2.02 |
| **Raw RCA** | 0.37 | 0.05 | 1.31 | 2.31 | -0.49 | 1.9 |
| **Linear BC CCLM** | 0.58, * | 0 | 1.07 | 2.47 | -0.32 | 2.07 |
| **Linear BC RACMO** | 0.51, * | 0.16 | 1.1 | 2.42 | -0.3 | 2.09 |
| **Linear BC RCA** | 0.44, * | 0.16 | 1.17 | 2.39 | -0.34 | 2.05 |
| **Variance BC CCLM** | 0.74, * | 0.17 | 0.81 | 2.41 | -0.28 | 2.11 |
| **Variance BC RACMO** | 0.79, * | 0.22 | 0.72 | 2.48 | -0.19 | 2.2 |
| **Variance BC RCA** | 0.66, * | 0.29 | 0.9 | 2.4 | -0.26 | 2.13 |
| **QQ BC CCLM** | 0.56, * | 0.11 | 1.25 | 2.37 | -0.5 | 1.89 |
| **QQ BC RACMO** | 0.53, * | 0.15 | 1.1 | 2.54 | -0.2 | 2.19 |
| **QQ BC RCA** | 0.42, * | 0.15 | 1.57 | 2.44 | -0.56 | 1.82 |

*Table A16: PCM yield simulations, DEA (West Germany).*

| Model and BC method | Correlation | | RMSE | | Mean bias | |
|---|---|---|---|---|---|---|
| | EOBS | Obs | EOBS | Obs | EOBS | Obs |
| **Raw CCLM** | 0.37 | 0.05 | 1.7 | 1.7 | -0.18 | 0.07 |
| **Raw RACMO** | 0.16 | 0.11 | 1.81 | 1.57 | 0.56 | 0.8 |
| **Raw RCA** | -0.07 | -0.11 | 2.19 | 1.98 | 1.16 | 1.41 |
| **Linear BC CCLM** | 0.33 | -0.04 | 1.81 | 1.87 | -0.08 | 0.17 |
| **Linear BC RACMO** | 0.23 | 0.09 | 1.81 | 1.67 | 0.4 | 0.65 |
| **Linear BC RCA** | 0.15 | 0.07 | 1.71 | 1.43 | 0.37 | 0.61 |
| **Variance BC CCLM** | 0.92, * | -0.18 | 0.58 | 1.71 | -0.06 | 0.19 |
| **Variance BC RACMO** | 0.9, * | -0.12 | 0.65 | 1.64 | 0.06 | 0.31 |
| **Variance BC RCA** | 0.89, * | -0.11 | 0.69 | 1.54 | 0.09 | 0.34 |
| **QQ BC CCLM** | 0.29 | 0.01 | 1.95 | 1.93 | -0.21 | 0.04 |
| **QQ BC RACMO** | -0.02 | 0.38, * | 2.21 | 1.59 | 0.31 | 0.55 |
| **QQ BC RCA** | 0.03 | 0.15 | 1.91 | 1.48 | 0.27 | 0.51 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $r$, smaller $RMSE$ or bias) relative to the yield hindcast or observations.*

Table A17: PCM yield simulations, DED (East Germany).

| Model and BC method | Correlation | | RMSE | | Mean bias | |
|---|---|---|---|---|---|---|
| | EOBS | Obs | EOBS | Obs | EOBS | Obs |
| **Raw CCLM** | 0.51, * | 0.06 | 1.21 | 2.18 | 0.11 | 1.65 |
| **Raw RACMO** | 0.19 | 0.14 | 1.59 | 2.51 | 0.61 | 2.16 |
| **Raw RCA** | 0.23 | 0 | 1.86 | 3.1 | 1.31 | 2.86 |
| **Linear BC CCLM** | 0.47, * | 0.01 | 1.56 | 2.11 | -0.44 | 1.11 |
| **Linear BC RACMO** | 0.42, * | 0.4, * | 1.3 | 1.78 | -0.19 | 1.36 |
| **Linear BC RCA** | 0.22 | 0 | 1.97 | 2.27 | -0.4 | 1.15 |
| **Variance BC CCLM** | 0.86, * | 0.44, * | 0.75 | 1.69 | -0.34 | 1.21 |
| **Variance BC RACMO** | 0.84, * | 0.47, * | 0.68 | 1.67 | -0.22 | 1.33 |
| **Variance BC RCA** | 0.69, * | 0.41, * | 0.93 | 1.58 | -0.33 | 1.22 |
| **QQ BC CCLM** | 0.42, * | -0.03 | 1.64 | 2.14 | -0.43 | 1.11 |
| **QQ BC RACMO** | 0.32 | 0.41, * | 1.55 | 1.81 | -0.3 | 1.24 |
| **QQ BC RCA** | 0.24 | 0.08 | 1.59 | 1.95 | -0.29 | 1.26 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $r$, smaller $RMSE$ or bias) relative to the yield hindcast or observations.*

Table A18: PCM yield simulations, DEF (North Germany).

| Model and BC method | Correlation | | RMSE | | Mean bias | |
|---|---|---|---|---|---|---|
| | EOBS | Obs | EOBS | Obs | EOBS | Obs |
| **Raw CCLM** | 0.11 | 0.13 | 1.85 | 1.79 | -0.19 | -0.97 |
| **Raw RACMO** | 0.26 | -0.23 | 1.53 | 1.38 | 0.54 | -0.24 |
| **Raw RCA** | 0.27 | -0.16 | 1.92 | 1.54 | 1.15 | 0.37 |
| **Linear BC CCLM** | 0.01 | 0.21 | 2.11 | 1.87 | -0.09 | -0.87 |
| **Linear BC RACMO** | 0.23 | -0.03 | 1.68 | 1.69 | 0.15 | -0.63 |
| **Linear BC RCA** | 0.31 | 0.04 | 1.61 | 1.61 | 0.26 | -0.52 |
| **Variance BC CCLM** | 0.91, * | -0.2 | 0.59 | 1.73 | 0.16 | -0.62 |
| **Variance BC RACMO** | 0.92, * | -0.26 | 0.57 | 1.84 | 0.1 | -0.68 |
| **Variance BC RCA** | 0.83, * | -0.14 | 0.83 | 1.75 | 0.21 | -0.57 |
| **QQ BC CCLM** | -0.06 | 0.25 | 2.25 | 1.99 | -0.24 | -1.02 |
| **QQ BC RACMO** | 0.06 | 0.05 | 1.99 | 1.91 | -0.14 | -0.92 |
| **QQ BC RCA** | 0.19 | 0.08 | 1.87 | 1.81 | 0.14 | -0.64 |

*($*$) indicates statistical significance ($p$ <0.05). A green color indicates an improvement (larger $r$, smaller $RMSE$ or bias) relative to the yield hindcast or observations.*

# Bibliography


Acevedo, E., Silva, P., and Silva, H. (2002). Wheat growth and physiology. Technical report, United Nations Food and Agriculture Organization.

Albers, H., Gornott, C., and Hüttel, S. (2017). How do inputs and weather drive wheat yield volatility? The example of Germany. *Food Policy*, 70:50–61.

Angulo, C., Gaiser, T., Rötter, R. P., Børgesen, C. D., Hlavinka, P., Trnka, M., and Ewert, F. (2014). "Fingerprints" of four crop models as affected by soil input data aggregation. *European Journal of Agronomy*, 61:35–48.

Angulo, C., Rötter, R., Lock, R., Enders, A., Fronzek, S., and Ewert, F. (2013). Implication of crop model calibration strategies for assessing regional impacts of climate change in Europe. *Agricultural and Forest Meteorology*, 170:32–46.

Arguez, A. and Vose, R. S. (2011). The definition of the standard WMO climate normal: The key to deriving alternative climate normals. *Bulletin of the American Meteorological Society*, 92(6):699–704.

Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., Kimball, B. A., Ottman, M. J., Wall, G. W., White, J. W., Reynolds, M. P., Alderman, P. D., Prasad, P. V. V., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Challinor, A. J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurralde, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Koehler, A.-K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J. E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A. C., Semenov, M. A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P. J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., and Zhu, Y. (2014). Rising temperatures reduce global wheat production. *Nature Climate Change*, 5(2):143–147.

Asseng, S., Ewert, F., and Rosenzweig, C. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, 3(9):627–632.

Asseng, S., Foster, I., and Turner, N. C. (2011). The impact of temperature variability on wheat yields. *Global Change Biology*, 17(2):997–1012.

Atkinson, M., Kettlewell, P., Hollins, P., Stephenson, D., and Hardwick, N. (2005). Summer climate mediates UK wheat quality response to winter North Atlantic Oscillation. *Agricultural and Forest Meteorology*, 130(1-2):27–37.

Balkovič, J., Skalský, R., Folberth, C., Khabarov, N., Schmid, E., Madaras, M., Obersteiner, M., and van der Velde, M. (2018). Impacts and Uncertainties of +2°C of Climate Change and Soil Degradation on European Crop Calorie Supply. *Earth's Future*.

Balkovič, J., van der Velde, M., Schmid, E., Skalský, R., Khabarov, N., Obersteiner, M., Stürmer, B., and Xiong, W. (2013). Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation. *Agricultural Systems*, 120:61–75.

Barlow, K., Christy, B., O'Leary, G., Riffkin, P., and Nuttall, J. (2015). Simulating the impact of extreme heat and frost events on wheat crop production: A review. *Field Crops Research*, 171:109–119.

Barsugli, J. J., Guentchev, G., Horton, R. M., Wood, A., Mearns, L. O., Liang, X.-Z., Winkler, J. a., Dixon, K., Hayhoe, K., Rood, R. B., Goddard, L., Ray, A., Buja, L., and Ammann, C. (2013). The Practitioner's Dilemma: How to Assess the Credibility of Downscaled Climate Projections. *Eos, Transactions American Geophysical Union*, 94(46):424–425.

Beauregard, F. and De Blois, S. (2014). Beyond a climate-centric view of plant distribution: Edaphic variables add value to distribution models. *PLoS ONE*, 9(3).

Bindoff, N., Stott, P., AchutaRao, K., Allen, M., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I., Overland, J., Perlwitz, J., Sebbari, R., and Zhang, X. (2013). Detection and Attribution of Climate Change: from Global to Regional. In Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 867–952. Cambridge University Press.

Boote, K. J., Jones, J. W., White, J. W., Asseng, S., and Lizaso, J. I. (2013). Putting mechanisms into crop production models. *Plant, Cell and Environment*, 36(9):1658–1672.

Booth, J. F., Kwon, Y. O., Ko, S., Small, R. J., and Msadek, R. (2017). Spatial patterns and intensity of the surface storm tracks in CMIP5 models. *Journal of Climate*, 30(13):4965–4981.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., and Schär, C. (2013). Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. *Water Resources Research*, 49(3):1523–1536.

Brent, R. P. (1971). An algorithm with guaranteed convergence for finding a zero of a function. *The Computer Journal*, 14(4):422–425.

Brisson, N., Gate, P., Gouache, D., Charmet, G., Oury, F.-X., and Huard, F. (2010). Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *Field Crops Research*, 119(1):201–212.

Brown, I., Thompson, D., Bardgett, R., Berry, P., Crute, I., Morison, J., Morecroft, M., Pinnegar, J., Reeder, T., and Topp, K. (2016). UK Climate Change Risk Assessment Evidence Report: Chapter 3, Natural Environment and Natural Assets. Technical report, Report prepared for the Adaptation Sub-Committee of the Committee on Climate Change, London.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, 28(17):6938–6959.

Casanueva, A., Kotlarski, S., Herrera, S., Fernández, J., Gutiérrez, J. M., Boberg, F., Colette, A., Christensen, O. B., Goergen, K., Jacob, D., Keuler, K., Nikulin, G., Teichmann, C., and Vautard, R. (2016). Daily precipitation statistics in a EURO-CORDEX RCM ensemble: added value of raw and bias-corrected high-resolution simulations. *Climate Dynamics*, 47(3-4):719–737.

CDO (2018). Climate data operators. Max-Planck-Instituts für Meteorologie. Available online at `http://www.mpimet.mpg.de/cdo`.

Challinor, A., Martre, P., Asseng, S., Thornton, P., and Ewert, F. (2014). Making the most of climate impacts ensembles. *Nature Climate Change*, 4(2):77–80.

Challinor, A. and Wheeler, T. (2008). Use of a crop model ensemble to quantify CO2 stimulation of water-stressed and well-watered crops. *Agricultural and Forest Meteorology*, 148(6-7):1062–1077.

Challinor, A., Wheeler, T., Craufurd, P., Slingo, J., and Grimes, D. (2004). Design and optimisation of a large-area process-based model for annual crops. *Agricultural and Forest Meteorology*, 124(1-2):99–120.

Challinor, A. J., Ewert, F., Arnold, S., Simelton, E., and Fraser, E. (2009). Crops and climate change: progress, trends, and challenges in simulating impacts and informing adaptation. *Journal of Experimental Botany*, 60(10):2775–2789.

Challinor, A. J., Müller, C., Asseng, S., Deva, C., Nicklin, K. J., Wallach, D., Vanuytrecht, E., Whitfield, S., Ramirez-Villegas, J., and Koehler, A. K. (2017). Improving the use of crop models for risk assessment and climate change adaptation. *Agricultural Systems*, 159(June 2017):296–306.

Challinor, A. J., Simelton, E. S., Fraser, E. D. G., Hemming, D., and Collins, M. (2010). Increased crop failure due to climate change: assessing adaptation options using models and socio-economic data for wheat in China. *Environmental Research Letters*, 5(3):034012.

Challinor, A. J., Smith, M. S., and Thornton, P. (2013). Use of agro-climate ensembles for quantifying uncertainty and informing adaptation. *Agricultural and Forest Meteorology*, 170:2–7.

Challinor, A. J., Wheeler, T. R., Slingo, J. M., Craufurd, P. Q., and Grimes, D. I. F. (2005). Simulation of Crop Yields Using ERA-40: Limits to Skill and Nonstationarity in Weather-Yield Relationships. *Journal of Applied Meteorology*, 44(4):516–531.

Chen, J., Brissette, F. P., Chaumont, D., and Braun, M. (2013). Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two North American river basins. *Journal of Hydrology*, 479:200–214.

Chen, J., Brissette, F. P., and Lucas-Picher, P. (2015). Assessing the limits of bias-correcting climate model outputs for climate change impact studies. *Journal of Geophysical Research: Atmospheres*, 120(3):1123–1136.

Chenu, K., Porter, J. R., Martre, P., Basso, B., Chapman, S. C., Ewert, F., Bindi, M., and Asseng, S. (2017). Contribution of Crop Models to Adaptation in Wheat. *Trends in Plant Science*, 22(6):472–490.

Cho, K., Falloon, P., Gornall, J., Betts, R., and Clark, R. (2012). Winter wheat yields in the UK: uncertainties in climate and management impacts. *Climate Research*, 54(1):49–68.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A., and Woodward, S. (2011). Development and evaluation of an Earth-system model - HadGEM2. *Geoscientific Model Development Discussions*, 4(2):997–1062.

Cubasch, U., Wuebbles, D., Chen, D., Facchini, M., Frame, D., Mahowald, N., and Winther, J.-G. (2013). Introduction. In Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Curtis, B. C. (2012). Wheat in the world. United Nations Food and Agriculture Organization. Available online from `http://www.fao.org/docrep/006/y4011e/y4011e04.htm`.

de Elía, R., Laprise, R., Biner, S., and Merleau, J. (2017). Synchrony between reanalysis-driven RCM simulations and observations: variation with time scale. *Climate Dynamics*, 48(7-8):2597–2610.

Dee, D. P., Uppala, S. M., Simmons, a. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. a., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, a. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, a. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

Destatis (2018). Statistisches Bundesamt Land- und Forstwirtschaft, Fischerei. Available online at `https://www.destatis.de/DE/Publikationen/Thematisch/LandForstwirtschaft/ThemaLandForstwirtschaft.html`.

Dettori, M., Cesaraccio, C., Motroni, A., Spano, D., and Duce, P. (2011). Using CERES-Wheat to simulate durum wheat production and phenology in Southern Sardinia, Italy. *Field Crops Research*, 120(1):179–188.

Di Luca, A., Argüeso, D., Evans, J. P., de Elía, R., and Laprise, R. (2016). Quantifying the overall added value of dynamical downscaling and the contribution from different spatial scales. *Journal of Geophysical Research: Atmospheres*, 121(4):1575–1590.

Di Luca, A., de Elía, R., and Laprise, R. (2015). Challenges in the Quest for Added Value of Regional Climate Dynamical Downscaling. *Current Climate Change Reports*, 1(1):10–21.

Dobor, L., Barcza, Z., Hlásny, T., Árendás, T., Spitkó, T., and Fodor, N. (2016). Crop planting date matters: Estimation methods and effect on future yields. *Agricultural and Forest Meteorology*, 223:103–115.

Dosio, A. (2016). Projections of climate change indices of temperature and precipitation from an ensemble of bias-adjusted high-resolution EURO-CORDEX regional climate models. *Journal of Geophysical Research*, 121(10):5488–5511.

Dufresne, J. L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J. P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J. Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M. P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N. (2013). *Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5*, volume 40.

Eden, J. M. and Widmann, M. (2014). Downscaling of GCM-simulated precipitation using model output statistics. *Journal of Climate*, 27(1):312–324.

Eden, J. M., Widmann, M., Maraun, D., and Vrac, M. (2014). Comparison of GCM- and RCM-simulated precipitation following stochastic postprocessing. *Journal of Geophysical Research*, 119(19):11,040–11,053.

Eisner, S., Voss, F., and Kynast, E. (2012). Statistical bias correction of global climate projections – consequences for large scale modeling of flood flows. *Advances in Geosciences*, 31:75–82.

Eitzinger, J., Thaler, S., Schmid, E., Strauss, F., Ferrise, R., Moriondo, M., Bindi, M., Palosuo, T., Rotter, R., Kersebaum, K. C., Olesen, J. E., Patil, R. H., Åđaylan, L., Çaldag, B., and Çaylak, O. (2013). Sensitivities of crop models to extreme weather conditions during flowering period demonstrated for maize and winter wheat in Austria. *The Journal of Agricultural Science*, 151(06):813–835.

Eitzinger, J., Trnka, M., Hösch, J., Žalud, Z., and Dubrovský, M. (2004). Comparison of CERES, WOFOST and SWAP models in simulating soil water content during growing season under different soil conditions. *Ecological Modelling*, 171(3):223–246.

Ekström, M., Grose, M. R., and Whetton, P. H. (2015). An appraisal of downscaling methods used in climate change research. *Wiley Interdisciplinary Reviews: Climate Change*, 6(3):301–319.

Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C., Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C., and Sheffield, J. (2015). The Global Gridded Crop Model Intercomparison: Data and modeling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, 8(2):261–277.

Erda, L., Wei, X., Hui, J., Yinlong, X., Yue, L., Liping, B., and Liyong, X. (2005). Climate change impacts on crop yield and quality with CO2 fertilization in China. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1463):2149–2154.

Ewert, F., van Ittersum, M. K., Heckelei, T., Therond, O., Bezlepkina, I., and Andersen, E. (2011). Scale changes and model linking methods for integrated assessment of agri-environmental systems. *Agriculture, Ecosystems and Environment*, 142(1-2):6–17.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958.

FAO (2014). Food and Agriculture Organization of the United Nations (FAOSTAT). Available online at `http://www.fao.org/faostat/en/#home`. Accessed in 2014.

Farina, R., Seddaiu, G., Orsini, R., Steglich, E., Roggero, P. P., and Francaviglia, R. (2011).

Soil carbon dynamics and crop productivity as influenced by climate change in a rainfed cereal system under contrasting tillage using EPIC. *Soil and Tillage Research*, 112(1):36–46.

Farooq, M., Bramley, H., Palta, J. A., and Siddique, K. H. (2011). Heat Stress in Wheat during Reproductive and Grain-Filling Phases. *Critical Reviews in Plant Sciences*, 30(6):491–507.

Ferrara, R. M., Trevisiol, P., Acutis, M., Rana, G., Richter, G. M., and Baggaley, N. (2010). Topographic impacts on wheat yields under climate change: Two contrasted case studies in Europe. *Theoretical and Applied Climatology*, 99(1-2):53–65.

Feser, F., Rockel, B., von Storch, H., Winterfeldt, J., and Zahn, M. (2011). Regional Climate Models Add Value to Global Model Data: A Review and Selected Examples. *Bulletin of the American Meteorological Society*, 92(9):1181–1192.

Fischer, E. M., Seneviratne, S. I., Vidale, P. L., Lüthi, D., and Schär, C. (2007). Soil moisture-atmosphere interactions during the 2003 European summer heat wave. *Journal of Climate*, 20(20):5081–5099.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M. (2013). Evaluation of Climate Models. In Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assess- ment Report of the Intergovernmental Panel on Climate Change*, pages 741–866. Cambridge University Press.

Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L. B., Obersteiner, M., and Van Der Velde, M. (2016). Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. *Nature Communications*, 7(May):1–13.

Fowler, H., Blenkinsop, S., and Tebaldi, C. (2007). Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, 1578(September):1547–1578.

Frank, C. W., Wahl, S., Keller, J. D., Pospichal, B., Hense, A., and Crewell, S. (2018). Bias correction of a novel European reanalysis data set for solar energy applications. *Solar Energy*, 164(December 2017):12–24.

García-Herrera, R., Díaz, J., Trigo, R. M., Luterbacher, J., and Fischer, E. M. (2010). A Review of the European Summer Heat Wave of 2003. *Critical Reviews in Environmental Science and Technology*, 40(4):267–306.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, 5(3):572–597.

Giorgi, F. and Gutowski, W. J. (2015). Regional Dynamical Downscaling and the CORDEX Initiative. *Annual Review of Environment and Resources*, 40(1):467–490.

Glotter, M., Elliott, J., McInerney, D., Best, N., Foster, I., and Moyer, E. J. (2014). Evaluating the utility of dynamical downscaling in agricultural impacts projections. *Proceedings of the National Academy of Sciences*, 111(24):8776–8781.

Gobiet, A., Suklitsch, M., and Heinrich, G. (2015). The effect of empirical-statistical correction of intensity-dependent model errors on the temperature climate change signal. *Hydrology and Earth System Sciences*, 19(10):4055–4066.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., and Toulmin, C. (2010). Food Security: The Challenge of Feeding 9 Billion People. *Science*, 327(5967):812–818.

Godwin, D. and Jones, C. (1991). Nitrogen dynamics in soil-plant systems. In Hanks, J. and Ritchie, J., editors, *Modeling plant and soil systems. Agronomy Monograph No. 31, ASA CSSA and SSSA*, pages 287–321. Madison, WI.

Gornott, C. and Wechsung, F. (2016). Statistical regression models for assessing climate impacts on crop yields: A validation study for winter wheat and silage maize in Germany. *Agricultural and Forest Meteorology*, 217:89–100.

Greeves, C. Z., Pope, V. D., Stratton, R. A., and Martin, G. M. (2007). Representation of Northern Hemisphere winter storm tracks in climate models. *Climate Dynamics*, 28(7-8):683–702.

Grillakis, M. G., Koutroulis, A. G., Daliakopoulos, I. N., and Tsanis, I. K. (2017). A method to preserve trends in quantile mapping bias correction of climate modeled temperature. *Earth System Dynamics*, 8(3):889–900.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Catherine, A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., Von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J. J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and Von Storch, J. S. (2016). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, 9(11):4185–4208.

Hamilton, N. (2018). Package 'ggtern'. CRAN. Available online at `https://cran.r-project.org/package=ggtern`.

Hansen, J. W. and Jones, J. W. (2000). Scaling-up crop models for climate variability applications. *Agricultural Systems*, 65(1):43–72.

Hattermann, F. F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., Fekete, B., Florke, F., Gosling, S. N., Hoffmann, P., Liersch, S., Masaki, Y., Motovilov, Y., Muller, C., Samaniego, L., Stacke, T., Wada, Y., Yang, T., and Krysnaova, V. (2018). Sources of uncertainty in hydrological climate impact assessment: A cross-scale study. *Environmental Research Letters*, 13(1).

Hawkins, E., Fricker, T. E., Challinor, A. J., Ferro, C. A. T., Ho, C. K., and Osborne, T. M. (2013a). Increasing influence of heat stress on French maize yields from the 1960s to the 2030s. *Global Change Biology*, 19(3):937–947.

Hawkins, E., Osborne, T. M., Ho, C. K., and Challinor, A. J. (2013b). Calibration and bias correction of climate projections for crop modelling: An idealised case study over Europe. *Agricultural and Forest Meteorology*, 170:19–31.

Hawkins, E. and Sutton, R. (2009). The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bulletin of the American Meteorological Society*, 90(8):1095–1107.

Hawkins, E. and Sutton, R. (2011). The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, 37(1-2):407–418.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *Journal of Geophysical Research*, 113(D20):D20119.

Hazeleger, W., Severijns, C., Semmler, T., Ştefănescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J. M., Bintanja, R., Bougeault, P., Caballero, R., Ekman, A. M. L., Christensen, J. H., van den Hurk, B., Jimenez, P., Jones, C., Kållberg, P., Koenigk, T., McGrath, R., Miranda, P., van Noije, T., Palmer, T., Parodi, J. A., Schmith, T., Selten, F., Storelvmo, T., Sterl, A., Tapamo, H., Vancoppenolle, M., Viterbo, P., and Willén, U. (2010). EC-Earth. *Bulletin of the American Meteorological Society*, 91(10):1357–1364.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F. (2013). A trend-preserving bias correction – The ISI-MIP approach. *Earth System Dynamics*, 4(2):219–236.

Herrera, S., Fita, L., Fernández, J., and Gutiérrez, J. M. (2010). Evaluation of the mean and extreme precipitation regimes from the ENSEMBLES regional climate multimodel simulations over Spain. *Journal of Geophysical Research*, 115(D21):D21117.

Ho, C. K., Stephenson, D. B., Collins, M., Ferro, C. A. T., and Brown, S. J. (2012). Calibration Strategies: A Source of Additional Uncertainty in Climate Change Projections. *Bulletin of* *the American Meteorological Society*, 93(1):21–26.

Hoffmann, H., Zhao, G., Asseng, S., Bindi, M., Biernath, C., Constantin, J., Coucheney, E., Dechow, R., Doro, L., Eckersten, H., Gaiser, T., Grosz, B., Heinlein, F., Kassie, B. T., Kersebaum, K. C., Klein, C., Kuhnert, M., Lewan, E., Moriondo, M., Nendel, C., Priesack, E., Raynal, H., Roggero, P. P., Rötter, R. P., Siebert, S., Specka, X., Tao, F., Teixeira, E., Trombi, G., Wallach, D., Weihermüller, L., Yeluripati, J., and Ewert, F. (2016). Impact of spatial soil and climate input data aggregation on regional Yield Simulations. *PLoS ONE*, 11(4):1–23.

Högy, P. and Fangmeier, A. (2008). Effects of elevated atmospheric CO2 on grain quality of wheat. *Journal of Cereal Science*, 48(3):580–591.

Huffman, T., Qian, B., De Jong, R., Liu, J., Wang, H., McConkey, B., Brierley, T., and Yang, J. (2015). Upscaling modelled crop yields to regional scale: A case study using DSSAT for spring wheat on the Canadian prairies. *Canadian Journal of Soil Science*, 95(1):49–61.

Hurrell, J. W. and Deser, C. (2009). North Atlantic climate variability: The role of the North Atlantic Oscillation. *Journal of Marine Systems*, 78(1):28–41.

Hwang, S., Graham, W. D., Geurink, J. S., and Adams, A. (2014). Hydrologic implications of errors in bias-corrected regional reanalysis data for west central Florida. *Journal of Hydrology*, 510:513–529.

Iizumi, T. and Ramankutty, N. (2015). How do weather and climate influence cropping area and intensity? *Global Food Security*, 4:46–50.

Iizumi, T., Sakuma, H., Yokozawa, M., Luo, J.-J., Challinor, A. J., Brown, M. E., Sakurai, G., and Yamagata, T. (2013). Prediction of seasonal climate-induced variations in global food production. *Nature Climate Change*, 3(10):904–908.

Iizumi, T., Takikawa, H., Hirabayashi, Y., Hanasaki, N., and Nishimori, M. (2017). Contributions of different bias-correction methods and reference meteorological forcing data sets to uncertainty in projected temperature and precipitation extremes. *Journal of Geophysical Research*, 122(15):7800–7819.

Intergovernmental Panel on Climate Change (IPCC) (2012). Summary for Policymakers. In Intergovernmental Panel on Climate Change, editor, *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, pages 3–21. Cambridge University Press, Cambridge.

Intergovernmental Panel on Climate Change (IPCC) (2013). Summary for Policymakers. In Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. B., and (eds.), P. M., editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and

New York, NY, USA.

Intergovernmental Panel on Climate Change (IPCC) (2014). Summary for policymakers. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report.

IPCC (2014). Synthesis Report. In Pachauri, R. and Meyer, L., editors, *Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and Geneva, Switzerland.

IRI et al. (2015). Soil grids for DSSAT. International Research Institute for Climate and Society (IRI), Michigan State University (MSU), HarvestChoice, and the International Food Policy Research Institute. Available online from the Harvard Dataverse, `https://doi:10.7910/DVN/1PEEY0`.

J, L. (2006). Plotrix: a package in the red light district of r. *R-News*, 6(4):8–12.

Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O. B., Bouwer, L. M., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.-F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., and Yiou, P. (2014). EURO-CORDEX: new high-resolution climate change projections for European impact research. *Regional Environmental Change*, 14(2):563–578.

Jaeger, E. B., Anders, I., Lüthi, D., Rockel, B., Schär, C., and Seneviratne, S. I. (2008). Analysis of ERA40-driven CLM simulations for Europe. *Meteorologische Zeitschrift*, 17(4):349–367.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*, volume 103 of *Springer Texts in Statistics*. Springer New York, New York, NY.

Jamieson, P., Porter, J., Goudriaan, J., Ritchie, J., van Keulen, H., and Stol, W. (1998). A comparison of the models AFRCWHEAT2, CERES-Wheat, Sirius, SUCROS2 and SWHEAT with measurements from wheat grown under drought. *Field Crops Research*, 55(1-2):23–44.

Jones, J., Hoogenboom, G., Porter, C., Boote, K., Batchelor, W., Hunt, L., Wilkens, P., Singh, U., Gijsman, A., and Ritchie, J. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18(3-4):235–265.

Jury, M. W., Prein, A. F., Truhetz, H., and Gobiet, A. (2015). Evaluation of CMIP5 models in the context of dynamical downscaling over Europe. *Journal of Climate*, 28(14):5575–5582.

Keeley, S. P. E., Sutton, R. T., and Shaffrey, L. C. (2012). The impact of North Atlantic sea surface temperature errors on the simulation of North Atlantic European region climate. *Quarterly Journal of the Royal Meteorological Society*, 138(668):1774–1783.

Kendon, E. J., Ban, N., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., Evans, J. P., Fosser, G., and Wilkinson, J. M. (2017). Do convection-permitting regional climate models improve projections of future precipitation change? *Bulletin of the American Meteorological Society*, 98(1):79–93.

Kersebaum, K. and Nendel, C. (2014). Site-specific impacts of climate change on wheat production across regions of Germany using different CO2 response functions. *European Journal of Agronomy*, 52:22–32.

Kettlewell, P. S., Stephenson, D. B., Atkinson, M. D., and Hollins, P. D. (2003). Summer rainfall and wheat grain quality: Relationships with the North Atlantic Oscillation. *Weather*, 58(4):155–164.

Kirtman, B. P., Bitz, C., Bryan, F., Collins, W., Dennis, J., Hearn, N., Kinter, J. L., Loft, R., Rousset, C., Siqueira, L., Stan, C., Tomas, R., and Vertenstein, M. (2012). Impact of ocean model resolution on CCSM climate simulations. *Climate Dynamics*, 39(6):1303–1328.

Kjellström, E., Bärring, L., Nikulin, G., Nilsson, C., Persson, G., and Strandberg, G. (2016). Production and use of regional climate model projections – A Swedish perspective on building climate services. *Climate Services*, 2-3:15–29.

Knutti, R. and Sedláček, J. (2012). Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, 3(4):369–373.

Koehler, A.-K., Challinor, A. J., Hawkins, E., and Asseng, S. (2013). Influences of increasing temperature on Indian wheat: quantifying limits to predictability. *Environmental Research Letters*, 8(3):034016.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V. (2014). Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geoscientific Model Development*, 7(4):1297–1333.

Kristensen, K., Schelde K., and Olesen J. E. (2011). Winter wheat yield response to climate variability in Denmark. *Journal of Agricultural Science*, 149:33–47.

Kumar, D., Kodra, E., and Ganguly, A. R. (2014). Regional and seasonal intercomparison of CMIP3 and CMIP5 climate model ensembles for temperature and precipitation. *Climate Dynamics*, 43(9-10):2491–2518.

Langensiepen, M., Hanus, H., Schoop, P., and Gräsle, W. (2008). Validating CERES-wheat under North-German environmental conditions. *Agricultural Systems*, 97(1-2):34–47.

Lauer, A. and Hamilton, K. (2013). Simulating clouds with global climate models: A comparison of CMIP5 results with CMIP3 and satellite data. *Journal of Climate*, 26(11):3823–3845.

Lawless, C. and Semenov, M. A. (2005). Assessing lead-time for predicting wheat growth using a crop simulation model. *Agricultural and Forest Meteorology*, 135(1-4):302–313.

Leff, B., Ramankutty, N., and Foley, J. A. (2004). Geographic distribution of major crops across the world. *Global Biogeochemical Cycles*, 18(1).

Lemos, M. C. and Rood, R. B. (2010). Climate projections and their impact on policy and practice. *Wiley Interdisciplinary Reviews: Climate Change*, 1(5):670–682.

Li, G., Yu, M., Fang, T., Cao, S., Carver, B. F., and Yan, L. (2013). Vernalization requirement duration in winter wheat is controlled by Ta VRN-A1 at the protein level. *The Plant Journal*, 76(5):742–753.

Li, S., Tompkins, A., Lin, E., and Ju, H. (2016). Simulating the impact of flooding on wheat yield – Case study in East China. *Agricultural and Forest Meteorology*, 216:221–231.

Li, Y., Huang, H., Ju, H., Lin, E., Xiong, W., Han, X., Wang, H., Peng, Z., Wang, Y., Xu, J., Cao, Y., and Hu, W. (2015). Assessing vulnerability and adaptive capacity to potential drought for winter-wheat under the RCP 8.5 scenario in the Huang-Huai-Hai Plain. *Agriculture, Ecosystems & Environment*, 209:125–131.

Liu, B., Asseng, S., Müller, C., Ewert, F., Elliott, J., Lobell, D. B., Martre, P., Ruane, A. C., Wallach, D., Jones, J. W., Rosenzweig, C., Aggarwal, P. K., Alderman, P. D., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A., Deryng, D., Sanctis, G. D., Doltra, J., Fereres, E., Folberth, C., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurralde, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Kimball, B. A., Koehler, A.-K., Kumar, S. N., Nendel, C., O'Leary, G., Olesen, J. E., Ottman, M. J., Palosuo, T., Prasad, P. V. V., Priesack, E., Pugh, T. A. M., Reynolds, M., Rezaei, E. E., Rötter, R. P., Schmid, E., Semenov, M. A., Shcherbak, I., Stehfest, E., Stöckle, C. O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wall, G. W., Wang, E., White, J. W., Wolf, J., Zhao, Z., and Zhu, Y. (2016). Similar estimates of temperature impacts on global wheat yield by three independent methods. *Nature Climate Change*, 1(September):1–8.

Lobell, D. B. and Asseng, S. (2017). Comparing estimates of climate change impacts from process- based and statistical crop models. *Environmental Research Letters*, 12:1–12.

Lobell, D. B. and Burke, M. B. (2008). Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation. *Environmental Research Letters*, 3(3):034007.

Lobell, D. B. and Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, 150(11):1443–1452.

Lobell, D. B. and Field, C. B. (2007). Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, 2(1):014002.

Lobell, D. B., Hammer, G. L., McLean, G., Messina, C., Roberts, M. J., and Schlenker, W. (2013). The critical role of extreme heat for maize production in the United States. *Nature Climate Change*, 3(5):497–501.

Lobell, D. B., Sibley, A., and Ivan Ortiz-Monasterio, J. (2012). Extreme heat effects on wheat senescence in India. *Nature Climate Change*, 2(3):186–189.

Luo, Q. and Yu, Q. (2012). Developing higher resolution climate change scenarios for agricultural risk assessment: progress, challenges and prospects. *International Journal of Biometeorology*, 56(4):557–568.

Lüttger, A. B. and Feike, T. (2018). Development of heat and drought related extreme weather events and their effect on winter wheat yields in Germany. *Theoretical and Applied Climatology*, 132(1-2):15–29.

Ma, X., Jing, Z., Chang, P., Liu, X., Montuoro, R., Small, R. J., Bryan, F. O., Greatbatch, R. J., Brandt, P., Wu, D., Lin, X., and Wu, L. (2016). Western boundary currents regulated by interaction between ocean eddies and the atmosphere. *Nature*, 535(7613):533–537.

Macadam, I., Argüeso, D., Evans, J. P., Liu, D. L., and Pitman, A. J. (2016). The effect of bias correction and climate model resolution on wheat simulations forced with a regional climate model ensemble. *International Journal of Climatology*, 36(14):4577–4591.

Maharjan, G. R., Hoffmann, H., Webber, H., Srivastava, A. K., Weihermüller, L., Villa, A., Coucheney, E., Lewan, E., Trombi, G., Moriondo, M., Bindi, M., Grosz, B., Dechow, R., Kuhnert, M., Doro, L., Kersebaum, K. C., Stella, T., Specka, X., Nendel, C., Constantin, J., Raynal, H., Ewert, F., and Gaiser, T. (2019). Effects of input data aggregation on simulated crop yields in temperate and Mediterranean climates. *European Journal of Agronomy*, 103(December 2018):32–46.

Maiorano, A., Martre, P., Asseng, S., Ewert, F., Müller, C., Rötter, R. P., Ruane, A. C., Semenov, M. A., Wallach, D., Wang, E., Alderman, P. D., Kassie, B. T., Biernath, C., Basso, B., Cammarano, D., Challinor, A. J., Doltra, J., Dumont, B., Rezaei, E. E., Gayler, S., Kersebaum, K. C., Kimball, B. A., Koehler, A. K., Liu, B., O'Leary, G. J., Olesen, J. E., Ottman, M. J., Priesack, E., Reynolds, M., Stratonovitch, P., Streck, T., Thorburn, P. J., Waha, K., Wall, G. W., White, J. W., Zhao, Z., and Zhu, Y. (2017). Crop model improvement reduces the uncertainty of the response to temperature of multi-model ensembles. *Field Crops Research*, 202:5–20.

Maraun, D. (2013). Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue. *Journal of Climate*, 26(6):2137–2143.

Maraun, D. (2016). Bias Correcting Climate Change Simulations - a Critical Review. *Current Climate Change Reports*, 2(4):211–220.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M., Hall, A., and Mearns, L. O. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7(11):764–773.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48(3):RG3003.

Maraun, D. and Widmann, M. (2015). The representation of location by a regional climate model in complex terrain. *Hydrology and Earth System Sciences*, 19(8):3449–3456.

Maraun, D., Widmann, M., Gutiérrez, J. M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R., and Wilcke, R. A. (2015). VALUE: A framework to validate downscaling approaches for climate change studies. *Earth's Future*, 3(1):1–14.

Martin, G. M., Bellouin, N., Collins, W. J., Culverwell, I. D., Halloran, P. R., Hardiman, S. C., Hinton, T. J., Jones, C. D., McDonald, R. E., McLaren, A. J., O'Connor, F. M., Roberts, M. J., Rodriguez, J. M., Woodward, S., Best, M. J., Brooks, M. E., Brown, A. R., Butchart, N., Dearden, C., Derbyshire, S. H., Dharssi, I., Doutriaux-Boucher, M., Edwards, J. M., Falloon, P. D., Gedney, N., Gray, L. J., Hewitt, H. T., Hobson, M., Huddleston, M. R., Hughes, J., Ineson, S., Ingram, W. J., James, P. M., Johns, T. C., Johnson, C. E., Jones, A., Jones, C. P., Joshi, M. M., Keen, A. B., Liddicoat, S., Lock, A. P., Maidens, A. V., Manners, J. C., Milton, S. F., Rae, J. G., Ridley, J. K., Sellar, A., Senior, C. A., Totterdell, I. J., Verhoef, A., Vidale, P. L., and Wiltshire, A. (2011). The HadGEM2 family of Met Office Unified Model climate configurations. *Geoscientific Model Development*, 4(3):723–757.

Martín, M. M.-S., Olesen, J. E., and Porter, J. R. (2015). Can crop-climate models be accurate and precise? A case study for wheat production in Denmark. *Agricultural and Forest Meteorology*, 202:51–60.

Maurer, E. P. and Pierce, D. W. (2014). Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. *Hydrology and Earth System Sciences*, 18(3):915–925.

McIntosh, M. S. (2015). Can Analysis of Variance Be More Significant? *Agronomy Journal*, 107(2):706.

McSweeney, C. F., Jones, R. G., Lee, R. W., and Rowell, D. P. (2015). Selecting CMIP5 GCMs for downscaling over multiple regions. *Climate Dynamics*, 44(11-12):3237–3260.

Meehl, G. A., Washington, W. M., Arblaster, J. M., Hu, A., Teng, H., Tebaldi, C., Sanderson, B. N., Lamarque, J.-F., Conley, A., Strand, W. G., and White, J. B. (2012). Climate System Response to External Forcings and Climate Change Projections in CCSM4. *Journal of Climate*, 25(11):3661–3683.

Mehan, S., Guo, T., Gitau, M., and Flanagan, D. C. (2017). Comparative Study of Different Stochastic Weather Generators for Long-Term Climate Data Simulation. *Climate*, 5(2):26.

Mehran, A., AghaKouchak, A., and Phillips, T. J. (2014). Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations. *Journal of Geophysical Research: Atmospheres*, 119(4):1695–1707.

Menut, L., Tripathi, O. P., Colette, A., Vautard, R., Flaounas, E., and Bessagnet, B. (2013). Evaluation of regional climate simulations for air quality modelling purposes. *Climate Dynamics*, 40(9-10):2515–2533.

Merbach, I. and Schulz, E. (2013). Long-term fertilization effects on crop yields, soil fertility and sustainability in the Static Fertilization Experiment Bad Lauchstädt under climatic conditions 2001–2010. *Archives of Agronomy and Soil Science*, 59(8):1041–1057.

Michel, L. and Makowski, D. (2013). Comparison of Statistical Models for Analyzing Wheat Yield Time Series. *PLoS ONE*, 8(10):e78615.

Mitchell, D., James, R., Forster, P. M., Betts, R. A., Shiogama, H., and Allen, M. (2016). Realizing the impacts of a 1.5 °C warmer world. *Nature Climate Change*, 6(8):735–737.

Moore, F. C., Baldos, U. L. C., and Hertel, T. (2017). Economic impacts of climate change on agriculture: a comparison of process-based and statistical yield models. *Environmental Research Letters*, 12(6):065008.

Moore, F. C. and Lobell, D. B. (2014). Adaptation potential of European agriculture in response to climate change. *Nature Climate Change*, 4(7):610–614.

Moore, F. C. and Lobell, D. B. (2015). The fingerprint of climate trends on European crop yields. *Proceedings of the National Academy of Sciences*, 112(9):2670–2675.

Moss, R. H., Edmonds, J. a., Hibbard, K. a., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. a., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., and Wilbanks, T. J. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, 463(7282):747–756.

Muggeo, V. M. R. (2008). segmented: An R Package to Fit Regression Models with Broken-Line Relationships. In *R News*, volume 8, pages 20–25.

Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane,

A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., De Wit, A., and Yang, H. (2017). Global gridded crop model evaluation: Benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, 10(4):1403–1422.

Nain, A. S. and Kersebaum, K. C. (2007). Calibration and validation of CERES model for simulating water and nutrients in Germany. In *Modelling water and nutrient dynamics in soil-crop systems*, pages 161–181. Springer.

Northrop, P. J. and Chandler, R. E. (2014). Quantifying sources of uncertainty in projections of future climate. *Journal of Climate*, 27(23):8793–8808.

Oettli, P., Sultan, B., Baron, C., and Vrac, M. (2011). Are regional climate models relevant for crop yield prediction in West Africa? *Environmental Research Letters*, 6(1):014008.

Olesen, J., Trnka, M., Kersebaum, K., Skjelvåg, A., Seguin, B., Peltonen-Sainio, P., Rossi, F., Kozyra, J., and Micale, F. (2011). Impacts and adaptation of European crop production systems to climate change. *European Journal of Agronomy*, 34(2):96–112.

Olsson, J., Arheimer, B., Borris, M., Donnelly, C., Foster, K., Nikulin, G., Persson, M., Perttu, A.-M., Uvo, C., Viklander, M., and Yang, W. (2016). Hydrological Climate Change Impact Assessment at Small and Large Scales: Key Messages from Recent Progress in Sweden. *Climate*, 4(3):39.

Ortiz, R., Sayre, K. D., Govaerts, B., Gupta, R., Subbarao, G., Ban, T., Hodson, D., Dixon, J. M., Iván Ortiz-Monasterio, J., and Reynolds, M. (2008). Climate change: Can wheat beat the heat? *Agriculture, Ecosystems & Environment*, 126(1-2):46–58.

Osborne, T., Rose, G., and Wheeler, T. (2013). Variation in the global-scale impacts of climate change on crop productivity due to climate model uncertainty and adaptation. *Agricultural and Forest Meteorology*, 170:183–194.

Palosuo, T., Kersebaum, K. C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J. E., Patil, R. H., Ruget, F., Rumbaur, C., Takáč, J., Trnka, M., Bindi, M., Çaldağ, B., Ewert, F., Ferrise, R., Mirschel, W., Åđaylan, L., Šiška, B., and Rötter, R. (2011). Simulation of winter wheat yield and its variability in different climates of Europe: A comparison of eight crop growth models. *European Journal of Agronomy*, 35(3):103–114.

Pasten Zapata, E. (2017). *An Assessment of the Impact of Climate Change on Hydroelectric Power*. Doctoral thesis, University of Sheffield, Sheffield, United Kingdom.

Piani, C., Haerter, J. O., and Coppola, E. (2010). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology*, 99(1-2):187–192.

Pierce, D. (2017). Package 'ncdf4'. CRAN. Available online at `http://cirrus.ucsd.edu/~pierce/ncdf`.

Porter, J., Xie, L., Challinor, A., Cochrane, K., Howden, S., Iqbal, M., Lobell, D., and

Travasso, M. (2014). Food Security and Food Production Systems. In Field, C., Barros, V., Dokken, D., Mach, K., Mastrandrea, M., Bilir, T., Chatterjee, M., Ebi, K., Estrada, Y., Genova, R., Girma, B., Kissel, E., Levy, A., MacCracken, S., Mastrandrea, P., and White, L., editors, *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, chapter 7, pages 485–533. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Porter, J. R. and Gawith, M. (1999). Temperatures and the growth and development of wheat: a review. *European Journal of Agronomy*, 10(1):23–36.

Porter, J. R. and Semenov, M. A. (2005). Crop responses to climatic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1463):2021–2035.

Priestley, C. H. B. and Taylor, R. J. (1972). On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. *Monthly Weather Review*, 100(2):81–92.

Ramarohetra, J., Pohl, B., and Sultan, B. (2015). Errors and uncertainties introduced by a regional climate model in climate impact assessments: example of crop yield simulations in West Africa. *Environmental Research Letters*, 10(12):124014.

Ramirez-Villegas, J., Challinor, A. J., Thornton, P. K., and Jarvis, A. (2013). Implications of regional improvement in global climate models for agricultural impact research. *Environmental Research Letters*, 8(2):024018.

Ray, D. K., Gerber, J. S., MacDonald, G. K., and West, P. C. (2015). Climate variation explains a third of global crop yield variability. *Nature Communications*, 6:5989.

Reidsma, P., Ewert, F., Oude Lansink, A., and Leemans, R. (2009). Vulnerability and adaptation of European farmers: a multi-level analysis of yield and income responses to climate variability. *Regional Environmental Change*, 9(1):25–40.

Rezaei, E. E., Siebert, S., and Ewert, F. (2015). Intensity of heat stress in winter wheat-phenology compensates for the adverse effect of global warming. *Environmental Research Letters*, 10(2):024012.

Rezaei, E. E., Siebert, S., Hüging, H., and Ewert, F. (2018). Climate change effect on wheat phenology depends on cultivar change. *Scientific Reports*, 8(1):1–10.

Rial-Lovera, K., Davies, W. P., and Cannon, N. D. (2017). Implications of climate change predictions for UK cropping and prospects for possible mitigation: a review of challenges and potential responses. *Journal of the Science of Food and Agriculture*, 97(1):17–32.

Ritchie, J. and Otter, S. (1985). Description and Performance of CERES-Wheat: A User-Oriented Wheat Yield Model. In *ARS Wheat Yield Project*, pages 159–175. ARS-38, US Department of Agriculture-Agricultural Research Service, Springfield, VA.

Röder, M., Thornley, P., Campbell, G., and Bows-Larkin, A. (2014). Emissions associated with meeting the future global wheat demand: A case study of UK production under climate change constraints. *Environmental Science and Policy*, 39:13–24.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T. a. M., Schmid, E., Stehfest, E., Yang, H., and Jones, J. W. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, 111(9):3268–3273.

Rosenzweig, C., Jones, J., Hatfield, J., Ruane, A., Boote, K., Thorburn, P., Antle, J., Nelson, G., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., and Winter, J. (2013). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology*, 170:166–182.

Ruane, A. C., Phillips, M. M., and Rosenzweig, C. (2018). Climate shifts within major agricultural seasons for +1.5 and +2.0 °C worlds: HAPPI projections and AgMIP modeling scenarios. *Agricultural and Forest Meteorology*, 259(October 2017):329–344.

Ruane, A. C., Rosenzweig, C., Asseng, S., Boote, K. J., Elliott, J., Ewert, F., Jones, J. W., Martre, P., McDermid, S. P., Müller, C., Snyder, A., and Thorburn, P. J. (2017). An AgMIP framework for improved agricultural representation in integrated assessment models. *Environmental Research Letters*, 12(12):125003.

Ruiz-Ramos, M. and Mínguez, M. (2010). Evaluating uncertainty in climate change impacts on crop productivity in the Iberian Peninsula. *Climate Research*, 44(1):69–82.

Ruiz-Ramos, M., Rodríguez, A., Dosio, A., Goodess, C. M., Harpham, C., Mínguez, M. I., and Sánchez, E. (2016). Comparing correction methods of RCM outputs for improving crop impact projections in the Iberian Peninsula for 21st century. *Climatic Change*, 134(1-2):283–297.

Rummukainen, M. (2016). Added value in regional climate modeling. *Wiley Interdisciplinary Reviews: Climate Change*, 7(1):145–159.

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J., and Williams, A. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, 41(7):2514–2519.

Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106(37):15594–15598.

Schleussner, C.-F., Deryng, D., Müller, C., Elliott, J., Saeed, F., Folberth, C., Liu, W., Wang, X., Pugh, T. A. M., Thiery, W., Seneviratne, S. I., and Rogelj, J. (2018). Crop productivity changes in 1.5℃ and 2℃ worlds under climate sensitivity uncertainty. *Environmental Research Letters*, 13(6):064007.

Semenov, M. and Barrow, E. (1997). Use of a stochastic Weather Generator in the development of climate change scenarios. *Climatic Change*, 35(4):397–414.

Semenov, M., Stratonovitch, P., Alghabari, F., and Gooding, M. (2014). Adapting wheat in Europe for climate change. *Journal of Cereal Science*, 59(3):245–256.

Semenov, M. A. (2009). Impacts of climate change on wheat in England and Wales. *Journal of the Royal Society Interface*, 6(33):343–350.

Semenov, M. A., Mitchell, R. A. C., Whitmore, A. P., Hawkesford, M. J., Parry, M. A. J., and Shewry, P. R. (2012). Shortcomings in wheat yield predictions. *Nature Climate Change*, 2(6):380–382.

Semenov, M. A. and Shewry, P. R. (2011). Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in Europe. *Scientific Reports*, 1:66.

Shi, W., Tao, F., and Zhang, Z. (2013). A review on statistical models for identifying climate contributions to crop yields. *Journal of Geographical Sciences*, 23(3):567–576.

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., and Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres*, 118(4):1716–1733.

Singh, A. K., Tripathy, R., and Chopra, U. K. (2008). Evaluation of CERES-Wheat and CropSyst models for water-nitrogen interactions in wheat crop. *Agricultural Water Management*, 95(7):776–786.

Sippel, S., Otto, F. E., Forkel, M., Allen, M. R., Guillod, B. P., Heimann, M., Reichstein, M., Seneviratne, S. I., Thonicke, K., and Mahecha, M. D. (2016). A novel bias correction methodology for climate impact simulations. *Earth System Dynamics*, 7(1):71–88.

Soltani, A., Bakker, M. M., Veldkamp, A., and Stoorvogel, J. J. (2016). Comparison of three modelling approaches to simulate regional crop yield: A case study of winter wheat yield in western Germany. *Journal of Agricultural Science and Technology*, 18(1):191–206.

Sørland, S. L., Schär, C., Lüthi, D., and Kjellström, E. (2018). Bias patterns and climate change signals in GCM-RCM model chains. *Environmental Research Letters*, 13(7):074017.

Staffell, I. and Pfenninger, S. (2016). Using bias-corrected reanalysis to simulate current and future wind power output. *Energy*, 114:1224–1239.

Strer, M., Svoboda, N., and Herrmann, A. (2018). Abundance of adverse environmental conditions during critical stages of crop production in Northern Germany. *Environmental Sciences Europe*, 30(1).

Supit, I., van Diepen, C., de Wit, A., Wolf, J., Kabat, P., Baruth, B., and Ludwig, F.

(2012). Assessing climate change effects on European crop yields using the Crop Growth Monitoring System and a weather generator. *Agricultural and Forest Meteorology*, 164:96–111.

Takayabu, I., Kanamaru, H., Dairaku, K., Benestad, R., von Storch, H., and Christensen, J. H. (2016). Reconsidering the Quality and Utility of Downscaling. *Journal of the Meteorological Society of Japan. Ser. II*, 94A(0):31–45.

Tan, I., Storelvmo, T., and Zelinka, M. D. (2016). Observational constraints on mixed-phase clouds imply higher climate sensitivity. *Science*, 352(6282):224–227.

Tanaka, A., Takahashi, K., Masutomi, Y., Hanasaki, N., Hijioka, Y., Shiogama, H., and Yamanaka, Y. (2015). Adaptation pathways of global wheat production: Importance of strategic adaptation to climate change. *Scientific reports*, 5(August):14312.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7):7183–7192.

Taylor, K. E., Stouffer, R. J., and Meehl, G. a. (2012). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, 93(4):485–498.

Teutschbein, C. and Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, 456-457(August 2015):12–29.

Thaler, S., Eitzinger, J., Trnka, M., and Dubrovsky, M. (2012). Impacts of climate change and alternative adaptation options on winter wheat yield and water productivity in a dry climate in Central Europe. *The Journal of Agricultural Science*, 150(05):537–555.

Therond, O., Hengsdijk, H., Casellas, E., Wallach, D., Adam, M., Belhouchette, H., Oomen, R., Russell, G., Ewert, F., Bergez, J. E., Janssen, S., Wery, J., and Van Ittersum, M. K. (2011). Using a cropping system model at regional scale: Low-data approaches for crop management information and model calibration. *Agriculture, Ecosystems and Environment*, 142(1-2):85–94.

Thorup-Kristensen, K., Cortasa, M. S., and Loges, R. (2009). Winter wheat roots grow twice as deep as spring wheat roots, is this important for N uptake and N leaching losses? *Plant and Soil*, 322(1):101–114.

Trnka, M., Eitzinger, J., Kapler, P., Dubrovský, M., Semerádová, D., Žalud, Z., and Formayer, H. (2007). Effect of Estimated Daily Global Solar Radiation Data on the Results of Crop Growth Models. *Sensors*, 7(10):2330–2362.

Trnka, M., Hlavinka, P., and Semenov, M. A. (2015). Adaptation options for wheat in Europe will be limited by increased adverse weather events under climate change. *Journal of The Royal Society Interface*, 12(112):20150721.

Trnka, M., Olesen, JE, Kersebaum, K. C., Skjelvag, A., Eitzinger, J., Seguin, B., Peltonen-

Sainio, P., Rotter, R., Iglesias, A., Orlandini, S., Dubrovský, M., Hlavinka, P., Balek, J., Eckesten, H., Cloppet, E., Calanca, P., Gobin, A., Vucetic, V., Nejedlik, P., Kumar, S., Lalic, B., Mestre, A., Rossi, F., Kozyra, J., Alexandrov, V., Semeradova, D., and Zalud, Z. (2011). Agroclimatic conditions in Europe under climate change. *Global Change Biology*, 17(7):2298–2318.

Trnka, M., Rötter, R. P., Ruiz-Ramos, M., Kersebaum, K. C., Olesen, J. E., Žalud, Z., and Semenov, M. a. (2014). Adverse weather conditions for European wheat production will become more frequent with climate change. *Nature Climate Change*, 4(7):637–643.

Tsvetsinskaya, E. and Mearns, L. (2003). The effect of spatial scale of climatic change scenarios on simulated maize, winter wheat, and rice production in the Southeastern United States. *Climatic Change*, pages 37–71.

Urraca, R., Gracia-Amillo, A. M., Huld, T., Martinez-de Pison, F. J., Trentmann, J., Lindfors, A. V., Riihelä, A., and Sanz-Garcia, A. (2017a). Quality control of global solar radiation data with satellite-based products. *Solar Energy*, 158(September):49–62.

Urraca, R., Gracia-Amillo, A. M., Koubli, E., Huld, T., Trentmann, J., Riihelä, A., Lindfors, A. V., Palmer, D., Gottschalg, R., and Antonanzas-Torres, F. (2017b). Extensive validation of CM SAF surface radiation products over Europe. *Remote Sensing of Environment*, 199:171–186.

van der Velde, M., Tubiello, F. N., Vrieling, A., and Bouraoui, F. (2012). Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. *Climatic Change*, 113(3-4):751–765.

van Meijgaard, E., Ulft, L. H. V., Bosveld, F. C., Lenderink, G., and Siebesma, a. P. (2008). The KNMI regional atmospheric climate model RACMO version 2.1. *Technical report; TR - 302*, 17(4):43.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K. (2011). The representative concentration pathways: an overview. *Climatic Change*, 109(1-2):5–31.

Vanuytrecht, E., Raes, D., Willems, P., and Geerts, S. (2012). Quantifying field-scale effects of elevated carbon dioxide concentration on crops. *Climate Research*, 54(1):35–47.

Vermeulen, S. J., Challinor, A. J., Thornton, P. K., Campbell, B. M., Eriyagama, N., Vervoort, J. M., Kinyangi, J., Jarvis, A., Laderach, P., Ramirez-Villegas, J., Nicklin, K. J., Hawkins, E., and Smith, D. R. (2013). Addressing uncertainty in adaptation planning for agriculture. *Proceedings of the National Academy of Sciences*, 110(21):8357–8362.

Vetter, T., Reinhardt, J., Flörke, M., van Griensven, A., Hattermann, F., Huang, S., Koch, H., Pechlivanidis, I. G., Plötner, S., Seidou, O., Su, B., Vervoort, R. W., and Krysanova,

V. (2017). Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins. *Climatic Change*, 141(3):419–433.

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., Déqué, M., Deshayes, J., Douville, H., Fernandez, E., Madec, G., Maisonnave, E., Moine, M.-P., Planton, S., Saint-Martin, D., Szopa, S., Tyteca, S., Alkama, R., Belamari, S., Braun, A., Coquart, L., and Chauvin, F. (2013). The CNRM-CM5.1 global climate model: description and basic evaluation. *Climate Dynamics*, 40(9-10):2091–2121.

Wang, B., Liu, D. L., Asseng, S., Macadam, I., and Yu, Q. (2015). Impact of climate change on wheat flowering time in eastern Australia. *Agricultural and Forest Meteorology*, 209-210:11–21.

Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Rötter, R. P., Kimball, B. A., Ottman, M. J., Wall, G. W., White, J. W., Reynolds, M. P., Alderman, P. D., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A. J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurralde, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Koehler, A. K., Liu, L., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J. E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ripoche, D., Ruane, A. C., Semenov, M. A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wallach, D., Wang, Z., Wolf, J., Zhu, Y., and Asseng, S. (2017). The uncertainty of crop yield projections is reduced by improved temperature response functions. *Nature Plants*, 3(July):1–11.

Wang, X., Zhao, C., Li, C., Liu, L., Huang, W., and Wang, P. (2009). Use of CERES-wheat model for wheat yield forecast in Beijing. *IFIP International Federation for Information Processing*, 293(Fao 2004):29–37.

Waongo, M., Laux, P., and Kunstmann, H. (2015). Adaptation to climate change: The impacts of optimized planting dates on attainable maize yields under rainfed conditions in Burkina Faso. *Agricultural and Forest Meteorology*, 205:23–39.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J. (2014). The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework. *Proceedings of the National Academy of Sciences*, 111(9):3228–3232.

Watson, J. and Challinor, A. (2013). The relative importance of rainfall, temperature and yield data for a regional-scale crop model. *Agricultural and Forest Meteorology*, 170:47–57.

Watson, J., Challinor, A. J., Fricker, T. E., and Ferro, C. A. T. (2015). Comparing the effects of calibration and climate errors on a statistical crop model and a process-based crop model. *Climatic Change*, 132(1):93–109.

Webber, H., Martre, P., Asseng, S., Kimball, B., White, J., Ottman, M., Wall, G. W., De Sanctis, G., Doltra, J., Grant, R., Kassie, B., Maiorano, A., Olesen, J. E., Ripoche, D., Rezaei, E. E., Semenov, M. A., Stratonovitch, P., and Ewert, F. (2015). Canopy temperature for simulation of heat stress in irrigated wheat in a semi-arid environment: A multi-model comparison. *Field Crops Research*, (November).

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Learning in Doing: Social, Cognitive and Computational Perspectives. Cambridge University Press.

White, J. W. (2009). Comments on a report of regression-based evidence for impact of recent climate change on winter wheat yields. *Agriculture, Ecosystems & Environment*, 129(4):547–548.

White, J. W., Hoogenboom, G., Kimball, B. A., and Wall, G. W. (2011). Methodologies for simulating impacts of climate change on crop production. *Field Crops Research*, 124(3):357–368.

Wilby, R. L. and Dessai, S. (2010). Robust adaptation to climate change. *Weather*, 65(7):180–185.

Wilby, R. L., Troni, J., Biot, Y., Tedd, L., Hewitson, B. C., Smith, D. M., and Sutton, R. T. (2009). A review of climate risk information for adaptation and development planning. *International Journal of Climatology*, 29(9):1193–1215.

Wilcke, R. A. and Bärring, L. (2016). Selecting regional climate scenarios for impact modelling studies. *Environmental Modelling and Software*, 78:191–201.

Wolkowski, R. and Lowery, B. (2008). Soil comparison: causes, concerns, and cures. University of Wisconsin Cooperative Extension, Available online at `http://www.soils.wisc.edu/extension/pubs/A3367.pdf`.

Wong, G., Maraun, D., Vrac, M., Widmann, M., Eden, J. M., and Kent, T. (2014). Stochastic Model Output Statistics for Bias Correcting and Downscaling Precipitation Including Extremes. *Journal of Climate*, 27(18):6940–6959.

Woollings, T. (2010). Dynamical influences on European climate: an uncertain future. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1924):3733–3756.

Wrigley, C., Corke, H., Seetharaman, K., and Faubion, J. (2016). Agronomy of Grain Growing: Wheat Agronomy. In Wrigley, C., Corke, H., Seetharaman, K., and Faubion, J., editors, *Encyclopedia of Food Grains*, pages 176–185. Elsevier, Kidlington, Oxford.

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., and Dickinson, R. (2013). The role of satellite remote sensing in climate change studies. *Nature Climate Change*, 3(10):875–883.

Yip, S., Ferro, C. A. T., Stephenson, D. B., and Hawkins, E. (2011). A Simple, Coherent

Framework for Partitioning Uncertainty in Climate Predictions. *Journal of Climate*, 24(17):4634–4643.

Yoo, C. and Cho, E. (2018). Comparison of GCM precipitation predictions with their RMSEs and pattern correlation coefficients. *Water (Switzerland)*, 10(1).

Zampieri, M., Ceglar, A., Dentener, F., and Toreti, A. (2017). Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales. *Environmental Research Letters*, 12(6):064008.

Zappa, G., Masato, G., Shaffrey, L., Woollings, T., and Hodges, K. (2014). Linking Northern Hemisphere blocking and storm track biases in the CMIP5 climate models. *Geophysical Research Letters*, 41(1):135–139.