UNIVERSITY OF LEEDS

The Assessment of Surface Precipitation Type Measurements

by

Ben Stuart Pickering

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
School of Earth and Environment

October 2020

Declaration of authorship

The candidate confirms that the work submitted is their own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to work of others.

The publication Pickering et al., 2019, *The Disdrometer Verification Network (DiVeN): a UK network of laser precipitation instruments*, *Atmospheric Measurement Techniques* 12, p5845–5861, `https://doi.org/10.5194/amt-12-5845-2019`, jointly authored with Ryan R. Neely III and Dawn Harrison, is included as Chapter 2 of this thesis. The text was solely written by the candidate, with comments from co-authors. The candidate performed all data analysis and produced all figures. Neely and Harrison contributed to method refinement and interpretation of results. Harrison was responsible for the acquisition of the instruments from the Met Office.

Chapter 3: Pickering et al., *Evaluation of multiple precipitation sensor designs for precipitation rate and depth, drop size and velocity distribution, and precipitation type.*, jointly authored with Ryan R. Neely III, Judith Jeffery, David Dufton and Maryna Lukach, has been prepared for submission to *AMS Journal of Hydrometeorology.* The text was solely written by the candidate, with comments from co-authors. The candidate performed all data analysis and produced all figures. The candidate installed and maintained the two DiVeN instruments while Jeffery was responsible for the maintenance and calibration of the five non-DiVeN instruments. All co-authors contributed to interpretation of results. The Chilbolton Atmospheric Observatory (CAO) group published the non-DiVeN observational data to the CEDA Archive.

Chapter 4: Pickering et al., *Improving Observations of Precipitation Type at the Surface: A 5-year Verification of a Radar-derived Product from the United Kingdom Met Office. Part I: Rain, Mixed-phase and Snow*, jointly authored with Steven Best, David Dufton, Maryna Lukach, Darren Lyth and Ryan R. Neely III, has been prepared for submission to *AMS Journal of Hydrometeorology.* The text was solely written by the candidate, with comments from co-authors. The candidate performed all data analysis and produced all figures. Steven Best supplied the product data in a converted format to the candidate. All co-authors contributed to interpretation of results.

Chapter 5: Pickering et al., *Improving Observations of Precipitation Type at the Surface: A 5-year Verification of a Radar-derived Product from the United Kingdom Met Office. Part II: Hail*, jointly authored with Steven Best, David Dufton, Maryna Lukach, Darren Lyth and Ryan R. Neely III, has been prepared for submission to *AMS Journal of Hydrometeorology.* The text was solely written by the candidate, with comments from co-authors. The candidate performed all data analysis and produced all figures. Steven Best supplied the product data in a converted format to the candidate. All co-authors contributed to interpretation of results.

Acknowledgements

I would like to thank all of my supervisors for their support over the last 4 years: Ryan Neely III, Steven Best, Dawn Harrison, David Dufton, Maryna Lukach and Alan Blyth. In particular, I cannot fully express my gratitude to Neely for all of the opportunities that he enabled for me throughout my PhD. Our supervisor (and grant writing) sessions were enjoyable, memorable and inspiring. I respect Neely for his ability to not deconstruct an idea, but to instead ask what needs to happen for it to succeed. His constructive thinking fostered my creativity and it is a mindset that I will take forward with me in my future career.

This research was funded by a NERC Industrial CASE studentship (NE/N008359/1). Thank you to the academic partners at the UK Met Office for their support throughout the project. The visits to the Met Office were productive and enjoyable, and the research would not have been possible without their continued efforts. Thank you to the whole radar group at the Met Office for making me feel welcome during my visits. Thank you to Dawn Harrison for supervising the start-up of the research, and for sourcing the Thies LPM instruments, which was fundamental to the success of the Disdrometer Verification Network. Thank you to Steven Best for supervising and hosting me at the Met Office. In particular, his assistance and insight into the Met Office systems and products has been invaluable. Darren Lyth has been another exceptional colleague at the Met Office who has provided helpful expertise on the ground instruments that the Met Office operate. Thanks goes to Steve Willington at the Met Office and Thomas Schreiner at the European Severe Storms Laboratory for assistance retrieving the European Severe Weather Database (ESWD) data.

Particular thanks goes to the entire team at the CAO, who have been a fantastic group of people to work with both on my PhD research and other research we have collaborated on. In particular, thank you to Judith Jeffery who managed the precipitation instruments at the site and for providing insightful knowledge during the research evaluating the DiVeN instruments in Chapter 3. To Dave Hazzard, Chris Walden, Darcy Ladd and Judith Jeffery for assisting with the installation of the DiVeN instruments.

I wish to extend my gratitude to the following people and institutions for contributing to the creation of the Disdrometer Verification Network. Morwenna Cooper (Met Office), Dan Walker (NCAS), James Groves (NCAS), and Darren Lyth (Met Office) for technical advice regarding the data acquisition design of DiVeN. The contacts at each site hosting a disdrometer for DiVeN: Judith Jeffery (CAO), Andrew Lomas (University of Reading), Rebecca Carling (Facility for Atmospheric Measurements), Grant Forster (University of East Anglia), David Hooper (NFARR), James Heath (University of Lancaster), Richard Essery (University of Edinburgh), Geoff Monk (Mountain Weather Information Service), Michael Flynn (University of Manchester), Louise Parry (Scottish Environment Protection Agency), Jim Cornfoot (Natural Retreats), Chris Taylor (Natural Retreats), Andrew Black (University of Dundee), Darren Lyth (Met Office), Megan Klaar (University of Leeds), and Stephen Mawle (Coverhead Farm). For providing accommodation and sanity during the month-long installation of DiVeN, I thank Jack Giddings, Ashley Nelis, Scott Duncan, and Daniel Page. I thank Philip Rosenberg (NCAS) for advice on statistical tests, and I also thank Stephen Best (Met Office), James Bowles (Met Office), Dave Hazard (NFARR), Darcy Ladd (NFARR), Stephen Burt (University of Reading), and Chris Westbrook (University of Reading) for other assistance and advice.

The National Centre for Atmospheric Science (NCAS) has welcomed me in throughout my PhD and provided additional support through training, resources and funding for which I am immensely grateful. The NCAS radar group has been wonderful and provided many interesting science conversations over the years. Furthermore, conferences with the NCAS radar group have been great fun. Thank you to Harriett Richardson for all the yoga classes keeping me zen both in the office and whilst working from home.

The Royal Meteorological Society (RMetS) has afforded me many opportunities over the last 4 years. Being part of several organising committees—special shout out to the Yorkshire Local Centre—has been rewarding and enjoyable. I look forward to working with RMetS in the future.

Funding from the NCAS Visiting Science Programme and the Royal Meteorological Society Legacies Award enabled me to partake in fieldwork with the RELAMPAGO field campaign in Argentina, an unforgettable experience working on mobile radar and MesoNet vehicles in late 2018. Special thanks to Professor Stephen Nesbitt and Karen Kosiba for encouraging me to get involved and also to the entire RELAMPAGO team for being incredibly friendly and welcoming on a project so far from home.

To Dongqi Lin and Jonathan Coney—thank you for being brilliant masters students.

It was a great pleasure to supervise you on the MRes programme, and I am incredibly proud of all the work you have achieved. I wish you all the very best with your PhD projects, and your future careers.

To Farran: I dedicate this thesis to you. You have been by my side throughout my entire PhD, you've supported me through the happy times and the difficult times, and you've also inspired me with your own plight back into education. For believing in me and my crazy ideas, and for listening to them—but most importantly, for knowing when I need to breathe and talk.

The Met Gang: Sabrina, Hristo, Dan, Richard, Ashley, Scott, Sarah, Jack, Becca, Ben. For all the crazy adventures, the video calls during the COVID-19 lockdown, for keeping me sane in general and for knowing that we'll always be there to support each other. Special thanks to Joshua, Freya, Dean, Beth and Anya for suffering with me in Fairbairn from the start. We got through it stronger, together. For all of my friends in the community who have kept me going, with a particular shout out to the wonderful Kris and Jim. To Elon Musk, for inspiring me to think bigger, and for giving me hope for humanity, especially in 2020.

Finally, some credit is necessary to the MathPix team for making LaTeX equations so easy, to the Google Docs team and the Overleaf team for making writing and getting feedback easy, and a big thank you to Beth Woodhams for creating the superb LaTeX thesis template which has saved me hours of coding and debugging when writing this thesis.

Abstract

The type of precipitation reaching the surface (SPT) has a significant impact on society, yet both SPT observations and forecast products have poor skill. Improving any observation or forecast requires a framework of suitable assessment techniques. The assessment of SPT data is difficult because the commonly used categorisation standards lack specificity; slight differences exist between the standards, and some categories contain several SPTs, obfuscating the reference data. Furthermore, the rarity of some SPTs makes capturing a statistically significant and climatologically representative reference dataset challenging. Finally, assessing the skill of spatial SPT products from radar or forecasts, against a single point reference SPT dataset introduces representativity errors. Within the literature, the statistical metrics used to perform verification of any SPT data differ, obscuring the comparison of SPT diagnosis techniques and thereby suppressing the advancement of more accurate techniques into widespread operational use.

This thesis tackles these assessment issues with a focus on observational techniques in the United Kingdom. Firstly, a new network of low-cost, real-time disdrometers were deployed to increase the number, quality and represented climatologies of SPT observations. These instruments were validated using case studies, followed by a 12-month evaluation with existing precipitation instruments. To facilitate the assessment, a new SPT classification standard, based on microphysical processes, is produced to reduce the ambiguity in SPT observations. In addition, an algorithm to amalgamate SPT observations over time is developed and applied successfully. Finally, a radar-based spatial SPT product from the UK Met Office is verified against all available ground-based datasets for five years. A novel tolerance technique is developed to address the representativity issues between spatial and point datasets. This technique is used to quantify the skill of the spatial SPT product and highlights several areas for improvement, which are being implemented by the UK Met Office.

Contents

List of Figures

List of Tables

Abbreviations

B bias

CAO Chilbolton Atmospheric Observatory

CCN cloud condensation nuclei

CSI critical success index

DiVeN Disdrometer Verification Network

DSD drop size distribution

DVD drop velocity distribution

ESWD European Severe Weather Database

F false alarm rate

FAR false alarm ratio

HK Hanssen-Kuipers discriminant

HSS Heidke Skill Score

INP ice nucelating particle

METAR Meteorological Aerodrome Report

NWP numerical weather prediction

POD probability of detection

PSS Peirce Skill Score

PW present weather

PWA present weather arbiter

QPE quantitative precipitation estimation

SPT surface precipitation type

TS threat score

TSS true skill statistic

UK United Kingdom

WMO World Meteorological Organization

Chapter 1.

Introduction

1. Motivation

Meteorological conditions influence almost all aspects of daily life. Singleton (1989) states that: *"there are few activities upon which meteorology does not have a bearing or the execution of which cannot benefit in some way by meteorological advice."* The goal of meteorology is to increase our scientific understanding, improve our ability to observe the current state of the atmosphere, and subsequently to improve forecasts of a future state of the atmosphere. By improving atmospheric observations and forecasts, the informedness of societal decision-making is also improved and therefore socioeconomic value is realised. The atmosphere is a fluid and behaviour within it is complex; patterns and features exist across broad spatiotemporal scales simultaneously, from seconds to weeks and from sub-metre to planetary, making numerical weather prediction (NWP) inherently difficult without accurate and timely observations. The skill in NWP models for slowly-evolving, broad-scale features such as the 500 hPa geopotential height has steadily increased since the 1980's to greater than 95% correlation for a 3-day lead time (Bauer et al. 2015). However, small-scale and unstable atmospheric processes such as precipitation are less well predicted. Even with the advent of more accurate, high resolution precipitation observations and data assimilation techniques, the advection of radar-based products generally has more skill than NWP in the first 6 hours (Lin et al. 2005). Ensemble techniques (running multiple NWP models with equally likely perturbations in the initial state) have aided the predictability and informedness of precipitation intensity forecasts (Epstein 1969; Tracton and Kalnay 1993; Bauer et al. 2015), yet the variable of precipitation type (rain, snow, mixed-phase, freezing rain, ice pellets, graupel, hail etc.) reaching the surface remains poorly resolved by NWP (Wandishin et al. 2005; Ikeda et al. 2013; Elmore et al. 2015; Gascón et al. 2018). In part, this is because the type of precipitation reaching the surface is not widely or accurately observed (Schmid and Mathis 2004; Bloemink and Lanzinger 2005; Chen et al. 2016; Reeves 2016; Boudala et al. 2017) which makes improvements to NWP difficult. Furthermore, there is a lack of statistically rigorous verification techniques for surface precipitation type (SPT) data. The lack of rigor is due to the subjective nature of many decisions in the assessment process—a problem which this thesis aims to address with a technology-led approach for SPT observations and products in the United Kingdom (UK).

Knowledge of the SPT can often have greater socioeconomic impact than knowledge of the amount and the precise timing of the precipitation. Broadly speaking, the majority of precipitation in the UK falls as rain, so there is a lack of preparedness for less common precipitation types (snow, mixed-phase, ice pellets, freezing rain, graupel and hail) and so the impact is increased.

The safety of vehicular transportation is dependent upon the ability of the vehicle to have sufficient grip on the road surface, and the visibility of the driver, both of which are lowered in any type of precipitation (Thornes 1992). However, certain precipitation types are more detrimental than others. Mixed-phase precipitation occurs at surface temperatures between 0–4 °C (Langleben 1954; Lumb 1961, 1963) which can be more dangerous than snow. The temperature at which tyres have the least amount of friction on ice is at $T = 0$ °C (Moore 1975), a temperature frequently experienced over large areas in the UK winter season (Parker et al. 1992). Both snow and mixed-phase precipitation can substantially reduce visibility which is critical to aviation and ground transport. Agnew and Thornes (1995) demonstrate an impact of snow to agriculture and retail: while cold temperatures have been shown to increase the demand for root vegetables, the presence of lying snow inhibits the harvesting of those crops. In this scenario, the demand increases at the same time that the supply is diminished. Skilful predictions of SPT would allow farmers to confidently over-harvest beforehand, in anticipation of increased demand. Notably, mixed-phase precipitation serves as the boundary between the accumulation of snow locally and the runoff of melting precipitation into bodies of water, which has importance for flood forecasting and hydrology (Berne and Krajewski 2013). For these reasons, being able to distinguish between the mixed-phase and snow SPTs is important. However, the processes controlling the mixed-phase to snow boundary are complex, depend on several atmospheric feedbacks, and require suitably dense spatial SPT observations or high forecast resolution and accuracy to be well-resolved. The rain–freezing rain–snow transition region has been identified as one of five key research areas for winter quantitative precipitation forecasts by the U.S. Weather Research Program (Ralph et al. 2005).

Freezing rain is one of the most impactful precipitation types; supercooled liquid drops which freeze upon contact with the ground and form a glaze of ice over objects. A sufficient mass of the freezing rain glaze can cause trees and structures to collapse. During the event later shown in Figure 1.12, 4.5 cm of freezing rain accumulated in Champaign, Illinois, U.S., causing more than $12 million in unadjusted damages (Rauber et al. 1994). Freezing rain and ice pellets occur less than 0.2% of the time in the UK (Carrière et al. 2000) but a recent and notable freezing rain event in early March 2018 impacted southwestern parts of the UK in Cornwall and Devon (Jameson 2018). Leading up to the event, the highest Met Office warning (red) was issued for a part of southwestern England and southern Wales. Widespread freezing rain occurred leading to main road closures, and some passengers were stuck in their vehicles for over 12 hours as ice glaze caused numerous accidents. Rail networks were also disrupted, with one train being stranded for 11 hours with passengers on-board. Across the UK, 10 weather-related deaths were reported.

The graupel and hail SPTs are only associated with convective precipitation. Graupel signifies the electrification of a storm (Carey and Rutledge 1996; Deierling et al. 2008; Courtier et al. 2019) but has little direct impact upon precipitating, unless in large quantities. Hail (defined as having a diameter greater than 5 mm (AMS 2012b)) transfers a significant amount of kinetic energy upon impact (Hohl et al. 2002), so it is important to distinguish between graupel and hail. In the United States, 60% of property losses from convective weather are attributed to hail (Guntur and Tippett 2017), raising the overall price of insurance premiums (Brown et al. 2015). Even individual hail-producing cells have resulted in damages on the order of 1 billion USD (Kaspar et al. 2009; Punge et al. 2014; Brown et al. 2015; Schemm et al. 2016). Although hail is rare in the UK with fewer than 3 events per 1000 km^2 per 100 years (Punge and Kunz 2016; Webb et al. 2009), costly events do occur (Webb et al. 2001; Clark 2011; Clark et al. 2018). Hail is particularly difficult to forecast with NWP models because it precipitates in narrow swaths of a few kilometres (Frisby 1961; Webb et al. 2001) which results in observation-based nowcasting being the most skillful warning system. Radar-based products such as the Met Office spatial SPT product (see Chapter 5) cover a large spatial area at high spatial and temporal resolution which is ideal for hail detection, if the diagnosis is accurate and skilful.

The need for improved SPT forecasts is evident from their socioeconomic impacts, but to improve the skill of SPT forecasts we must first examine the skill of our observations of SPT. Observations of SPT aid NWP forecasts of SPT in three ways. Firstly, by improving our understanding of the atmosphere, the processes which lead to different SPTs can be more accurately represented in NWP models. Secondly, if accurate enough, observations could be ingested by NWP directly through data assimilation schemes to reduce the NWP model initialisation errors. Finally, and most relevant to this thesis, when the skill of NWP forecasts of SPT are evaluated, observations are used as the truth—however, this is not always a good assumption. In many cases the diagnosis of SPT can be subjective. Many fundamental biases and artefacts can exist with automated ground observations, which differ between instruments (see Chapter 3). The maximum theoretical predictive skill of SPT from NWP is effectively limited by the maximum skill in our ability to observe SPT.

Our ability to observe the different SPTs is poor, especially for rarer types and at air temperatures of around $T \approx 0$ °C (Schmid and Mathis 2004; Bloemink and Lanzinger 2005; Chen et al. 2016; Boudala et al. 2017). Verification of SPT skill is challenging because the observations are discrete (assigned a category rather than being a continuous scale) and are typically non-probabilistic, which limits the number of applicable statistical techniques (see Section 4.4). Additionally, the standard format for recording SPT observations is both complex and ambiguous (see Section 3.1). The overarching goal of this thesis is to create a consistent verification framework that is applicable to observations of SPT. By doing so, the skill of existing instruments and products can be quantified and there will be sufficient evidence to motivate the operational implementation of the most accurate SPT diagnosis techniques. Furthermore, using the verification framework, new SPT diagnosis techniques can be rapidly and rigorously assessed, which should aid their future development.

The introduction to this thesis covers three topics:

1. What are the microphysical causes and characteristics of different SPTs?
2. How are SPTs measured, both at single points and spatially?
3. How can the skill of SPT measurements and products be verified?

2. Precipitation Types and Their Origins

This section defines a framework of precipitation types justified by the literature, details the microphysical conditions which lead to the nucleation and growth of each hydrometeor that ultimately becomes the SPT, and highlights misnomers and transitional boundaries between the precipitation types.

The following definitions are used in this thesis:

> A **hydrometeor** is defined as any liquid or solid phase water-based particle suspended in the atmosphere (Rogers and Yau 1996; Straka 2009). A hydrometeor may be descending, ascending, and may be inside or outside of saturated regions of the atmosphere (i.e. clouds) but upon labelling, exists suspended within the fluid of the atmosphere and does not touch the ground.
>
> A hydrometeor is defined to have **'precipitated'** when it comes into contact with the ground. Therefore, a hydrometeor exiting a region of saturated air (a cloud) is not defined to have 'precipitated', since it may evaporate or sublimate before it reaches the ground.
>
> **Precipitation** is defined in this thesis as the collective term for a group of hydrometeors which have come into contact with the ground. The terms SPT and precipitation type are used interchangeably in this thesis but have the same meaning. Precipitation can be given a classification or a type, which is the dominant type of precipitation by flux through the surface. In cases where near-equal fractions of precipitation are made of multiple types, a combined classification is required, such as mixed-phase. This is where some ambiguity exists in the measurement and categorisation of different precipitation types.

Before any further discussion on the subject of SPT, a reference list of precipitation types must be defined, since some are distinct and others have gradual boundaries from one type to another. Furthermore, there is no rigorously agreed-upon table of precipitation types; different organisations, countries, and research papers define SPTs differently. present weather (PW) codes described in Tables 4677, 4678, 4680 (WMO 1988, 2017) and Meteorological Terminal Air Report (METAR) codes (U.S. NWS 2020) all have slightly different classifications and naming conventions for SPTs (see Section 3.1). In Chapter 3, the classification of precipitation types shown in Table 3.3 is introduced, which forms the framework used throughout this thesis. These SPTs can be grouped into the following categories where similar atmospheric microphysical processes are involved:

1. Drizzle and Rain

2. Ice Crystals and Snow

3. Mixed-phase
4. Ice Pellets and Freezing Rain
5. Graupel and Hail

The following subsections describe these groups of SPTs that can be produced in the atmosphere and also describe the microphysical processes which lead to each type.

2.1. Drizzle and Rain

There are two processes which can lead to drizzle and rain SPT occurring at the surface. These are referred to as the **warm rain** process and the **cold rain** process. The warm rain process is when the hydrometeors have been liquid for their entire lifecycle from nucleation to growth to precipitation (Lau and Wu 2003; Seinfeld and Pandis 2006). The cold rain process is where at some stage the hydrometeors have been solid in phase and have subsequently melted before reaching the surface. Liquid precipitation is the dominant SPT by mass in most regions, including the UK (Fairman et al. 2015). The following section first describes the liquid cloud droplet nucleation and growth process by diffusion, followed by a description of the behaviours and characteristics which apply both warm rain drops and to cold rain drops (after melting has occurred). These behaviours have implications for the measurement techniques of SPT, which are described in Section 3.

Water molecules in vapour phase cannot condense homogeneously together into liquid droplets under natural atmospheric vapour pressure conditions due to the free energy barrier, which must be overcome. Liquid cloud droplets instead form heterogeneously in the warm rain process, with water molecules condensing onto the surface of particles known as cloud condensation nuclei (CCN), which reduce the required free energy barrier because the nucleus is orders of magnitude larger than the water molecules. The reduced curvature of larger CCN means that the required bonding energy is also reduced, such that condensed molecules are less likely to evaporate back into vapour than compared with homogeneous nucleation (Thomson 1870). The chemical composition of the CCN can also reduce the required vapour pressure for equilibrium to be reached (Raoult 1887; Seinfeld and Pandis 2006). Subsequently, equilibrium can be achieved at the much lower vapour pressures observed in Earth's atmosphere (Rogers and Yau 1996). After initial nucleation (known as nucleus activation), vapour continues to condense onto, or evaporate from, the surface of the CCN dependent on the atmospheric vapour pressure. Water molecules diffuse through the air to conserve the local atmospheric vapour pressure (Brown 1828; Einstein 1905; Montgomery 1947) and heat produced by condensation of vapour onto the droplet conducts into the surrounding atmosphere (Maxwell 1890; Mason 1971), keeping the air parcel buoyant. Cloud droplets can undergo this diffusional growth by condensation until their mass and therefore their terminal velocity, increases compared to smaller cloud droplets. The differential velocity of cloud droplets leads to collisions. At droplet radii of around 20–30 μm, the dominant growth process changes from diffusional condensation to collision coalescence (Rogers and Yau 1996). The efficiency of coalescence is maximised when the ratio between the radii of two colliding droplets is approximately 0.6 (Klett and Davis 1973; Lin and Lee 1975; Schlamp et al. 1976). At droplet ratios of less than 0.6 in laminar flow (neglecting atmospheric turbulence), the small drops have insufficient inertia and are swept around the airflow of the larger droplet; at similar radii, droplets have similar terminal velocities and a collision is thus less likely to occur. Collision and coalescence is a process which also occurs with liquid droplets formed through the cold rain process (solid phase nucleation followed by melting, see section 2.2), after the melting has occurred. The following descriptions of liquid droplet characteristics apply to droplets which have formed in either the warm or the cold rain process.

The terminal velocity of a liquid droplet increases and is strongly dependent on the increasing size of the droplet. Gunn and Kinzer (1949) built upon the work of Lenard (1904); Schmidt (1909); Laws (1941) by reducing the experimental uncertainty in the measurement of the terminal fall velocity of liquid droplets by their diameter. Their result is referred to as the Gunn-Kinzer curve, which is now the most commonly applied raindrop diameter–velocity relationship. Gunn and Kinzer (1949) also noted that once the drops grew larger than 90 μm, they did not obey Stokes' Law (Stokes 1851) due to aerodynamic drag. Table 1.1 shows a subset of the Gunn-Kinzer curve data, which was obtained at 1013 hPa, 20 °C. Note from Table 1.1 that the rate of the increase in terminal velocity with increasing diameter, decreases as the drop size increases. The Gunn-Kinzer curve can be applied to measurements of the precipitation diameter and velocity distribution at the surface, which is one method for identifying SPT (See section 3.1).

Droplets cannot grow in size indefinitely. As the diameter and terminal velocity increases, the aerodynamic flow around the droplet distorts the shape of the droplet from a nominal sphere into a horizontally oblate spheroid as shown in Figure 1.1. The flow of the air around the droplet can be described by the Reynolds number (Re; Stokes 1851; Reynolds 1883; Sommerfeld 1908). When droplets are small, $Re \ll 1$ and the

Table 1.1.: The terminal velocity of liquid droplets of distilled water from the laboratory experiments of Gunn and Kinzer (1949), obtained at 1013 hPa, 20 °C. Droplets larger than 5.8 mm diameter were unstable and experienced breakup. Droplets smaller than 90 μm (not shown) obeyed Stokes' Law (Stokes 1851). The number of repeat measurements for each diameter class was ≥ 50. Also included for each droplet diameter class is the droplet mass and the calculated Reynolds number.

Equivalent diameter (mm)	Terminal velocity (m s^{-1})	Mass of drop (mg)	Reynolds number (calculated)
0.2	0.72	0.004	9.6
0.4	1.62	0.034	43
0.6	2.47	0.113	99
0.8	3.27	0.268	175
1.0	4.03	0.524	269
1.4	5.17	1.44	483
1.8	6.09	3.05	731
2.2	6.90	5.58	1,013
2.6	7.57	9.20	1,313
3.0	8.06	14.1	1,613
3.4	8.44	20.6	1,915
3.8	8.72	28.7	2,211
4.2	8.92	38.8	2,500
4.6	9.03	51.0	2,772
5.0	9.09	65.5	3,033
5.4	9.14	82.4	3,293
5.8	9.17	102	3,549

flow is laminar, such that the air detaches around the droplet and merges again with no distortion. As the droplet increases in size and terminal velocity, the Reynolds number increases. At $Re \approx 10$, the airflow detaches from the rearward side of the droplet, and at $Re \approx 10^4$ the drop wake is turbulent. When the droplet is large enough it falls in an unstable regime and oscillations can overcome surface tension, resulting in the droplet breaking apart. The droplet can experience breakup in three categorised ways: filament/neck, sheet, or disk (McTaggart-Cowan and List 1975; Low and List 1982). Figure 1.2 shows a thin, horizontally oblate droplet which experiences disk breakup, where the surface tension in the centre of the original droplet is overcome. Aerodynamic processes effectively limit the size to which droplets can grow which, for Earth's atmosphere at sea-level pressure, has been shown to be a limit of approximately 6 mm diameter (Marshall and Palmer 1948; Gunn and Kinzer 1949; Villermaux and Bossa

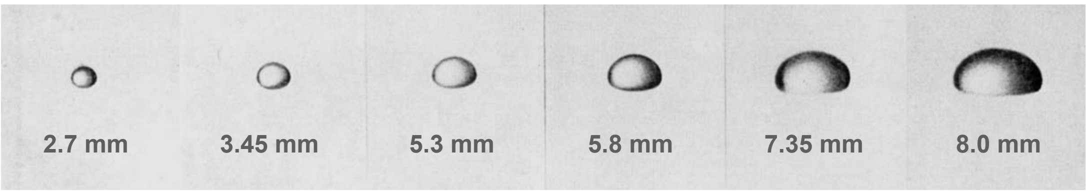


Figure 1.1.: Figure 5 from Pruppacher and Beard (1970) showing the shape of droplets falling at terminal velocity perpendicular to the imaging device. As the droplets grow in size, their base is flattened. The vertical wind tunnel was at a pressure of 1,007 ± 5 hPa but turbulence was minimised which allowed drops up to 9 mm in diameter to exist for a short time—larger than those seen in nature due to turbulence.

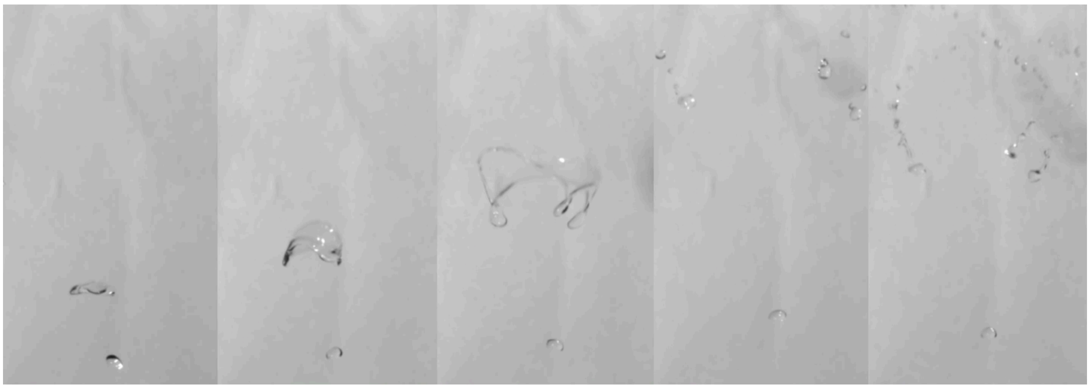


Figure 1.2.: A timeseries of five high-speed video frames showing a liquid hydrometeor droplet at terminal velocity undergoing disk breakup due to aerodynamic drag. No size scale is provided. ©Dr Christopher Emersic, University of Manchester, 2010.

2009). These results are useful for the verification of instruments which measure the diameter of precipitation particles in order to infer their SPT. The maximum size of measured raindrops is used as a benchmark in Chapter 3 where particles larger than 6 mm are noted as unrealistic.

Inevitably, a range of drop sizes exist in the rainfall which reaches the Earth's surface. The drop size distribution (DSD) is defined as a distribution of the number of drops as a function of diameter per unit volume (Jameson and Kostinski 2001):

$$DSD = \int_{0}^{\infty} N(D) dD \tag{1.1}$$

where N is the number of drops as a function of the drop diameter, D. The DSD is an important measure of precipitation to consider since it is a direct consequence of the processes which influence precipitation over its lifetime (Testud et al. 2001; Das et al. 2017; Lin et al. 2020, and many more). Note that size distributions also occur for other

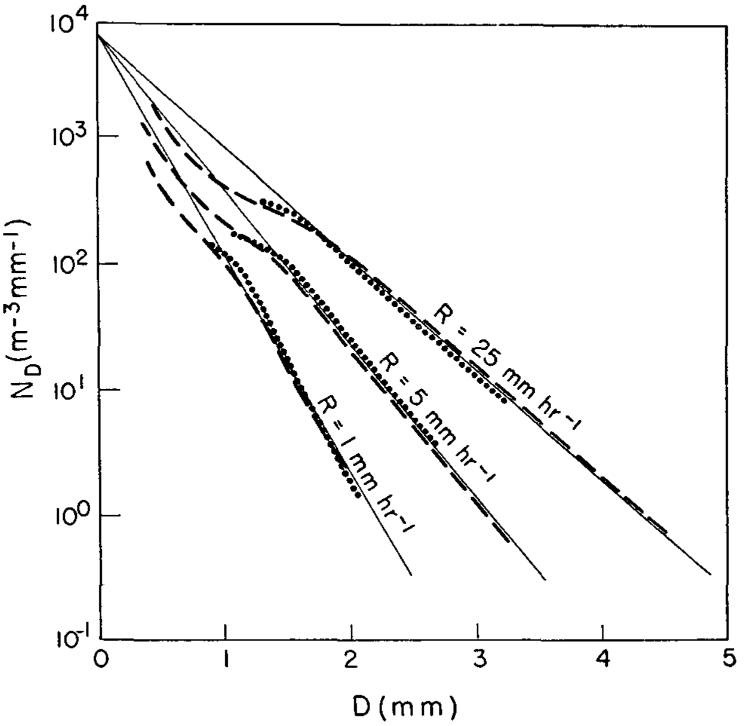


Figure 1.3.: Figure 2 from Marshall and Palmer (1948). The drop size distribution as a function of drop diameter per unit volume. Dashed lines are observational data from Washington D.C., U.S.A. Laws and Parsons (1943), dotted lines are observational data from Ottawa, Canada Marshall et al. (1947), and solid lines denote the distribution functions of Marshall and Palmer (1948).

SPTs but that DSD specifically refers to liquid precipitation of drizzle or rain. Drizzle drops are $D < 0.5$ mm in diameter and rain drops are $D \geq 0.5$ mm (Baker and Friday 1996; Met Office 2014). Marshall and Palmer (1948) identified a relationship between the precipitation rate and the DSD, shown in Figure 1.3. For higher precipitation rates, the negative gradient of the drop size distribution on a logarithmic plot, decreases, whilst the zero intercept appears approximately constant. This discovery led to the ability to use radars for quantitative precipitation estimation (QPE) over large spatial domains and contributes to the estimation of spatial SPT diagnosis in a UK Met Office algorithm which will be described in Section 3.2.

Since a distribution of drop diameters exists in liquid precipitation, a drop velocity distribution also exists. The drop velocity distribution (henceforth DVD) is defined as:

$$DVD = \int_0^\infty N(V)dV \tag{1.2}$$

where N is the number of drops as a function of the drop velocity, V. The combination of drop size and velocity is important generally because the total kinetic energy of precipitation can then be measured. As mentioned in Chapter 3, the kinetic energy of precipitation controls soil, building and aerofoil (aircraft and wind turbine) erosion, as well as downbursts (Kinnell 1981; Rosewell 1986; Feingold et al. 1991; Tang and Davidson 2004; Erkal et al. 2012; Keegan et al. 2013; Slot et al. 2015; Eisenberg et al. 2018). Of greater importance to the wider objectives of this thesis are the differences between the fall velocities of different SPTs. These SPT velocity differences arise due to their density and aerodynamic drag differences, and can be measured by ground-based instruments to infer the SPT (see Section 3.1). The diameter-velocity relationship for drizzle and rain is described by the aforementioned Gunn-Kinzer curve (Gunn and Kinzer 1949) but other relationships exist for other SPTs (Langleben 1954; Locatelli and Hobbs 1974; Bohm 1989, and more) and will be described in the following sections.

2.2. Ice Crystals and Snow

The formation of ice crystals in clouds is arguably more complex than the formation of liquid droplets, and the exact processes of ice nucleation are not yet fully understood (Rogers and Yau 1996; David et al. 2019). On a broad scale the same growth by diffusion of water vapour occurs as for the liquid droplet nucleation described in Section 2.1. Homogeneous hydrometeors can exist in liquid phase at temperatures approaching -38 °C (Kanji et al. 2017) before the enthalpy of fusion or latent heat energy of freezing is overcome. The presence of various aerosols and CCN within heterogeneous liquid drops (heterogeneous being the typical droplet state, as discussed in Section 2.1) reduces the latent heat energy of fusion. Depending on the chemical composition of the droplet solute, freezing of nucleated liquid particles can occur at temperatures as high as -4 °C (Houghton 1985). The natural variability of hydrometeor chemistry means that liquid hydrometeors can exist between $-38\ ^{\circ}\mathrm{C} < T < 0\ ^{\circ}\mathrm{C}$ but occur less frequently with decreasing temperature (Houghton 1985; Rogers and Yau 1996; Whale et al. 2015). Figure 1.4 summarises 4 processes in which an ice particle can form. Firstly, water vapour can directly undergo deposition onto an ice nucelating particle (INP), known as heterogeneous deposition. The remaining 3 processes involve first the nucleation by condensation of a liquid droplet and then freezing nucleation. An INP can catalyse both nucleations; first acting as a CCN as described in Section 2.1 and then when the latent heat of fusion can be overcome, the particle can also act as a freezing nuclei, reducing the free energy barrier in both instances. Alternatively, an INP can encounter an existing CCN-nucleated liquid droplet and either freeze the droplet immediately

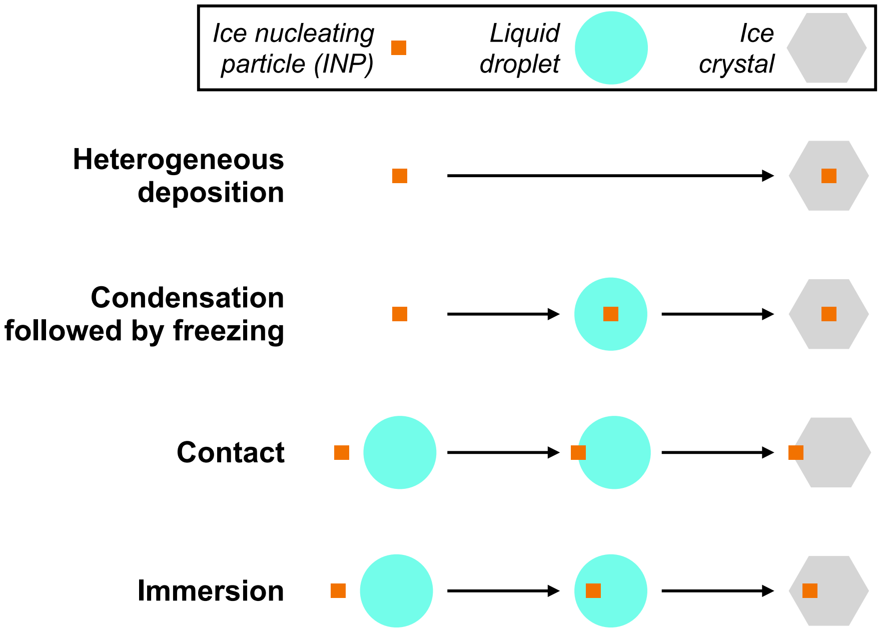


Figure 1.4.: Redrawn version of Figure 9.1 from Rogers and Yau (1996). Schematic illustration of the different processes in which atmospheric INPs can result in ice formation, through deposition or freezing.

upon contact or some time thereafter.

Saturation vapour pressure is lower over ice than over water, making nucleation more probable than liquid droplet nucleation at lower values of atmospheric vapour pressure, shown in Figure 1.5 (Murphy and Koop 2005). However, due to the sparsity of INPs in the atmosphere compared to the density of condensation nuclei, ice nucleation is typically less probable than liquid droplet nucleation when both effects are combined (Seinfeld and Pandis 2006). On the contrary, ice crystal concentrations observed in the atmosphere are sometimes higher than the concentrations of INPs. Therefore a secondary production process must be occurring. Several theories exist for the precise mechanism of secondary ice production which include nucleated ice particles emitting fragments of ice, which can then themselves serve as INPs (Hallett and Mossop 1974; Field et al. 2016).

Once the initial ice nucleation has occurred through any of the aforementioned processes, the growth rate of an ice particle by deposition can occur at any temperature less than 0 °C (Houghton 1985). Due to the vapour pressure difference between liquid and ice (Figure 1.5; Murphy and Koop 2005), deposition growth of ice can occur at a faster rate than condensational growth of droplets for the same atmospheric vapour

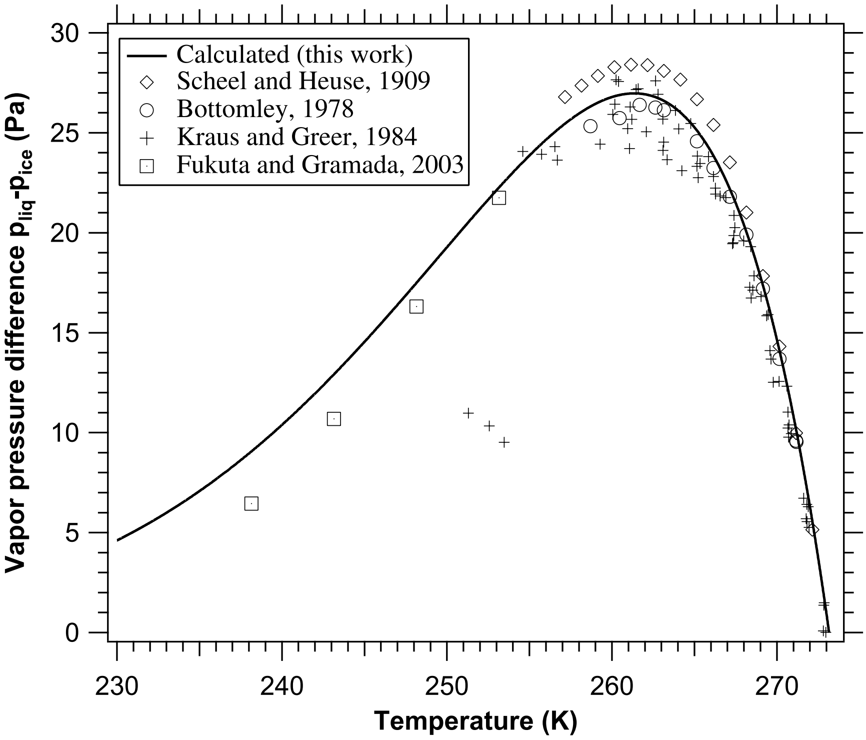


Figure 1.5.: Figure 9 from Murphy and Koop (2005). The difference between vapour pressure over water (p_{liq}) and over ice (p_{ice}) in Pa, for temperatures of -43 °C to 0 °C. Data shown are from this work, Scheel and Heuse (1909); Bottomley (1978); Kraus and Greer (1984); Fukuta and Gramada (2003).

pressure. In mixed-phase clouds without strong updrafts, ice particles can grow (deposition) at the expense (evaporation) of liquid droplets due to the difference in saturation vapour pressure shown in Figure 1.5 (Wegener 1911; Bergeron 1935; Findeisen 1938). The growth of an ice particle by deposition is both non-spherical and non-linear, and can occur either laterally across a plane (plate, dendrite) or principally, increasing in ‘depth’ (column). The wide range of ice crystal growth habits have been shown to depend on both the temperature of the air and the supersaturation vapour pressure, from laboratory experiments (Nakaya 1954; Hallett and Mason 1958; Kobayashi 1958, 1961). The broad habits are summarised with the Kobayashi (1958) habit diagram shown in Figure 1.6.

Whilst pristine ice crystal precipitation is not unheard of, the dominant form of solid SPT in the UK is snow, or aggregate hydrometeors, where pristine ice crystals collide and aggregate in a similar process to collision and coalescence (see Section 2.1). However, there are some complex, non-linear behaviours within the aggregation of ice

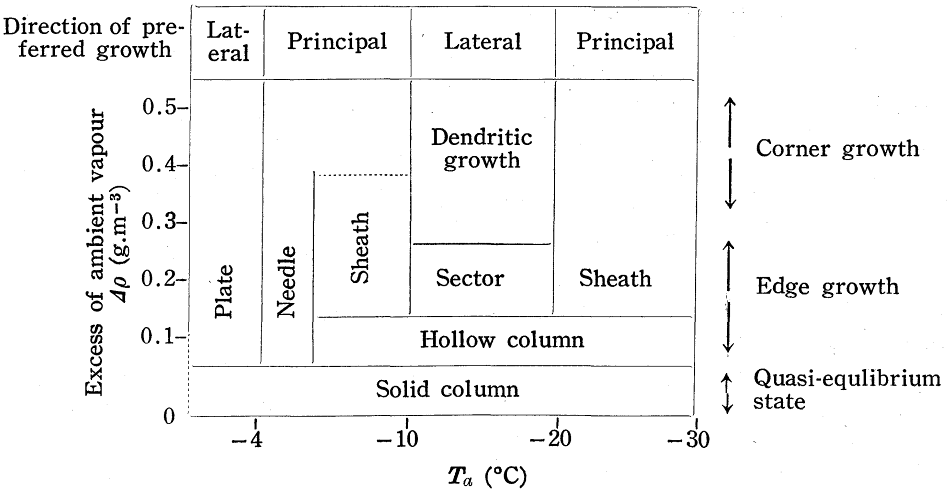


Figure 1.6.: From Kobayashi (1958). Ice crystal growth habits as a function of air temperature T_a and supersaturation vapour pressure over ice $\Delta\rho$. Sheath refers to columns with hollow cones starting from the nucleation point.

crystals. Hosler et al. (1957) found that temperature and vapour pressure were correlated with adhesion efficiency of ice spheres, which increased towards 0 °C. Secondly, it has been shown that the 'openness' of an ice crystal (the ratio of surface area to volume) increases the aggregation efficiency when contact occurs (Westbrook et al. 2004). Dendrites have the highest surface area to volume ratio and therefore have the highest aggregation efficiency. There is broad evidence that air temperatures of approximately −11 to −17 °C (consistent with the dendritic growth habit shown in Figure 1.6) support the highest aggregation efficiencies (Hosler and Hallgren 1960; Mitchell 1988; Connolly et al. 2012; Phillips et al. 2015). Finally, several studies have also shown a relationship between the strength or presence of an electric field in the atmosphere and aggregation efficiency (Latham and Saunders 1970; Saunders and Wahab 1975; Connolly et al. 2005; Phillips et al. 2015). Once two ice crystals have collided and adhered, the mass and terminal velocity of the aggregate is increased, and so the differential velocity between the aggregate and the pristine ice crystals increases the probability of more collisions.

The terminal velocity of ice crystals is generally less than that of liquid drops for an equivalent diameter (Nakaya and Terada 1935; Langleben 1954; Locatelli and Hobbs 1974; Bohm 1989). The density of ice is approximately 920 kg m^{-3} (dependent on temperature) which is less than the 1000 kg m^{-3} density of water (Bader 1964; Yen 1981), so the mass of an ice particle is less than a liquid drop of the same volume. In addition, ice crystals are typically non-spherical (dendrites, columns, needles, see Figure 1.6) which means their largest dimension is significantly larger than a liquid droplet

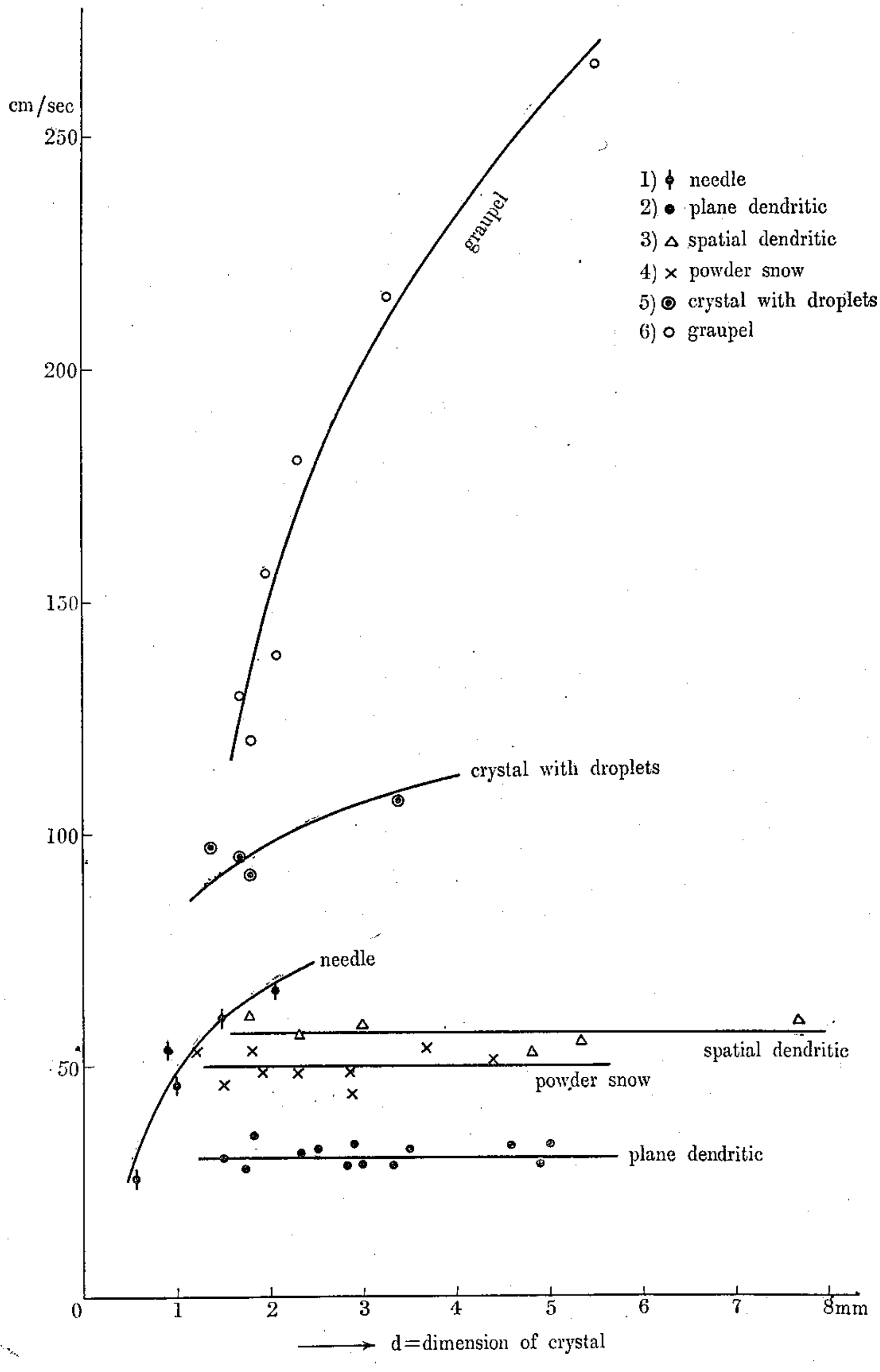


Figure 1.7.: Figure 3 from Nakaya and Terada (1935). The diameter (mm) and the fall velocity (cm s^{-1}) of different solid hydrometeor types. The diameter is referred to as the longest dimension of the particle, so for needles the diameter is the length of the needle.

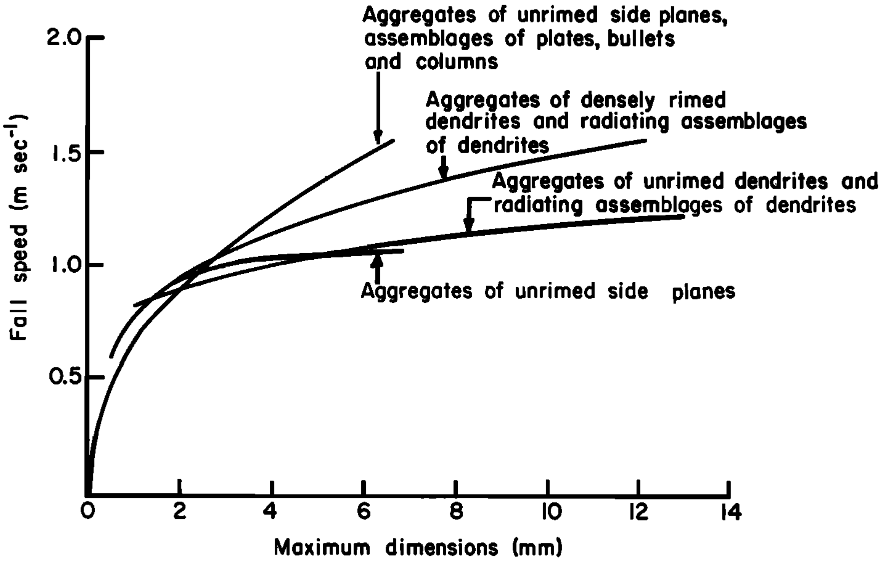


Figure 1.8.: Figure 24 from Locatelli and Hobbs (1974). The relationship between diameter (mm) and terminal velocity (m s^{-1}) for aggregates of different ice crystal habits.

of the same mass, increasing aerodynamic drag. Figure 1.7 shows the diameter-velocity relationships of several pristine ice crystal types and other solid hydrometeors measured by Nakaya and Terada (1935). Several later studies produced diameter velocity relationships for pristine ice using larger samples and with different habit classifications (Magono 1953; Langleben 1954; Locatelli and Hobbs 1974; Bohm 1989). Aggregates have a lower overall density due to the air trapped between the interlaced ice crystals. However, the larger mass and the rounded shape of aggregates results in a greater terminal fall velocity compared to pristine ice crystals. Figure 1.8 from Locatelli and Hobbs (1974) shows some experimental data for the diameter-velocity relationship for aggregates of varying ice crystal habits. Bohm (1989) produced similar results.

Throughout this thesis the term ‘aggregate’ refers to the hydrometeor as it exists in the atmosphere, and the term ‘snow’ refers to aggregates which precipitate to the ground, making snow an SPT class. The snow SPT class may also contain a low concentration of pristine ice crystals, since aggregates can experience breakup from aerodynamic forces, but these are neglected for classification simplicity. The SPT class of ‘ice crystals’ refers to pristine ice crystals falling to the ground which have not experienced aggregation, a rare phenomenon in the UK due to the comparatively slow terminal fall velocity of individual crystals and their tendency to aggregate at temperatures warmer than −20 °C. The ice crystals SPT class refers to all of the habits shown in Figure 1.6. While the great variations in crystal habits have implications for studies of cloud microphysics, the habits have negligible societal impact when precipitating and are therefore not considered in the SPT classification used throughout this thesis.

2.3. Mixed-phase

The term mixed-phase in this thesis refers to pristine ice crystals and aggregated ice crystal hydrometeors which are partially liquid upon precipitating to the surface due to the air temperature being greater than 0 °C. Throughout this thesis they are also referred to as 'wet' ice crystals and 'wet' snow individually. Graupel and hail do typically coexist with liquid precipitation but are distinct and are not included within the 'mixed-phase' SPT moniker (see Section 2.5). Note from Section 2.2 that the efficiency of ice crystal aggregation increases with increasing temperature from approximately -20 °C towards 0 °C. Hence, mixed-phase precipitation usually consists of melting aggregates, except for very low precipitation rates where the collisions between hydrometeors are improbable. The melting of ice crystals and aggregates is not an instantaneous process. Melting layers can vary in the time taken to melt and therefore the vertical depth in the atmosphere. Figure 1.9 shows results from Tabary et al. (2007) where the depth of the melting layer (inferred from the radar bright band depth, see Figure 1.21 in Section 3.2) was shown to vary between 20–1,000 metres over a 12-month period of non-convective precipitation cases in France. The most common melting layer depth from Tabary et al. (2007) was 200–400 m which accounted for approximately 55% of the cases, whereas the 20 m and 1 km deep melting layers accounted for 7% and 15% of cases respectively. Fabry and Zawadzki (1995) showed that the melting layer thickness varied from 250 m to 750 m with precipitation rates from $PR < 1.0$ increasing to $PR \approx 11.5$ mm h^{-1} in a 4-year study conducted near Montreal, Canada.

There are several factors which can control the melting layer depth, including the environmental lapse rate, the environmental saturation ratio, the hydrometeor size distribution and the hydrometeor mass concentration. A greater increase of temperature with decreasing height (the environmental lapse rate) results in faster melting as more sensible heat energy is conducted into the hydrometeors. Note that this process acts to cool the surrounding air parcel. If the environmental vapour pressure is below saturation, the melting particles will also lose latent heat energy through the evaporation of any liquid, until the surrounding air reaches saturation. Note that an unsaturated air parcel has the ability to cool further than a saturated air parcel, due to the additional heat lost through evaporation. Lumb (1963) derived a method to quantify the evaporative cooling effect, which is applied to a radar-based spatial SPT product described

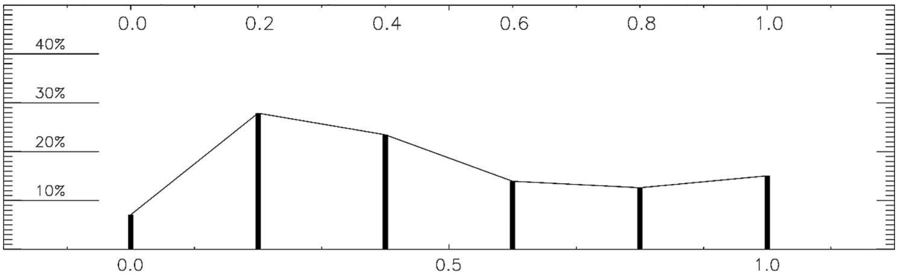


Figure 1.9.: Figure 3b from Tabary et al. (2007). The climatology of melting layer depths derived from eight scanning radars in France. The y-axis is frequency of occurrence (%) and the x-axis is the bright band thickness (km). The bright band thickness of 0 km actually refers to the lowest resolution increment of the radar processing method which was 20 m thick. Only stratiform precipitation over 452 days of independent observations from eight radars in the years 2003 and 2004 were considered.

in Section 3.2 and is verified in Chapter 4 of this thesis.

The prior aggregate size distribution is also important for the mixed-phase SPT. Compared to smaller diameter particles, larger diameter aggregates have a second-power larger surface area (increasing sensible heat flux) but a third-power larger volume and mass (increasing the total required energy to melt). Therefore the larger aggregates take more time to melt and fall further into the $T > 0$ °C layer than smaller aggregates or ice crystals. Furthermore, larger aggregates have a faster terminal fall velocity and thus reach further into the $T > 0$ °C layer for a given amount of time when compared with smaller aggregates. Lumb (1963) determined with the equations of Wexler (1955) that mixed-phase SPT could penetrate to maximum temperatures of 2.5 °C and 3.5 °C for stratiform and convective precipitation respectively, while Lumb (1961) observed mixed-phase SPT at a temperature of 4.2 °C. Finally, the mass concentration of hydrometeors per unit volume matters; a larger total mass demands more sensible heat energy from the surrounding atmosphere and can therefore reduce the surrounding atmospheric temperature. On the contrary, a high mass concentration can cause subsidence which adiabatically warms the air parcel. This effect is typically counteracted by the aforementioned processes such that an increase to the amount of precipitation typically reduces the height of the 0 °C isotherm. The subsidence of the 0 °C isotherm is visible with high resolution precipitation radar (see Section 3.2) and is referred to as bright band sagging.

For mixed-phase precipitation, the terminal fall velocity of the hydrometeors depends on both the size and the degree to which melting has occurred. Wexler (1955) determined with the velocity measurements of Langleben (1954) that melting aggregates

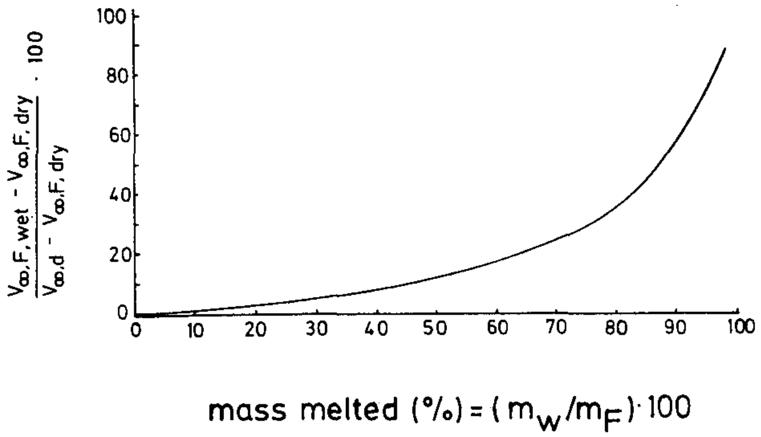


Figure 1.10.: Figure 3 from Mitra et al. (1990). The relationship between percentage of velocity change and percentage of mass melted for 5 and 10 mm aggregates in a vertical wind tunnel which was warmed at realistic rates to those experienced by hydrometeors in the atmosphere. A total of 45 aggregates were measured.

retain their shape until 90% of the hydrometeor has melted. However, as aggregates melt, their density (diameter) increases (decreases), and thus their aerodynamic drag decreases for an overall increase in their terminal velocity. Zenith-pointing Doppler radars are useful in order to determine the vertical fall velocities of melting hydrometeors, which are difficult to simulate in laboratories. Fabry and Zawadzki (1995) show a change in vertical Doppler velocity from between 0-3 m s^{-1} before melting occurred to between 4–9 m s^{-1} after melting occurred, in a 4-year study near Montreal, Canada. Similarly, Lin et al. (2020) show a change in vertical Doppler velocity from between 0.5–1.5 m s^{-1} before melting occurred to between 3.5–6.5 m s^{-1} after melting occurred, in a 12-month study at Chilbolton Atmospheric Observatory (CAO), UK. Over the whole atmospheric column the measured velocity change has a non-linear, sharp S-curve characteristic (Fabry and Zawadzki 1995; Lin et al. 2020). However, as shown in Section 2.2, aggregation increases the terminal velocity of hydrometeors at temperatures less than 0 °C before melting occurs. The transition from aggregates to liquid drops is a continuous process rather than a discrete change, which complicates the use of diameter-velocity relationships for mixed-phase SPT diagnosis (see Section 3.1).

Mitra et al. (1990) used a vertical wind tunnel to measure aggregates of 5 and 10 mm diameter and found that the increase in fall velocity change as a function of mass melted was exponential (Figure 1.10), and that the velocity of the aggregate rapidly increased once 70% of the hydrometeor mass was liquid. Yuter et al. (2006) showed

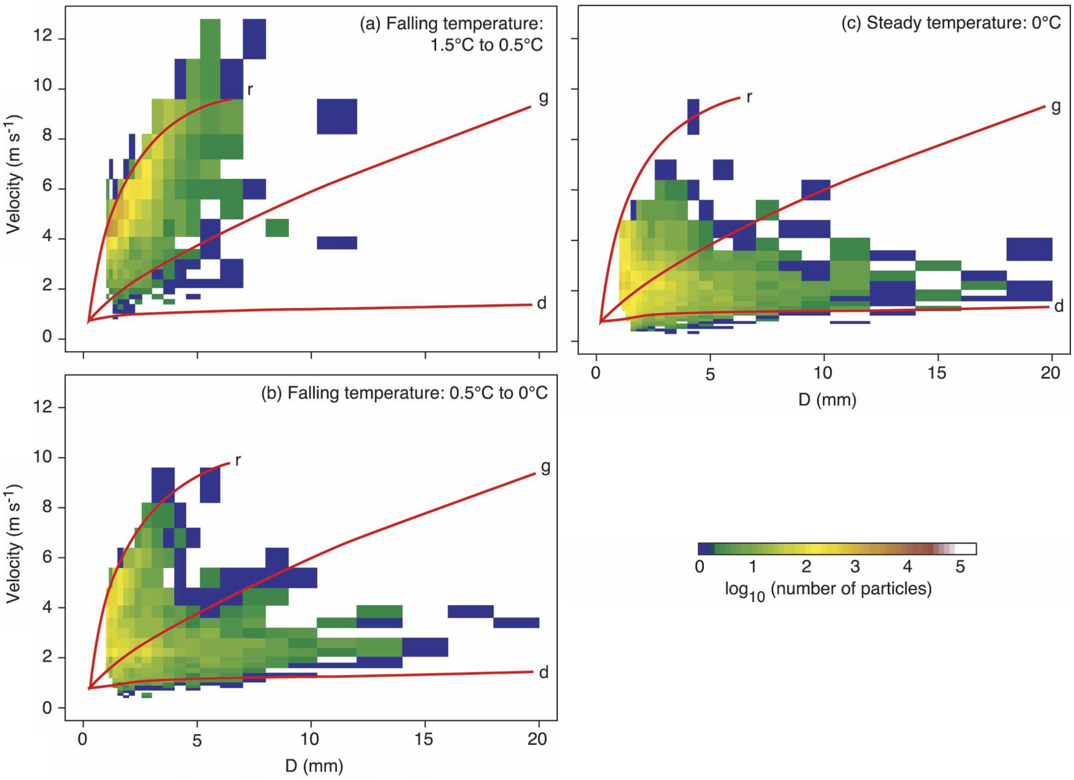


Figure 1.11.: Figure 3 from Yuter et al. (2006). Diameter (mm) and velocity (m s^{-1}) relationships for aggregates in three temperature regimes: a) falling temperature from 1.5–0.5 °C; b) falling temperature from 0.5–0 °C; c) steady temperature at 0 °C. The curves represent the empirical diameter–velocity distributions from Locatelli and Hobbs (1974) for r: rain, g: graupel, d: dendrites.

with a laser occultation method (see Section 3.1) that diameter-velocity relationships for mixed-phase precipitation are broad due to the degree of melting to which each hydrometeor has experienced (Figure 1.11). 5% of the total particle count (1% of total volume) are small liquid drops at 0 °C, which increases to 23% (4% volume) at 0–0.5 °C and 93% (74% volume) at 0.5–1.5 °C. The maximum size of the liquid drops also increased with increasing temperature in the study of Yuter et al. (2006). The lack of constraint in the diameter-velocity relationship for melting precipitation makes the observation of the mixed-phase SPT difficult for some ground based instruments (see Section 3.1).

When the hydrometeors have fully melted, they exhibit the same behaviours as the falling droplets described in Section 2.1, where collision coalescence and breakup control the overall drop size distribution. As previously mentioned, rain produced by solid hydrometeors melting before reaching the ground are referred to as cold rain.

2.4. Ice Pellets and Freezing Rain

Frozen or supercooled liquid water droplets are common within convective clouds but the atmospheric conditions which are conducive to precipitation reaching the surface in the form of ice pellets or freezing raindrops are rare, especially in the UK (Carrière et al. 2000). Synoptically speaking, the hydrometeors will typically exist as solid-phase crystals or aggregates first (see Section 2.2), fall into an environment with $T > 0$ °C to begin melting (see Section 2.3), and then fall into a region with $T < 0$ °C which should extend to the surface. Atmospheric temperatures usually decrease with increasing height, so it is uncommon for warmer air to exist between two colder layers (a temperature inversion). An example is shown in Figure 1.12 of three vertical temperature profiles recorded during an atmospheric inversion event. If the warm ($T > 0$ °C) layer is of a depth and temperature to allow only partial melting, then hydrometeors which fall into the cold ($T < 0$ °C) layer retain some mass of ice. The presence of ice acts as an INP at $T = 0$ °C (see Section 2.2; Houghton 1985; Seinfeld and Pandis 2006) which enables the hydrometeors to re-freeze into **ice pellets** with a shallow cold layer. If the warm ($T > 0$ °C) layer is of a depth and temperature to allow total melting, then the liquid hydrometeors which fall into the cold ($T < 0$ °C) layer lack the ice INPs which allow freezing at temperatures of $T = 0$ °C. They are also unlikely to contain the INPs which allow freezing at temperatures as high as -4 °C (see Section 2.2; Houghton 1985; Seinfeld and Pandis 2006). These hydrometeors therefore remain liquid and if they exist in the cold layer for enough time, the liquid drops can supercool into **freezing rain**. When a supercooled liquid hydrometeor falls to the ground, the impinged surface triggers ice nucleation and hydrometeors freeze immediately upon, or shortly after impact. The outcome is that, over time, surfaces and objects are covered in a dense, translucent glaze of ice, making freezing rain one of the most disruptive SPTs. If the lower altitude sub-zero temperature layer is sufficiently cold, the supercooled drops may also freeze again into ice pellets before reaching the ground.

Czys et al. (1996) derived a non-dimensional parameter based on the ratio of the time available for melting to the time required for total melting to occur. The required and available time durations were estimated based on microphysical models of melting physics (see Section 2.3) on the largest hydrometeors which are the last hydrometeors to melt. From Figure 1.13, several scenarios for the ice pellets to freezing rain boundary can be suggested: a warm layer depth of 2 km with an average temperature of approximately 0.8 °C; a warm layer depth of 1 km with an average temperature of approximately 1.7 °C; a warm layer depth of 500 m with an average temperature of approximately 3.2 °C. Czys et al. (1996) also noted that during the 1990 event shown

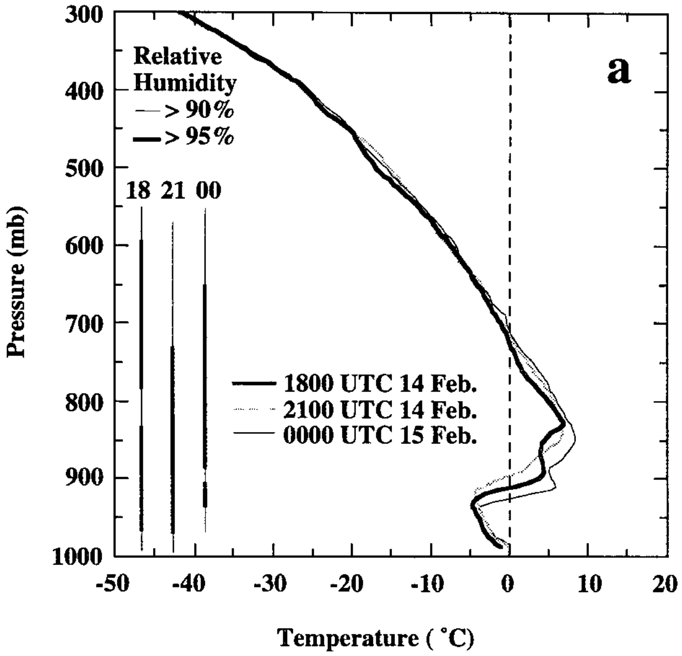


Figure 1.12.: Figure 8 from Rauber et al. (1994). Temperature and relative humidity measured during three rawinsonde ascents at 1800 and 2100 UTC 14 February and 0000 15 February 1990. The event caused more than \$12 million in damages in Champaign, Illinois (Rauber et al. 1994) and over \$120 million in property damages across 35 states (Czys et al. 1996).

in Figure 1.12, that freezing rain was observed in air temperatures only a few tenths of a degree below 0 °C.

Freezing rain exhibits the same fall velocity and diameter microphysics as with non-supercooled liquid rain, namely following the Gunn-Kinzer curve (Gunn and Kinzer 1949). The temperature or the glaze which occurs upon reaching the ground are the only discernible features which can be used to detect freezing rain with in-situ measurements (Section 2.1). For ice pellets, Nagumo and Fujiyoshi (2015) show that there are two regimes of ice pellets which they describe as fast and slow. Figure 1.14 shows the diameter-velocity distribution of ice pellets during a prolonged event in Japan. The slow-falling regime of ice pellets is closely related to that of hailstones (rough surface) of average empirical density (Knight and Heymsfield 1983), and to smooth ice spheres with densities of 160 kg m^{-3}. The fast-falling regime is close to the empirical diameter-velocity relationship for raindrops. The similarity of these two ice pellet regimes to other SPT diameter-velocity relationships makes their distinction difficult for instruments which discern SPTs using diameter and velocity measurements. The difficulty cannot

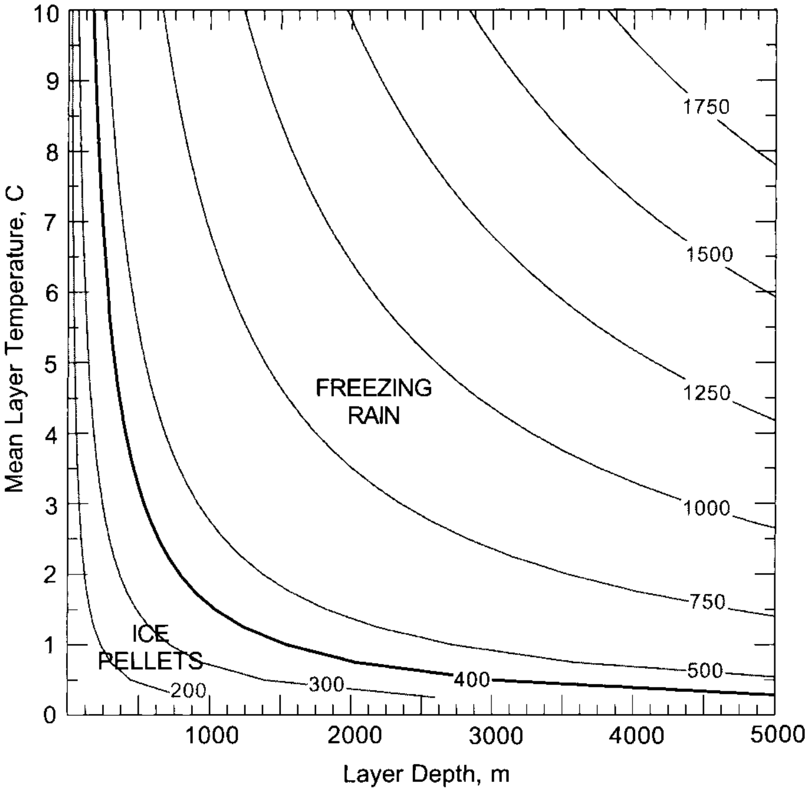


Figure 1.13.: Figure 8 from Czys et al. (1996). An isonomogram of boundaries between ice pellets and freezing rain based on the ratio of the time taken and the time available for the largest hydrometeors to melt. The curved isopleths denote the magnitude of the radius of the largest hydrometeor (μm), which is related to the precipitation rate. The highlighted isopleth is for a maximum hydrometeor radius of 400 μm (0.8 mm) which the study found to be most accurate in the 1990 event shown in Figure 1.12.

necessarily be overcome with temperature measurements. While ice pellets usually occur at surface temperatures below 0 °C the same as freezing rain, Nagumo and Fujiyoshi (2015) demonstrate that ice pellets can also occur at surface temperatures up to 2.7 °C with partial melting.

Carrière et al. (2000) examined PW code (see Section 3.1) reports between 1995–1998 across Europe and found that ice pellets or freezing rain made up between 0-0.2% of the UK reports. However, there was a recent notable ice pellet and freezing rain event in the UK in 2018 (Jameson 2018). While both ice pellets and freezing rain are relatively rare in the current climate of the UK (due to the mild maritime geography), they are included here for the completeness of the generalised SPT classification list. In the United States and other parts of the world, ice pellets are also referred to as sleet whereas in the UK, sleet is the nomenclature for mixed-phase (melting snow)

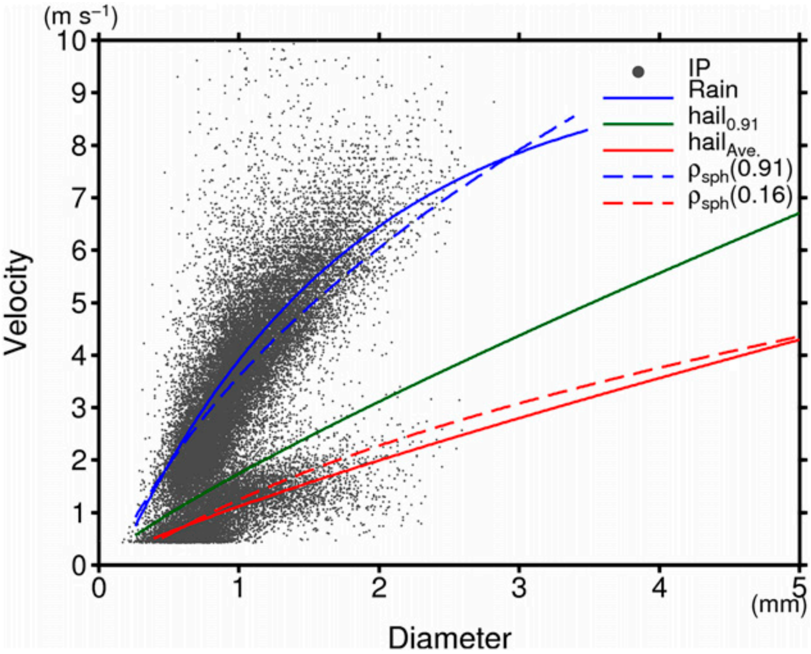


Figure 1.14.: Figure 9 from Nagumo and Fujiyoshi (2015). The relationship between diameter and velocity of particles during an ice pellet event, with drawn empirical or calculated relationships for different particles. The solid blue line is for empirical liquid raindrops, the dashed blue line is for calculated smooth ice spheres with a density of 910 kg m^{-3}, the solid green line is for calculated hailstones with a density of 910 kg m^{-3}, the dashed red line is for calculated smooth ice spheres with a density of 160 kg m^{-3}, the solid red line is for calculated hailstones with a density of 440 kg m^{-3} (Knight and Heymsfield 1983).

precipitation (WMO 1992; AMS 2012c; Met Office 2019a). Therefore the term 'sleet' is not used in this thesis to avoid confusion.

2.5. Graupel and Hail

In storm updraft environments, solid hydrometeors (pristine ice crystals, ice aggregates and ice pellets) can co-exist with supercooled liquid hydrometeors. When supercooled liquid drops collide with solid hydrometeors, the solid hydrometeor acts as an INP and drops will immediately freeze as the latent heat of fusion is overcome, in a process known as riming (Rogers and Yau 1996; Seinfeld and Pandis 2006). Where riming is the dominant form of hydrometeor growth, and the hydrometeor is less than 5 mm in diameter, the hydrometeor is referred to as **graupel** (AMS 2012a). If the riming continues and the hydrometeor grows to a diameter of $D \geq 5$ mm, it is referred to

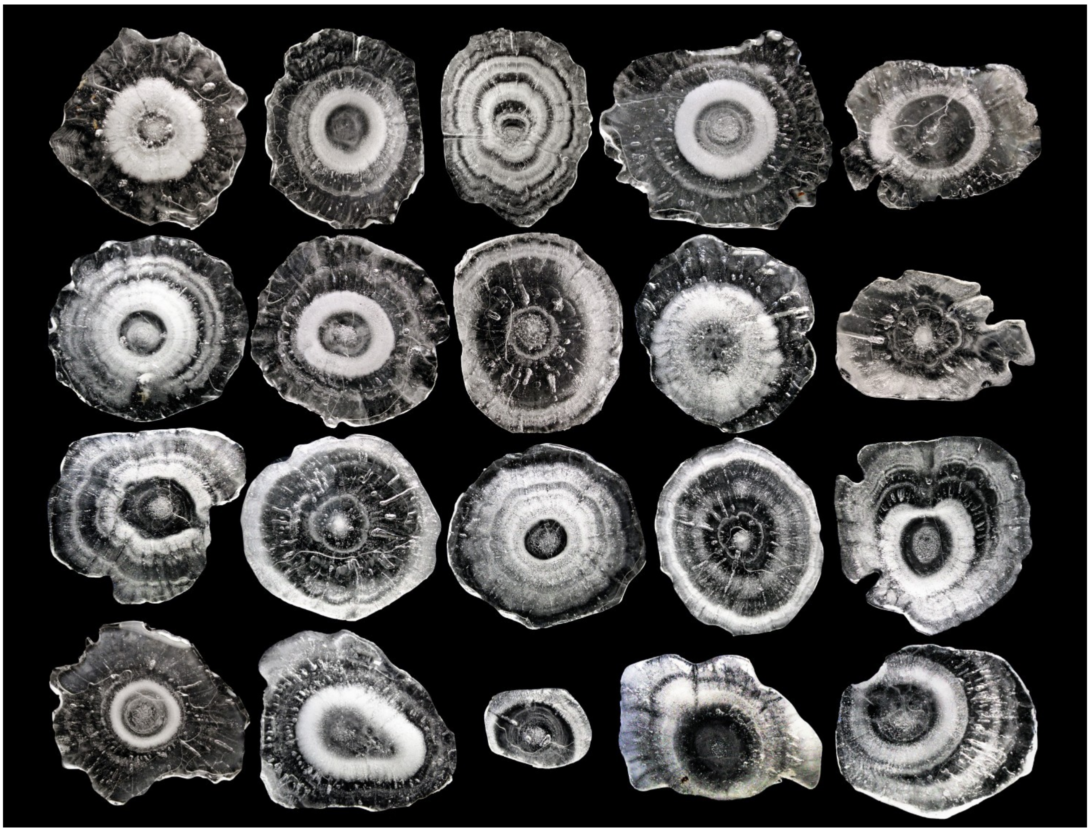


Figure 1.15.: Cross sections of 20 hailstones collected on 19 January 2020 in Camberwell, Melbourne, Australia. The hailstones were all collected from within a 5 m^2 area within 5 minutes, around 16:00 AEDT (UTC+11). The slices were created with a hot-wire cutter and photographed on a glass plate over a dark background. ©Joshua Soderholm (Australian Bureau of Meteorology), used with permission.

as **hail** (AMS 2012b). When riming occurs, latent heat of fusion is released, and the graupel or hail hydrometeor warms, gradually conducting the heat away to the atmosphere over time. If the flux of supercooled droplets riming onto the surface of the hydrometeor exceeds a critical value, the hydrometeor surface will be warmed by the latent heat release to a temperature $T \geq 0$ °C and a liquid layer will exist on the surface (Pruppacher and Klett 1997). In this state, supercooled drops accrete rather than rime onto the surface because freezing does not occur upon contact (Rogers and Yau 1996). Over time the hydrometeor loses heat through conduction to the air and evaporation of the liquid layer, freezing the liquid. When it freezes, the ice is translucent or clear, whereas for rime, air gaps exist between the rimed droplets and so the ice is opaque. The layers of larger hailstones (shown in Figure 1.15) are caused by the alternation between opaque (dry growth) and translucent/clear (wet growth) due to the change in the flux of supercooled liquid water droplets experienced by the hail as it grows (Allen

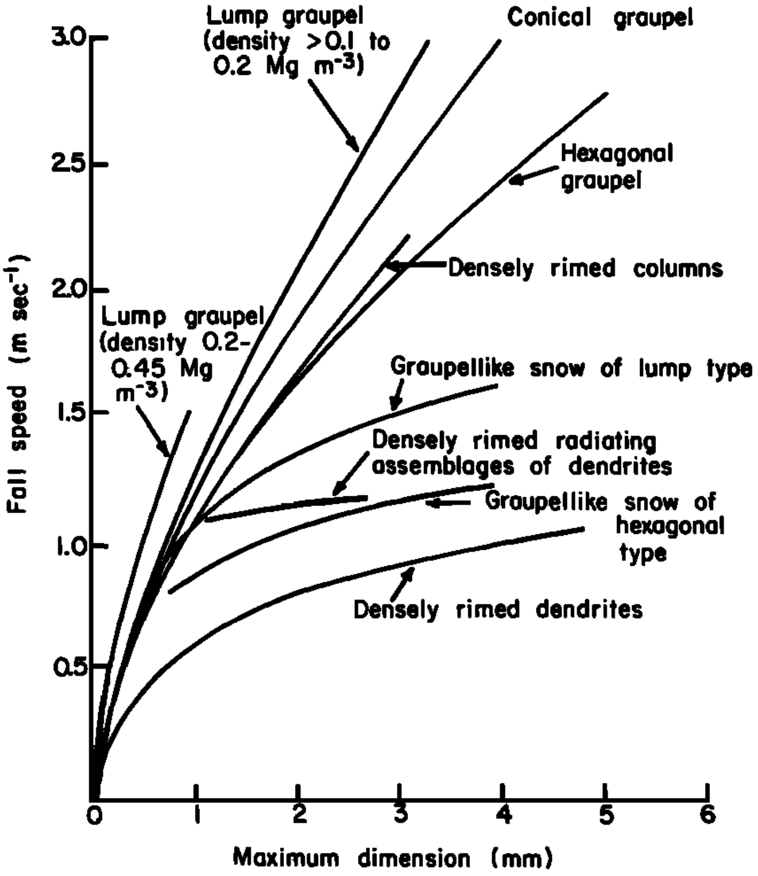


Figure 1.16.: Figure 17 from Locatelli and Hobbs (1974). The relationship between diameter (mm) and fall velocity (m s^{-1}) for sub-species of graupel.

et al. 2020). The notion that hailstones take multiple ascent and descent cycles within a convective updraft is a myth (Knight et al. 1975; Nelson 1983; Ziegler et al. 1983; Foote 1984; Conway and Zrnić 1993; Dennis and Kumjian 2017). There is some evidence that the layers of the atmosphere where temperatures of −10 to −25 °C exist are the most conducive to hail formation (Knight et al. 1975; Nelson 1983; Ziegler et al. 1983). Some radar-based criterion, such as the one used in the spatial SPT product verified in this thesis (see Chapter 5), use a reflectivity threshold at an altitude above the $T = 0$ °C level which would also correspond to temperatures of approximately $T = -10$ °C or lower. Finally, the spiked shape of larger hailstones is due to the preferential flow of liquid water towards any raised lumps on the surface of a hailstone during wet growth (Knight and Knight 1970). A more detailed description of the formation and convective dynamics of hail exists in Chapter 5.

A few notable studies present empirically derived diameter–velocity relationships for

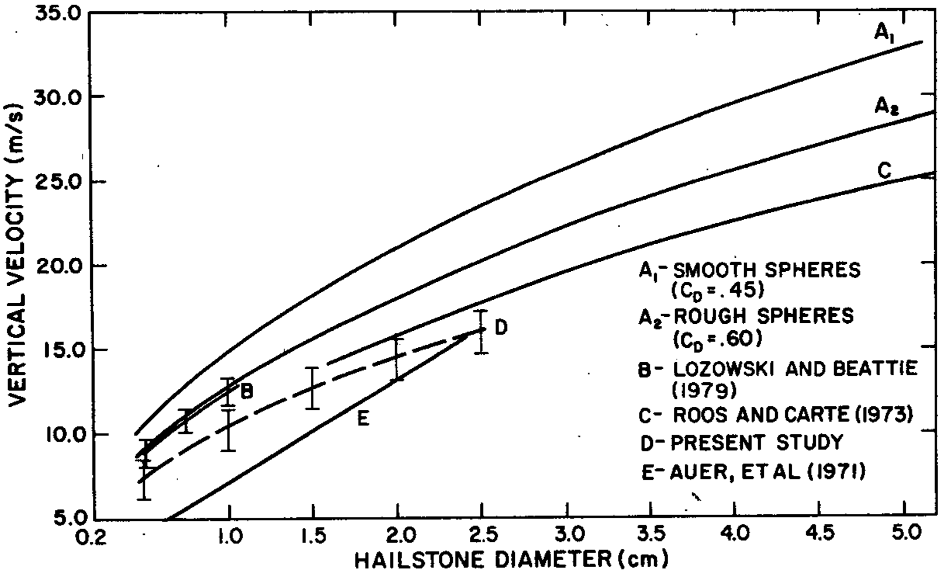


Figure 1.17.: Figure 12 from Matson and Huggins (1980). Theoretical and empirical relationships between diameter (mm) and fall velocity (m s^{-1}) for hailstones.

the various habits for the growth of graupel (Locatelli and Hobbs 1974; Bohm 1989; Heymsfield et al. 2018), and hail (Roos and Carte 1973; Locatelli and Hobbs 1974; Matson and Huggins 1980; Heymsfield et al. 2018). Some of these relationships are shown in Figures 1.16, 1.17, and 1.18. The terminal fall velocity of graupel and hail is dependent on the density of the hydrometeor. Macklin (1962) showed that the density of rimed particles decreased with decreasing wind speed, air temperature, supercooled droplet diameter, liquid-water concentration, and decreased with increasing graupel/hail hydrometeor size. In addition, Garrett and Yuter (2014) show that the presence of turbulence affects the density of graupel hydrometeors to broaden the range of velocities that graupel of a given diameter can exhibit. The spreading of the diameter-velocity distribution increases the uncertainty for ground-based instruments to identify graupel using diameter-velocity measurements compared with empirical relationships (see Section 3.1). For hail, the consistency of the ratio of dry to wet growth between hailstones will also result in a broad diameter-velocity relationship, the controls of which were described previously.

2.6. Summary

The processes that determine the SPT for any given location are complex but, as discussed above, the vertical temperature profile is a dominant factor (see also: Thériault

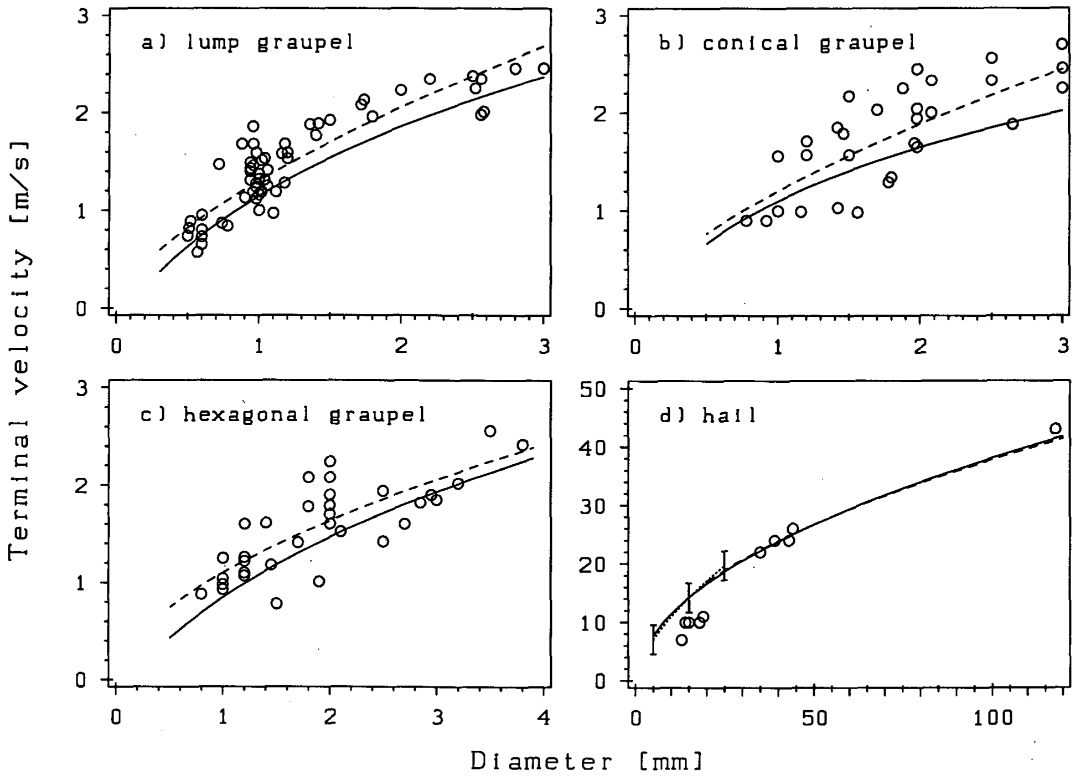


Figure 1.18.: Figure 3 from Bohm (1989). The diameter–velocity relationships for four hydrometeor types listed within each subplot. The solid lines represent the study distributions, the dashed lines in a), b) and c) represent the graupel relationships derived in Locatelli and Hobbs (1974), the dotted line with error bars in d) represents the hail relationship derived in Matson and Huggins (1980), and the dashed line in d) represents the hail relationship derived in Roos and Carte (1973).

et al. 2010). A simplified schematic demonstrating the broad atmospheric differences between SPTs is shown in Figure 1.19. There are some distinct differences between the SPTs as they reach the surface. Any (freezing) non-freezing liquid precipitation predominantly < 0.5 mm in drop size is labelled as **(freezing) drizzle**, and any (freezing) non-freezing liquid precipitation with drops ≥ 0.5 mm is labelled as **(freezing) rain**. Precipitation which remains liquid from nucleation is labelled as warm rain and precipitation which at some point contained solid hydrometeors followed by melting before reaching the surface are labelled as cold rain. For solid hydrometeors reaching the ground, there are five distinct categories. If deposition alone is the primary growth mode, hydrometeors will be pristine **ice crystals** and the SPT is labelled as such. If aggregation between ice crystals is the dominant growth mode, then the SPT is labelled as **snow**. If riming is the dominant form of hydrometeor growth, and if the hydrometeors are < 5 mm in diameter, then the SPT is labelled as **graupel**. If riming or accretion

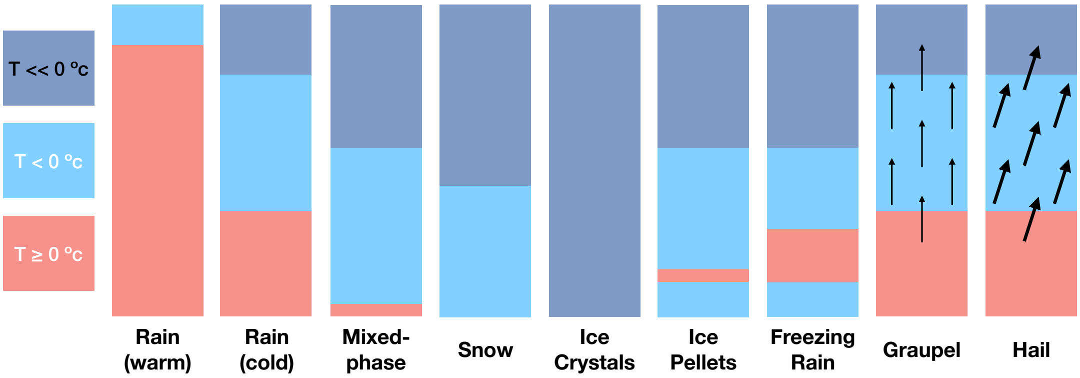


Figure 1.19.: A simplified schematic of the differences between SPTs. The vertical profile of temperature is the main differentiator between most SPTs but in the case of graupel and hail, a moderate to strong updraft (usually tilted for hail) is typically required, associated with convective clouds. A major control between snow and ice crystals, and between rain and drizzle, is the amount of atmospheric water content available for precipitation. Additional processes which were described throughout Section 2 are neglected here for the purposes of simplification. The columns reflect the height from the highest altitude the hydrometeors experience but do not reflect the height of the tropopause nor are the columns comparable between SPTs.

is the dominant form of hydrometeor growth, and critically if the hydrometeors are $\geq$ 5 mm in diameter, then the SPT is labelled as **hail**. Accretion is not impossible on graupel hydrometeors but is very unlikely to occur because a rapid growth rate is required for accretion (see Section 2.5). Finally, if the hydrometeors partially or fully melted on descent but then re-froze, the SPT is labelled as **ice pellets**. Note that a weak to moderate updraft is usually required in instances of graupel in order to advect supercooled droplets past solid hydrometeors, and that a moderate to strong updraft is usually required to support the growth of hailstones. Note also that a tilted updraft encourages larger hailstones to be produced. The **mixed-phase** precipitation class is defined in this thesis as when both liquid and solid aggregate hydrometeors are impacting the surface. Mixed-phase includes partially melted ice crystals (**wet ice crystals**) and partially melted snow (**wet snow**). While graupel and hail precipitation are typically accompanied by some liquid component, they are distinct atmospheric processes associated with convection and are therefore excluded from the mixed-phase SPT label.

An objective of this thesis (Section 5) is to develop a standard set of SPTs because the existing classification schemes are ambiguous, do not contain all SPTs and conflate SPT with other properties such as event duration or non-precipitating atmospheric phenomenon. The controlling microphysical processes on SPT can be used to assist with the classification of SPTs. Table 1.2 is duplicated from Chapter 3 here to emphasise

Table 1.2.: Duplicate of Table 3.3. The master table of precipitation types measured by a ground-based instrument.

ABC	PT	Description	PW Codes	Comments
Er	-2	Instrument Error	—	Instrument offline/data corrupt
Un	-1	Unidentified	40-42	Particles detected but type unknown
No	00	No Precipitation	0, 4, 5, 10, 30-34, 94	Includes fog, mist, haze, smoke
Dr	01	Drizzle	50-53	All drops < 0.5 mm diameter
DrRa	02	Rain & Drizzle	57, 58	Changing between each over time
Ra	03	Rain	60-63	
Ic	04	Ice Crystals	78	Pristine ice crystals; no aggregation
WIc	05	Wet Ice Crystals	—	Presence of liquid water
Sn	06	Snow	70-73	Aggregated ice crystals
WSn	07	Wet Snow	67, 68	Presence of liquid water
Pl	08	Ice Pellets	74-76	Partially melted aggregates, subsequently re-frozen
Gr	09	Graupel	—	Rimed and accreted ice particles, < 5 mm diameter
Ha	10	Hail	89, 96	Rimed and accreted ice particles, ≥ 5 mm diameter
FzDr	11	Freezing Drizzle	54-56	Supercooled below 0 °C All drops < 0.5 mm diameter
FzDrRa	12	Freezing Drizzle & Freezing Rain	—	Supercooled below 0 °C Changing between each over time
FzRa	13	Freezing Rain	64-66	Supercooled below 0 °C

the universal classification system for SPTs proposed in this thesis.

3. Measurement Techniques for SPT

Section 2 described, in detail, the origins and processes which lead to the different precipitation types. The understanding of the origins and processes of SPTs is based on observational, laboratory, theoretical, and computer modelling studies. However, there are several additional factors which specifically complicate our ability to measure the different precipitation types, depending on the technique used. This thesis assesses some of the techniques for observing SPT, and moreover examines the approaches to statistically verify measurements of precipitation type for evidence of skill. The following section provides some additional background material to the measurement techniques which are employed in Chapters 2, 3, 4 and 5. In section 4.6, existing studies of the assessment of these instrument techniques will be discussed, highlighting the statistical metrics used, which themselves will be outlined in Sections 4.4 and 4.5.

The terms ‘point’ and ‘spatial’ are used throughout this thesis to describe data representing limited locales and data representing broad locales. One of the challenges of this thesis is the verification of spatial data using point data as the reference, since the representativity differs by several orders of magnitude between datasets. Instruments are typically categorised as either in-situ or remote sensing techniques. In-situ refers to measurements made within the confines of the instrument, including those instruments which emit and/or receive radiation. Human-made observations either by trained observers or members of the public are also referred to as in-situ, if the SPT being observed is precipitating at the location of the observer. In-situ measurements are **point** data, including those by human observers who may be able to perceive a wider area. Remotely sensed measurements (and NWP model forecasts which are not assessed in this thesis) are **spatial** data. Remote sensing refers to measurements made where electromagnetic radiation is emitted and/or received, and typically covers a very large spatial locale. The following two sections describe the techniques for recording point and spatial SPT observations.

3.1. Point SPT Observations

Fundamentally, all SPTs are inferred into categories through observations of other variables. There are four fundamental categories that can be used to describe point SPT measurement techniques: vision, air temperature, hydrometeor scattering behaviour and hydrometeor diameter-velocity distribution. Human observers primarily use their vision over a short period of time and either *a priori* or *a posteriori* knowledge of the properties of different SPTs such as shape, translucency, and how hydrometeors interact with the surface (splashing, bouncing, settling etc.). Automated instruments in use today utilise some combination of air temperature, scattering behaviour and the diameter-velocity distribution. The following section describes the observations which are currently available in the UK and some of the mechanisms behind automated instrument diagnoses of SPT. Chapter 2 of this thesis introduces a new network of SPT-measuring instruments and the observation technique is also described here.

The primary source of SPT observations in the UK is from the Met Office, which operates and maintains a network of surface stations. During the study period examined in this thesis, there were 176 Met Office surface stations providing SPT information; 160 stations were automated and 16 were manual. Manual stations have a human observer reporting all aspects of the present weather at the site and are typically not operational for 24 hours a day. The human observer has been the most trusted source for present weather observations since the dawn of meteorology. Observers are capable of reliably differentiating between SPTs as each observer is trained in the rarer types such as being able to distinguish hail from graupel from ice pellets. Note that the observations are encoded into the PW code system (WMO 1988, 2017) described below, which ambiguates the SPT measurements. Since present weather is reported hourly, the human observer may note down significant weather that occurred in the prior hour but that is no longer presently occurring; this is a source of error that should be considered when performing verification of spatial products.

The automatic Met Office stations use a method of determining precipitation type by combining the output of several instruments. The Vaisala FD12P present weather sensors emit a beam of near-infrared light at a 16.5° angle below the horizontal which is then received by a diode at 16.5° angle below the horizontal for a total of a 33° angle forward scatter of the light beam. The voltage of the diode depends upon the amount of received light which can be interpreted as the visibility. The peaks of scattered light caused by hydrometeors falling through the beam are related to both the precipitation type and the precipitation rate. A capacitive plate measures precipitation rate (with high uncertainty, see Chapter 3) and the ratio of precipitation rate and diode peak intensity is used to infer the precipitation type. The secondary nature of the derived information makes the technique susceptible to high uncertainty measurements. However, the Vaisala FD12P instrument is not solely relied upon. The addition of a thermometer and a rain gauge are combined into the Met Office present weather arbiter (PWA), a scoring algorithm which relies heavily on the value of air temperature to support the SPT classification (Green 2010). Internal Met Office testing showed that the automatic stations can provide more accurate SPT diagnosis than the Vaisala visiometer alone (Lyth and Molyneux 2006). However, when compared with a human observer, there is still more overlap between SPT signals in the automatic station measurements. Boudala et al. (2017) found that the performance of automated present weather instruments was poorest at temperatures around $T \approx 0$ °C. For example, the SPTs which could occur at an air temperature slightly less than $T = 0$ °C include: snow, freezing drizzle, freezing rain, ice pellets, graupel and hail. For the Met Office automatic station, the tipping-bucket rain gauge would not register any precipitation until melting occurs; the visiometer might detect some changes between these SPTs but would do so with high uncertainty. Furthermore, Merenti-Välimäki et al. (2001) highlights low precipitation intensity ($PR < 0.05$ mm h^{-1}) as one of the reasons for disagreement between human observers and the FD12P instrument used in the Met Office automatic present weather system.

Due to the nature of the Met Office automatic stations and the lack of manual stations (both the overall number of stations and the number of reports made at night), a more accurate point SPT-measuring automated instrument was sought, in order to have higher confidence in the verification of spatial SPT products. Chapter 2 describes the development, installation and maintenance of a new network (called the Disdrometer Verification Network (DiVeN)) of Thies LPM instruments which create a precipitation diameter-velocity distribution from observing individual hydrometeors. To measure the diameter and fall velocity of an individual hydrometeor, a laser occultation technique is used, shown in Figure 1.20. On one end of the instrument, an infrared laser is emitted in a thin horizontal beam and on the other, a diode receives the infrared beam inducing a constant voltage into a circuit. When a hydrometeor falls through the laser beam the amount of infrared light received by the diode decreases and so does the circuit voltage. The amplitude of the voltage decrease is related to the diameter of the hydrometeor, and the duration of the voltage decrease is related to the fall velocity of the hydrometeor. Once a diameter-velocity distribution has been established (1 minute for the DiVeN sensors) the SPT can be inferred using the empirical relationships described in Section 2 (Nakaya and Terada 1935; Gunn and Kinzer 1949; Langleben 1954; Macklin 1962; Locatelli and Hobbs 1974; Matson and Huggins 1980; Bohm 1989; Mitra et al. 1990; Yuter et al. 2006; Garrett and Yuter 2014; Nagumo and Fujiyoshi 2015; Heymsfield et al. 2018).

Since mixed-phase SPT undergoes a continuous diameter–velocity transition, point measurements which use this technique to diagnose SPT can struggle to infer the mixed-phase SPT class if the ground instrument has any uncertainty or if turbulence around the instrument induces vertical air velocities, spreading the distribution of precipitation velocities. For example, Section 2.4 showed that the diameter–velocity relationship for the two regimes of ice pellets are very similar to those of rain and graupel. The addition of temperature would allow for further separation of the SPT classes than diameter–velocity alone. Also, note that the aerodynamic drag effect which determines the terminal fall velocity of a hydrometeor, is multiplicative (Section 2.1) and as such, the linear difference between two large hydrometeors of different SPT is much greater than two small hydrometeors of different SPT (see Chapter 2). The result is that

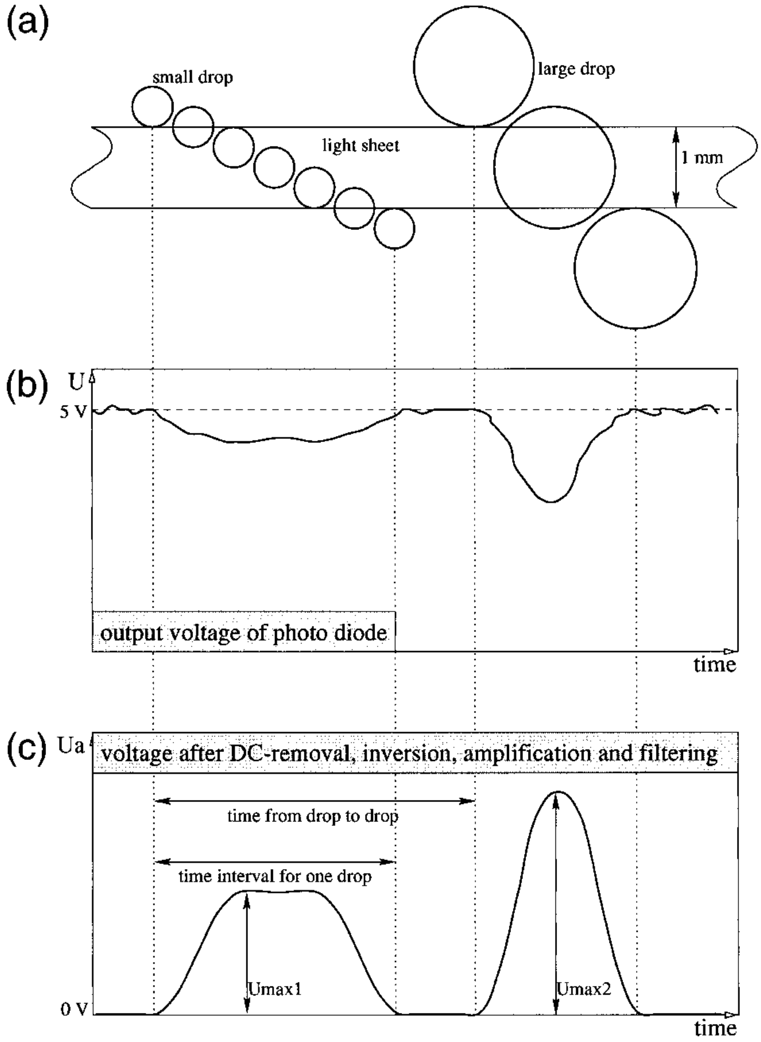


Figure 1.20.: Figure 1 from Löffler-Mang and Joss (2000). A schematic of the technique of laser occultation for precipitation diameter and fall velocity measurement. a) particles falling through the laser beam; b) the raw voltage of the receiving diode; c) the inverted voltage after post-processing.

SPTs are more accurately diagnosed when precipitation rates are higher since larger hydrometeors exist. Despite these issues, there is evidence that the instrument used in DiVeN as part of this thesis research, has higher accuracy when compared to a human than the instrument used at automated Met Office stations (Bloemink and Lanzinger 2005; Lyth and Molyneux 2006). All of the measurement techniques described here are used in this thesis and are subject to further description in the individual chapters where they are used.

For rare SPTs which automated instruments struggle to capture such as hail, the human vision method can be utilised by crowdsourcing reports from members of the public. The coverage of locations is increased with this method but will still not capture all events. In addition, the downside to this method is that the accuracy of elements of the report may differ from trained meteorological observers. For example, Elmore et al. (2015) use a crowdsourced dataset from the public to verify NWP forecasts of precipitation type. A specifically-designed mobile phone application was developed for volunteers to submit their reports (Elmore et al. 2014). In Chapter 5 a crowdsourced dataset from the European Severe Weather Database (Dotzek et al. 2009) is used as the verification reference for the Met Office spatial SPT product for the hail SPT. Since this product specifically neglects graupel and only detects hail, the ground stations (manual, automatic and DiVeN) are not appropriate due to the format of the SPT observations which do not unambiguously differentiate between graupel and hail. The crowdsourced dataset is described in greater detail including benefits and limitations, in Chapter 5.

The format of SPT observations is critical to the research conducted throughout this thesis. The most widely used format for SPT data is the World Meteorological Organization (WMO) PW code (WMO 1988, 2017). The PW code is a reporting system that describes not only the SPT but any meteorological features like fog and dust, includes the duration of SPTs with descriptors such as ‘showers’ or‘ ‘intermittent’, and even encompasses trends in the preceding hour; all distilled into a two-digit number. There are slight differences in the PW coding system for manual observers (Table 4677 WMO 2017) and for automated stations (Table 4680 WMO 2017) which make them slightly incompatible. There is also a third PW code table (Table 4678 WMO 2017) for aircraft pilots, referred to as the Meteorological Aerodrome Report (METAR). Note that some SPTs such as graupel (Section 2.5) are not explicitly mentioned, and that some PW codes contain two or several of the SPTs described in Section 2. In addition, not all sensors are capable of observing all PW codes. Without a temperature or icing sensor, PW instruments cannot distinguish between freezing and non-freezing liquid precipitation. Some automated instruments that are relevant to this study are the Thies LPM, the Campbell Scientific PWS100, the Vaisala PWD21, and the Vaisala FD12P (excluding the Met Office arbiter), each of which are able to report 21, 58, 42 and 52 PW codes from the 93 available in the WMO Table 4680 standard. However, the ability to report more codes does not correlate with more SPTs or a more accurate SPT diagnosis. The Thies LPM does not measure visibility which excludes many PW code discriminators. The manual observer has no restriction on the PW code choice but must interpret the meaning of the PW codes from training. For example, graupel is not explicitly listed in the PW code system and therefore the observer must choose between 20, 77 (contains the term 'snow grains'), 87, 88, 93–97, 99 (contains the terms 'snow pellets' and also 'small hail'), all with varying intensities and accompanying features. Note that the asterisked term 'hail' in codes 93–97 and 99 actually refers to either/or 'hail', 'small hail' and 'snow pellets'. Two examples of the most incoherent and ambiguous PW codes from Table 4677 from the 2017 update to the standard, are provided below (WMO 2017).

> Code 88: *"Shower(s) of snow pellets or small hail, with or without rain or rain and snow mixed—moderate or heavy"*.
>
> Code 93: *"Slight snow, or rain and snow mixed or hail at time of observation—thunderstorm during the preceding hour but not at time of observation"*.

For these reasons, a new standard SPT classification list is defined in this thesis, determined by the microphysical processes which lead to each type. The standard table is shown in Figure 1.19 of Section 2.6, and is developed and applied in Chapter 3, along with the development of a new algorithm (Figure 3.4 in Chapter 3) to merge short-timescale observations of SPT into longer timescales (e.g. 5× 1-minute observations into 1× 5-minute observation). In Chapter 4, the PW codes are converted to match the SPT classes defined by the Met Office radar-based spatial SPT product, the details of which are shown in Figure 4.6 of Chapter 4.

3.2. Spatial SPT Observations

Radars are the most suitable measurement technique for spatial SPT observations because of their high spatial and temporal resolution; on the order of 1 km^2 every 5 minutes in the UK (Harrison et al. 2000). In addition, they can cover a large area on the order of 100,000 km^2. Radars emit a pulse of polarised microwave energy which is scattered by hydrometeors in the beam; the backscattered energy is then received by the radar and measured (Figure 2.2 in Fabry 2015). The power received back to the radar at a single polarisation is described by the radar equation:

$$P_r = \frac{\pi^3 \lambda^2 P_t g^2 \theta \phi h |k|^2 l}{1024 \ln(2) r^2 \pi^2} Z, \tag{1.3}$$

from Collier (1989), where P_t is the power transmitted, g is the gain factor of the antenna, θ, ϕ is the beamwidth (horizontal and vertical), h is the pulse length, $|k|^2$ is the dielectric factor (backscattering efficiency of hydrometeor) which is 0.93 for liquid water and ~0.176 for solid ice (Collier 1989; Fabry 2015; Ryzhkov and Zrnic 2019), l is the loss of power through propagation which is not always known, λ is the beam wavelength, r is the distance from the radar to the target, and Z is the radar reflectivity factor:

$$Z = \int_0^\infty D^6 n(D) dD, \tag{1.4}$$

which is the sum of the sixth-powers of all hydrometeor diameters in a unit volume. Equation 1.4 assumes all hydrometeors are liquid spheres and are small compared to the wavelength of the radar beam so that scattering is within the Rayleigh regime (Bringi and Chandrasekar 2001; Ryzhkov and Zrnic 2019). The amount of energy received by the radar can be used to compute the radar reflectivity factor, and the time taken between emit and receive gives the range at which the hydrometeors exist since the speed of the beam is assumed to be constant. The exact relationship between the radar reflectivity factor and the precipitation rate is incalculable, since multiple DSDs can produce the same reflectivity due to the sixth-power of the diameter of drops. Fabry (2015) gives an example: *"one 2-mm diameter drop falling at 7 m s^{-1} has a reflectivity factor similar to 64 drops of 1-mm diameter falling at 4 m s^{-1}"*. However, the DSD has been shown to vary depending on the rainfall rate (see Section 2.1; Marshall and Palmer 1948) and so approximations of Z–R relationships exist, typically in the form:

$$Z = aR^b, \tag{1.5}$$

where a and b are coefficients based on the local climatology and the type of rainfall. Warm rain, cold rain, convective and stratiform are just some examples of precipitation regimes which have been shown to exhibit different Z–R relationships (Marshall and Palmer 1948; Wexler and Atlas 1963; Joss and Waldvogel 1970; Battan 1976; Austin 1987; Lin et al. 2020). These relationships form the basis of QPE from radar.

Of importance to this thesis is that different SPTs, or hydrometeors as they are viewed by the radar, exhibit vastly different radar properties. If it is snowing on the ground then the radar must also be observing solid hydrometeors, and will observe a much

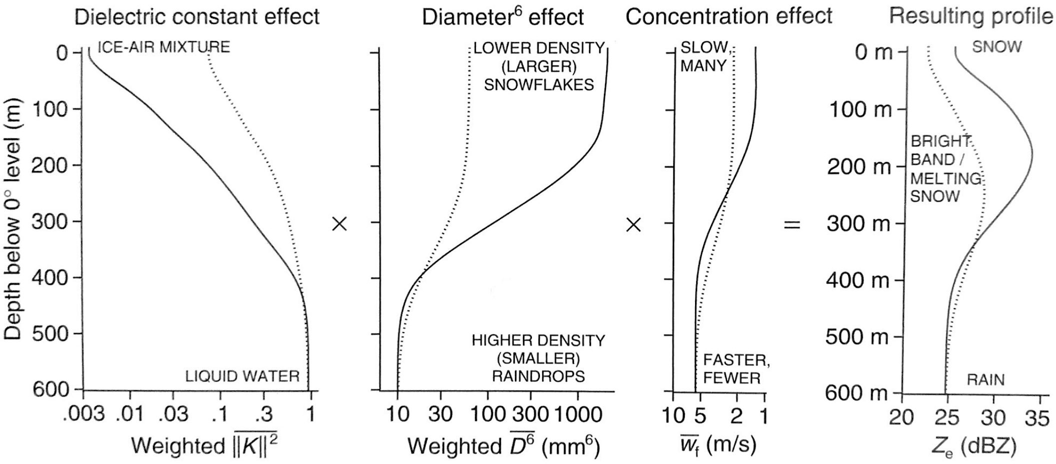


Figure 1.21.: Figure 4.8 from Fabry (2015). The properties of aggregates melting into liquid hydrometeors which result in the enhancement of radar reflectivity known as the bright band. The solid line represents a typical snowflake (grown by aggregation) density, and the dashed line represents a graupel (grown by riming and accretion) density.

weaker reflectivity for the same mass of water due to the dielectric constant ($|k|^2$) of different hydrometeors. Figure 1.21 shows the properties of hydrometeors which affect the radar reflectivity factor, as the hydrometeors change from large aggregates to mixed-phase (partially melted aggregates with a liquid surface), to raindrops.

Knowledge of the precipitation type within the radar beam can therefore aid more accurate QPE from radar. Dual-polarisation radar technology provides numerous polarimetric variables that provide more information about the scattering process and thus reflect the characteristics of the target population of the hydrometeors (Seliga and Bringi 1978; Hall et al. 1984; Aydin et al. 1986; Chandrasekar et al. 1990; Cobb 2004). Polarimetric variables can then be ingested into logic algorithms to estimate the dominant hydrometeor type within the radar beam (Straka and Zrnic 1993; Liu and Chandrasekar 2000; Park et al. 2009; Al-Sakka et al. 2013). For the estimation of SPT, microphysical effects below the radar beam must still be considered for either single-polarisation or dual-polarisation radars since they are incapable of observing at ground level. The lowest usable radar beam is a function of distance from the radar dish, beam blockage from terrain and anthropogenic structures, ground or sea clutter, anomalous propagation, and more (Wilson and Brandes 1979; Collier 1989; Berne and Krajewski 2013; Fabry 2015). If these artefacts and errors are not considered, the quality of the radar QPE can be degraded, which also affects SPT diagnosis. Note that the removal of non-meteorological echoes is improved with dual polarisation radars using

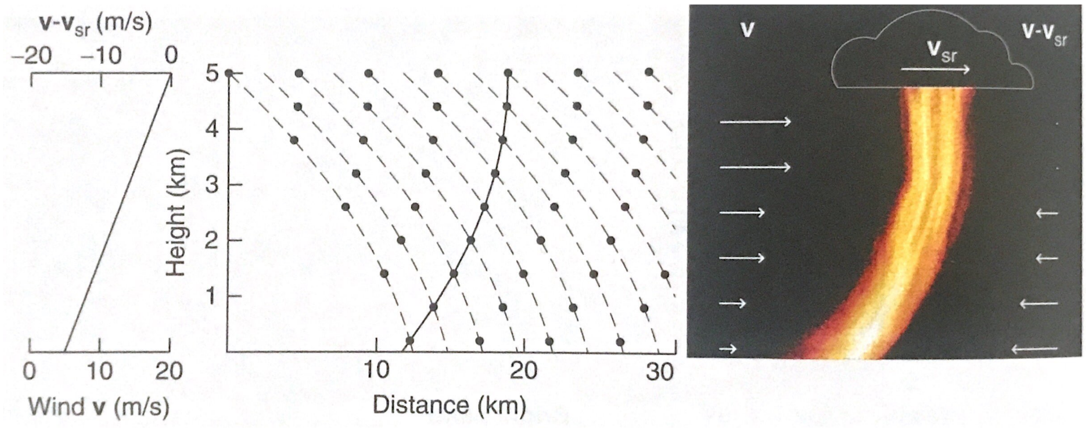


Figure 1.22.: Figure 4.6 from Fabry (2015). A vertical wind profile (left) pushes the precipitation falling from a height of 5 km along the dashed lines (middle) to the ground, advecting several kilometres horizontally. A vertically pointing radar (right) first observes the 5 km height followed by lower altitudes, which over time generates a time–height plot which appears to show the precipitation being advected in the opposite direction (the solid line in the middle figure), a geometric illusion.

fuzzy logic classifiers such as those described in Gourley et al. (2007); Rico-Ramirez and Cluckie (2008); Dufton and Collier (2015). Chapter 4 and Chapter 5 contain long-term frequency plots of SPTs over 5 years from a radar-based product (see below) which are useful to visualise these artefacts and biases in radar observations. Finally, precipitation may advect horizontally between its location when observed by the lowest-usable radar beam and its location when the precipitation impacts the ground (see Figure 1.22), which can deteriorate a verification metric score deceptively. Chapter 4 introduces a verification method to address the issue of representativity due to precipitation advection.

The UK Met Office operates a network of 15 C-band (5.6 GHz frequency) radars and collaborates closely with Met Éireann and the Channel Islands Meteorological Department who operate an additional two and one C-band radar(s) respectively. Hence, there are a total of 18 radars contributing to the coverage over the UK, as shown in Figure 1.23. Harrison et al. (2000, 2015) describes some of the design choices and the processing techniques used in the network.

The placement of UK radars shown in Figure 1.23 are chosen for optimal scanning with low partial beam blockage, minimal ground clutter, and appropriate land permissions. Low elevation scans of 0.5° reduce the below-beam area and thus the uncertainty of below-beam effects and thus aids QPE and SPT diagnosis. The scan strategy of the Met Office radars begins with the highest elevation scan first and then tilts down, ending

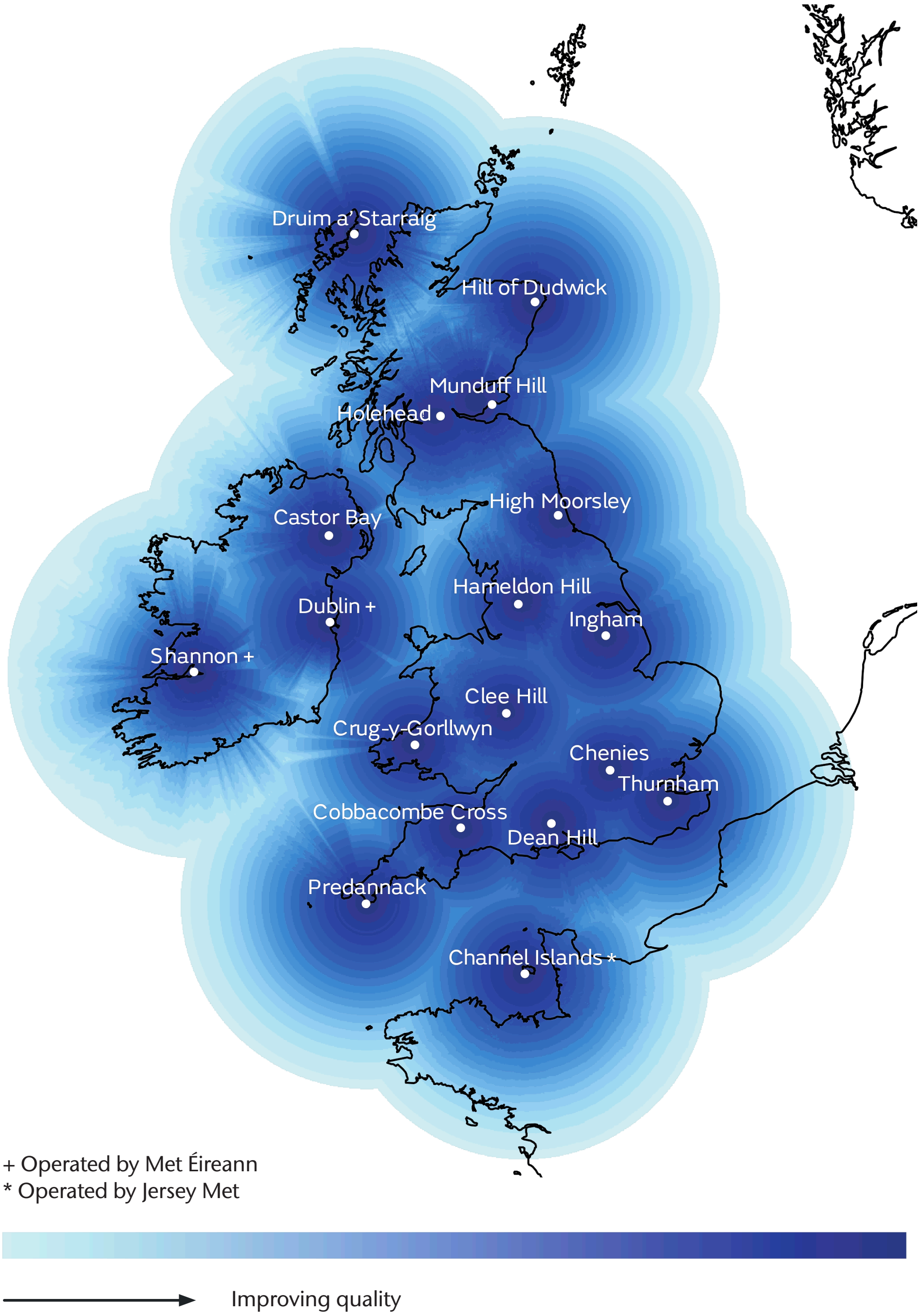


Figure 1.23.: Figure 3 from Met Office (2019b). A map of the UK radar network consisting of 15 Met Office-owned sites and 3 third-party owned sites denoted by a '+' symbol for Met Éireann and a '*' for the Channel Islands Meteorological Department. Radar beams increase with height and are susceptible to blockage by terrain and anthropogenic objects, which lowers their QPE and SPT diagnosis skill. An indicator of increasing quality is shown by the colours light blue to dark blue.

with the lowest elevation scan last, which is beneficial for surface product diagnoses because the scan closest to the ground is used for the QPE and SPT calculation and is therefore conducted adjacent to the denoted timestamp of the surface products.

The Met Office owned sites have recently been upgraded to dual-polarisation capability and were fully operational in January 2018, which should improve both QPE and SPT diagnosis in the long-term over the UK. The focus of the research in this thesis was the development of a verification framework for SPT-datasets in general. However, in Chapter 4 and Chapter 5, the verification framework is applied to an existing SPT product which has been operational since late 2013 using only the radar reflectivity factor and the derived surface rainfall rate data. The results of this verification show that the framework provides reasonable results and provides a baseline value for the skill of a spatial SPT product which motivated improvements to a dual-polarisation equivalent product. The single-polarisation Met Office spatial SPT product is described in more detail in Chapter 4 which diagnoses SPT classes of rain, mixed-phase, snow, and hail encompassing the SPTs described in Sections 2.1, 2.3, 2.2, and 2.5 respectively (note that graupel is explicitly neglected). It uses Lumb's critical rate (Lumb 1963, see section 2.3) for mixed-phase diagnosis and a reflectivity with temperature criterion for hail (Waldvogel et al. 1979). There were several intermediary objectives regarding the enhancement of the available surface point SPT reference observations which were required prior to the development of a verification framework for the spatial SPT product. The next section describes the statistical challenges of the assessment of SPT data generally, as well as specific considerations for the verification of spatial SPT products such as the Met Office radar-based product.

4. Statistical Techniques Employed in this Thesis

This thesis approaches the quantitative assessment of SPT measurements. A brief introduction into the statistical methods typically used for discrete, non-probabilistic data is necessary, along with specific mentions of those methods which are used in this research. Here a wider statistical background is introduced, as well as highlighting some of the statistical difficulties in categorising transitional precipitation types. The section concludes with a summary of the statistical approaches used in the literature to assess point SPT observations and spatial SPT products.

4.1. Assessment Terminology

Firstly, the terms evaluation, validation and verification have different meanings. Despite this, their use in the literature is often inconsistent (STC 2013; Sparrius 2016), so a standard terminology to follow for this thesis must be chosen. The fields of engineering and medical science have more strict definitions (ECS 2002; IEEE 2011) which are useful considerations to follow. The following definitions used are:

> **Validation** is checking that a system meets the needs of an application or an end user: does the system function as intended? ISO 15189:2012 defines validation as *"confirmation, through the provision of objective evidence that the requirements for a specific intended use or application have been fulfilled"* (ECS 2012). Chapter 2 of this thesis is primarily a validation of an SPT observing system.
>
> **Evaluation** is comparing multiple methods of achieving the same output. Metrics to quantify an evaluation can include accuracy, practicality, cost etc. (ECS 2002; Rossi et al. 2018). A more accurate system may be less desirable if the practicality or cost is inhibitive. Here the choice of evaluation criteria should be simplified; attempting to meet too many, or poorly defined criteria, can result in an unclear outcome. Chapter 3 of this thesis is an evaluation of SPT instruments.
>
> **Verification** is checking that a product or system is accurate against a known or accepted standard. Metrics such as accuracy, precision, linearity etc. can be used to quantify verification. ISO 15189:2012 defines verification as *"the confirmation, through provision of objective evidence that specified requirements have been fulfilled"* (ECS 2012). Chapters 4 and 5 of this thesis are both classified as the verification of a spatial SPT product.

All three of these definitions are used in this thesis to differentiate the research conducted in each of the results chapters.

4.2. Discrete Non-probabilistic Data

Section 2 introduced the processes and behaviours of hydrometeors in the atmosphere from growth to precipitation at the surface. SPT data are discrete and in this research the SPT data are also non-probabilistic. Discrete and non-probabilistic datasets are such that the values are split into categories and are counted by integers, which limits the available statistical methods that can be applied. The most basic discrete variable is dichotomous; a simple yes or no outcome for the occurrence of an event. SPT is a more complex discrete variable, where the number of categories is subjective. The most basic distinction of precipitation type is the phase: liquid or solid. However, even here the border between liquid and solid phase is not well defined. The melting layer is a region of the atmosphere of non-negligible depth, where falling hydrometeors exist at temperatures greater than 0 °C but are not yet fully liquid (Section 2.3). The hydrometeors exist as a mixture of both solid and liquid phase particles. Solid hydrometeors melt from the outer edge inwards, such that ice can be coated with a surface layer of water. For ice aggregates which contain large volumes of air and are thus significantly less dense than liquid water, the existence of a liquid surface appears to precipitation radar instruments as an extremely large raindrop (Fabry 2015; Lin et al. 2020). Note that the enhancement of radar reflectivity in the melting layer is known as the bright band, which must be corrected for when equating reflectivity into precipitation rate (Smyth and Illingworth 1998; Harrison et al. 2000; Villarini and Krajewski 2010).

Further still, within the solid hydrometeor types there are hail, graupel, aggregates, and pristine ice crystals. Within ice crystals there are the crystal habits shown by Kobayashi (1958) in Figure 1.6. Each extra level of precipitation class specificity requires higher skill, specifically of discrimination (in statistical terms; Wilks 2011), in the observation or forecast. Figure 1.24 demonstrates an example of a hierarchical tree of precipitation types with increasing complexity with each subclass. If the discrimination of the observation or forecast is not sufficient, then increasing the number of SPT classes will diminish its useful skill.

4.3. Confusion Matrices and Contingency Tables

The verification of dichotomous discrete non-probabilistic datasets is typically visualised in a 2 × 2 table known as a contingency table (Wilks 2011). Figure 1.25 is an example of the layout of the contingency table, where the rows (y-axis) are the dataset under examination, and the columns (x-axis) are the reference dataset which is considered the truth. The quadrants have the associated labels of a = hit, b = false alarm, c = miss and d = correct null. The quadrants can be represented in normalised proba-

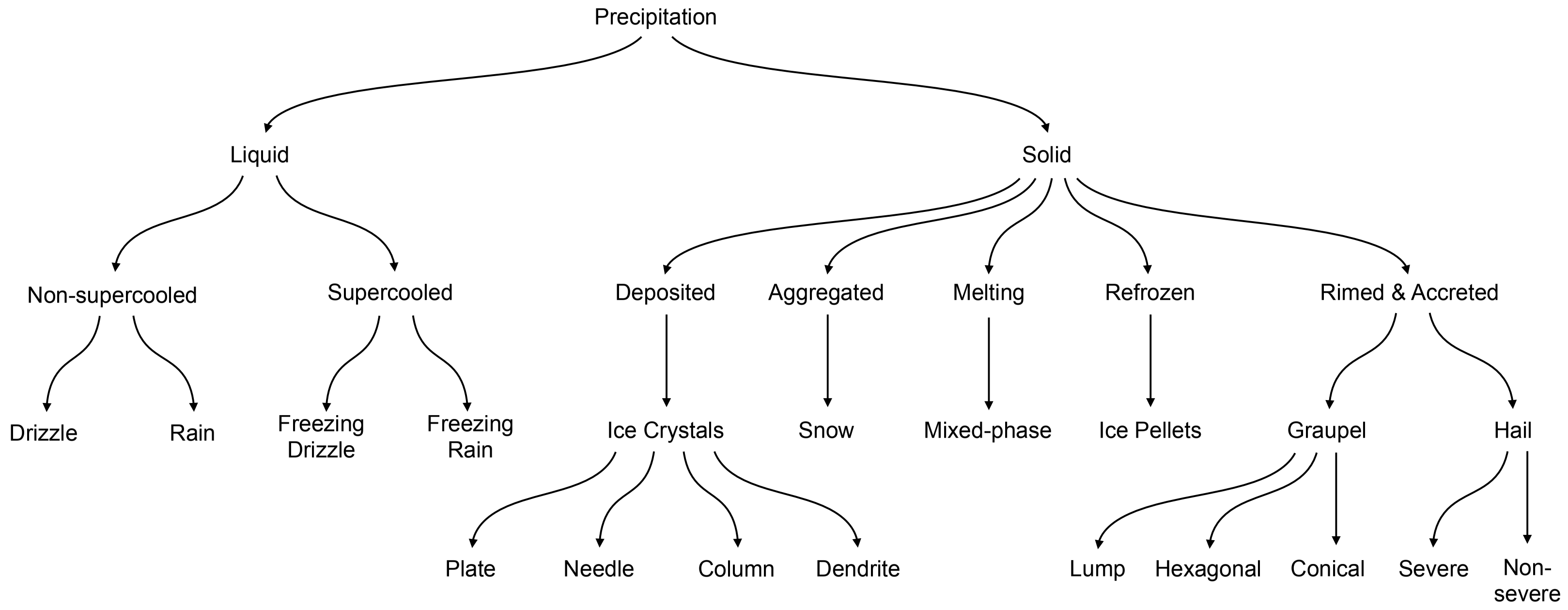
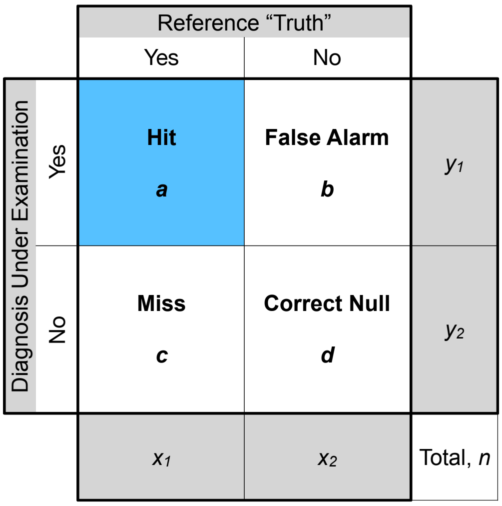


Figure 1.24.: A non-exhaustive hierarchical tree of precipitation classes, with each level a subset of the latter and increasing the overall complexity.



		Reference "Truth"		
		Yes	No	
Diagnosis Under Examination	Yes	**Hit** ***a***	**False Alarm** ***b***	y_1
	No	**Miss** ***c***	**Correct Null** ***d***	y_2
		x_1	x_2	Total, *n*

Figure 1.25.: The contingency table for dichotomous discrete non-probabilistic verification. Dimensions are 2 × 2 with quadrants referred to as a) hit, b) false alarm, c) miss, and d) correct null.

bilistic form:

$$\mathrm{a} = \mathrm{p}\left(\mathrm{y}_1, \mathrm{x}_1\right) \tag{1.6}$$

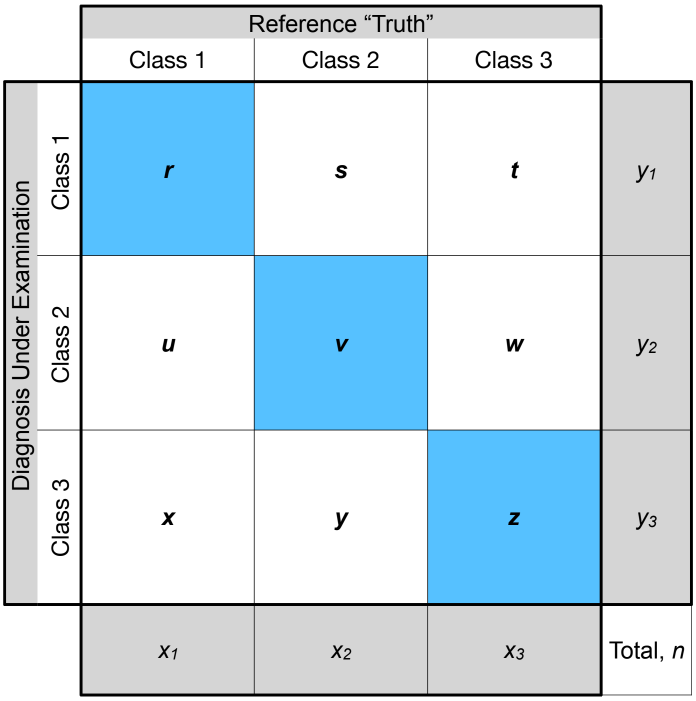
$$\mathrm{b} = \mathrm{p}\left(\mathrm{y}_1, \mathrm{x}_2\right) \tag{1.7}$$

$$\mathrm{c} = \mathrm{p}\left(\mathrm{y}_2, \mathrm{x}_1\right) \tag{1.8}$$

$$\mathrm{d} = \mathrm{p}\left(\mathrm{y}_2, \mathrm{x}_2\right) \tag{1.9}$$

where y_1 and y_2 are the probability of a diagnosed event being a 'yes' and a 'no' respectively, and x_1 and x_2 are the probability of the true event being a 'yes' and a 'no' respectively. For the contingency table there are various metrics which are described in the following section.

SPT data are rarely dichotomous—as shown in Figure 1.24. Verification of non-dichotomous discrete non-probabilistic datasets are typically visualised in an higher dimensional table known as a confusion matrix. Figure 1.26 is an example of the layout of a 3 × 3 confusion matrix, where the rows (y-axis) are the classes of the dataset under examination, and the columns (x-axis) are the classes of the reference dataset which is considered the truth. The cells have the associated labels of r–z, where the top-left to bottom-right diagonal cells of the matrix are the hits. The remaining cells can be false alarms, misses, or correct nulls depending on the class under consideration.



		Reference “Truth”			
		Class 1	Class 2	Class 3	
Diagnosis Under Examination	Class 1	***r***	***s***	***t***	y_1
	Class 2	***u***	***v***	***w***	y_2
	Class 3	***x***	***y***	***z***	y_3
		x_1	x_2	x_3	Total, *n*

Figure 1.26.: The structure of the 3 × 3 confusion matrix applied in this study.

Cells can also be represented in the normalised probabilistic form of Eqs. 1.6–1.9, such as $\mathrm{w} = \mathrm{p}\left(\mathrm{y}_2, \mathrm{x}_3\right)$ and so on. The difficulty with SPT data is that the borders between the classes can be sequential in the case of solid to mixed-phase to liquid, or can be statistically independent such as hail and snow.

The higher dimensional confusion matrix can be reformulated into dichotomous, 2 × 2 contingency tables for each of the I classes (Wilks 2011) as shown in Figure 1.27. For example, the quadrants of the dichotomous 2 × 2 contingency table which is created for “Class 1” can be surmised in terms of the elements in Figure 1.26. The hit quadrant (a) is r, the false alarm quadrant (b) is $s + t$, the miss quadrant (c) is $u + x$, and the correct null (d) is $v + w + y + z$. This is repeated as appropriate for “Class 2” and “Class 3”.

4.4. Scalar Metrics

There are many ways to assign a single numeric value to a diagnosis from the dichotomous 2 × 2 contingency table, but all encounter a loss of information since the dimensionality of the table is $I \times J - 1 = 3$ (Murphy 1991). The following section

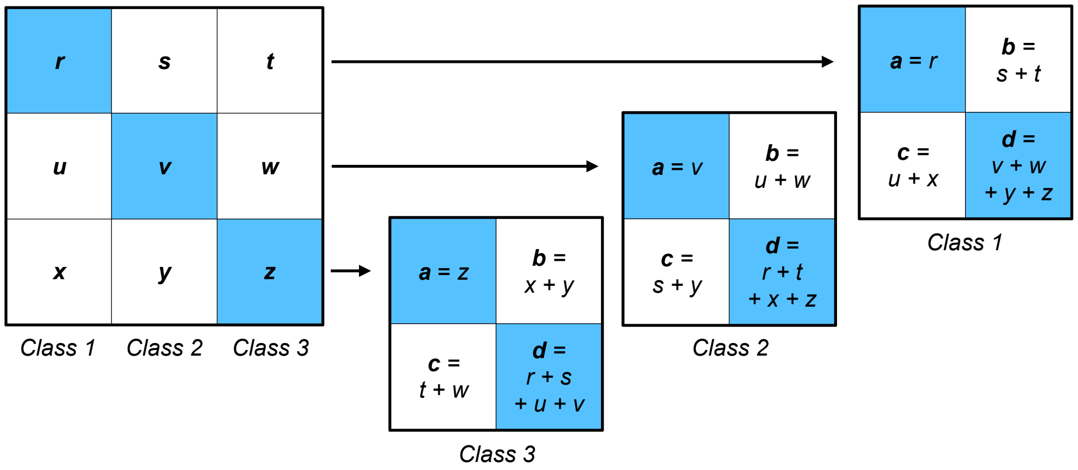


Figure 1.27.: Redrawn version of Figure 8.3 from Wilks (2011). A 3 × 3 confusion matrix being collapsed into three 2 × 2 contingency tables, one for each class in the confusion matrix. The collapsing arithmetic extends to a confusion matrix of any square dimensions.

describes some of the scalar metrics used on 2 × 2 contingency tables as well as highlighting their strengths and weaknesses. In this thesis the notation of Wilks (2011) is followed, where more details on the scalar metrics can be found.

Since the perfect diagnosis in a contingency table would be $a + d = n$, the proportion correct (PC; Finley 1884) is:

$$PC = \frac{(a+d)}{n} \tag{1.10}$$

where the metric ranges from 0 (always incorrect) to 1 (always correct). There is no differentiation between correct diagnoses of the event a and correct diagnoses of the non-event d, which may be undesirable for high-impact events where a hit (a) is more valuable than a correct null (d). For rare events where $c + d$ is large, the PC will be larger for more diagnoses of ‘no’ which increases the value of d.

Attempting to resolve the issue with hit and correct null events receiving equivalent weighting in the proportion correct metric, the critical success index (CSI) is an accuracy metric which neglects the correct null events:

$$CSI = \frac{a}{(a+b+c)} \tag{1.11}$$

CSI ranges from 0 (no hit cases) to 1 (only hit cases) and is also referred to as the threat score (TS) (Gilbert 1884). Here the false alarm and miss cases are given equal weighting, which is not always desirable. A miss for a high-impact event may be a worse outcome than a false alarm.

A metric which only considers the diagnosis when the event does occur and measures the probability that the diagnosis will be correct, is called the probability of detection (POD):

$$POD = \frac{a}{(a+c)} \tag{1.12}$$

which also ranges from 0 (always a miss) to 1 (always a hit), and is also referred to as the hit rate. In some instances however, it may be more detrimental to have a false alarm b than a miss c, which is not quantified with POD.

There are two similarly named equivalent scalar metrics which include the false alarm quadrant. Firstly the fraction of non-event occurrences $(b+d)$ that were false alarms called the false alarm rate (F), and secondly the fraction of diagnoses $(a+b)$ that were false alarms called the false alarm ratio (FAR):

$$F = \frac{b}{(b+d)} \tag{1.13}$$

$$FAR = \frac{b}{(a+b)} \tag{1.14}$$

F ranges from 0 (always a correct null) to 1 (always a false alarm). FAR ranges from 0 (always a hit) to 1 (always a false alarm). The POD and F/FAR are usually interdependent, such that changing the bias (see Eq. 1.15) of a product to reduce the number of miss events typically results in an increase to the number of false alarm events.

Knowing the ratio of ‘yes’ diagnoses $(a+b)$ to the ‘yes’ occurrences $(a+c)$ is useful for determining whether a product or forecast is overzealous and can be the simplest metric to fix with a post-correction. The bias (B) is defined as:

$$B = \frac{(a+b)}{(a+c)} \tag{1.15}$$

ranging from 0 (events are never diagnosed) to ∞ (events never occur), with $B = 1$ being the perfect score since the number of diagnoses is equal to the number of events. Any bias value less than 1 is an underdiagnosis, and any bias value greater than 1 is an overdiagnosis. A product can have the perfect bias score of 1 and have a POD of 0, by having equal numbers of false alarms and miss cases. In some cases, if there is a preference for miss cases over false alarms, or vice versa, then $B \neq 1$ will be desired.

Again, no single scalar metric is able to represent the behaviour of a 2 × 2 contingency table which has 3 degrees of freedom (Murphy 1991). Ultimately the choice of scalar metrics on a 2 × 2 contingency table should depend on the desired goal of the product but a combination of turning false alarms into hits and misses into correct nulls, is the overall goal of any verification.

4.5. Skill Score Metrics

Skill is a term used to define how much improvement a diagnosis or forecast has over some reference. The choice of the reference differs depending on the exact intentions of the verification, and ultimately determines the outcome of the value of skill. Skill scores are additional scalar metrics which set the reference skill value to zero and the perfect skill value to 1. As such, skill scores are defined as:

$$SS = \frac{V - V_{ref}}{V_{perf} - V_{ref}} \tag{1.16}$$

where V is the verification metric, V_{ref} is the verification metric for a reference diagnosis, and V_{perf} is the verification metric for a perfect diagnosis, set equal to 1. While many skill score formatted metrics exist for the contingency table, a subset of these metrics are higher-dimensional and are therefore also applicable to the confusion matrix. In this thesis it is desirable to use skill scores which apply to both the confusion matrix and the compressed contingency tables which originate from the confusion matrix. With this approach, the contribution (be that the advantage or detriment) of each precipitation class to the overall classifier skill score can be quantified. The classifications can also be adjusted, grouped or divided such as those described in Figure 1.24 to determine whether another classification structure warrants higher skill.

Therefore, only a subset of skill scores are considered. The higher-dimensional Heidke Skill Score (HSS) is defined following the structure of Eq. 1.16 as:

$$HSS = \frac{\sum_{i=1}^{I} p\left(y_i, x_i\right) - \sum_{i=1}^{I} p\left(y_i\right) p\left(x_i\right)}{1 - \sum_{i=1}^{I} p\left(y_i\right) p\left(x_i\right)} \tag{1.17}$$

where $\sum_{i=1}^{I} p\left(y_i, x_i\right)$ is the proportion correct (the normalised sum of all diagonal confusion matrix terms), $\sum_{i=1}^{I} p\left(y_i\right) p\left(x_i\right)$ is the random proportion correct (the product of diagnosed and observed normalised probabilities summed over each class), 1 is the perfect score, I is the length of the confusion matrix $I \times J$, y_i is the ith row and x_i is the ith column (Doolittle 1888; Heidke 1926). The HSS indicates the fractional improvement in diagnosis over the probability of a correct diagnosis by chance, which would score zero. The highest score (V_{perf}) is 1, and the lowest possible score is $-\infty$; negative values therefore indicate that a random guess would have been more skilful. For a dichotomous 2 × 2 contingency table the HSS collapses to:

$$HSS = \frac{2(a \times d - b \times c)}{(a+c)(c+d) + (a+b)(b+d)} \tag{1.18}$$

where a, b, c, and d have the same interpretation as in Figure 1.25

The higher-dimensional Peirce Skill Score (PSS) (also referred to in the literature as the true skill statistic (TSS) and Hanssen-Kuipers discriminant (HK)) is defined as:

$$PSS = \frac{\sum_{i=1}^{1} p\left(y_i, x_i\right) - \sum_{i=1}^{1} p\left(y_i\right) p\left(x_i\right)}{1 - \sum_{j=1}^{J} \left[p\left(x_j\right)\right]^2} \tag{1.19}$$

where $\sum_{i=1}^{I} p\left(y_i, x_i\right)$ and $\sum_{i=1}^{I} p\left(y_i\right) p\left(x_i\right)$ are the proportion correct and the random proportion correct respectively (identical to the HSS in Eq. 1.17), and the remaining symbols have the same meaning as in Eq. 1.17 (Peirce 1884; Hansen and Kuipers 1965). One term differs from the HSS—on the denominator of Eq. 1.17, $\sum_{j=1}^{J} \left[p\left(x_j\right)\right]^2$ is a modified random proportion correct which is the square of the observed normalised probabilities summed over each class, meaning that the diagnosed normalised probabilities are excluded. Again the highest score (V_{perf}) is 1, and the lowest possible score is $-\infty$, and zero means no skill, so values of PSS < 0 indicate that a random guess would have been more skilful. The PSS also indicates the fractional improvement in diagnosis over the probability of a correct diagnosis by chance (with a different definition of "chance"), which would score zero. For a dichotomous 2 × 2 contingency table the PSS collapses to:

$$PSS = \frac{(a \times d) - (b \times c)}{(a+c) \times (b+d)} \tag{1.20}$$

which can be represented as the sum of the hit rate (POD, Eq. 1.12) minus the false alarm rate (F, Eq. 1.13):

$$PSS = \frac{a}{(a+c)} - \frac{b}{(b+d)} \tag{1.21}$$

The PSS denominator modification compared with the HSS, where the denominator V_{ref} term is $\sum_{j=1}^{J} \left[p\left(x_j\right)\right]^2$, is equivalent to the squared climatology distribution of events. The score must be applied to each climatological subset if significant differences exist in the dataset under verification (Wilks 2011). The fundamental difference between the HSS and the PSS is the definition of a random chance diagnosis. For the

HSS, the random chance is the combined product of the diagnosed and the true event probabilities. For the PSS, the random chance (in the denominator only) is the square of the climatological probability. In other words, the HSS uses the random chance that the final confusion or contingency table would reach the same diagnosed and observed probabilities, whereas the PSS uses the random chance that the diagnosed probability would match the observed (climatological) probability.

Doswell III et al. (1990) showed that the value of the PSS for a 2 × 2 contingency table asymptotes toward the value of the POD scalar metric (Eq. 1.12), as the weighting of the correct null (d) quadrant increases—in other words, as the event under scrutiny increases in climatological rarity. In this instance, the observation or forecast under examination can appear to have a higher PSS value by increasing the frequency of the 'no' diagnosis. Furthermore, if the bias of the dataset under examination is approximately 1 then the HSS and PSS have the same value, since if $B = 1$, then $y_i = o_i$ such that:

$$\sum_{j=1}^{J} \left[p\left(x_j\right)\right]^2 = \sum_{i=1}^{I} p\left(y_i\right) p\left(x_i\right) \tag{1.22}$$

More skill scores do exist for the 2 × 2 contingency table (Gilbert 1884; Yule 1900; Clayton 1927, 1934; Stephenson 2000; Hogan et al. 2009; Ferro and Stephenson 2011) but are not scalable to the higher-dimensional confusion matrix and are thus not desirable. Spatial skill scores such as the Fractions Skill Score (FSS; Roberts and Lean 2008; Mittermaier et al. 2013) and the Localised FSS (Woodhams et al. 2018), compare two spatial datasets which must exist on the same resolution grid. The FSS combines spatial and temporal inaccuracies into a single score and avoids the so-called 'double penalty' problem where a convective cell in the wrong location results in both a false alarm and a miss. However, in this thesis the comparisons made are first point-to-point evaluations, and then point-to-grid verifications, not grid-to-grid. A novel method for the comparison of gridded spatial data against point-based reference data, using several of the metrics introduced in this section, will be introduced in Chapter 4.

4.6. Metrics Used in the Literature

For the verification of point SPT observations, a manual observer is the most commonly used reference data. Sheppard and Joe (2000) use POD, FAR and HSS (no justification is provided for the choice of metrics) to verify the Vaisala FD12P, two more forward-scattering sensors and a close range (2 m) vertically-pointing radar (with and without multi-parameter temperature/icing sensor) against human observers over two 6-month winter periods in Toronto, Canada. Merenti-Välimäki et al. (2001) use POD and a proprietary correlation model (which they claim is *"fascinating in its flexibility and simplicity"*) based on confusion matrices of PW codes to verify a Vaisala FD12P sensor against a manual observer for two winters at the Jokioinen Observatory of the Finnish Meteorological Institute. Bloemink and Lanzinger (2005) also use POD, FAR, and HSS (without justification) and verify the Thies LPM and the Vaisala FD12P against manual observers over a six-week period in central Germany (at 950 m a.m.s.l) where liquid and solid precipitation occurred for 7% and 5% of the study period respectively. Boudala et al. (2017) simply compare the frequencies (no justification is given for the lack of scalar metrics) of observed SPTs from a range of automated instruments and manual observers for 12 months in Northeastern Alberta, Canada. In all cases, some level of interpretation of the PW coding system is required.

Several studies perform verification of SPT diagnosis from spatial observations. In Schmid and Mathis (2004), dry-bulb and dew-point temperature from 72 ground stations between 200–2,000 m altitude in the Swiss Alps are combined with a radar-based precipitation mask into a rain or snow SPT algorithm (Koistinen and Saltikoff 1998) at 10-minute frequency. When verified against 52 automatic stations (Vaisala PWD11) the inferred metrics used were the bias and POD but no rationale was provided for the lack of other scalar metrics. Reeves (2016) used automated stations and crowdsourced reports of SPT to verify atmospheric sounding-based SPT algorithms. The main metric used in the verification is the POD, justified with the claim that the interpretation of the FAR, HSS and PSS *"is the same as what can be gleaned through consideration of only the POD"*. Of note is that the crowdsourced verification datasets almost always gave lower POD values, which was attributed to the skill of the members of the public submitting reports. Chen et al. (2016) verify an SPT algorithm by Zhang et al. (2011) which uses a 3D radar mosaic plus model temperature surface data, against the same crowdsourced reports dataset in Reeves (2016) over 4 winter months of 2012/13. The POD, FAR and CSI were applied to this verification but no reasoning is provided.

More studies perform a verification of SPT forecasts from NWP models. Gascón et al. (2018) describe two probabilistic forecast products from the European Centre for Medium Range Weather Forecasts (ECMWF) ensemble model. Since the product is probabilistic, different verification metrics can apply and are well-reasoned. However, for the deterministic portion of the forecasts, POD, FAR and CSI are used because their mathematical relatability allows geometrical representation in a visualisation called a performance diagram. Elmore et al. (2015) evaluated 3 NWP short-range (0–24h) SPT models against crowdsourced observations over 2 winter seasons. The bias and PSS metrics are applied on 2 × 2 contingency tables (PSS gives similar values as HSS if the bias is minimal; Wilks 2011) but the Gerrity Skill Score (GSS; Gerrity Jr. 1992) is applied on higher-dimensional confusion matrices, which gives weighted influence to precipitation classes based on their frequency more than the HSS does. However, the skill contribution from each precipitation class toward the higher-dimension score is not tangible because the PSS and the higher-dimensional GSS are incomparable. Ikeda et al. (2013) verified the High-Resolution Rapid Refresh (HRRR) 1–8h lead-time SPT diagnosis through winter 2010/11 with a fractional confusion matrix approach to overcome the “double penalty” problem for NWP forecast verification (Roberts and Lean 2008; Mittermaier and Roberts 2010). The model value (the column) of the confusion matrix is spread across the precipitation types based on what fraction of each type exists in a neighbourhood region around the ground data location. The neighbourhood used is 18 × 18 km ± 6-min, significantly larger than the neighbourhood used in Chapters 4 and 5, however, an NWP forecast has codependent spatiotemporal errors. POD is the main verification metric but misses and false alarms are also presented on maps for two case studies. Wandishin et al. (2005) compare 5 algorithms using short-range (0–48h) ensemble forecast data against manual reports for the first 3 months of 2002 in the United States. Many probabilistic metrics are applied but for the deterministic output, the POD is the main scalar metric used. Of note is that Wandishin et al. (2005) remove sample events when surface temperatures greater than $T > 5$ °C because these are cases where a rain diagnosis is ‘easy’.

The metrics used in the current literature to perform verification of SPT data vary widely between studies, which makes comparison difficult. Furthermore, there is typically an absence of justification for the choice of scalar metrics used in the verification. However, the most commonly used metrics are bias, POD, FAR, and HSS. Comparison between studies is critical in order to assess which measurement or forecasting techniques are more accurate in space, time and by SPT. Knowing where some methods succeed and where others fail is useful to combine and improve methods for rapid improvement of SPT observation or forecast diagnosis. This thesis addresses the ambiguities of SPT verification in several ways, which are outlined in the next section.

5. Thesis Objectives and Structure

Surface precipitation type has a high impact on society but is poorly observed and poorly forecast, due, in part, to the lack of an unambiguous measurement standard and the absence of a suitable verification framework to assess the skill of measurement techniques.

The objectives and questions of this thesis are:

1. Collect a dataset of low-cost, high quality in-situ surface precipitation type measurements in the UK.
    a) Build and validate a low-cost network of precipitation type instruments.
    b) Make the observations available in real-time to end users.
    c) Make the observational dataset available to the wider research community.
2. Evaluate the new and the existing measurement techniques for in-situ surface precipitation type measurements.
    a) Are the new instruments comparable to those sensing methods already used?
    b) How should SPTs be classified?
    c) What is a suitable approach to evaluate discrete, non-probabilistic SPT observations?
    d) How does the evaluation result differ at various timescales?
3. Perform a verification of a radar-based surface precipitation type product in the UK.
    a) What is a suitable statistical method to perform verification of spatial aerial products with surface-based point reference data?
    b) What is the skill of surface precipitation type diagnoses from existing spatial, gridded products?
    c) How does the diagnosis skill vary between precipitation types?
    d) When and where does the skill vary—and what underlying processes are causing discrepancies?
    e) Using a consistent verification skill methodology, how does the result differ between reference datasets?

f) Can the skill of existing SPT products be improved by modifying the classification structure or the underlying algorithm?

Chapter 2 describes the development, deployment, and continued maintenance of a new network of 14 precipitation sensors in the UK which was named DiVeN (objective 1a). The deployment of these sensors aimed to address the lack of high-quality SPT observations available through the existing surface station network, which rely on a visiometer, rain gauge and thermometer to determine SPT. The deployed sensors (Thies LPMs) instead use the occultation technique described in Section 3.1 to estimate the SPT, and their network forms the first of its kind in the UK. The research highlights the difficulties that were overcome in creating such a network on a restricted budget, and the challenges that were faced in maintaining a continuous dataset. A real-time website (objective 1b) was created to encourage the regular usage of the instrument network for forecasting. The publication of a community-accessible dataset in a compliant format (objective 1c) aimed to encourage further research beyond the scope of the research conducted in this thesis. A qualitative validation through three case studies against manual trained observers is described which motivates the study performed in Chapter 3.

Chapter 3 describes the evaluation of six precipitation measurement techniques using seven sensors at the CAO. Two of the sensors were co-located Thies LPMs from DiVeN described in Chapter 2. The purpose of this research is to evaluate the instruments to determine if they meet or exceed the quality of existing SPT-measuring instruments similar to those used in the existing surface station network from the Met Office (objective 2a). All instruments are compared across a 12-month period for four precipitation variables: precipitation rate/depth, drop diameter, drop velocity, and SPT diagnosis. The wider evaluation, in addition to the SPT comparisons, is undertaken primarily to support the SPT results since some instruments use the diameter and velocity measurements to diagnose SPT. Furthermore, the results for other variables were evaluated with the aim of informing decisions on which sensors to use for future sensor deployments. A new SPT classification standard is developed (objective 2b) and statistical metrics are chosen which are suitable for SPT data (objective 2c). Time scales were varied with the aim of showing the limit of high temporal resolution where some measurement techniques lose quality (objective 2d), which is important since higher temporal resolution allows for faster response to an SPT event and more accurate quantification at high precipitation rates.

Chapter 4 describes the verification of a spatial SPT product for rain, mixed-phase and snow, which uses an empirical relationship between radar-derived rainfall rate and the height of the 0 °C wet-bulb isotherm from an NWP model. The product is operational within the Met Office and this research is the first time that a spatial, radar-based SPT product has received verification in the UK. A total of 5 years of the product are sampled and three ground datasets are used as a reference, including the Met Office surface stations and the DiVeN instruments from Chapter 2. The verification is performed with the aim of informing users when the product is uncertain, and to motivate improvements to a future version of the product. The choice of statistical metrics and skill scores, as well as the method of verification are specifically appropriate for SPT data (objective 3a). The approach taken is novel compared to the literature and the aim is for these methods to be applied to future verifications of spatial SPT products. Chapter 5 describes the verification of a hail-specific spatial SPT product from the UK Met Office. The methods from Chapter 4 are used in a limited spatial SPT verification using crowdsourced hail reports as the point reference because the automated ground point instruments have poor skill for hail diagnosis. The main purpose of the research in Chapter 4 and Chapter 5 was to provide a baseline skill level (objective 3b) and to identify weaknesses with the current product (objectives 3c, 3d) in order to motivate improvements for a future SPT product (objective 3f).

The results of Chapters 2, 3, 4 and 5 are summarised in Chapter 6 along with a discussion and synthesis of the results for their wider impact to the scientific field. Recommendations for future research in a research sense are made as well as recommendations for changes to operational observing practices.

References

Agnew, M. D., and J. E. Thornes, 1995: The weather sensitivity of the UK food retail and distribution industry. *Meteorological Applications*, **2 (2)**, 137–147, doi:10.1002/met.5060020207.

Al-Sakka, H., A. A. Boumahmoud, B. Fradon, S. J. Frasier, and P. Tabary, 2013: A new fuzzy logic hydrometeor classification scheme applied to the french X-, C-, and S-band polarimetric radars. *Journal of Applied Meteorology and Climatology*, **52 (10)**, 2328–2344, doi:10.1175/JAMC-D-12-0236.1.

Allen, J. T., I. M. Giammanco, M. R. Kumjian, H. Jurgen Punge, Q. Zhang, P. Groenemeijer, M. Kunz, and K. Ortega, 2020: Understanding Hail in the Earth System. *Reviews of Geophysics*, **58 (1)**, 1–49, doi:10.1029/2019RG000665.

AMS, 2012a: Graupel. Glossary of Meteorology. URL https://glossary.ametsoc.org/wiki/Graupel.

AMS, 2012b: Hail. Glossary of Meteorology. URL https://glossary.ametsoc.org/wiki/Hail.

AMS, 2012c: Sleet. Glossary of Meteorology. URL https://glossary.ametsoc.org/wiki/Sleet.

Austin, P. M., 1987: Relation between measured radar reflectivity and surface rainfall. *Monthly Weather Review*, **115 (5)**, 1053–1070.

Aydin, K., T. A. Seliga, and V. Balaji, 1986: Remote sensing of hail with a dual linear polarization radar. *Journal of climate and applied meteorology*, **25 (10)**, 1475–1484.

Bader, H., 1964: Density of ice as a function of temperature and stress.

Baker, J. D., and E. W. Friday, 1996: Observing Handbook No.8 Aviation Weather Observations for Supplementary Aviation Weather Reporting Stations (SAWRS). Tech. Rep. 8, The United States National Weather Service.

Battan, L. J., 1976: Vertical air motions and the ZR relation. *Journal of Applied Meteorology*, **15 (10)**, 1120–1121.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525 (7567)**, 47–55, URL http://dx.doi.org/10.1038/

nature14956http://10.0.4.14/nature14956.

Bergeron, T., 1935: On the Physics of Clouds and Precipitation. *International Union of Geodesy and Geophysics*, 156–178.

Berne, A., and W. F. Krajewski, 2013: Radar for hydrology: Unfulfilled promise or unrecognized potential? *Advances in Water Resources*, **51**, 357–366, doi:10.1016/j.advwatres.2012.05.005, URL http://dx.doi.org/10.1016/j.advwatres.2012.05.005.

Bloemink, H., and E. Lanzinger, 2005: Precipitation type from the Thies disdrometer. *WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO-2005)*, 1–7, URL http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-82-TECO_2005/Papers/3(11)_Netherlands_4_Bloemink.pdf.

Bohm, H. P., 1989: A general equation for the terminal fall speed of solid hydrometeors. 2419–2427 pp., doi:10.1175/1520-0469(1989)046⟨2419:AGEFTT⟩2.0.CO;2.

Bottomley, G. A., 1978: The vapour pressure of supercooled water and heavy water. *Australian Journal of Chemistry*, **31 (6)**, 1177–1180.

Boudala, F. S., G. A. Isaac, P. Filman, R. Crawford, D. Hudak, and M. Anderson, 2017: Performance of emerging technologies for measuring solid and liquid precipitation in cold climate as compared to the traditional manual gauges. *Journal of Atmospheric and Oceanic Technology*, **34 (1)**, 167–185, doi:10.1175/JTECH-D-16-0088.1.

Bringi, V. N., and V. Chandrasekar, 2001: *Polarimetric Doppler weather radar: principles and applications*. Cambridge university press.

Brown, R., 1828: XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *The Philosophical Magazine*, **4 (21)**, 161–173.

Brown, T. M., W. H. Pogorzelski, and I. M. Giammanco, 2015: Evaluating hail damage using property insurance claims data. *Weather, Climate, and Society*, **7 (3)**, 197–210, doi:10.1175/WCAS-D-15-0011.1.

Carey, L. D., and S. A. Rutledge, 1996: A multiparameter radar case study of the microphysical and kinematic evolution of a lightning producing storm. *Meteorology and Atmospheric Physics*, **59 (1-2)**, 33–64, doi:10.1007/BF01032000.

Carrière, J. M., C. Lainard, C. Le Bot, and F. Robart, 2000: A climatological study of surface freezing precipitation in Europe. *Meteorological Applications*, **7 (3)**, 229–238, doi:10.1017/S1350482700001560.

Chandrasekar, V., V. N. Bringi, N. Balakrishnan, and D. S. Zrnić, 1990: Error structure of multiparameter radar and surface measurements of rainfall. Part III: Specific differential phase. *Journal of Atmospheric and Oceanic Technology*, **7 (5)**, 621–629.

Chen, S., J. J. Gourley, Y. Hong, Q. Cao, N. Carr, P. E. Kirstetter, J. Zhang, and Z. Flamig, 2016: Using Citizen Science Reports to Evaluate Estimates of Surface Precipitation Type. *Bulletin of the American Meteorological Society*, **97 (2)**, 187–193, doi:10.1175/BAMS-D-13-00247.1.

Clark, M. R., 2011: An observational study of the exceptional 'Ottery St Mary'thunderstorm of 30 October 2008. *Meteorological Applications*, **18 (2)**, 137–154.

Clark, M. R., J. D. C. Webb, and P. J. Kirk, 2018: Fine-scale analysis of a severe hailstorm using crowd-sourced and conventional observations. *Meteorol. Appl.*, **492 (April)**, 472–492, doi:10.1002/met.1715.

Clayton, H. H., 1927: A method of verifying weather forecasts. *Bulletin of the American Meteorological Society*, **8 (10)**, 144–146.

Clayton, H. H., 1934: Rating weather forecasts. *Bulletin of the American Meteorological Society*, **15 (12)**, 279–283.

Cobb, S., 2004: Weather radar development highlight of national severe storms laboratory's first 40 years. *NOAA Magazine. National Oceanic and Atmospheric Administration. Retrieved 2017-03-07.*

Collier, C. G., 1989: *Applications of weather radar systems.* Wiley.

Connolly, P. J., C. Emersic, and P. R. Field, 2012: A laboratory investigation into the aggregation efficiency of small ice crystals. *Atmospheric Chemistry and Physics*, **12 (4)**, 2055.

Connolly, P. J., C. P. R. Saunders, M. W. Gallagher, K. N. Bower, M. J. Flynn, T. W. Choularton, J. Whiteway, and R. P. Lawson, 2005: Aircraft observations of the influence of electric fields on the aggregation of ice crystals. *Quarterly Journal of the Royal Meteorological Society*, **131 (608)**, 1695–1712.

Conway, J. W., and D. S. Zrnić, 1993: A study of embryo production and hail growth using dual-Doppler and multiparameter radars. *Monthly weather review*, **121 (9)**, 2511–2528.

Courtier, B. M., T. H. Stein, R. G. Harrison, K. E. Hanley, and J. M. Wilkinson, 2019: Intensification of single cell storms prior to lightning onset. *Atmospheric Science Letters*, **20 (4)**, 1–7, doi:10.1002/asl.873.

Czys, R. R., R. W. Scott, K. C. Tang, R. W. Przybylinski, and M. E. Sabones, 1996: A physically based, nondimensional parameter for discriminating between locations of freezing rain and ice pellets. *Weather and Forecasting*, **11 (4)**, 591–598, doi:10.1175/1520-0434(1996)011⟨0591:APBNPF⟩2.0.CO;2.

Das, S. K., M. Konwar, K. Chakravarty, and S. M. Deshpande, 2017: Raindrop size distribution of different cloud types over the Western Ghats using simultaneous measurements from Micro-Rain Radar and disdrometer. *Atmospheric Research*, **186**, 72–82, doi:10.1016/j.atmosres.2016.11.003, URL http://dx.doi.org/10.1016/j.atmosres.2016.11.003.

David, R. O., and Coauthors, 2019: Pore condensation and freezing is responsible for ice formation below water saturation for porous particles. *Proceedings of the National Academy of Sciences of the United States of America*, **116 (17)**, 8184–8189, doi:10.1073/pnas.1813647116.

Deierling, W., W. A. Petersen, J. Latham, S. Ellis, and H. J. Christian, 2008: The relationship between lightning activity and ice fluxes in thunderstorms. *Journal of Geophysical Research Atmospheres*, **113 (15)**, 1–20, doi:10.1029/2007JD009700.

Dennis, E. J., and M. R. Kumjian, 2017: The impact of vertical wind shear on hail growth in simulated supercells. *Journal of the Atmospheric Sciences*, **74 (3)**, 641–663, doi:10.1175/JAS-D-16-0066.1.

Doolittle, M. H., 1888: Association ratios. *Bull. Philos. Soc. Washington*, **7**, 122–127.

Doswell III, C. A., R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Weather and Forecasting*, **5 (4)**, 576–585.

Dotzek, N., P. Groenemeijer, B. Feuerstein, and A. M. Holzer, 2009: Overview of ESSL's severe convective storms research using the European Severe Weather

Database ESWD. *Atmospheric Research*, **93 (1-3)**, 575–586, doi:10.1016/j.atmosres.2008.10.020, URL http://dx.doi.org/10.1016/j.atmosres.2008.10.020.

Dufton, D. R. L., and C. G. Collier, 2015: Fuzzy logic filtering of radar reflectivity to remove non-meteorological echoes using dual polarization radar moments. *Atmospheric Measurement Techniques*, **8 (10)**, 3985–4000, doi:10.5194/amt-8-3985-2015.

ECS, 2002: European Committee for Standardization. BS EN 13612:2002 Incorporating Corrigendum No.1 - Performance evaluation of in vitro diagnostic medical devices. Tech. rep., British Standards Institute.

ECS, 2012: European committee on Standardization. Medical laboratories - Requirements for quality and competence (ISO 15189:2012). Tech. rep., British Standards Institution, 1–50 pp.

Einstein, A., 1905: On the electrodynamics of moving bodies. *Annalen der physik*, **17 (10)**, 891–921.

Eisenberg, D., S. Laustsen, and J. Stege, 2018: Wind turbine blade coating leading edge rain erosion model: Development and validation. *Wind Energy*, **21 (10)**, 942–951, doi:10.1002/we.2200.

Elmore, K. L., Z. L. Flamig, V. Lakshmanan, B. T. Kaney, V. Farmer, H. D. Reeves, and L. P. Rothfusz, 2014: MPING: Crowd-sourcing weather reports for research. *Bulletin of the American Meteorological Society*, **95 (9)**, 1335–1342, doi:10.1175/BAMS-D-13-00014.1.

Elmore, K. L., H. M. Grams, D. Apps, and H. D. Reeves, 2015: Verifying forecast precipitation type with mPING. *Weather and Forecasting*, **30 (3)**, 656–667, doi:10.1175/WAF-D-14-00068.1.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, **21 (6)**, 739–759.

Erkal, A., D. D'Ayala, and L. Sequeira, 2012: Assessment of wind-driven rain impact, related surface erosion and surface strength reduction of historic building materials. *Building and Environment*, **57**, 336–348, doi:10.1016/j.buildenv.2012.05.004.

Fabry, F., 2015: *Radar Meteorology: Principles and Practice.* Cambridge University Press, 254 pp., doi:10.1017/CBO9781107707405.

Fabry, F., and I. Zawadzki, 1995: Long-term radar observations of the melting layer of precipitation and their interpretation. *Journal of the atmospheric sciences*, **52 (7)**, 838–851.

Fairman, J. G., D. M. Schultz, D. J. Kirshbaum, S. L. Gray, and A. I. Barrett, 2015: A radar-based rainfall climatology of Great Britain and Ireland. *Weather*, **70 (5)**, 153–158, doi:10.1002/wea.2486.

Feingold, G., Z. Levin, and S. Tzivion, 1991: The evolution of raindrop spectra. Part III: downdraft generation in an axisymmetrical rainshaft model. 315–330 pp., doi: 10.1175/1520-0469(1991)048⟨0315:TEORSP⟩2.0.CO;2.

Ferro, C. A., and D. B. Stephenson, 2011: Extremal dependence indices: Improved Verification measures for deterministic forecasts of rare binary events. *Weather and Forecasting*, **26 (5)**, 699–713, doi:10.1175/WAF-D-10-05030.1.

Field, P. R., and Coauthors, 2016: Chapter 7. Secondary Ice Production - current state of the science and recommendations for the future. *Meteorological Monographs*, doi:10.1175/amsmonographs-d-16-0014.1.

Findeisen, W., 1938: Kolloid-meteorologische Vorgänge bei Neiderschlags-bildung. *Meteor. Z*, **55**, 121–133.

Finley, J. P., 1884: Tornado predictions. *American Meteorological Journal. A Monthly Review of Meteorology and Allied Branches of Study (1884-1896)*, **1 (3)**, 85.

Foote, G. B., 1984: A study of hail growth utilizing observed storm conditions. *Journal of climate and applied meteorology*, **23 (1)**, 84–101.

Frisby, E. M., 1961: Relationship of Ground Hail Damage Patterns to Features of the Synoptic Map in the Upper Great Plains of the United States. *Journal of Applied Meteorology*, **1 (3)**, 348–352.

Fukuta, N., and C. M. Gramada, 2003: Vapor pressure measurement of supercooled water. *Journal of the atmospheric sciences*, **60 (15)**, 1871–1875.

Garrett, T. J., and S. E. Yuter, 2014: Observed influence of riming, temperature, and turbulence on the fallspeed of solid precipitation. *Geophysical Research Letters*, **41 (18)**, 6515–6522, doi:10.1002/2014GL061016.

Gascón, E., T. Hewson, and T. Haiden, 2018: Improving Predictions of Precipitation Type at the Surface: Description and Verification of Two New Products from the ECMWF Ensemble. *Weather and Forecasting*, **33 (1)**, 89–108, doi:10.1175/WAF-D-17-0114.1, URL http://journals.ametsoc.org/doi/10.1175/WAF-D-17-0114.1.

Gerrity Jr., J. P., 1992: A Note on Gandin and Murphy's Equitable Skill Score. *Monthly Weather Review*, **120 (11)**, 2709–2712, doi:10.1175/1520-0493(1992)120⟨2709:ANOGAM⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0493(1992)120%3C2709:ANOGAM%3E2.0.CO.

Gilbert, G. K., 1884: Finley's tornado predictions. *American Meteorological Journal. A Monthly Review of Meteorology and Allied Branches of Study (1884-1896)*, **1 (5)**, 166.

Gourley, J. J., P. Tabary, and J. Parent du Chatelet, 2007: A fuzzy logic algorithm for the separation of precipitating from nonprecipitating echoes using polarimetric radar observations. *Journal of Atmospheric and Oceanic Technology*, **24 (8)**, 1439–1451, doi:10.1175/JTECH2035.1.

Green, A., 2010: From Observations to Forecasts – Part 7 . A new meteorological monitoring system for the United Kingdom's Met Office. *Weather*, **65 (10)**, 272–277.

Gunn, R., and G. D. Kinzer, 1949: The Terminal Velocity of Fall for Water Droplets in Stagnant Air. *Journal of Meteorology*, 243–248, doi:10.1175/1520-0469(1949)006⟨0243:TTVOFF⟩2.0.CO;2.

Guntur, P., and M. Tippett, 2017: Impact of ENSO on U.S. Tornado and Hail frequencies. Tech. Rep. March, Willis Re, Minneapolis, MN, USA. doi:10.1175/JAMC-D-16-0249.1.

Hall, M. P. M., J. W. F. Goddard, and S. M. Cherry, 1984: Identification of hydrometeors and other targets by dual-polarization radar. *Radio Science*, **19 (1)**, 132–140.

Hallett, J., and B. J. Mason, 1958: The influence of temperature and supersaturation on the habit of ice crystals grown from the vapour. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **247 (1251)**, 440–453.

Hallett, J., and S. C. Mossop, 1974: Production of secondary ice particles during the riming process. *Nature*, **249 (5452)**, 26–28, doi:10.1038/249026a0, URL https://doi.org/10.1038/249026a0.

Hansen, A. W., and W. J. A. Kuipers, 1965: *On the Relationship Between the Frequency of Rain and Various Meteorological Parameters*, Vol. 81. 2–15 pp.

Harrison, D., S. J. Driscoll, and M. Kitchen, 2000: Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorol. Appl.*, **6**, 135–144, doi:10.1017/S1350482700001468.

Harrison, D., K. Norman, T. Darlington, D. Adams, N. Husnoo, C. Sandford, and S. Best, 2015: The evolution of the Met Office radar data quality control and product generation system: Radarnet. *Proceedings of the 37th AMS Conference on Radar Meteorology, Norman, OK, USA*, 14–18.

Heidke, P., 1926: Berechnung Des Erfolges Und Der Güte Der Windstärkevorhersagen Im Sturmwarnungsdienst. *Geografiska Annaler*, **8 (4)**, 301–349, doi:10.1080/20014422.1926.11881138, URL https://doi.org/10.1080/20014422.1926.11881138.

Heymsfield, A., M. Szakáll, A. Jost, I. Giammanco, and R. Wright, 2018: A comprehensive observational study of graupel and hail terminal velocity, mass flux, and kinetic energy. *Journal of the Atmospheric Sciences*, **75 (11)**, 3861–3885, doi:10.1175/JAS-D-18-0035.1.

Hogan, R. J., E. J. O'Connor, and A. J. Illingworth, 2009: Verification of cloud-fraction forecasts. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **135 (643)**, 1494–1511.

Hohl, R., H. H. Schiesser, and D. Aller, 2002: Hailfall: The relationship between radar-derived hail kinetic energy and hail damage to buildings. *Atmospheric Research*, **63 (3-4)**, 177–207, doi:10.1016/S0169-8095(02)00059-5.

Hosler, C. L., and R. E. Hallgren, 1960: The aggregation of small ice crystals. *Discussions of the Faraday Society*, **30**, 200–207.

Hosler, C. L., D. C. Jensen, and L. Goldshlak, 1957: On the aggregation of ice crystals to form snow. *Journal of Meteorology*, **14 (5)**, 415–420.

Houghton, H. G., 1985: *Physical meteorology.* MIT Press, Cambridge, MA, 442 pp.

IEEE, 2011: Institute of Electrical and Electronics Engineers. Adoption of the Project Management Institute (PMI) Standard: A Guide to the Project Management Body of Knowledge (PMBOK Guide) (4th edition). 1–505, doi:10.1109/IEEESTD.2011.

5937011.

Ikeda, K., M. Steiner, J. Pinto, and C. Alexander, 2013: Evaluation of cold-season precipitation forecasts generated by the hourly updating high-resolution rapid refresh model. *Weather and Forecasting*, **28 (4)**, 921–939, doi:10.1175/WAF-D-12-00085.1.

Jameson, A. R., and A. B. Kostinski, 2001: What is a raindrop size distribution? *Bulletin of the American Meteorological Society*, **82 (6)**, 1169–1177, doi:10.1175/1520-0477(2001)082⟨1169:WIARSD⟩2.3.CO;2.

Jameson, S., 2018: Forecasting Freezing Rain in the UK – March 1st and 2nd 2018. *euroforecaster*, **(23)**, 16–23, URL http://www.euroforecaster.org/newsletter23/freezing_rain.pdf.

Joss, J., and A. Waldvogel, 1970: Raindrop size distributions and Doppler velocities, 14th Radar Meteor. *Conf., Tuscon, Ariz., AMS, Boston*, 153–155.

Kanji, Z. A., L. A. Ladino, H. Wex, Y. Boose, M. Burkert-Kohn, D. J. Cziczo, and M. Krämer, 2017: Overview of Ice Nucleating Particles. *Meteorological Monographs*, **58**, 1.1–1.33, doi:10.1175/amsmonographs-d-16-0006.1.

Kaspar, M., M. Müller, V. Kakos, D. Rezacova, and Z. Sokol, 2009: Severe storm in Bavaria, the Czech Republic and Poland on 12-13 July 1984: A statistic- and model-based analysis. *Atmospheric Research*, **93 (1-3)**, 99–110, doi:10.1016/j.atmosres.2008.10.004, URL http://dx.doi.org/10.1016/j.atmosres.2008.10.004.

Keegan, M. H., D. H. Nash, and M. M. Stack, 2013: On erosion issues associated with the leading edge of wind turbine blades. *Journal of Physics D: Applied Physics*, **46 (38)**, doi:10.1088/0022-3727/46/38/383001.

Kinnell, P. I., 1981: Rainfall intensity-kinetic energy relationships for soil loss prediction. *Soil Science Society of America Journal*, **45 (1)**, 153–155, doi:10.2136/sssaj1981.03615995004500010033x.

Klett, J. D., and M. H. Davis, 1973: Theoretical collision efficiencies of cloud droplets at small Reynolds numbers. *Journal of the Atmospheric Sciences*, **30 (1)**, 107–117.

Knight, C. A., D. H. Ehhalt, N. Roper, and N. C. Knight, 1975: Radial and tangential variation of deuterium in hailstones. *Journal of the Atmospheric Sciences*, **32 (10)**, 1990–2000.

Knight, C. A., and N. C. Knight, 1970: Lobe structures of hailstones. *Journal of the Atmospheric Sciences*, **27 (4)**, 667–671.

Knight, N. C., and A. J. Heymsfield, 1983: Measurement and interpretation of hailstone density and terminal velocity. *Journal of the atmospheric sciences*, **40 (6)**, 1510–1516.

Kobayashi, T., 1958: On the Habit of Snow Crystals Artificially Produced at Low Pressures. *Journal of the Meteorological Society of Japan. Ser. II*, **36 (5)**, 193–208, doi:10.2151/jmsj1923.36.5_193.

Kobayashi, T., 1961: The growth of snow crystals at low supersaturations. *Philosophical Magazine*, **6 (71)**, 1363–1370.

Koistinen, J., and E. Saltikoff, 1998: Experience of customer products of accumulated snow, sleet and rain. *COST75 Advanced Weather Radar Systems*, 397–406.

Kraus, G. F., and S. C. Greer, 1984: Vapor pressures of supercooled water and deuterium oxide. *The Journal of Physical Chemistry*, **88 (20)**, 4781–4785.

Langleben, M. P., 1954: The terminal velocity of snowflakes. *Quarterly Journal of the Royal Meteorological Society*, **80 (344)**, 174–181, doi:10.1002/qj.49708034404.

Latham, J., and C. P. R. Saunders, 1970: Experimental measurements of the collection efficiencies of ice crystals in electric fields. *Quarterly Journal of the Royal Meteorological Society*, **96 (408)**, 257–265.

Lau, K. M., and H. T. Wu, 2003: Warm rain processes over tropical oceans and climate implications. *Geophysical Research Letters*, **30 (24)**, doi:10.1029/2003GL018567.

Laws, J. O., 1941: Measurements of the fall-velocity of water-drops and raindrops. *Eos, Transactions American Geophysical Union*, **22 (3)**, 709–721.

Laws, J. O., and D. A. Parsons, 1943: The relation of raindrop-size to intensity. *Eos, Transactions American Geophysical Union*, **24 (2)**, 452–460.

Lenard, P., 1904: Über Regen. *Meteorol. Z*, **6 (249)**, 62–92.

Lin, C., S. Vasić, A. Kilambi, B. Turner, and I. Zawadzki, 2005: Precipitation forecast skill of numerical weather prediction models and radar nowcasts. *Geophysical Research Letters*, **32 (14)**, 1–4, doi:10.1029/2005GL023451.

Lin, C. L., and S. C. Lee, 1975: Collision efficiency of water drops in the atmosphere. *Journal of the Atmospheric Sciences*, **32 (7)**, 1412–1418.

Lin, D., B. Pickering, and R. R. Neely, 2020: Relating the radar bright band and its strength to surface rainfall rate using an automated approach. *Journal of Hydrometeorology*, **21 (2)**, 335–353, doi:10.1175/JHM-D-19-0085.1.

Liu, H., and V. Chandrasekar, 2000: Classification of hydrometeors based on polarimetric radar measurements: Development of fuzzy logic and neuro-fuzzy systems, and in situ verification. *Journal of Atmospheric and Oceanic Technology*, **17 (2)**, 140–164, doi:10.1175/1520-0426(2000)017⟨0140:COHBOP⟩2.0.CO;2.

Locatelli, J. D., and P. V. Hobbs, 1974: Fall speeds and masses of solid precipitation particles. *Journal of Geophysical Research*, **79 (15)**, 2185–2197, doi:10.1029/JC079i015p02185.

Löffler-Mang, M., and J. Joss, 2000: An optical disdrometer for measuring size and velocity of hydrometeors. *Journal of Atmospheric and Oceanic Technology*, **17 (2)**, 130–139, doi:10.1175/1520-0426(2000)017⟨0130:AODFMS⟩2.0.CO;2.

Low, T. B., and R. List, 1982: Collision, coalescence and breakup of raindrops. Part I: Experimentally established coalescence efficiencies and fragment size distributions in breakup. *Journal of the Atmospheric Sciences*, **39 (7)**, 1591–1606.

Lumb, F. E., 1961: The problem of forecasting the downward penetration of snow. *Meteor. Mag*, **90**, 310–319.

Lumb, F. E., 1963: Downward penetration of snow in relation to the intensity of precipitation. *Meteorology Magazine*, **92 (1086)**, 1–14.

Lyth, D., and M. Molyneux, 2006: Results of Using Present Weather Instruments in the United Kingdom. Tech. rep., The Met Office, 46 pp.

Macklin, W. C., 1962: The density and structure of ice formed by accretion. *Quarterly Journal of the Royal Meteorological Society*, **88 (375)**, 30–50.

Magono, C., 1953: On the growth of snow flake and graupel. *Science Reports of the Yokohama National University*, **1 (2)**, 18–40.

Marshall, J. S., R. C. Langille, and W. M. K. Palmer, 1947: Measurement of Rainfall By Radar. *Journal of Meteorology*, **4 (6)**, 186–192, doi:10.1175/1520-0469(1947)

004⟨0186:MORBR⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0469%281947%29004%3C0186%3AMORBR%3E2.0.CO%3B2.

Marshall, J. S., and W. M. Palmer, 1948: The distribution of raindrops with size. *Journal of Meteorology*, **5**, 165–166, doi:10.1002/qj.49707632704.

Mason, B. J., 1971: *The Physics of Clouds*. Clarendon Press, Oxford, 671 pp.

Matson, R. J., and A. W. Huggins, 1980: The Direct Measurement of the Sizes, Shapes and Kinematics of Falling Hailstones. 1107–1125 pp., doi:10.1175/1520-0469(1980)037⟨1107:tdmots⟩2.0.co;2.

Maxwell, J. C., 1890: *The Scientific Papers of James Clerk Maxwell*, Vol. 2. University Press.

McTaggart-Cowan, J. D., and R. List, 1975: Collision and breakup of water drops at terminal velocity. *Journal of the Atmospheric Sciences*, **32 (7)**, 1401–1411.

Merenti-Välimäki, H. L., J. Lönnqvist, and P. Laininen, 2001: Present weather: Comparing human observations and one type of automated sensor. *Meteorological Applications*, **8 (4)**, 491–496, doi:10.1017/S1350482701004108.

Met Office, 2014: Fact Sheet 7 - Climate. Tech. rep., National Meteorological Library and Archive, 1–22 pp. URL http://www.metoffice.gov.uk/media/pdf/c/n/MetLIB_13_013_FactSheet_7_Final.pdf.

Met Office, 2019a: Definition of Sleet. URL https://www.metoffice.gov.uk/weather/learn-about/weather/types-of-weather/snow/sleet.

Met Office, 2019b: Fact Sheet 15 – Weather Radar. Tech. rep., National Meteorological Library and Archive, 22 pp. URL https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/library-and-archive/library/publications/factsheets/factsheet_15-weather-radar.pdf.

Mitchell, D. L., 1988: Evolution of snow-size spectra in cyclonic storms. Part I: Snow growth by vapor deposition and aggregation. *Journal of the atmospheric sciences*, **45 (22)**, 3431–3451.

Mitra, S. K., O. Vohl, M. Ahr, and H. R. Pruppacher, 1990: A wind tunnel and theoretical study of the melting behavior of atmospheric ice particles. IV: Experiment and theory for snow flakes. *Journal of the Atmospheric Sciences*, **47 (5)**, 584–591.

Mittermaier, M., and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the fractions skill score. *Weather and Forecasting*, **25 (1)**, 343–354, doi:10.1175/2009WAF2222260.1.

Mittermaier, M., N. Roberts, and S. A. Thompson, 2013: A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteorological Applications*, **20 (2)**, 176–186, doi:10.1002/met.296.

Montgomery, R. B., 1947: Viscosity and thermal conductivity of air and diffusivity of water vapor in air. *Journal of Meteorology*, **4 (6)**, 193–196.

Moore, D. F., 1975: The friction of pneumatic tyres.

Murphy, A. H., 1991: Forecast verification: Its complexity and dimensionality. *Monthly Weather Review*, **119 (7)**, 1590–1601.

Murphy, D. M., and T. Koop, 2005: Review of the vapour pressures of ice and supercooled water for atmospheric applications. *Quarterly Journal of the Royal Meteorological Society*, **131 (608)**, 1539–1565, doi:10.1256/qj.04.94.

Nagumo, N., and Y. Fujiyoshi, 2015: Microphysical properties of slow-falling and fast-falling ice pellets formed by freezing associated with evaporative cooling. *Monthly Weather Review*, **143 (11)**, 4376–4392, doi:10.1175/MWR-D-15-0054.1.

Nakaya, U., 1954: Snow Crystals, Natural and Artificial. Harvard University Press, Cambridge, MA, 510 pp.

Nakaya, U., and T. Terada, 1935: Simultaneous Observations of the Mass, Falling Velocity and Form of Individual Snow Crystals. *Journal of the Faculty of Science, Hokkaido Imperial University. Series 2, Physics*, **1 (7)**, 191–200.

Nelson, S. P., 1983: The Influence of Storm Flow Structure on Hail Growth. *Journal of the Atmospheric Sciences*, **40 (8)**, 1965–1983, doi:10.1175/1520-0469(1983)040⟨1965:TIOSFS⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0469(1983)040%3C1965:TIOSFS%3E2.0.CO;2.

Park, H. S., A. V. Ryzhkov, D. S. Zrnić, and K.-E. Kim, 2009: The Hydrometeor Classification Algorithm for the Polarimetric WSR-88D: Description and Application to an MCS. *Weather and Forecasting*, **24 (3)**, 730–748, doi:10.1175/2008WAF2222205.1.

Parker, D. E., T. P. Legg, and C. K. Folland, 1992: A new daily central England temperature series, 1772–1991. *International Journal of Climatology*, **12 (4)**, 317–342, doi:10.1002/joc.3370120402.

Peirce, C. S., 1884: The numerical measure of the success of predictions. *Science*, **(93)**, 453–454.

Phillips, V. T., M. Formenton, A. Bansemer, I. Kudzotsa, and B. Lienert, 2015: A parameterization of sticking efficiency for collisions of snow and graupel with ice crystals: Theory and comparison with observations. *Journal of the Atmospheric Sciences*, **72 (12)**, 4885–4902, doi:10.1175/JAS-D-14-0096.1.

Pruppacher, H. R., and K. V. Beard, 1970: A wind tunnel investigation of the internal circulation and shape of water drops falling at terminal velocity in air. *Quarterly Journal of the Royal Meteorological Society*, **96 (408)**, 247–256, doi:10.1002/qj.49709640807.

Pruppacher, H. R., and J. D. Klett, 1997: *Microphysics of Clouds and Precipitation.* 2nd ed., Reidel, Dordrecht.

Punge, H. J., K. M. Bedka, M. Kunz, and A. Werner, 2014: A new physically based stochastic event catalog for hail in Europe. *Natural Hazards*, **73 (3)**, 1625–1645, doi:10.1007/s11069-014-1161-0.

Punge, H. J., and M. Kunz, 2016: Hail observations and hailstorm characteristics in Europe: A review. *Atmospheric Research*, **176-177**, 159–184, doi:10.1016/j.atmosres.2016.02.012, URL http://dx.doi.org/10.1016/j.atmosres.2016.02.012.

Ralph, F. M., and Coauthors, 2005: Improving short-term (0-48 h) cool-season quantitative precipitation forecasting: Recommendations from a USWRP Workshop. *Bulletin of the American Meteorological Society*, **86 (11)**, 1619–1632, doi:10.1175/BAMS-86-11-1619.

Raoult, F.-M., 1887: Loi générale des tensions de vapeur des dissolvants. *CR Hebd. Seances Acad. Sci*, **104**, 1430–1433.

Rauber, R. M., M. K. Ramamurthy, and A. Tokay, 1994: Synoptic and mesoscale structure of a severe freezing rain event: The St. Valentine's Day ice storm. *Weather and forecasting*, **9 (2)**, 183–208.

Reeves, H. D., 2016: The uncertainty of precipitation-type observations and its effect on the validation of forecast precipitation type. *Weather and Forecasting*, **31 (6)**, 1961–1971, doi:10.1175/WAF-D-16-0068.1.

Reynolds, O., 1883: XXIX. An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels. *Philosophical Transactions of the Royal society of London*, **(174)**, 935–982.

Rico-Ramirez, M. A., and I. D. Cluckie, 2008: Classification of ground clutter and anomalous propagation using dual-polarization weather radar. *IEEE Transactions on Geoscience and Remote Sensing*, **46 (7)**, 1892–1904, doi:10.1109/TGRS.2008.916979.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, **136 (1)**, 78–97.

Rogers, R. R., and M. K. Yau, 1996: *A short course in cloud physics*. Elsevier, 290 pp.

Roos, D. S., and A. E. Carte, 1973: The falling behavior of oblate and spiky hailstones. *Journal de Recherches Atmospheriques*, **7**, 39–52.

Rosewell, C. J., 1986: Rainfall kinetic energy in eastern Australia. 1695–1701 pp., doi:10.1175/1520-0450(1986)025⟨1695:RKEIEA⟩2.0.CO;2.

Rossi, P. H., M. W. Lipsey, and G. T. Henry, 2018: *Evaluation: A systematic approach*. Sage publications.

Ryzhkov, A. V., and D. S. Zrnic, 2019: *Radar Polarimetry for Weather Observations*. 1st ed., Springer International Publishing, 486 pp., doi:10.1007/978-3-030-05093-1.

Saunders, C. P. R., and N. M. A. Wahab, 1975: The influence of electric fields on the aggregation of ice crystals. *Journal of the Meteorological Society of Japan.*, **53 (2)**, 121–126.

Scheel, K., and W. Heuse, 1909: Bestimmung des Sättigungsdruckes von Wasserdampf unter 0. *Annalen der Physik*, **334 (9)**, 723–737.

Schemm, S., L. Nisi, A. Martinov, D. Leuenberger, and O. Martius, 2016: On the link between cold fronts and hail in Switzerland. *Atmospheric Science Letters*, **17 (5)**, 315–325, doi:10.1002/asl.660.

Schlamp, R. J., S. N. Grover, H. R. Pruppacher, and A. E. Hamielec, 1976: A Numerical

Investigation of the Effect of Electric Charges and Vertical External Electric Fields and the Collision Efficiency of Cloud Drops. *Journal of Atmospheric Science*, **33**, 1747–1755.

Schmid, W., and A. Mathis, 2004: Validation of methods to detect winter precipitation and retrieve precipitation type. *Annalen der Meteorologie*, **(January)**, 1–8, URL http://www.iac.ethz.ch/staff/wueest/sirwec/conferences/bingen2004/topic_iv/sirwec-iv-3-schmid.pdf.

Schmidt, W., 1909: Eine unmittelbare bestimmung der fallgeschwindigkeit von regentropfen. *Sitz. Akad. Wiss. Wien. Mathem.-naturw. Klasse*, **118**, 71–84.

Seinfeld, J. H., and S. N. Pandis, 2006: *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. 2nd ed., Wiley, 1248 pp.

Seliga, T. A., and V. N. Bringi, 1978: Differential reflectivity and differential phase shift: Applications in radar meteorology. *Radio Science*, **13 (2)**, 271–275.

Sheppard, B. E., and P. I. Joe, 2000: Automated precipitation detection and typing in winter: A two-year study. *Journal of Atmospheric and Oceanic Technology*, **17 (11)**, 1493–1507, doi:10.1175/1520-0426(2000)017⟨1493:APDATI⟩2.0.CO;2.

Singleton, F., 1989: Commercial aspects of the application of Meteorology. *Meteorological Magazine*, **118 (1407)**, 217–222.

Slot, H. M., E. R. Gelinck, C. Rentrop, and E. Van der Heide, 2015: Leading edge erosion of coated wind turbine blades: Review of coating life models. *Renewable Energy*, **80**, 837–848, doi:10.1016/j.renene.2015.02.036, URL http://dx.doi.org/10.1016/j.renene.2015.02.036.

Smyth, T. J., and A. J. Illingworth, 1998: Radar estimates of rainfall rates at the ground in bright band and non-bright band events. *Quarterly Journal of the Royal Meteorological Society*, **(124)**, 2417–2434.

Sommerfeld, A., 1908: *Ein Beitrag zur hydrodynamischen Erklärung der turbulenten Flüssigkeitsbewegungen*. Rome, 116–124 pp.

Sparrius, A., 2016: Everything You Thought You Knew about Validation and Verification is Probably Dodgy. *12th INCOSE SA Systems Engineering Conference*, June.

STC, 2013: Software Testing Class: Difference between software Verification and Validation. URL http://www.softwaretestingclass.com/difference-between-verification-and-validation/.

Stephenson, D. B., 2000: Use of the 'odds ratio' for diagnosing forecast skill. *Weather and Forecasting*, **15 (2)**, 221–232, doi:10.1175/1520-0434(2000)015⟨0221:UOTORF⟩2.0.CO;2.

Stokes, G. G., 1851: *On the effect of the internal friction of fluids on the motion of pendulums*, Vol. 9. Pitt Press Cambridge.

Straka, J. M., 2009: *Cloud and precipitation microphysics: principles and parameterizations*. Cambridge University Press.

Straka, J. M., and D. S. Zrnic, 1993: An algorithm to deduce hydrometeor types and contents from multiparameter radar data. *Preprints, 26th Conf. on Radar Meteorology, Norman, OK, Amer. Meteor. Soc*, 513–515.

Tabary, P., J. Desplats, K. Do Khac, F. Eideliman, C. Gueguen, and J. C. Heinrich, 2007: The New French operational radar rainfall product. Part II: Validation. *Weather and Forecasting*, **22 (3)**, 409–427, doi:10.1175/WAF1005.1.

Tang, W., and C. I. Davidson, 2004: Erosion of limestone building surfaces caused by wind-driven rain: 2. Numerical modeling. *Atmospheric Environment*, **38 (33)**, 5601–5609, doi:10.1016/j.atmosenv.2004.06.014.

Testud, J., S. Oury, R. A. Black, P. Amayenc, and X. Dou, 2001: The concept of "normalized" distribution to describe raindrop spectra: A tool for cloud physics and cloud remote sensing. *Journal of Applied Meteorology*, **40 (6)**, 1118–1140, doi:10.1175/1520-0450(2001)040⟨1118:TCONDT⟩2.0.CO;2.

Thériault, J. M., R. E. Stewart, and W. Henson, 2010: On the dependence of winter precipitation types on temperature, precipitation rate, and associated features. *Journal of Applied Meteorology and Climatology*, **49 (7)**, 1429–1442, doi:10.1175/2010JAMC2321.1.

Thomson, W., 1870: On the Size of the Molecule. *Proceedings of the Royal Society of Edinburgh.*

Thornes, E., 1992: The impact of weather and climate on transport in the UK. *Process in Physical Geography*, **16 (2)**, 187–208.

Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Weather and Forecasting*, **8 (3)**, 379–398.

U.S. NWS, 2020: Weather Symbols. URL https://www.aviationweather.gov/metar/symbol.

Villarini, G., and W. F. Krajewski, 2010: Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. *Surveys in Geophysics*, **31 (1)**, 107–129, doi:10.1007/s10712-009-9079-x.

Villermaux, E., and B. Bossa, 2009: Single-drop fragmentation determines size distribution of raindrops. *Nature Physics*, **5 (9)**, 697–702, doi:10.1038/nphys1340.

Waldvogel, A., B. Federer, and P. Grimm, 1979: Criteria for detection of hail cells. *Journal of Applied Meteorology*, **18 (12)**, 1521–1525, doi:10.1175/1520-0450(1979)018⟨1521:CFTDOH⟩2.0.CO;2.

Wandishin, M. S., M. E. Baldwin, S. L. Mullen, and J. V. Cortinas, 2005: Short-range ensemble forecasts of precipitation type. *Weather and Forecasting*, **20 (4)**, 609–626, doi:10.1175/WAF871.1.

Webb, J. D., D. M. Elsom, and G. T. Meaden, 2009: Severe hailstorms in Britain and Ireland, a climatological survey and hazard assessment. *Atmospheric Research*, **93 (1-3)**, 587–606, doi:10.1016/j.atmosres.2008.10.034, URL http://dx.doi.org/10.1016/j.atmosres.2008.10.034.

Webb, J. D., D. M. Elsom, and D. J. Reynolds, 2001: Climatology of severe hailstorms in Great Britain. *Atmospheric Research*, **56 (1-4)**, 291–308, doi:10.1016/S0169-8095(00)00081-8.

Wegener, A., 1911: *Thermodynamik der atmosphäre*. JA Barth.

Westbrook, C. D., R. C. Ball, P. R. Field, and A. J. Heymsfield, 2004: Universality in snowflake aggregation. *Geophysical Research Letters*, **31 (15)**, 1–3, doi:10.1029/2004GL020363.

Wexler, R., 1955: *The melting layer*, Vol. 3. Harvard University.

Wexler, R., and D. Atlas, 1963: Radar reflectivity and attenuation of rain. *Journal of Applied Meteorology*, **2 (2)**, 276–280.

Whale, T. F., and Coauthors, 2015: A technique for quantifying heterogeneous ice nucleation in microlitre supercooled water droplets. *Atmospheric Measurement Techniques*, **8 (6)**, 2437–2447, doi:10.5194/amt-8-2437-2015.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*, Vol. 100. Third edit ed., Academic press, 306 pp.

Wilson, J. W., and E. A. Brandes, 1979: Radar measurement of rainfall—A summary. *Bulletin of the American Meteorological Society*, **60 (9)**, 1048–1060.

WMO, 1988: Manual on Codes. *WMO Publ. 306*, **1**, 203 pp.

WMO, 1992: *International Meteorological Vocabulary.* 2nd ed., World Meteorological Organization, Geneva.

WMO, 2017: WMO Guide to Meteorological Instruments and Methods of Observation (the CIMO Guide). Chapter 6. Masurement of Precipitation. Tech. rep., World Meteorological Organization.

Woodhams, B. J., C. E. Birch, J. H. Marsham, C. L. Bain, N. M. Roberts, and D. F. Boyd, 2018: What is the added value of a convection-permitting model for forecasting extreme rainfall over tropical East Africa? *Monthly Weather Review*, **146 (9)**, 2757–2780, doi:10.1175/MWR-D-17-0396.1.

Yen, Y.-C., 1981: *Review of thermal properties of snow, ice, and sea ice*, Vol. 81. US Army, Corps of Engineers, Cold Regions Research and Engineering Laboratory.

Yule, G. U., 1900: On the association of attributes in statistics. *Philosophical Transactions of the Royal Society of London*, **194**, 257–319.

Yuter, S. E., D. E. Kingsmill, L. B. Nance, and M. Löffler-Mang, 2006: Observations of precipitation size and fall speed characteristics within coexisting rain and wet snow. *Journal of Applied Meteorology and Climatology*, **45 (10)**, 1450–1464.

Zhang, J., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans. *Bulletin of the American Meteorological Society*, **92 (10)**, 1321–1338.

Ziegler, C. L., P. S. Ray, and N. C. Knight, 1983: Hail growth in an Oklahoma multicell storm. *Journal of the atmospheric sciences*, **40 (7)**, 1768–1791.

Chapter 2.

The Disdrometer Verification Network (DiVeN): a UK network of laser precipitation instruments

Published in *Atmospheric Measurement Techniques* (2019)

Atmos. Meas. Tech., 12, 5845–5861, 2019
https://doi.org/10.5194/amt-12-5845-2019

Atmospheric Measurement Techniques
Open Access

The Disdrometer Verification Network (DiVeN): a UK network of laser precipitation instruments

Ben S. Pickering[1], **Ryan R. Neely III**[2], **and Dawn Harrison**[3]

[1]Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds,
Leeds, Yorkshire, LS2 9JT, UK
[2]National Centre for Atmospheric Science, 71-75 Clarendon Rd, Leeds,
Yorkshire, LS2 9PH, UK
[3]United Kingdom Meteorological Office, Fitzroy Rd, Exeter, EX1 3PB, UK

Correspondence: Ben S. Pickering (eebp@leeds.ac.uk)

Received: 7 September 2018 – Discussion started: 17 October 2018
Revised: 22 April 2019 – Accepted: 28 June 2019 – Published: 8 November 2019

Abstract. Starting in February 2017, a network of 14 Thies laser precipitation monitors (LPMs) were installed at various locations around the United Kingdom to create the Disdrometer Verification Network (DiVeN). The instruments were installed for verification of radar hydrometeor classification algorithms but are valuable for much wider use in the scientific and operational meteorological community. Every Thies LPM is able to designate each observed hydrometeor into one of 20 diameter bins from ≥ 0.125 to > 8 mm and one of 22 speed bins from > 0.0 to $> 20.0\,\mathrm{m\,s^{-1}}$. Using empirically derived relationships, the instrument classifies precipitation into one of 11 possible hydrometeor classes in the form of a present weather code, with an associated indicator of uncertainty. To provide immediate feedback to data users, the observations are plotted in near-real time (NRT) and made publicly available on a website within 7 min. Here we describe the Disdrometer Verification Network and present specific cases from the first year of observations. Cases shown here suggest that the Thies LPM performs well at identifying transitions between rain and snow, but struggles with detection of graupel and pristine ice crystals (which occur infrequently in the United Kingdom) inherently, due to internal processing. The present weather code quality index is shown to have some skill without the supplementary sensors recommended by the manufacturer. Overall the Thies LPM is a useful tool for detecting hydrometeor type at the surface and DiVeN provides a novel dataset not previously observed for the United Kingdom.

1 Introduction

Precipitation in all its various forms is one of the most important meteorological variables. In the UK, severe precipitation events cause millions of pounds worth of damage every year (Thornes, 1992; Penning-Rowsell and Wilson, 2006; Muchan et al., 2015). The phase of precipitation is also important. In winter, limited resources such as flood defences, ploughs, and grit will be allocated differently based on forecasts of hydrometeor type (Elmore et al., 2015; Gascón et al., 2018, and references therein). Accurate observations and forecasts of precipitation amount and type are therefore essential.

1.1 Motivation for DiVeN

Observations of precipitation are traditionally conducted with networks of tipping-bucket rain gauges (henceforth TBRs) such as the UK Met Office network described in Green (2010). TBR gauges funnel precipitation into a bucket, which tips and empties when a threshold volume is reached. The threshold volume is typically equivalent to 0.2 mm depth of rainfall, which means the TBR has a coarse resolution and struggles to measure low rainfall rates over short intervals. For example, a rain rate of $2.4\,\mathrm{mm\,h^{-1}}$ would only tip a TBR once every 5 min. Moreover, TBRs cannot detect hydrometeor type, only the liquid equivalent when the solid hydrometeors in the funnel melt naturally or from a heating element. Even liquid precipitation is poorly measured by TBRs. Ciach (2003) analysed 15 collocated TBRs and showed that

Published by Copernicus Publications on behalf of the European Geosciences Union.

considerable errors occur between the instruments, inconsistent across time and intensity scales. Finally, TBRs are easily blocked by debris and bird droppings, and the airflow around the instrument has been shown to influence the measurement (Groisman et al., 1994).

Weather radar can observe a large area at high spatial and temporal resolution. Since 1979 the United Kingdom Meteorological Office has operated and maintained a network of weather radars at C-band frequency (5.60–5.65 GHz) which, as of March 2018, consists of 15 radars. The 5 min frequency volume data from each radar are quality controlled and corrected before an estimate of surface precipitation rate is derived. Surface precipitation rate estimates from each radar are then composited into a 1 km resolution product (Harrison et al., 2000).

The first operational weather radars only observed a single polarization (Fabry, 2015). An issue with single-polarization weather radar is that it only provides the radar reflectivity factor for the sample volume. Deriving an accurate quantitative estimate of the equivalent rainfall rate from radar reflectivity factor requires additional knowledge about the size distribution and type of hydrometeors being observed.

Dual-polarimetric weather radars are better able to estimate the type of hydrometeor within a sample volume. Thus, variables derived from the dual-polarimetric returns provide information about the shape, orientation, oscillation, and homogeneity of observed particles (Seliga and Bringi, 1978; Hall et al., 1984; Chandrasekar et al., 1990). This information may be used to infer the hydrometeor type through hydrometeor classification algorithms (HCAs). HCAs combine observed polarimetric variables using prior knowledge of typical values for each hydrometeor type, to identify the most likely hydrometeor species within a sample volume (Liu and Chandrasekar, 2000). Chandrasekar et al. (2013) give an overview of recent work on HCAs.

Starting in mid-2012 and completing early 2018, every radar in the UK Met Office network was upgraded from single to dual-polarization using in-house design and off-the-shelf components, reusing the pedestal and reflector from the original radar systems. To take advantage of the new information and to improve precipitation estimates, an operational HCA was developed within the Met Office, based on work at Météo France (Al-Sakka et al., 2013). While significant amounts of literature have been published on the technical improvement of HCAs (Chandrasekar et al., 2013), the verification of HCA skill has not been discussed as widely. There is a need for more rigorous validation of HCAs and DiVeN was created specifically for the verification of the UK Met Office radar network HCA.

Typically in situ aircraft are used to verify radar HCA (Liu and Chandrasekar, 2000; Lim et al., 2005; Ribaud et al., 2016). Instrumented aircraft flights such as the Facility for Airborne Atmospheric Measurements (FAAM) take a swath volume using 20 Hz photographic disdrometer instruments (Abel et al., 2014). However there is no fall speed information, which distinguishes hydrometeor type with high skill due to distinct particle density differences (Locatelli and Hobbs, 1974). The lack of fall speed information on FAAM instruments means that the 1200 images collected in every minute of flight must be visually analysed manually or with complex image recognition algorithms. The major disadvantage with FAAM data is the sparsity of cases due to the expense of operating the aircraft.

Therefore, in situ surface observations must be utilized to expand the quantity of comparison data. A larger dataset allows bulk verification statistics to be performed on radar HCAs. Here we introduce a new surface hydrometeor type dataset and examine the skill of the dataset, independently of any radar instruments.

1.2 Precipitation measurement with disdrometers

A disdrometer is an instrument which measures the drop size distribution of precipitation over time. The drop size distribution (henceforth DSD) of precipitation is the function of drop size and drop frequency. Jameson and Kostinski (2001) provide an in-depth discussion on the definition of a DSD. Disdrometers typically record drop sizes into bins of non-linearly increasing widths due to the accuracy reducing with increasing values.

The disdrometer is also a useful tool for verifying radar hydrometeor classification algorithms. Hydrometeor type can be empirically derived using information about the diameter and fall speed of the particle, which the Thies laser precipitation monitor (LPM) instrument used in DiVeN is able to measure. The Gunn–Kinzer curve (Gunn and Kinzer, 1949) describes the relationship between raindrop diameter and fall speed. As diameter increases, the velocity of a raindrop increases asymptotically. Other velocity–diameter relations have been shown in the literature for snow, hail, and graupel, which are well described in Locatelli and Hobbs (1974).

At of the time of writing this publication, operational networks of disdrometers are uncommon, with the notable exceptions of Canada (Sheppard, 1990) and Germany. Networks of disdrometers solely for research purposes have been frequently deployed for short periods of time. From March 2009 to July 2010 (16 months), 16 disdrometers were placed on rooftops within 1 km by 1 km on the campus of the Swiss Federal Institute of Technology in Lausanne to study the inter-radar pixel variability in rainfall (Jaffrain et al., 2011). Another example of research using networked disdrometers is the Midlatitude Continental Convective Clouds Experiment (MC3E) (Jensen et al., 2016), which utilized 18 Parsivel-1 disdrometers and seven 2DVDs (two-dimensional video disdrometers) within a 6 km radius of a central facility near Ponca City, Oklahoma. The project lasted for 6 weeks (22 April through 6 June 2011). DiVeN has an initial deployment phase of 3 years with a high expectation of renewal, which enables unique long-term research to be conducted with the data.

1.3 Paper structure

This paper describes DiVeN and demonstrates the data products of the Thies LPM instruments being used. The first part of the paper provides a technical description of the disdrometer instruments used in the network, the locations chosen to host the instruments, and data management in the network. Case studies from the first 12 months of DiVeN observations are then discussed. The case studies include rain–snow transitions in the 2017 winter storm named Doris, a convective rainfall event, and graupel observations. These events will provide an illustrative analysis of the observations being produced by all the individual disdrometer instruments within DiVeN. Enhanced scrutiny will be placed on the performance of the present weather code because this variable will be used to verify the Met Office radar HCAs.

2 Thies Clima laser precipitation monitor

2.1 Specification

The instruments used in DiVeN (see Fig. 1) are the Thies™ laser precipitation monitor (LPM), model number 5.4110.00.200, which is described in detail in Adolf Thies GmbH & Co. KG (2011). To make observations the instrument utilizes an infrared (785 nm) beam with dimensions of 228 mm × 20 mm × 0.75 mm, a total horizontal area of 45.6 cm^2. The infrared beam is emitted from one end of the instrument and is directed to the other. A photodiode and signal processor determine the optical characteristics including optical intensity, which is reduced as a particle falls through the beam. The diameter of the hydrometeor is inferred by the maximum amplitude of the signal reduction and the speed of the hydrometeor is estimated by the duration of the signal reduction. Figure 1 in Löffler-Mang and Joss (2000) describes a similar instrument (Parsivel-1) with the same observing principle and is an excellent visualization of the technique which is employed by the Thies LPM. The signal processing claims to detect and remove particles that fall on the edge of the beam: "the measured values are processed by a signal processor (DSP), and checked for plausibility (e.g. edge hits)." No further details are given by the manufacturer. The instrument is able to allocate individual hydrometeors into 20 diameter bins from 0.125 to > 8 mm and 22 speed bins from > 0.0 to > 20 $\mathrm{m\,s^{-1}}$.

The Thies disdrometer performs additional calculations on the incoming data which it attaches to the Telegram 4 serial output. Table 1 provides details of the variables and the range of possible values that the instrument is capable of recording. The quantity, intensity, and type of precipitation (drizzle, rain, snow, ice, grains, soft hail, and hail as well as combinations of multiple types) are calculated. Hydrometeor type is recorded as a present weather code. Table 2 lists all of the WMO Table 4680 present weather codes that the Thies laser precipitation monitor is capable of recording. The present weather code is encoded as a number between 1 and 99, which has a corresponding description of the weather using the standardized codes from the World Meteorological Organization Table 4860 (WMO, 1988). The present weather descriptors cover most hydrometeor types but not all; graupel is not explicitly mentioned, for example.

Figure 1. A DiVeN Thies LPM located at Weybourne Observatory in Weybourne, East Anglia, UK, which is an Atmospheric Measurement Facility (AMF) site, part of the National Centre for Atmospheric Science (NCAS).

Hydrometeor type is inferred by the instrument, using empirical relationships between hydrometeor size and fall speed. The diameter–fall speed relation described in Gunn and Kinzer (1949) is the only relationship cited in the instrument manual but it is expected that further relationships are used for solid precipitation, undisclosed by the manufacturer. Section 4 of this paper will qualitatively test the skill of the present weather code regardless of the algorithm it uses, since the exact method of derivation is not known.

Lastly, the present weather code quality index (Table 1) is calculated based on the number of particles within each hydrometeor class. Thies do not recommended using the quality index without additional temperature and wind sensors which can be added to the disdrometer (Marc Hillebrecht, Adolf Thies GmbH & Co. KG, personal communication, 2017). Although DiVeN does not employ the additional sensors, the quality index is still published and can be a useful indicator as shown in Sect. 4.1.

2.2 Limitations

Tapiador et al. (2016) performed a physical experiment with 14 laser disdrometers (Parsivel-1) placed in close proximity (within 6 m^2) on the roof of a building in Toledo, Spain. Precipitation characteristics were calculated for one disdrometer's data, then for two instruments' combined data, and so

Table 1. Variable output from the Thies laser precipitation monitor (LPM).

Output	Units	Resolution	Range
Particle diameter	mm	0.125 mm (max)	$\geq$ 0.125–> 8 mm
Particle velocity	m s^{-1}	0.2 m s^{-1} (max)	> 0–> 20 m s^{-1}
Particle count	Counts	1 count	0–99 999
Rainfall rate	mm h^{-1}	0.001 mm h^{-1}	0.000–999.999 mm h^{-1}
Precipitation visibility	m	1 m	0–99 999 m
Radar reflectivity factor	dBZ	0.1 dBZ	−9.9–99.9 dBZ
PW code quality index	%	1 %	0 %–100 %

Table 2. World Meteorological Organization (WMO) synoptic present weather codes (Table 4680) output by the Thies laser precipitation monitor (LPM).

SYNOP (Table 4680)	Description
−1	Sensor error
41	Light/moderate unknown precipitation
42	Heavy unknown precipitation
0	No precipitation
51, 52, 53	Light/moderate/heavy drizzle
57	Light drizzle with rain
58	Moderate/heavy drizzle with rain
61, 62, 63	Light/moderate/heavy rain
67	Light rain and/or drizzle with snow
68	Moderate/heavy rain and/or drizzle with snow
71, 72, 73	Light/moderate/heavy snow fall
74, 75, 76	Light/moderate/heavy soft hail/ice grains
77	Snow grains
89	Hail

on until all 14 disdrometers' data were used. The aim was to test how many disdrometers' data were needed for the precipitation parameters to asymptote towards a stable value. It was found that a single disdrometer could underestimate instantaneous rain rate by 70 %. Tapiador et al. (2016) proposed that large drops contribute disproportionately to the rain rate and that instantaneous measurements have a lower chance of measuring large drops because they are sparsely populated. The DiVeN disdrometers have a shortest temporal resolution of 1 min, which alleviates some of the sampling issues by allowing time for larger droplets to be observed.

Hydrometeor type observations are less affected by the aforementioned sample size limitations as the dominant type can be estimated from a relatively small sample of the total precipitation. Theoretically only one hydrometeor needs to be sampled by the disdrometer to determine hydrometeor type. The hydrometeor type accuracy is only as good as the diameter and fall speed measurements. In reality, the accuracy of the diameter and fall velocity measurements for a single particle are not accurate enough to determine the dominant hydrometeor phase from an instantaneous measurement. Furthermore, the fall velocity and diameter of small hydrometeors may be indistinguishably similar for several hydrometeor types when observed by the disdrometer. Similar to the results of Smith (2016) for rainfall rate, the largest particles also give the strongest indication of hydrometeor type. This is because fall velocity is related to the density of the particle multiplicatively (Gunn and Kinzer, 1949); i.e. the difference in fall speed for a 5 mm raindrop and a 5 mm snow aggregate is large compared with the difference between a 0.5 mm raindrop and 0.5 mm ice crystal. Therefore the disdrometer can determine with greater confidence the type of hydrometeor when the hydrometeors are larger.

If the sample size of the instrument were larger and thus could count more particles at a faster rate, other limitations would occur. The instrument relies on observing one particle in the beam at any given time; the optical intensity of the beam must return to normal (no obstruction) for maximum confidence of speed observations. If two hydrometeors partially overlap vertically as they fall through the beam, the disdrometer will observe a double dipped reduction in optical intensity which the signal processor must account for. Similarly for diameter, if two hydrometeors fall through the beam simultaneously, the disdrometer will observe a hydrometeor twice as large at the same speed. The sample area is thus limited to reduce the possibility of overlapping particles. Again, Fig. 1 in Löffler-Mang and Joss (2000) is an excellent diagram to aid the understanding of this limitation.

The chance of two drops being in the disdrometer at the same time is unlikely except at extremely high precipitation rates. To examine this, a Poisson distribution test is applied using the sampling volume of the disdrometer with increasing drop concentrations. Figure 2 shows that precipitation rates of greater than 10 000 drops min^{-1} are required before the probability of simultaneous drops in the beam occurring becomes non-negligible. There is a 0.09 % chance of two or more drops in the beam simultaneously for 10^4 drops min^{-1} observed by the disdrometer; one in every 1075 drops. For 10^5 drops min^{-1} observed by the disdrometer there is a 7 % chance of two or more drops in the beam simultaneously; one in every 14 drops. For context, a drop count of 12 000 observed by a disdrometer located at NFARR Atmospheric Observatory, Chilbolton, England, in

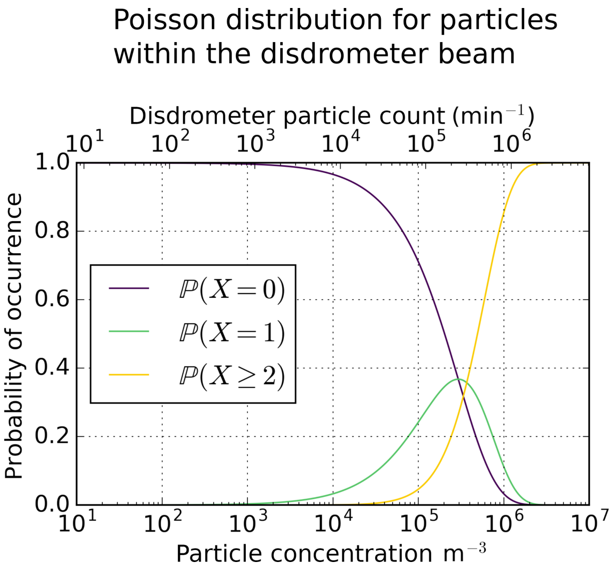


Figure 2. Probability of X number of drops residing within the disdrometer beam for a given drop concentration. If two or more drops are within the beam simultaneously, data quality can be reduced. More than 12 000 drops m^{-3} (equivalent to 10 000 drops min^{-1} recorded by the disdrometer*) are required before the probability of two or more drops occurring in the beam simultaneously becomes non-negligible. As such, any events with more than 10 000 drops observed per minute should be treated as less reliable. * Drops falling through the disdrometer beam assume a 3 m s^{-1} fall velocity, which from Gunn and Kinzer (1949) is a particle of approximately 0.8 mm diameter, typically the average size observed for a moderate rainfall event. Droplet breakup on the housing of the Thies LPM is not factored into this test.

March 2017 (see Sect. 4.2) was equivalent to 22 mm h^{-1}. Rain rates approaching 100 mm h^{-1} would be necessary for the chance of two drops existing in the beam simultaneously to be non-negligible. Such rainfall rates are extremely rare in the UK.

3 Description of the network

3.1 DiVeN locations

Disdrometers have similar site specification requirements as other precipitation instruments. Ideally a flat site with no tall objects or buildings nearby that can cause shadowing, and steps taken to minimize the splash of liquid droplets from the surrounding ground into the instrument. To this end, Thies recommends that the instrument be mounted on a 1.5 m pole above a grassy surface. A grassy surface also minimizes convective upwelling from solar heating of the ground – a particular problem for concrete surfaces – which can slow hydrometeor fall speeds and create turbulence. Turbulence from buildings should also be avoided if possible since it acts to break larger particles into smaller particles, resulting in skewed drop size distributions.

Figure 3. Instrument locations that make up the Disdrometer Verification Network (DiVeN) as of September 2018. Grey icons are the operational Met Office radars as well as the Met Office research radar at Wardon Hill. Map data © 2018 GeoBasis-DE/BKG (© 2009), Google, Inst. Geogr. Nacional.

The locations chosen for DiVeN cover a variety of geophysical conditions such as mountain peaks, valleys, and flat regions, as well as inland and coastal sites. The locations also cover the full breadth of the climatology of precipitation totals and hydrometeor types in the UK (Fairman et al., 2015) with sites in wetter (Wales) and drier (East Anglia) regions as well as sites in warmer (southern England) and colder (northern Scotland) climates.

The typical range at which the Met Office radar HCA product will need to perform is $<$ 120 km (maximum range used to produce surface rainfall rate composite). For the disdrometers to be representative when verification work is performed, the instruments in DiVeN are located at varying ranges from Met Office radars. Figure 3 shows the DiVeN site locations and the Met Office radar locations for comparison. Table 3 gives an overview of each site in DiVeN, including the coordinates, height above mean sea level, and terrain characteristics.

Two instruments are located 10 m apart at NFARR Atmospheric Observatory in Chilbolton. These two instruments form part of an extended observational period of 12 months

Table 3. Site location descriptions of disdrometers in the Disdrometer Verification Network.

Site name	Latitude, longitude (decimal °)	Altitude (m a.m.s.l.)	Installation date	Description
Chilbolton	51.1455,−1.4396	83	10 Feb 2017	NFARR Atmospheric Observatory. Two instruments, 10 m apart. Land type: flat, agricultural fields for > 500 m in all directions. Nearby objects: 25 m diameter radar dish antenna 100 m ESE; two-floor building 25 m SSW.
RUAO	51.4415, −0.9376	63	13 Feb 2017	Reading University Atmospheric Observatory. Land type: open grass in vicinity; campus with lake and trees situated within a wider suburban area. Lake 100 m W–NW, three-floor building 50 m SSE. Shed 30 m ENE.
Cranfield	52.0744, −0.6252	105	15 Feb 2017	Facility for airborne atmospheric measurements. Land type: two-floor rooftop observatory within a cluster of buildings at a university airport. Nearby objects: stairwell NW, hangar ESE. Above most nearby buildings.
Weybourne	52.9505, 1.1218	8	17 Feb 2017	NCAS Atmospheric Measurement Facility. Land type: military base, mostly grass. Sandy beach and ocean 100 m NNE. Nearby objects: small one-floor building ESE, four-floor scaffold tower E.
Aberystwyth	52.4248, −4.0045	44	20 Feb 2017	NFARR/NERC (Natural Environment Research Council) mesosphere–stratosphere–troposphere (MST) radar site. Land type: agricultural fields in a WSW–ENE valley. Nearby objects: single tree and one-floor building SSE, hedgerow N–SSE.
Lancaster	54.0138, −2.7749	94	22 Feb 2017	Hazelrigg Weather Station, University of Lancaster. Land type: agricultural fields. Nearby objects: 100 m tall wind turbine 150 m WSW, meteorological mast 10 m NW. Road and trees 30 m E.
Edinburgh	55.9217, −3.1745	105	24 Feb 2017	GeoSciences Weather Station, University of Edinburgh. Land type: roof of six-floor James Clark Maxwell Building. Urban campus W–N–E, with golf course S. Nearby objects: rooftop above all surrounding buildings.
Laurieston	54.9614, −4.0605	67	28 Feb 2017	Mountain Weather Information Service. Land type: rural village, undulating agricultural terrain beyond. Nearby objects: one-floor buildings 10 m SE, trees 30 m S–W.
Holme Moss	53.5335, −1.8574	522	10 Mar 2017	Holme Moss transmitting station. Land type: hilltop moorland. Nearby objects: 228 m transmitting mast 40 m SW with anchoring cables overhead. Cabin 10 m SW, wire mesh fence NNW.
Cairngorm	57.1269, −3.6628	781	12 Jun 2017	Cairngorm Mountain ski resort with Scottish Environment Protection Agency (SEPA) collaboration. Land type: arctic tundra, frequently snow-covered valley, facing NW. Nearby objects: road and power outbuilding uphill (SE) 20 m.
Feshie	57.0063, −3.8550	882	13 Jun 2017	Druim nam Bo weather station owned by University of Dundee. Land type: arctic tundra, frequently snow-covered, rounded mountain ridge oriented SW–NE, sloping SW. Nearby objects: weather station 10 m N.
Dunkeswell	50.8603, −3.2398	255	14 Jul 2017	Met Office official observatory at Dunkeswell Aerodrome. Land type: flat in all directions. Runway N–E–S with surrounding agricultural fields and forest SW–N. Nearby objects: one-floor building 20 m NW.
Coverhead	54.2038, −1.9849	316	15 Dec 2017	Coverhead Estate with Water@Leeds collaboration. Land type: NW slope of SW–NE valley, agricultural fields. Nearby objects: mounted on a small outhouse facing S. Telegraph pole 10 m NW and trees E–SW.

where their performance will be assessed against several other precipitation sensors located at the same site. A separate paper will be produced to address the results of this dual-instrument study.

3.2 Installation

The main installation campaign occurred in February 2017 for nine instruments. The Holme Moss site was installed shortly after in March, followed by Cairngorm and Feshie in June 2017. Dunkeswell is a Met Office site which was added to the network via a Raspberry Pi with 3G dongle being appended in July 2017. The last instrument to be installed was at Coverhead Estate in the Yorkshire Dales in December 2017, as a collaboration with Water@Leeds https://water.leeds.ac.uk/ (last access: 7 August 2019).

Installation took around 2 h at each site and consisted of anchoring the tripod to the ground, attaching the disdrometer and data logging box, plugging the disdrometer cables into the power strip and the Raspberry Pi, and cutting the power strip cable to length for the site. The installation was designed to be “as plug and play as possible”. Wiring of plugs, data, and power cables onto the disdrometer and coding of the Raspberry Pi were all completed in a lab before arriving at the site.

3.3 DiVeN costs and environmental impact

Each site required the following components to support the disdrometer: Davis Instruments® tripod (GBP 100, http://www.davisnet.com/product_documents/weather/manuals/07395-299_IM_07716.pdf, last access: 7 August 2019), IP67-rated box (GBP 25, https://www.timeguard.com/products/safety/weathersafe-outdoor-power/outdoor-multi-connector-box, last access: 7 August 2019), Raspberry Pi 3 Model B (GBP 30, https://www.raspberrypi.org/products/raspberry-pi-3-model-b/, last access: 7 August 2019), and a generic RS-485 to USB converter (GBP 12). Therefore the total cost per site for hardware was GBP 167. A total of 200 m of power/data cable and tools required for the installation cost an additional GBP 270 and GBP 60 respectively. Some sites rely on a 3G dongle to upload data. The dongles themselves were free when purchased with a single-use data allotment. The total cost of hardware and equipment to build DiVeN amounted to GBP 2500.

The Thies Clima instrument is power rated at a maximum of 750 mA at 230 v. No typical usage has been measured but should the maximum be continuous, then the annual consumption would be 1500 kWh per year, or GBP 190 per year at average UK electricity costs (valid March 2018). In reality the power consumed is subjectively known to be much less than the maximum rating.

Most sites use existing networks at their sites for uploading data to the NCAS server, but those with 3G dongles have an ongoing cost of GBP 75 per year for a yearly data plan. There are eight sites using 3G dongles; hence the ongoing annual cost is GBP 600.

The emissions from the first 3700 km journey in a diesel van were approximately 966 kg of CO_2 and 1.74 kg of $NO_x + PMs$ (nitrogen oxides + particulate matters). Ongoing power consumption for 13 sites (the Druim nam Bo (Feshie) site is powered off-grid by solar and wind) at the aforementioned maximum rating would be 7150 kg of CO_2 annually (using the UK average of 0.367 $kg\,kWh^{-1}$, valid October 2017). In reality the power consumption is less and the UK average $kg\,kWh^{-1}$ is gradually decreasing over time. Computational energy consumed by DiVeN is nearly unquantifiable; the data hosting, processing, and analysis were carried out on shared systems (National Centre for Atmospheric Science server, JASMIN server), so the fractional consumption is difficult to estimate.

3.4 Data acquisition and management

The disdrometer data are read through a serial port by a Raspberry Pi, which executes a Python script to receive and digest the Telegram 4 format data. The Python code performs file management with timestamps taken from the Raspberry Pi internal clock (set over IP) and backs up files to a memory card into a directory specific to the date. Separate programming triggers the uploading of new files in the “today” directory to an NCAS server every 5 min over Secure File Transfer Protocol (SFTP). At 01:00 UTC each day, the Raspberry Pi attempts to upload any remaining files in the directory of the previous day. At 02:00 UTC each day, the Raspberry Pi attempts to upload files from the directory for 7 d ago as a backup command in the event that no connection could be made at the time. Only new files that do not already exist on the NCAS server are uploaded to avoid duplication. The entire directory of data for a single day is compressed using tar gunzip, 8 d after it is recorded. A support script exists to keep the processing and uploading scripts running and self-regulating. The support script checks that the processing script is running; if not, it will issue a command to start the processing script again. This means that the data acquisition script will be reattempted if an exit error occurs. In the event of a power loss the Raspberry Pi will start up and initiate all of the required scripts itself when power is restored, without user intervention.

Each disdrometer produces 3.2 MB of ASCII .txt files per day but this can be compressed significantly. A total of 10 years of continuous minute-frequency disdrometer data (5.3 million minutes) can be compressed to as small as 400 MB.

3.5 Open-access website

Data are uploaded to an NCAS server every 5 min. Plotting scripts are initiated 1 min after the upload. An addi-

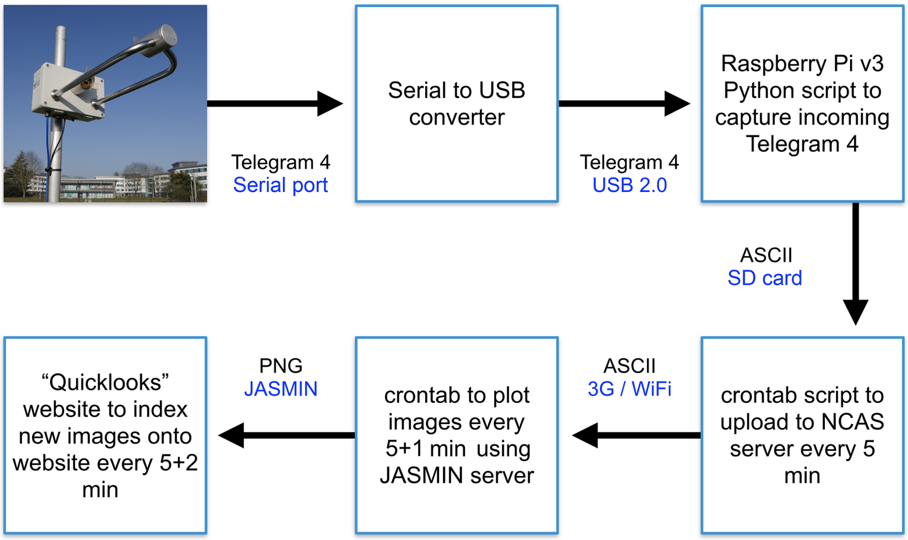


Figure 4. Flow chart of the sequence of data in the Disdrometer Verification Network. The instrument outputs a Telegram 4 format serial ping every minute, which is then captured by a Raspberry Pi (v3) running a Python script. The Python script then saves the file to the built-in SD card as an ASCII.txt. Separate BASH scripts upload the new files every 5 min (xx:05, xx:10, xx:15) to an NCAS server, which JASMIN then reads to plot the data (xx:06, xx:11, xx:16). The website indexes for new images at xx:07, xx:12, xx:17, and so on. Thus the time taken for the xx:00 to xx:05 data to reach the website is 2 min.

tional minute later, a QuickLook system indexes the target directories for new images and displays them on the public website. The public website can be accessed here: https://sci.ncas.ac.uk/diven/ (last access: 7 August 2019). Data can currently be downloaded from NCAS upon request to the lead author. At the end of the first DiVeN deployment phase (early 2020) all data collected by DiVeN will be archived into netCDF at the Centre for Environmental Data Analysis (CEDA).

3.6 DiVeN users

Although the data from DiVeN will be used for radar verification, there are many other uses for the data. Several stakeholders have used DiVeN data. Met Office operational forecasters are able to see live hydrometeor type data and compare with numerical weather prediction forecasts to adjust their guidance. Second, there are some research projects at the University of Leeds being carried out. This includes research on DSD characteristics in bright band and non-bright band precipitation, calibration work with the NCAS X-band polarimetric (NXPol) radar in Cumbria, England, for the Environment Agency (EA), and flood forecasting research with the Water@Leeds project. Other institutions have also used DiVeN data; The University of Dundee and the Scottish Environment Protection Agency (SEPA) are conducting work on snowmelt and the University of Reading may use DSD information from the Reading University Atmospheric Observatory (RUAO) disdrometer to study aerosol sedimentation rates. Finally, the wind turbine manufacturer Vestas has used annual DSD data to evaluate models of blade-tip drag to improve turbine efficiency. The applications of disdrometer data are broad and cover many fields. The authors intend that this publication combined with the open accessibility of data will inspire new uses of DiVeN observations.

3.7 Performance of DiVeN in the first year

Figure 5 shows the uptime of each site in DiVeN in the order that they were installed. Generally the uptime of the network has been good for the period shown, with most sites uploading more than 95 % each day. A few sites have not been as good but this was mostly anticipated. In particular the Druim nam Bo site at 900 m a.m.s.l. in the Scottish Highlands has poor upload percentages. The 3G signal is weak at the site and a signal booster was added in January 2018. Furthermore the site is powered by a small wind turbine and solar panel, which became rimed in ice during the winter (Fig. 6). Although these issues were anticipated, the site was still chosen because it can provide cases of solid hydrometeors nearly all year round, in a terrain which is notoriously difficult for radar performance. Radar hydrometeor classification will be particularly difficult at this location and thus the site will provide a "worst-case scenario" for radar HCA verification work.

Holme Moss is a remote site at relatively high altitude and uses satellite broadband, which has been somewhat unreliable; however the amount of data stored on the Raspberry Pi may be higher than depicted in Fig. 5, which was created based from data successfully uploaded to the NCAS server. Furthermore, the data are being archived on the University of Manchester's system at Holme Moss and this is known to be

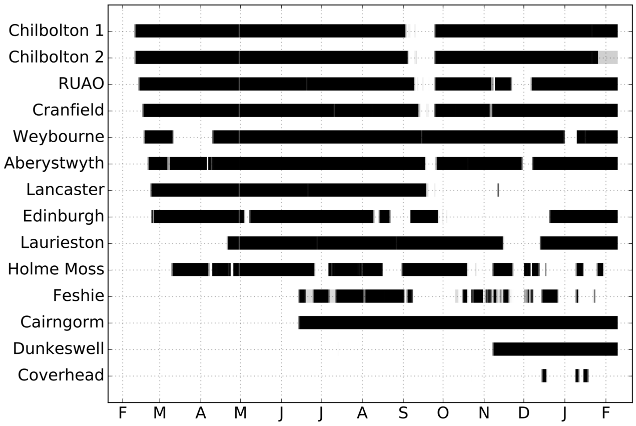


Figure 5. Daily upload performance of DiVeN in the first 365 d of operation. Black indicates 100 % upload (1440 files in a day), and white indicates 0 % upload.

Figure 6. Disdrometer at Druim nam Bo, Scotland, covered in rime in January 2018. The instrument was still receiving power and recording nullified (no beam received by optical diode) data, which it interpreted as a "sensor error" (−1) present weather code.

a much more complete dataset, which will be transferred to the NCAS servers in the future.

There were several unanticipated downtime periods. Weybourne had to be moved for construction work at the field site and was without power for approximately 1 month in March 2017. In late April 2017, the NCAS server blacklisted all disdrometer IP addresses and these had to be manually whitelisted. This was detected and resolved within 8 d. The 7 d backup upload filled in the majority of the missing data but the eighth day prior to the issue being fixed was never reattempted because of the design of the code discussed in Sect. 3.4.

The largest unanticipated downtime occurred in September 2017. An issue arose with the disdrometers being unable to record any new data, in the order that they were installed. A total of 2 GB of free space remained on the SD cards; however there was a (previously unknown) limit to the number of files that can be saved to certain card formats regardless of the space remaining. The issue was fixed by the creation of a new script which merged old files together. The script had to be added to all of the Raspberry Pi's in the network. The issue was detected after the first four DiVeN disdrometer installations failed sequentially, so the failure of other sites in the network was anticipated and mitigated. This can be seen in Fig. 5 as a stepped-failure starting with the Chilbolton 1 instrument in September 2017.

Some further issues occurred which were avoidable. Laurieston was disconnected from power whilst closing the data logger box after the installation, which meant it was offline for the first 2 months until the site could be visited again. Similarly during the Dunkeswell installation in July 2017 the serial data cable was damaged, which could not be fixed until November 2017. The Raspberry Pi at Lancaster was not reconnected after the aforementioned file number problem in September 2017.

Although several problems have arisen with the Disdrometer Verification Network in the first 12 months, the network manager and site owners have been, on the whole, quick to respond to these issues, which has minimized downtime. DiVeN is in an ideal state for long-term data collection as it has been designed with few potential failure points and with several backup methods in place in the event of a failure.

4 Case studies

The following sections subjectively analyse the skill of the disdrometer instrument for classifying hydrometeor type. Three types are discussed here: snow from winter storm Doris, an intense rainfall event at NFARR Atmospheric Observatory, and a graupel shower at the Reading University Atmospheric Observatory. NFARR Atmospheric Observatory instrument data were sourced from Science and Technology Facilities Council et al. (2003) and Agnew (2013).

4.1 Rain–snow transition

During the first disdrometer installation trip in February 2017, the Met Office-named winter storm Doris impacted the UK. The disdrometer at Lancaster was installed on 22 February, and Edinburgh was scheduled for installation on 24 February. Storm Doris was forecast to bring heavy snowfall to the central belt of Scotland on the morning of 23 February. Therefore a decision was made to leave Lancaster early on the evening of 22 February, to arrive in Gladhouse Reservoir before the expected snowfall. An opportunity arose to temporarily operate a disdrometer at Gladhouse Reservoir (55.7776, −3.1173). Observations began at 01:00 UTC, by which time light rain had begun precipitating.

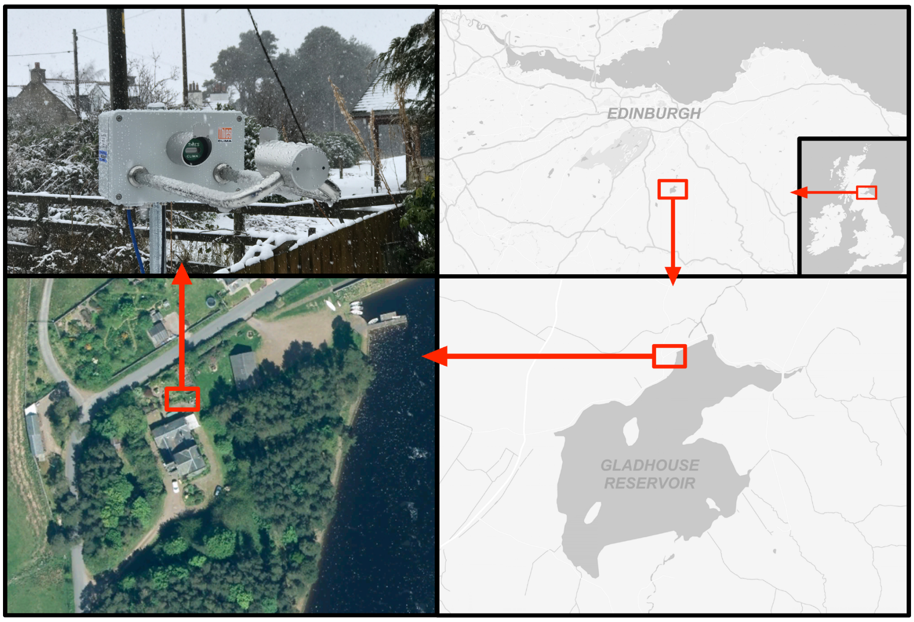


Figure 7. Maps, satellite images, and ground images of the disdrometer location and setup for winter storm Doris at Gladhouse Reservoir House, Scotland. Map data © 2018 GeoBasis-DE/BKE (© 2009), Google. Satellite image: copyright © 2012–2016 Apple Inc. All rights reserved.

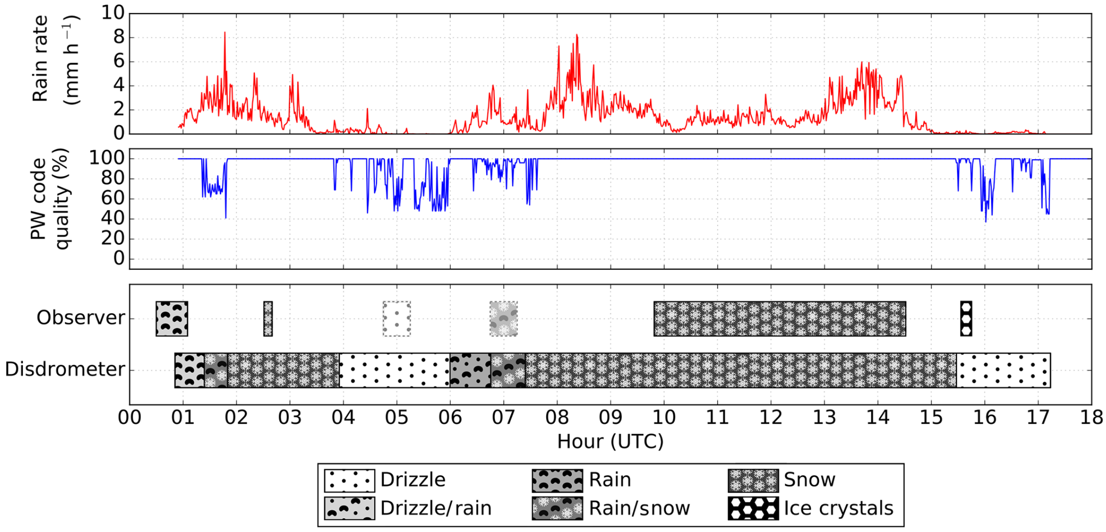


Figure 8. Rain rate, hydrometeor type, and present weather code quality index during the storm Doris event on 23 February 2017, which occurred over approximately 16 h at Gladhouse Reservoir, Scotland. Rain rate is liquid equivalent for periods of snow and is recorded by a Thies LPM disdrometer. Hydrometeor type is shown from both the disdrometer and impromptu from a trained meteorologist. The meteorologist observations at 05:00 and 07:00 UTC are approximate due to a lack of accurate time information. The disdrometer misidentified individual ice crystals at 15:39 as drizzle.

The opportunistic observations made during storm Doris provide a unique dataset by which to evaluate the skill of the disdrometer for prescribing hydrometeor type. Several transitions between rain and snow occurred that were also observed by a qualified meteorologist. The following section compares the disdrometer present weather codes and the eyewitness observations taken by the lead author during the event. An important consideration is the fact that the disdrometer was set up in a suboptimal observing environment, which had approximately 200° of tall objects in close proximity. Figure 7 shows the instrument operating at Gladhouse Reservoir. There were tall evergreen trees to the east and west and a two-floor building to the south. Telecom cables were also overhead and associated poles are visible to the NNE

Table 4. Present weather code evolution throughout the named winter storm Doris event on 23 February 2017. All times in UTC.

Time	Disdrometer present weather code	Time	Qualified meteorologist observation
00:55 to 01:24	Rain	00:30 to 01:05	Rain
01:24 to 01:50	Rain or mixed precipitation		
01:50 to 03:55	Snow	02:31 to 02:40	Snow
03:55 to 06:00	Light/moderate drizzle	Approx. 05:00	Drizzle
06:00 to 06:45	Drizzle or rain		
06:45 to 07:24	Rain or mixed precipitation	Approx. 07:00	Mixed precipitation
07:24 to 15:28	Moderate/heavy snow	09:49 to 14:31	Moderate/heavy snow
15:28 to 17:13	Light/moderate drizzle	15:39	Pristine ice crystals

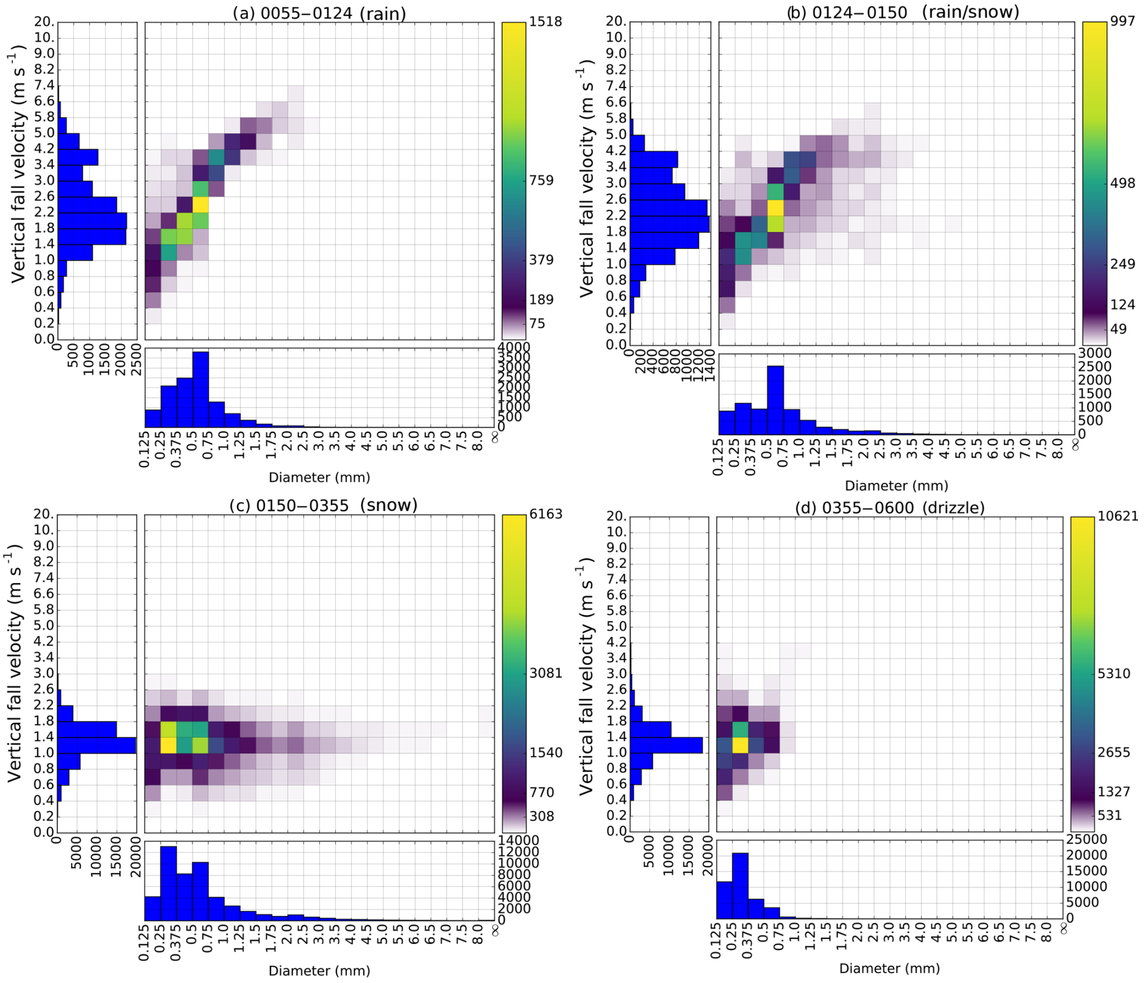


Figure 9. Accumulated particle information for each hydrometeor class period described in Fig. 8. The centre grid shows particle counts binned by size and fall velocity. The y-axis histogram shows particle velocity distribution (DVD) and the x-axis histogram shows particle size distribution (DSD) for the time period described. Since the time periods between each subplot are inconsistent in length, the colour scale and histograms have been normalized for the total precipitation over each period. The periods are as follows: **(a)** 00:55–01:24 UTC (rain), **(b)** 01:24–01:50 (rain/snow), **(c)** 01:50–03:55 (snow), and **(d)** 03:55–06:00 (drizzle).

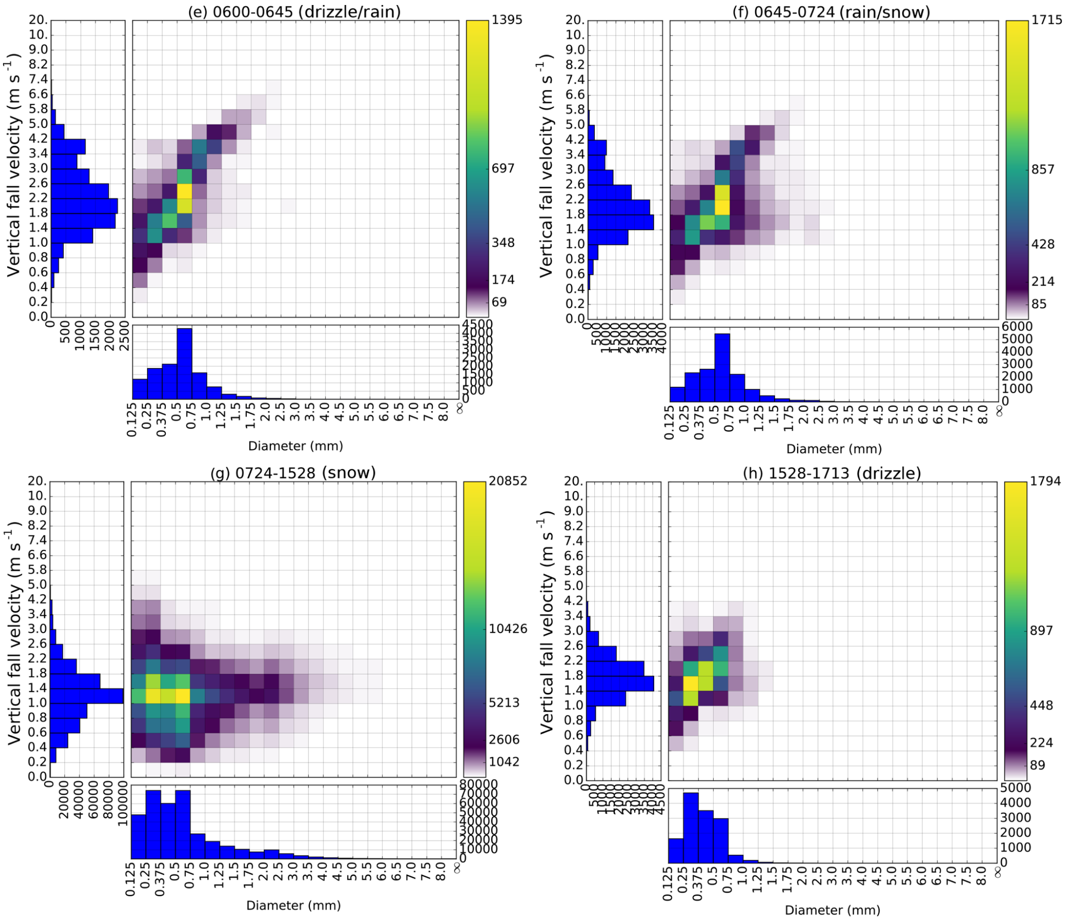


Figure 10. As in Fig. 9, but time periods are as follows: **(e)** 06:00–06:45 (drizzle/rain), **(f)** 06:45–07:24 (rain/snow), **(g)** 07:24–15:28 (snow), and **(h)** 15:28–17:13 (drizzle).

behind the disdrometer in Fig. 7. This was unavoidable given the impromptu circumstances of deployment.

Despite the suboptimal observing conditions, the disdrometer performed well at diagnosing the correct present weather code during the storm Doris event. Table 4 and Fig. 8 show that the disdrometer correctly output a present weather code of rain initially, followed by an unverified "mixed precipitation" from 01:24 to 01:50 UTC. From 01:50 onwards a consistent snowfall present weather (PW) code was observed, which agrees with visible observations made within 01:50–03:55. At 03:55 the precipitation became light and was described as drizzle by the disdrometer.

From 06:00 onwards the precipitation intensified and the present weather code changed between drizzle and rain. By 06:45 the PW code was switching between only rain and a rain–snow mix. From 07:24 onwards the present weather code was constant snow, which continued with varying intensity until 15:28. The eyewitness observation at 15:39 is of individual ice crystals, which the disdrometer perceived as low precipitation rates of 0.293 mm h^{-1} misclassified as drizzle. Weak precipitation continued until 17:13 where no precipitation is observed by the disdrometer, concluding the IOP.

Table 4 shows that the Thies LPM has good skill with regard to determining the present weather type. Every disdrometer-diagnosed present weather code is in agreement with the eyewitness observations throughout the IOP, with the exception of 15:39. The difference in fall velocity between drizzle particles and individual ice crystals is small and as such the disdrometer struggled to identify the precipitation correctly.

Figures 9 and 10 show the periods of constant hydrometeor type observed by the disdrometer in Fig. 8, normalized for particle count. There are clear differences between rain, snow, and rain–snow mix periods. Rain follows the curve shown by Gunn and Kinzer (1949). The rain–snow mix periods in (b) and (f) retain the Gunn–Kinzer relationship but with additional, larger particles with slower fall velocities. The snow categories in (c) and (g) are markedly different with broader distributions of particle size and a shifted fall velocity distribution. The drizzle and ice crystal periods, however, are very similar. Both are characterized by distributions of particle fall speed and diameter peaking at approximately 1.4 m s^{-1} and 0.375 mm respectively. The distribution similarities of drizzle and pristine ice crystals in Figs. 9 and 10 illustrate the difficulty in distinguishing between these two types by fall speed and diameter alone, without additional information. A temperature sensor added to the disdrometer may have aided the PW code classification. The misidentification described here is not a major concern since pristine ice crystal precipitation is (a) uncommon in the UK and (b) contributes negligible amounts to total rainfall as indicated during this event.

The present weather code quality index shown in Fig. 8 demonstrates that the Thies LPM is able to detect when recording conditions are challenging. The PW code quality index decreases, showing a poor quality measurement, during times of weak precipitation rates and in mixed precipitation phases.

The opportunistic data collected in the storm Doris event are unusual in their number of transitional periods and will be a valuable case by which to compare the performance of radar-derived surface hydrometeor classification schemes.

4.2 Intense convective rainfall

Storm Doris also brought an interesting event to another site; a high rainfall rate observed by the NFARR Atmospheric Observatory pair of disdrometers (Chilbolton 1 & 2). The event was synoptically characterized by a narrow swath of intense precipitation oriented meridionally. The high-intensity precipitation moved west to east across the UK, associated with a cold front originating from the low associated with winter storm Doris. About 30 km NE of NFARR Atmospheric Observatory in Stratfield Mortimer, a private weather station managed by Stephen Burt also observed the intense band of rainfall (Stephen Burt, personal communication, 2017). A high-resolution Lambrecht gauge (recorded resolution of 0.01 mm) on the site observed a 75.6 mm h^{-1} rain rate over 10 s at 07:51 UTC. The 1 min rain rate at 07:51 was 54.6 mm h^{-1} and the 5 min rain rate ending at 07:52 was 30.6 mm h^{-1}. The event was described by a trained observer as “rain quickly became heavy then torrential”.

The event was particularly outstanding from a DiVeN point of view due to the drop count measured by the Thies LPMs situated at NFARR Atmospheric Observatory, Chilbolton, which peaked at around 12 000 drops in a single minute (200 per second) at 07:39 UTC on 23 February 2017. Both disdrometers observed a similar evolution of drop count over the short 26 min rainfall event. This does not prove that the instruments are recording accurately; conversely it may be a signal of a systematic issue with the measurement technique used in every Thies LPM.

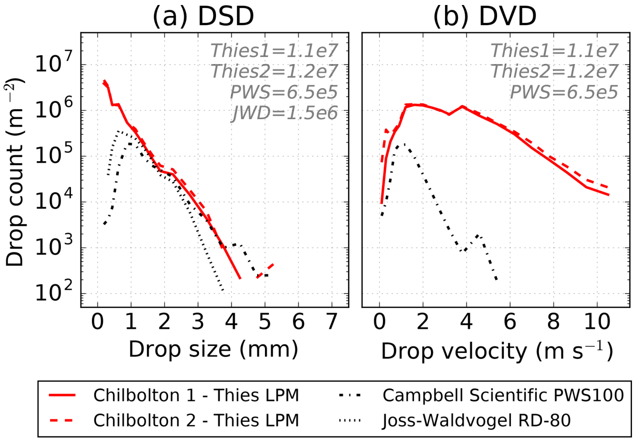


Figure 11. Drop characteristics of a heavy rain event at NFARR Atmospheric Observatory, Chilbolton, England, on 23 March 2017. Distributions are accumulated from 07:25 to 07:50 UTC inclusively for a 26 min summation. Panel **(a)** shows drop size distribution and panel **(b)** shows drop velocity distribution. The Joss-Waldvogel RD-80 (JWD) does not provide drop velocity information. Each instrument has been normalized for sampling area and bin widths. Total drop count is listed in the top right of each plot. Both of the Thies LPMs have a higher total drop count, as well as significantly higher counts of small and high-velocity particles compared with the PWS and JWD. The frame of the Thies LPM may be splashing droplets into the beam, leading to increased counts of small, fast-moving droplets.

Figure 11 shows an anomalously large left-tailed DSD from both of the Thies LPMs when compared against the Joss-Waldvogel RD-80 and Campbell Scientific PWS100 disdrometers. A high concentration of small drop sizes suggests that splashing is occurring, where larger drops breakup on impact with either the instrument itself or the surroundings. Earlier versions of the Thies LPM did not have shields on top of the sensor, which the manufacturer acknowledged were added because of splashing issues. It is possible that at very high rainfall rates, splashed droplets are still reaching the instrument beam and are being erroneously recorded. The drop velocity distribution (DVD) from the Thies LPM is also in disagreement with the PWS100. The PWS100 uses a similar optical technique to the Thies LPM with the addition of having four vertically stacked beams versus one on the Thies LPM, which should increase the accuracy of fall velocity measurements. Furthermore, the Thies LPM categorizes the highest velocity particles into the smallest diameter particle bins, which is unphysical. Finally, the total drop count per metre is significantly higher for both of the Thies LPMs.

The DVD during the event is very wide. A noteworthy observation from the Stratfield Mortimer observatory is the wind characteristics. Marking the passage of the cold front at 07:45, winds became increasingly gusty and 10 min wind mean ending at 07:40 was 20 knots. A strong surface wind is associated with turbulent eddies, which have some vertical component. The intermittent vertical wind acts to widen the drop velocity distribution. Furthermore, turbulence breaks up droplets, thus skewing the drop size distribution. Finally, winds tangent to the beam (N–S-oriented beam, westerly

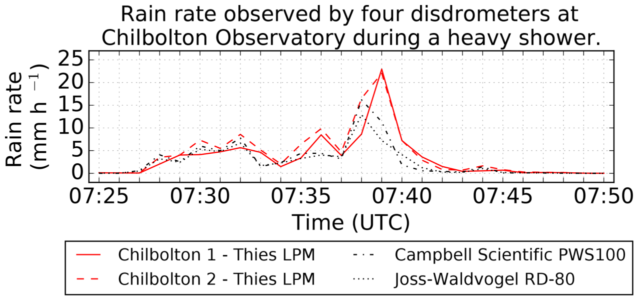


Figure 12. Rain rate measured by four instruments during a heavy rain event at NFARR Atmospheric Observatory, Chilbolton, England, on 23 March 2017. The total accumulated rain depth over the 26 min for each instrument is as follows: Chilbolton 1 (1.481 mm); Chilbolton 2 (1.847 mm); PWS100 (1.237 mm); JWD (1.090 mm). Each instrument has been normalized for sampling area and bin widths. Both of the Thies LPMs have a higher total rain rate than the PWS100 and JWD. The difference in rain rate between both of the Thies LPMs and the PWS100 and JWD is greatest during the most intense precipitation, which may be evidence of droplets splashing from the instrument housing into the measuring beam.

wind as was the case here) increase the number of beam-edge hits, which reduce the quality of the data.

Figure 12 shows that the two Thies LPMs have good agreement for rain rate from 07:25 to 07:35 where the rain rates are moderate, but that the Thies LPMs overestimate the rainfall from 07:35 to 07:40 where the rain rate is heavy. In total, Chilbolton 1 and Chilbolton 2 recorded 120 % and 149 % of the rainfall measured by the PWS100. The JWD is expected to underestimate slightly due to the range of observable diameters (0.3 to 5 mm) being smaller than true raindrop sizes and smaller drop sizes being undetectable in the presence of large droplets due to sensor oscillation.

It appears that in these conditions the hydrometeors were not correctly measured by the Thies LPM. However, the hydrometeor type is still correctly identified despite these shortcomings in rain rate, particle diameter, and particle velocity.

4.3 Graupel shower

Graupel (rimed ice crystals) is an important signature of convection for the UK, where hail is relatively uncommon. The Thies instrument does not have a graupel category because the category does not exist within the WMO Table 4680, which it uses to convey hydrometeor type. Codes 74, 75, and 76 (light/moderate/heavy soft hail/ice grains) are presumed to be equivalent to what is commonly described as graupel.

On 25 April 2017 a shower containing conical-shaped graupel passed over Reading University "between 16:30 and 16:45 UTC" as observed by Chris Westbrook (Chris Westbrook, personal communication, 2017). Figure 13 shows the temporal evolution of hydrometeor type identified by the DiVeN instrument during the event. The disdrometer observed only a single minute (16:36) of "soft hail/ice grains" PW code (indicating graupel) during the entire 21 min of precipitation detected. Between 16:30 and 16:50 UTC inclusively, the following codes were also observed: 7 min of code 68 (moderate/heavy rain and/or drizzle with snow), 12 min of codes 61/62 (light/moderate rain), and 1 min of code 72 (moderate snowfall). Clearly the instrument struggled to diagnose graupel in this particular event.

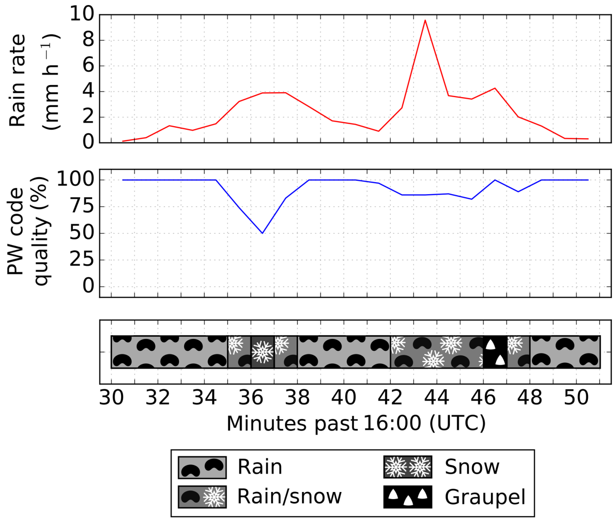


Figure 13. Rain rate, present weather code quality index, and hydrometeor type during a graupel shower in Reading, England, on 25 April 2017. The event was recorded by a Thies LPM at the Reading University Atmospheric Observatory. Conical graupel was also observed from a nearby building (approximately 500 m away) by a qualified meteorologist between 16:30 and 16:45 UTC. Rain rate is the liquid equivalent for periods of solid hydrometeors as recorded by a Thies LPM disdrometer. Hydrometeor type is shown based on the present weather code (WMO Table 4680) recorded by the Thies LPM. The instrument struggles to diagnose the graupel and instead outputs a present weather code of snow and mixed rain–snow precipitation.

Figure 14 shows the particle size and velocity information grouped by hydrometeor type prescribed by the Thies LPM. Throughout the graupel shower the instrument observed a bimodal distribution in both velocity and diameter for all hydrometeor types, which is indicative of both rain and graupel precipitating simultaneously. Furthermore in the rain/snow, snow, and graupel periods, a few hydrometeors exist below the Gunn–Kinzer curve, which are misidentified as snow. Although the accumulated drop characteristics for the rain and rain/snow minutes are indicative of a rain–graupel mixture, in a single minute only a few particles may fall through the disdrometer beam versus several hundred raindrops. The ratio of rain to graupel may therefore be insufficient for the PW code to change to graupel. No PW code exists in the WMO Table 4680 for a rain–graupel mixture or rain–soft hail mixture. The false detection of snow hydrometeors may be attributed to graupel particles bouncing off nearby surfaces or the instrument itself, slowing the fall velocity and thus ap-

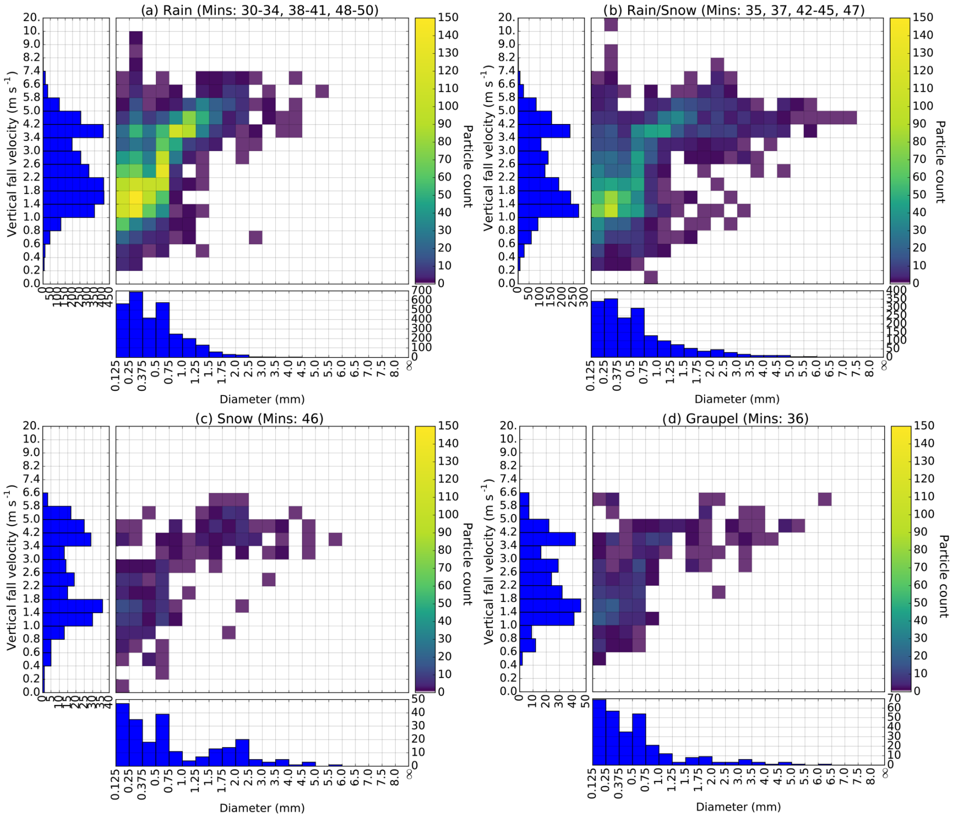


Figure 14. Accumulated particle information for each hydrometeor class period described in Fig. 13. The centre grid shows particle counts binned by size and fall velocity. The y-axis histogram shows particle velocity distribution (DVD) and the x-axis histogram shows particle size distribution (DSD) for the time period described. The periods are as follows: **(a)** rain (12 min), **(b)** rain/snow (7 min), **(c)** snow (1 min), and **(d)** graupel (1 min). The colour scale is identical in all plots despite the different time accumulations in order to highlight the rare particles.

pearing to the disdrometer as a lower-density particle such as an ice aggregate.

For future work with DiVeN data it is important to note 1 min observations of "soft hail/ice grain" PW codes when longer time periods are being analysed. For example, radar hydrometeor classification will be performed with DiVeN data at 5 min intervals. If in one of the 5 min soft hail or snow grains is observed, this must be highlighted. Graupel likely existed for longer than 1 min but it was either not the dominant hydrometeor or the instrument was unable to correctly identify it.

5 Summary

The Disdrometer Verification Network is the largest network of laser precipitation measurements in the UK. Here we have fully described the network and discussed three specific observation cases to subjectively discuss the accuracy of the Thies LPM with a focus on hydrometeor type diagnosis.

In summary, the instruments are able to correctly identify changes between snow and rain during storm Doris even with the suboptimal observing conditions. Snow is easily detected by the disdrometer and it is also able to accurately signal a mixture of hydrometeor types when transitioning between rain and snow.

Yet, the Thies LPM appears to have difficulty with measuring heavy rainfall events, where droplet breakup may be occurring due to instrument design. Distributions of drop size are skewed, such that small particle counts are significantly enhanced when compared with the Joss-Waldvogel RD-80 and the Campbell Scientific PWS100. The hydrometeor type variable was unaffected by the distribution discrepancies in the case studied.

The Thies LPM also struggled to detect graupel in the event studied here. This shortcoming can be somewhat compensated for by flagging individual minutes of present weather codes 74, 75, and 76 within larger datasets but there will be graupel cases that the Thies LPM fails to detect entirely.

A factor affecting the Thies LPM for hydrometeor classification is that empirical relationships do not account for instrument errors or the design of the instrument, which may interfere with the precipitation being measured. The hydrometeor type signatures should be derived using data from the instrument to which they will be applied. Furthermore, by using the present weather code to describe hydrometeor type, the Thies LPM is restricted in its ability to express the true nature of the observations being made, particularly noted in instances of graupel.

DiVeN offers open-access data in near-real time at 5 min updates. The 1 min frequency data are available upon request from the authors or via the Centre for Environmental Data Analysis (CEDA) from 2020. Data have been made publicly accessible in the hope that the Disdrometer Verification Network will be used for research beyond the original scope of the network.

Data availability. Data plots are available in near-real time here: https://sci.ncas.ac.uk/diven/ (last access: 7 August 2019). Original data are available through the Centre for Environmental Data Analysis (CEDA, http://www.ceda.ac.uk, last access: 7 August 2019) in NetCDF format (CF-1.6, NCAS-AMF-1.0) under the following DOI: https://doi.org/10.5285/602f11d9a2034dae9d0a7356f9aeaf45 (Natural Environment Research Council et al., 2019).

Author contributions. We used the taxonomy of CASRAI's CRediT definitions of contributor roles to describe the author contributions.

BSP contributed to the conceptualisation, data curation, formal analysis, investigation, methodology, project administration, resources, software, supervision, validation, writing of the original draft, and the review and editing of the writing.

RRNIII contributed to the conceptualisation, funding acquisition, methodology, project administration, resources, supervision, and the review and editing of the writing.

DH contributed to the conceptualisation, methodology, project administration, resources, supervision, and the review and editing of the writing.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The lead author wishes to thank the following people and institutions for contributing to the creation of the Disdrometer Verification Network.

We thank the United Kingdom Meteorological Office for loaning the Thies LPM instruments used in DiVeN, Thies for advice and communication regarding the instrument, and the National Centre for Atmospheric Science (NCAS) for all other supporting hardware.

We thank Morwenna Cooper (Met Office), Dan Walker (NCAS), James Groves (NCAS), and Darren Lyth (Met Office) for technical advice regarding the data acquisition design of DiVeN.

We thank the contacts at each site hosting a disdrometer for DiVeN: Judith Jeffery (NFARR), Andrew Lomas (University of Reading), Rebecca Carling (Facility for Atmospheric Measurements), Grant Forster (University of East Anglia), David Hooper (NFARR), James Heath (University of Lancaster), Richard Essery (University of Edinburgh), Geoff Monk (Mountain Weather Information Service), Michael Flynn (University of Manchester), Louise Parry (Scottish Environment Protection Agency), Jim Cornfoot (Natural Retreats), Chris Taylor (Natural Retreats), Andrew Black (University of Dundee), Darren Lyth (Met Office), Megan Klaar (University of Leeds), and Stephen Mawle (Coverhead Farm).

We thank Jack Giddings, Ashley Nelis, Scott Duncan, and Daniel Page for providing accommodation and sanity during the month-long installation trip.

We thank Philip Rosenberg (NCAS) for advice on statistical tests.

We also thank Stephen Best (Met Office), James Bowles (Met Office), Dave Hazard (NFARR), Darcy Ladd (NFARR), Stephen Burt (University of Reading), and Chris Westbrook (University of Reading).

Financial support. This research has been supported by the NERC (NERC Industrial CASE Studentship (grant no. NE/N008359/1)).

Review statement. This paper was edited by Szymon Malinowski and reviewed by Frederic Fabry and two anonymous referees.

References

Abel, S. J., Cotton, R. J., Barrett, P. A., and Vance, A. K.: A comparison of ice water content measurement techniques on the FAAM BAe-146 aircraft, Atmos. Meas. Tech., 7, 3007–3022, https://doi.org/10.5194/amt-7-3007-2014, 2014.

Adolf Thies GmbH & Co. KG: Laser Precipitation Monitor – Instruction for Use, Tech. rep., Adolf Thies GmbH & Co. KG, Hauptstraße 76, 37083 Göttingen, Germany, 2011.

Agnew, J.: Chilbolton Facility for Atmospheric and Radio Research (CFARR) Campbell Scientific PWS100 present weather sensor data, NCAS British Atmospheric Data Centre, available at: https://catalogue.ceda.ac.uk/uuid/e490cd13d86d832bd2d62f1650d7b265 (last access: 7 August 2019), 2013.

Al-Sakka, H., Boumahmoud, A. A., Fradon, B., Frasier, S. J., and Tabary, P.: A new fuzzy logic hydrometeor classification scheme applied to the french X-, C-, and S-band polarimetric radars, J. Appl. Meteorol. Climatol., 52, 2328–2344, https://doi.org/10.1175/JAMC-D-12-0236.1, 2013.

Chandrasekar, V., Bringi, V., Balakrishnan, N., and Zrnić, D.: Error structure of multiparameter radar and surface measurements of rainfall, Part III: Specific differential phase, J. Atmos. Ocean Tech., 7, 621–629, https://doi.org/10.1175/1520-0426(1990)007<0621:ESOMRA>2.0.CO;2, 1990.

Chandrasekar, V., Keranen, R., Lim, S., and Moisseev, D.: Recent advances in classification of observations from dual polarization weather radars, Atmos. Res., 119, 97–111, https://doi.org/10.1016/j.atmosres.2011.08.014, 2013.

Ciach, G. J.: Local random errors in tipping-bucket rain gauge measurements, J. Atmos. Ocean. Tech., 20, 752–759, https://doi.org/10.1175/1520-0426(2003)20<752:LREITB>2.0.CO;2, 2003.

Elmore, K. L., Grams, H. M., Apps, D., and Reeves, H. D.: Verifying Forecast Precipitation Type with mPING, Weather Forecast., 30, 656–667, https://doi.org/10.1175/WAF-D-14-00068.1, 2015.

Fabry, F.: Radar meteorology: principles and practice, Cambridge University Press, Cambridge, 2015.

Fairman, J. G., Schultz, D. M., Kirshbaum, D. J., Gray, S. L., and Barrett, A. I.: A radar-based rainfall climatology of Great Britain and Ireland, Weather, 70, 153–158, https://doi.org/10.1002/wea.2486, 2015.

Gascón, E., Hewson, T., and Haiden, T.: Improving Predictions of Precipitation Type at the Surface: Description and Verification of Two New Products from the ECMWF Ensemble, Weather Forecast., 33, 89–108, https://doi.org/10.1175/WAF-D-17-0114.1, 2018.

Green, A.: From Observations to Forecasts – Part 7, A new meteorological monitoring system for the United Kingdom's Met Office, Weather, 65, 272–277, https://doi.org/10.1002/wea.649, 2010.

Groisman, P. Y., Legates, D. R., Groisman, P. Y., and Legates, D. R.: The Accuracy of United States Precipitation Data, B. Am. Meteorol. Soc., 75, 215–227, https://doi.org/10.1175/1520-0477(1994)075<0215:TAOUSP>2.0.CO;2, 1994.

Gunn, R. and Kinzer, G. D.: The Terminal Velocity of Fall for Water Droplets in Stagnant Air, 6, 243–248, https://doi.org/10.1175/1520-0469(1949)006<0243:TTVOFF>2.0.CO;2, 1949.

Hall, M. P. M., Goddard, J. W. F., and Cherry, S. M.: Identification of hydrometeors and other targets by dual-polarization radar, Radio Sci., 19, 132–140, https://doi.org/10.1029/RS019i001p00132, 1984.

Harrison, D., Driscoll, S. J., and Kitchen, M.: Improving precipitation estimates from weather radar using quality control and correction techniques, Meteorol. Appl., 6, 135–144, https://doi.org/10.1017/S1350482700001468, 2000.

Jaffrain, J., Studzinski, A., and Berne, A.: A network of disdrometers to quantify the small-scale variability of the raindrop size distribution, Water Resour. Res., 47, 1–8, https://doi.org/10.1029/2010WR009872, 2011.

Jameson, A. R. and Kostinski, A. B.: What is a raindrop size distribution?, B. Am. Meteorol. Soc., 82, 1169–1177, https://doi.org/10.1175/1520-0477(2001)082<1169:WIARSD>2.3.CO;2, 2001.

Jensen, M. P., Petersen, W. A., Bansemer, A., Bharadwaj, N., Carey, L. D., Cecil, D. J., Collis, S. M., Genio, A. D. D., Dolan, B., Gerlach, J., Giangrande, S. E., Heymsfield, A., Heymsfield, G., Kollias, P., Lang, T. J., Nesbitt, S. W., Neumann, A., Poellot, M., Rutledge, S. A., Schwaller, M., Tokay, A., Williams, C. R., Wolff, D. B., Xie, S., and Zipser, E. J.: The Midlatitude Continental Convective Clouds Experiment (MC3E), B. Am. Meteorol. Soc., 97, 1667–1686, https://doi.org/10.1175/BAMS-D-14-00228.1, 2016.

Lim, S., Chandrasekar, V., and Bringi, N.: Hydrometeor Classification System Using Dual-Polarization Radar Measurements: Model Improvements and In Situ Verification, IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 43, 792–801, https://doi.org/10.1109/WARSD.2003.1295197, 2005.

Liu, H. and Chandrasekar, V.: Classification of hydrometeors based on polarimetric radar measurements: Development of fuzzy logic and neuro-fuzzy systems, and in situ verification, J. Atmos. Ocean. Tech., 17, 140–164, https://doi.org/10.1175/1520-0426(2000)017<0140:COHBOP>2.0.CO;2, 2000.

Locatelli, J. D. and Hobbs, P. V.: Fall speeds and masses of solid precipitation particles, J. Geophys. Res., 79, 2185–2197, https://doi.org/10.1029/JC079i015p02185, 1974.

Löffler-Mang, M. and Joss, J.: An optical disdrometer for measuring size and velocity of hydrometeors, J. Atmos. Ocean. Tech., 17, 130–139, https://doi.org/10.1175/1520-0426(2000)017<0130:AODFMS>2.0.CO;2, 2000.

Muchan, K., Lewis, M., Hannaford, J., and Parry, S.: The winter storms of 2013/2014 in the UK: hydrological responses and impacts, Weather, 70, 55–61, https://doi.org/10.1002/wea.2469, 2015.

Natural Environment Research Council, Met Office, Pickering, B. S., Neely III, R. R., and Harrison, D.: The Disdrometer Verification Network (DiVeN): particle diameter and fall velocity measurements from a network of Thies Laser Precipitation Monitors around the UK (2017–2019), Centre for Environmental Data Analysis, https://doi.org/10.5285/602f11d9a2034dae9d0a7356f9aeaf45, 2019.

Penning-Rowsell, E. and Wilson, T.: Gauging the impact of natural hazards: The pattern and cost of emergency response during flood events, T. I. Brit. Geogr., 31, 99–115, https://doi.org/10.1111/j.1475-5661.2006.00200.x, 2006.

Ribaud, J. F., Bousquet, O., Coquillat, S., Al-Sakka, H., Lambert, D., Ducrocq, V., and Fontaine, E.: Evaluation and application of hydrometeor classification algorithm outputs inferred from multi-frequency dual-polarimetric radar observations collected during HyMeX, Q. J. Roy. Meteor. Soc., 142, 95–107, https://doi.org/10.1002/qj.2589, 2016.

Science and Technology Facilities Council, Chilbolton Facility for Atmospheric and Radio Research, Natural Environment Research Council, and Wrench, C. L.: Chilbolton Facility for Atmospheric and Radio Research (CFARR) Disdrometer Data, Chilbolton Site, NCAS British Atmospheric Data Centre, available at: http://catalogue.ceda.ac.uk/uuid/aac5f8246987ea43a68e3396b530d23e (last access: 7 August 2019), 2003.

Seliga, T. and Bringi, V.: Differential reflectivity and differential phase shift: Applications in radar meteorology, Radio Sci., 13, 271–275, https://doi.org/10.1029/RS013i002p00271, 1978.

Sheppard, B. E.: Measurement of raindrop size distributions using a small Doppler radar, J. Atmos. Ocean. Tech., 7, 255–268, 1990.

Smith, P. L.: Sampling Issues in Estimating Radar Variables from Disdrometer Data, J. Atmos. Ocean. Tech., 33, 2305–2313, https://doi.org/10.1175/JTECH-D-16-0040.1, 2016.

Tapiador, F. J., Navarro, A., Moreno, R., Jiménez-Alcázar, A., Marcos, C., Tokay, A., Durán, L., Bodoque, J., Martín, R., Petersen, W., and de Castro, M.: On the Optimal Measuring Area for Pointwise Rainfall Estimation: A Dedicated Experiment with Fourteen Laser Disdrometers, J. Hydrometeorol., 18, 753–760, https://doi.org/10.1175/JHM-D-16-0127.1, 2016.

Thornes, E.: The impact of weather and climate on transport in the UK, Process in Physical Geography, 16, 187–208, https://doi.org/10.1177/030913339201600202, 1992.

WMO: Manual on Codes. Vol. 1, WMO Publ. 306, 203 pp., 1988.

Chapter 3.

Evaluation of multiple precipitation sensor designs for precipitation rate and depth, drop size and velocity distribution, and precipitation type.

In revision to *AMS Journal of Hydrometeorology.*

Contributions from: Ryan R. Neely III, Judith Jeffery, David Dufton and Maryna Lukach

Abstract

Observations of the drop-size distribution, drop-velocity distribution, rainfall rate and precipitation type are compared from six in-situ precipitation sensor designs over 12 months to assess their performance and provide a benchmark for future design and deployment. The designs considered are: tipping-bucket (TBR), drop-counting (RAL), acoustic (JWD), optical (LPM), single-angle visiometer with capacitor (PWD21) and dual-angle visiometer (PWS100). Precipitation rates are compared for multiple time resolutions over the study period, while drop size and velocity distributions are compared with cases at stable precipitation rates. To examine precipitation type a new index and a logic algorithm to amalgamate consecutive precipitation type observations consistently is introduced and applied. Overall the choice of instrument for deployment depends on the usage. For fast response (less than 15 minutes), the PWD21 and TBR should not be used. As precipitation rate or the duration of a sample increases, the correlation of the TBR with the majority of other instruments increases. However, the PWD21 consistently underestimates precipitation. The RAL, PWS100 and JWD are within ± 15% for precipitation depth over 12 months. All instruments are inconsistent in their ability to observe drop size and velocity distributions for differing precipitation rates. There is low agreement between the instruments for precipitation type estimation. The PWD21 and PWS100 rarely report some precipitation types, but the LPM reports more broadly. Meteorological stations should use several instrument designs for redundancy and to more accurately capture precipitation characteristics.

1. Introduction

All meteorologists agree that precipitation must be recorded accurately, yet there is no consensus on the best method to do so. There are many ways to measure precipitation, both in-situ or through remote sensing. For remote sensing techniques, the sample volume of any single remote sensing measurement contains a population of hydrometeors that must be derived statistically from the measurement. As such, spatial variability smaller than the measurement scale is lost and important details may be obscured. Meanwhile, though surface in-situ measurements are able to directly measure, they typically only sample a fraction of a square metre, which renders their observations as unrepresentative of the wider area. A combination—using surface precipitation measurements as a “ground truth” to calibrate remote sensing techniques—is commonly used in operational meteorological agencies (Fulton et al. 1998; Harrison et al. 2000; Rubel and Brugger 2009). Precipitation depth (mm) and rate (mm h^{-1}), distributions of drop size (mm) and velocity (m s^{-1}), and dominant precipitation type are 4 fundamental variables used to describe precipitation. Whilst WMO intercomparison experiments have focussed on the rate of liquid (Lanza and Vuerich 2009) or solid (Kochendorfer et al. 2017) precipitation, this study examines all of these variables for a single location over 12 months.

1.1. Precipitation Depth & Precipitation Rate

This study will consider measured precipitation depth and precipitation rate (henceforth PD and PR) performance from very weak to very intense precipitation (0.05 to 50.0 mm h^{-1}). Instruments must be capable of accurately detecting the heaviest precipitation since these events have the highest impact on society through flooding. At the other end of the scale, weak precipitation can also be important. Oppenheim and Shinar (2012) showed that drivers do not reduce their speed sufficiently on wet roads. Warning signage based on automatic rainfall sensors must be able to detect the weakest precipitation events that could otherwise endanger motorists. When studying the accuracy of precipitation gauges, a "ground truth" reference value is difficult to obtain. This is because multiple factors can either increase (splashing, non-hydrometeors, convergent airflow, condensation) or decrease (divergent airflow, shadowing, unfilled buckets, edge hits, leaves, insects, evaporation) measured rainfall rates. This study examines 12 months of real-world data and uses methods to compare the instruments which reveals strengths and weaknesses in different applications.

1.2. Drop Size Distribution

The drop size distribution (henceforth DSD) is a distribution of the number of drops as a function of diameter per unit volume (Jameson and Kostinski 2001) and is defined as $N(D)$. Instruments which are able to measure this variable are called disdrometers; for disdrometers, the DSD defines every subsequent calculated variable such as rainfall rate, equivalent radar reflectivity factor and precipitation type. The first DSDs were measured with either ink-dusted paper or trays of flour (Wiesner 1895; Bentley 1904; Laws and Parsons 1943) which led to the Marshall-Palmer relations of DSD to rainfall rate (R) and radar reflectivity (Z) (Marshall and Palmer 1948). DSD-derived Z–R relations are commonly used to retrieve radar-based quantitative precipitation estimates. Therefore, the importance of the accuracy of in-situ surface observations of DSD cannot be understated.

1.3. Drop Velocity Distribution

The drop velocity distribution (henceforth DVD) is defined as $N(V)$, the number of drops as a function of drop velocity. DVD is important because it allows the total kinetic energy of precipitation to be considered. Our understanding of soil erosion

(Kinnell 1981; Rosewell 1986), building erosion (Tang and Davidson 2004; Erkal et al. 2012), vertical winds and downbursts caused by mass loading (Feingold et al. 1991) and the erosion of aerofoils like planes and wind turbine blades (Keegan et al. 2013; Slot et al. 2015; Eisenberg et al. 2018) benefits from more accurate observations of DVD. Some instruments also rely on the fall velocity of the particle to distinguish its precipitation type, since different hydrometeors have different aerodynamic properties (Gunn and Kinzer 1949; Locatelli and Hobbs 1974).

1.4. Precipitation Type

Precipitation type (henceforth PT) describes the dominant phase, shape, and density of hydrometeors within a volume of the atmosphere that fall to the surface. The present weather (henceforth PW) code is used to represent PT, which presents problems for bulk statistical analysis. The codes are a qualitative description of the PW type. Table 4680 from the World Meteorological Organization is the standard for automatic sensors (WMO 1988, 2017), which has 99 entries. Some hydrometeor types are represented by several PW codes—typically variations of intensity and longevity—whereas this investigation only concerns the type of hydrometeor detected. The motivation for the PW code's existence was to reduce the bandwidth of descriptive information, only upheld today for consistency with existing data.

PT can have great impact on transportation, agriculture and infrastructure but is poorly forecasted (Ralph et al. 2005; Reeves 2016). PT has become more prominent in the field of operational meteorology in the last decade as operational radar networks have gained dual-polarisation capabilities (Park et al. 2009; Saltikoff and Nevvonen 2011; Al-Sakka et al. 2013). PT information within a radar sample can also be used to provide more accurate quantitative precipitation estimates (QPE) because Z–R relationships vary with hydrometeor class (Atlas and Ludlam 1961; Harimaya 1978; Fujiyoshi et al. 1990).

2. Experimental Conditions

The following sections briefly describe the instruments as summarised in Table 3.1, and recorded variables in Table 3.2. This is followed by a summary of the experimental conditions on-site, and finally the broader structure of this study.

Table 3.1.: Summary of the study instrumentation specifications. ^{*1} PWD21 Visiometer Sample Volume in litres. ^{*2} PWD21 Capacitive Plate Area. ^{*3} Calculated from 0.313 minimum drop diameter (Distromet Ltd. 2002). ^{*4} With a range 0.5–20 mm h^{-1} (Adolf Thies GmbH & Co. KG 2011). ^{*5} Rain and 0.5... 20mm h^{-1} (Campbell Scientific 2012). ^{*6} "accuracy will be degraded in windy conditions, for frozen precipitation and very high rainfall rates" (Vaisala Oyj 2001). ^{*7} ± 5% of measured drop diameter (Distromet Ltd. 2002), assuming spherical raindrops for ± 16% volume. ^{*8} at > 20 mm h^{-1} rain rate and at 10 second sampling period (Agnew 2014). ^{*9} Marshall-Palmer Distributed (Campbell Scientific 2012). ^{*10} For liquid precipitation (Vaisala Oyj 2001). ^{*11}

	Observing Design	Sample Region	Resolution	Accuracy	Rain Rate Range
LPM	Optical Occlusion Disdrometer	46.5 cm^2	1 x 10^{-3} mm	$\leq$ 15% ^{*4}	< 0.005...> 250 mm h^{-1}
PWS100	Dual-angle Visiometer	40.0 cm^2	1 x 10^{-4} mm	± 10% ^{*5}	0...400 mm h^{-1} ^{*9}
PWD21	Visiometer, Capacitor, Temp.	0.1 L ^{*1}, 7.2 cm^2 ^{*2}	0.01	± 30% ^{*6}	0.5...20 mm h^{-1} ^{*10}
JWD	Acoustic Disdrometer	50.0 cm^2	1.3 x 10^{-7} mm ^{*3}	± 16% ^{*7}	Undefined
RAL	Drop Counting	324 cm^2	1.89 x 10^{-3} mm	< ± 10% ^{*8}	5...200 mm h^{-1}
TBR	Tipping Bucket	324 cm^2	0.2 mm	Undefined	Undefined

Table 3.2.: Summary of the instruments, variables, and time aggregations used in this study. Precipitation data are scrutinized from 6 different instrument designs, 4 precipitation variables, and over several timescales. The precipitation variables each have their own methodologies for data analysis and are thus presented separately in each section. For the DSD and DVD variables, cases of 5 minutes, 15 minutes, and 1 hour are presented. 1-minute distributions are rarely used because of known sampling issues (Tapiador et al. 2016; Smith 2016), and precipitation events longer than 1 hour typically do not retain consistent distribution characteristics. 1-minute, 5-minute, 15-minute, 1-hour, and daily time durations of PR are considered, and PT analyses are performed for 1-minute, 5-minute and 1-hour intervals (bulk statistics, not case studies).

	Thies1	Thies2	PWS100	PWD21	JWD	RAL	TBR	Total	Durations
PR & PD	✓	✓	✓	✓	✓	✓	✓	7	1, 5, 15m, 1h, 1d
DSD	✓	✓	✓		✓			4	Cases of 5, 15m, 1h
DVD	✓	✓	✓					3	Cases of 5, 15m, 1h
PT	✓	✓	✓	✓				4	1, 5m, 1h

2.1. Instruments

2.1.1. Thies Laser Precipitation Monitors

Two identical Thies Laser Precipitation Monitors (LPMs) are used in this study (henceforth Thies1, Thies2, or the Thies LPMs). The Thies LPMs are both part of a wider network of 14 Thies LPM instruments called the Disdrometer Verification Network (DiVeN), described by Pickering et al. (2019). The Thies LPM emits an infrared beam which is received by a photodiode. As precipitation or other particles cross the path of the beam, the electrical signal produced by the photodiode is reduced (Adolf Thies GmbH & Co. KG 2011). The amplitude and duration of signal reduction is analogous to the size and speed of the particle. Löffler-Mang and Joss (2000) describe this optical occlusion method of observation. Observed particles are sorted into 20 diameter bins (from $\geq$ 0.125 mm to $>$ 8 mm) and 22 velocity bins ($>$ 0.0 to $>$ 20.0 m s^{-1}), and 21 out of 93 total PW codes are supported for 6 PTs (drizzle, rain, mixed rain/snow, snow grains, snow aggregates, hail). Non-hydrometeors (insects, debris) can appear as precipitation and the housing of the instrument is a surface on which precipitation can rebound into the beam. Partial beam-hits are accounted for with internal processing.

2.1.2. Campbell Scientific PWS100

The Campbell Scientific Present Weather Sensor (PWS100) uses a forward-scattering technique using 4 light beams and two receiving diodes; one diode at a vertical angle only and one diode at a combined horizontal and vertical angle from the emitted beams. Drop size distributions are measured in 0.1 mm bins from $>$ 0.0 to 30.0 mm but the manual states that the "proportion of particles detected will fall off significantly below about 0.5 mm diameter". Drop velocity distributions are recorded in non-linearly spaced bins from $>$ 0.0 to $>$ 25.6 m s^{-1}. Both diameter and velocity measurements have a quoted accuracy of ± 5% (for liquid particles $>$ 0.3 mm). Optical scattering characteristics differ between solid and liquid hydrometeors and integrated temperature and humidity sensors assist the determination of PT. Hydrometeor types are reported explicitly: drizzle, rain, snow grains, snowflakes, hail, ice pellets, graupel, and 58 out of a total 93 PW codes (Table 4680) are supported. The limitations of the PWS100 are similar to those of the Thies LPM.

2.1.3. Vaisala PWD21

The Vaisala Present Weather Detector PWD21 (Vaisala Oyj 2001) combines an optical forward-scattering sensor (875 nm peak, single emit and receive diode) with a resistive capacitive plate indicating water content, and a temperature sensor. Heating elements evaporate condensation and melt solid hydrometeors. PT and current weather are reported using Table 4680 (WMO 1988), of which 42 out of 93 descriptors are supported—7 types of precipitation (drizzle, rain, freezing drizzle, freezing rain, snow aggregates, mixed rain/snow, ice pellets). The capacitive plate is only used to discriminate between rain and snow between 0 °C and 6 °C. Outside of this range, the types are overruled by temperature ($T < 0$ °C = snow, $T > 6$ °C = rain).

2.1.4. Joss-Waldvogel RD-80

The Joss-Waldvogel Disdrometer (JWD) has an exposed Styrofoam cone atop a spring-loaded transducer. The kinetic energy applied to the transducer is related to the size of a hydrometeor (Joss and Waldvogel 1967), and is recorded into 127 non-linearly spaced size bins (from ≥ 0.313 to 5.145 mm with an accuracy of $\pm$ 5%). The terminal fall velocity of the particle species must be assumed to convert kinetic energy into mass and then diameter, so there are no DVD measurements from the JWD. Different hydrometeors have different diameter-velocity relationships, the JWD only functions accurately for one type of hydrometeor because it has no way to distinguish PT to then change the internal processing. By default, the JWD assumes all particles to be liquid. Snowflakes (slower terminal velocities) appear as small raindrops and hail (faster terminal velocities) appear as large raindrops.

2.1.5. RAL Drop Counting Gauge

The RAL (Rutherford Appleton Laboratory) Drop Counting Rain Gauge (henceforth RAL) funnels liquid precipitation into a reservoir, from which drops overflow through a precision tube, occluding an optical infrared sensor (Norbury and White 1971). Occlusions are tallied over time to generate a rain rate. A smaller minimum resolution than the TBR should provide greater sensitivity in light rainfall, but the funnel can still disturb the airflow and become blocked. Above 50 mm h^{-1} droplets begin to merge, so occlusions become non-linear with increasing rainfall rate. A continuous stream of fluid cannot be measured.

2.1.6. Munro 0.2 mm Tipping-bucket Rain Gauge

The Munro 0.2 mm tipping bucket rain gauge (TBR) funnels liquid precipitation into a pivoting double bucket, which limits the temporal representation of PR. A single tip over 1 minute represents a 12 mm h^{-1} PR; over 5 minutes 2.4 mm h^{-1}. Due to this, TBR PRs are typically only used for $\geq$ 15-minute durations (1 tip = 0.8 mm h^{-1}). In addition, TBRs are only able to funnel liquid precipitation (when sufficiently wet) into the tipping bucket; snow, hail and graupel must melt to be detected with delay. A heated funnel consumes significantly more energy and limits where such sensors may be deployed. The TBR alone cannot distinguish between PTs. Furthermore, the TBR can be easily rendered unusable if debris blocks the funnel. The Munro is similar in design to gauges used throughout the fields of hydrology and meteorology and thus represents this category of design.

2.2. Study Location

All instruments were installed at the Chilbolton Atmospheric Observatory (CAO), Hampshire, UK. The climate at CAO is temperate and maritime where the weather can fluctuate on a daily basis. Proximity to the polar front jet stream ensures that mid-latitude depressions impact the site regularly. Precipitation is therefore frequent, and intermittent rainfall is the modal type. Air temperature extremes are uncommon due to the UK being a small landmass, surrounded by ocean and situated on the north-eastern Atlantic, influenced by the warmth of the North Atlantic Current. The Met Office (2016) provides a climate summary for Southern England and shows that on average (1981-2010) 12 days per year have falling snow and 109.5 days have precipitation accumulation of $\geq$ 1.0 mm. Ventouras et al. (2006) contains more information on the climatology of PRs observed at CAO. The frequency and variation of precipitation events as well as the occurrence of stratiform, convective, and different PTs make CAO a suitable location to conduct this study because it covers a wide range of precipitation conditions seen worldwide.

2.3. Instrument Installations

Figure 3.1 shows the installation of each instrument at CAO. The JWD is situated inside a circular pit slightly below ground level to reduce turbulence over the instrument whereas the TBR and RAL are on the ground. Both Thies LPM beams are at 1.5 m,

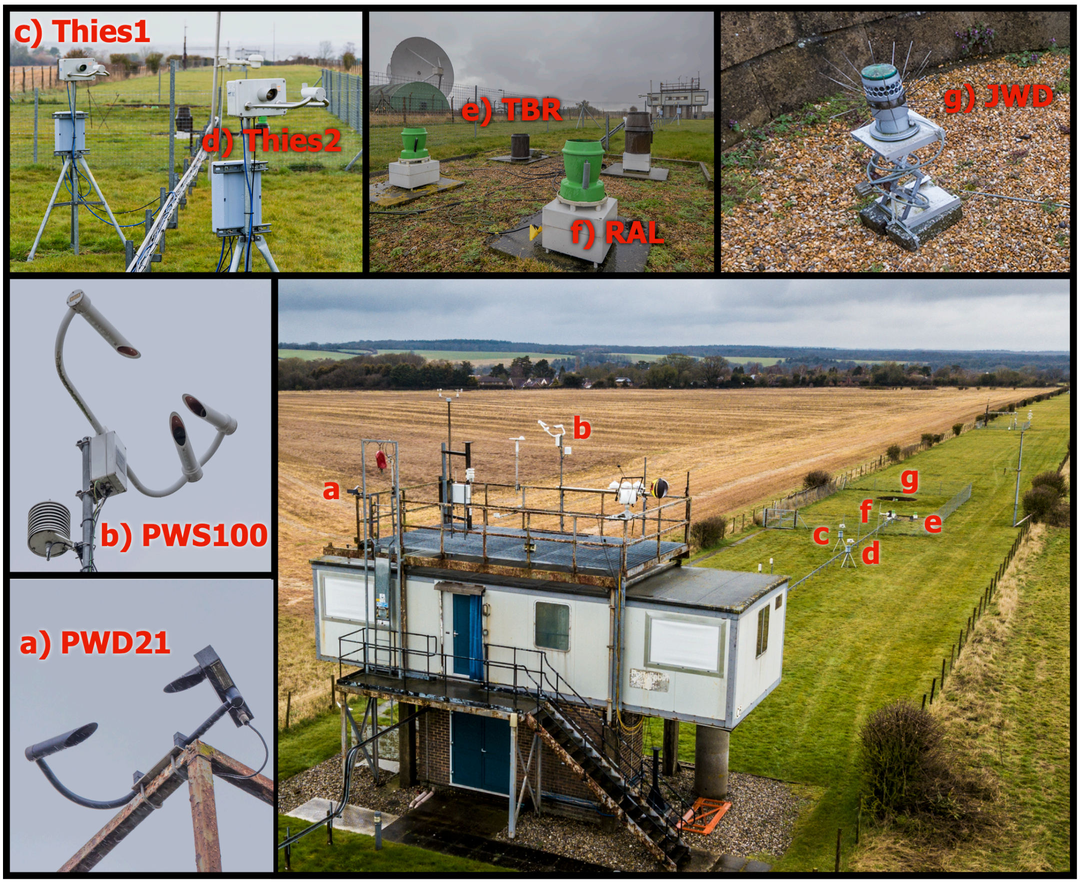


Figure 3.1.: Aerial view of the Chilbolton Atmospheric Observatory (CAO), showing the instruments used in this study and their proximity. a) Vaisala PWD21 b) Campbell Scientific PWS100 c) Thies Clima LPM #1 d) Thies Clima LPM #2 e) Munro 0.2 mm Tipping Bucket Gauge f) Joss-Waldvogel RD-80 g) RAL Drop Counting Gauge. The furthest distance between any two instruments is 80 m.

whereas the PWD21 and PWS100 are at 8 m and 9 m respectively, mounted on 1 & 2 m poles above a 7 m tall building. All of the instruments are spatially within 80 m, are installed within manufacturer recommendations and meet WMO standards where practical. The agricultural land surrounding CAO is flat for at least 500 m in all directions. The Chilbolton Advanced Meteorological Radar (CAMRa) dish (30 m tall) is 80 m east of the instrument enclosure; visible in Figure 3.1. The radar should not lead to bias of the observations since prevailing wind direction at the site is westerly and it is sufficiently distant.

Both the climatic conditions and the physical mounting conditions of the instruments at CAO makes the results of this study applicable to a wide range of precipitation and monitoring conditions at other mid-latitude locations. Therefore the performance of the instruments in this study should be comparable to similar instrumentation located at analogous locations.

2.4. Data Quality

All instruments used in this study have been calibrated as per the manufacturer standards. The TBR and RAL are calibrated dynamically (as described in Humphrey et al. (1997)) with a pump at different flow rates. The TBR was calibrated by the manufacturer in 2011, verified on site in 2018 and in 2020 both with no corrections needed. The RAL was last calibrated in 2013 and was checked in 2020 with no correction needed. We therefore conclude that the TBR and RAL instruments are well calibrated over the study period. The PWD21 and PWS100 have no CEN standardised calibration method for the optical scattering technique, and therefore the manufacturer calibration cannot be verified on site. The capacitive plate on the PWD21 also has no CEN standardised calibration method. The manufacturer calibration is therefore relied upon for the PWD21 and PWS100—which is also true of any deployment of these sensors and is not unique to this study. Therefore, this study is using a calibration comparable to typical installations of these two instruments. Similarly with the JWD, there is no CEN standardised calibration method for on-site verification. The JWD manufacturer recommends that calibration is repeated every 6 years (Distromet Ltd. 2012), which this study period is within, so the instrument is said to be calibrated. The Thies LPMs were calibrated by the manufacturer in 2011 and again no CEN standardised method exists for verification on-site. The Thies LPM manual also states that a calibration is then only necessary when a component is changed (Adolf Thies GmbH & Co. KG 2011), so the Thies LPMs are said to be calibrated during this study period. In addition, all rain gauges at CAO are monitored on an on-going basis using all rain data recorded at the site. Any faults or any relative changes in sensitivity of the gauges are investigated and suspect measurements are removed from the datasets before publication to the Centre for Environmental Data Analysis (CEDA). Figure 3.2 shows the availability of each instrument during the study period.

3. Methodology

The following sections explain the methodologies employed in this study, split by each precipitation variable, since each variable is unique and requires a unique approach.

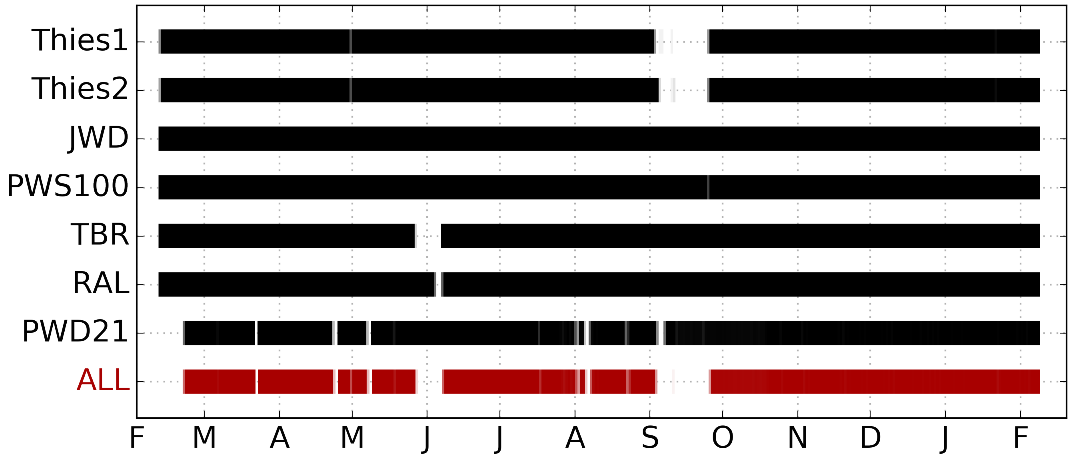


Figure 3.2.: Availability of usable, quality controlled data during the 12-month period under examination (10th February 2017 - 9th February 2018, totaling 365 days). Each day is a vertical strip with the opacity representative of the percentage of time successfully recorded. The lower bar ('ALL') represents the combination of all instruments above, and represents the data under study since all instruments must be operational for a fair comparison. Any time with one or more instruments down, are excluded. The Thies LPMs had a period of downtime in September 2017 resulting in approximately 20 days without data. The missing data from the TBR and RAL gauges was due to the gauge becoming temporarily blocked with debris. For the PWD21, the fault at the start of February 2017 was due to a failure in datalogging. The remaining missing files were due to temporary, intermittent logging or file formatting problems. The minutes where all instruments were simultaneously recording during the study sum up to 309.8 days, which is equivalent to 84.9% of the study period.

The overall goal is to compare the CAO instruments for all precipitation variables that are important to operational and research meteorology today.

3.1. Precipitation Depth & Precipitation Rate

PR measurement techniques are subject to random and systematic errors, affecting the exact determination of PR at ground level. As such, there is no measurement at the site that could be considered a reference. The instruments will be compared and discussed in context of their measurement design to compare and contrast the strengths and weaknesses of each. Sample duration is also an important consideration when examining PR measurements. Some uses and applications of PR data need only daily values while others require much shorter PR timescales, so several durations will be considered.

The comparison of PDs uses the full 12 month dataset described above. To examine bulk behaviours and general differences in the observations of the instruments, the cumulative precipitation measured over the study period is compared first. We compare both the raw accumulations and the deviation of the accumulations of each instrument from the mean of the instrument ensemble.

We then compare the PR from each instrument, performed using simple linear regression by least squares applied between two instruments. The observations should have a 1:1 relation through the origin if both instruments have similar characteristics. Gradient and intercept values of the regression reveal differences in behaviour of the instruments. Each instrument is compared with every other instrument for 5 time durations. In each regression, both of the instruments being compared must detect precipitation in order for that datum to be included, which removes major anomalies but favors insensitive instruments.

3.2. DSD & DVD

DSD and DVD measurements are difficult to analyse using bulk statistics because they are non-linear. To facilitate comparison, DSDs can be parameterized into a gamma or generalised-gamma model (Marshall and Palmer 1948; Ulbrich 1983; Thurai and Bringi 2018) in the form:

$$N(D) = N_0 \exp(-\Lambda D) \quad (0 \leqslant D \leqslant D_{\max}) \tag{3.1}$$

where N_0 is the intercept, Λ the slope function and D the drop diameter. The advantage of parameterization is that the comparison is simplified to a purely numerical one, but in doing so the artefacts or biases in the data are potentially concealed. For example, an overestimation of medium-diameter drops combined with an underestimation of small-diameter drops could result in the same slope function as for an instrument without biases. There is also uncertainty on the gamma model performance for drop-diameters < 1.0 mm. Chandrasekar et al. (2003) suggest that the reduction in counts of small drops in the gamma model is a consequence of a parametric fit on experimental data that underdiagnoses small drops. Therefore this study compares measured DSDs and DVDs without parameterization.

Here a case study approach is used, similar to that of Tokay et al. (2013) but with some improvements. In Tokay et al. (2013), the cases are all 1 hour in duration, the PR varies throughout, and the average PRs range from 1.8 to 12.6 mm h^{-1}, so less-common (very

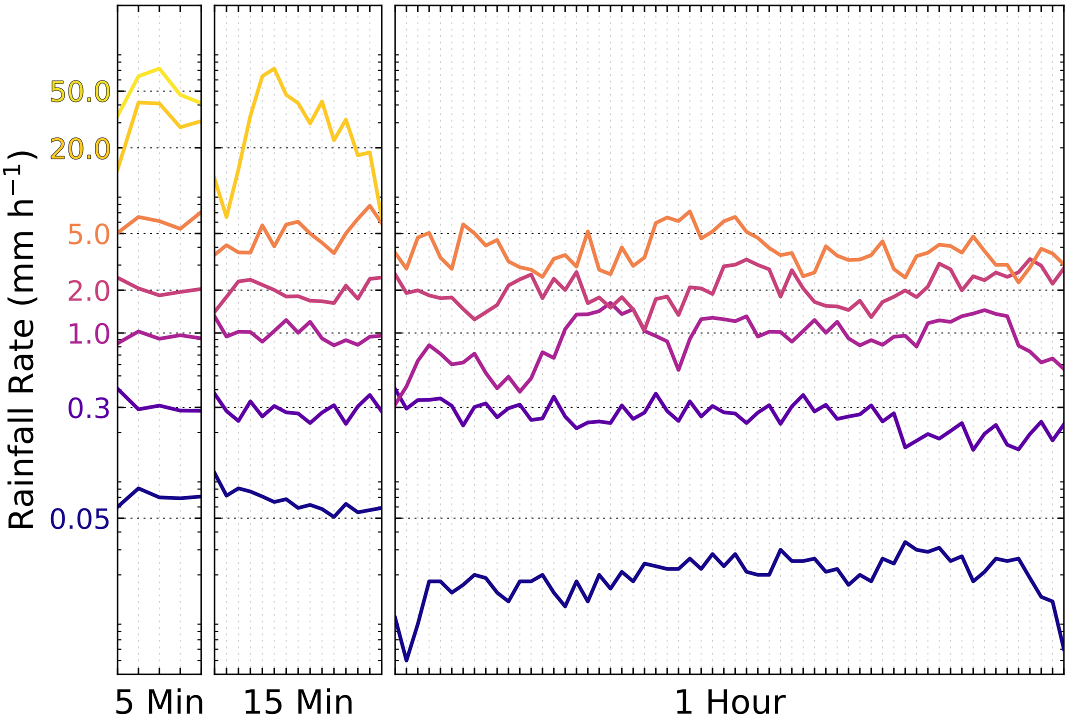


Figure 3.3.: PRs from Thies1 for the chosen a) 5 minute, b) 15 minute, and c) 60 minute cases.

weak and very heavy) PR-DSDs are not explicitly represented. Here cases are chosen based on constant PR throughout 5-minute, 15-minute and 1-hour periods so that the observed DSD/DVD remains consistent. The PRs used to identify cases are: 0.05, 0.3, 1.0, 2.0, 5.0, 20.0 and 50.0 mm h^{-1}.

Thies1 measurements are used to select cases. The absolute deviation from the desired PR is calculated in a moving window of the desired duration and the period with the least absolute deviation is selected as the case from the entire dataset. Figure 3.3 shows the PR over the cases selected. A 50 mm h^{-1} PR does not occur consistently for 15 minutes or 1 hour, and a 20 mm h^{-1} PR does not occur consistently for 1 hour. In total there are 18 cases. When averaged, the 1 hour weakest rainfall rate (case 15 in Figure 3.5 and 3.6) is approximately 0.02 mm h^{-1}.

Each instrument uses different DSD and DVD bin widths. The Thies LPMs have the broadest bin widths so the data from the PWS100 and JWD are mean-weight normalized into the Thies LPM bins. The upper and lower limits of the JWD are narrower than the Thies LPMs and result in partially filled bins after normalization that are discarded. The effective measurement range of the JWD in this analysis is

Table 3.3.: A master look up table (LuT) for precipitation type.

ABC	PT	Description	PW Codes	Comments
Er	-2	Instrument Error	—	Instrument offline/data corrupt
Un	-1	Unidentified	40-42	Particles detected but type unknown
No	00	No Precipitation	0, 4, 5, 10, 30-34, 94	Includes fog, mist, haze, smoke
Dr	01	Drizzle	50-53	All drops < 0.5 mm diameter
DrRa	02	Rain & Drizzle	57, 58	Changing between each over time
Ra	03	Rain	60-63	
Ic	04	Ice Crystals	78	Pristine ice crystals; no aggregation
WIc	05	Wet Ice Crystals	—	Presence of liquid water
Sn	06	Snow	70-73	Aggregated ice crystals
WSn	07	Wet Snow	67, 68	Presence of liquid water
Pl	08	Ice Pellets	74-76	Partially melted aggregates, subsequently re-frozen
Gr	09	Graupel	—	Rimed and accreted ice particles, < 5 mm diameter
Ha	10	Hail	89, 96	Rimed and accreted ice particles, ≥ 5 mm diameter
FzDr	11	Freezing Drizzle	54-56	Supercooled below 0 °C All drops < 0.5 mm diameter
FzDrRa	12	Freezing Drizzle & Freezing Rain	—	Supercooled below 0 °C Changing between each over time
FzRa	13	Freezing Rain	64-66	Supercooled below 0 °C

therefore shortened to 0.375 mm to 5.0 mm. The DSD and DVD data have also been normalized by instrument sampling area.

3.3. Precipitation Type

PT observations are recorded as PW codes which are difficult to compare because the data are categorical and their interpretation is ambiguous. PT is often not explicitly described by a PW code. To simplify the analysis, a new PT scheme has been created (shown in Table 3.3) with translations from PW codes, and is broadly ranked by the impact of the SPTs to society.

Wexler (1955) notes that empirical evidence by Langleben (1954) demonstrates that a snow aggregate retains its velocity characteristics until > 90% of the mass has melted. Lumb (1963) further showed empirically that snowflakes are reported by trained observers up to 1.5 to 3.0 °C, corresponding to approximately 300–500 m penetration below the 0 °C dry-bulb isotherm. Wet types, therefore, occur in a wide range of temperatures and are important to include.

All 1-minute PW code data from the study period are converted into PT as an initial step. Confusion matrices are produced which tally occurrences of each PT between instruments. For numerical comparison, a statistical score for evaluating instrument agreement is needed. Rather than comparing individual PTs between instruments, a multi-class score is desirable; the multi-class adaptation of the Heidke Skill Score (henceforth HSS) is used:

$$HSS = \frac{\sum_{i=1}^{I} p\left(y_i, x_i\right) - \sum_{i=1}^{I} p\left(y_i\right) p\left(x_i\right)}{1 - \sum_{i=1}^{I} p\left(y_i\right) p\left(x_i\right)} \tag{3.2}$$

where $\sum_{i=1}^{I} p\left(y_i, x_i\right)$ is the proportion correct, $\sum_{i=1}^{I} p\left(y_i\right) p\left(x_i\right)$ is the random proportion correct, 1 is the perfect forecast score, I is the length of the confusion matrix, y_i is the ith row and x_i is the ith column (Heidke 1926). HSS indicates the fractional improvement in agreement over a randomised observation set, which would score zero. The worst possible score is $-\infty$, and the best score is 1; negative values indicate that a random guess would have been more skillful. The HSS is symmetric, e.g. Thies1 vs. PWD21 yields the same score as PWD21 vs. Thies1. HSS tends to decrease for more complex PT schemes with more classes. The PT scheme introduced here has 13 classes which are explicit (i.e. more classes would not add value to a user). HSS is also calculated for a simplified scheme: none, liquid, mixed-phase, solid, to illustrate the range of HSS that can be obtained by using a simplified scheme that is more ambiguous to a user.

The instruments used in this study do not output PW codes at consistent time periods. 5, 10, 15 and 60 minute intervals are used by some but not by all instruments but all instruments do output a 1 minute interval. Evaluating longer timescales of PT is useful because radars often operate on a 5 minute (or longer) sampling strategy, and some weather reports are conducted hourly. The instruments in this study employ different and poorly documented methods to merge 1-minute observations.

For fair comparison between instruments, a new algorithm to merge 1-minute PT observations into PT assessments of longer duration is developed. A set of Boolean logic criteria are applied sequentially and are outlined in Figure 3.4.

4. Results

4.1. Precipitation Depth and Rate

Over the entire 12 month period, Figure 3.5 shows the variation in PD recorded by the instruments. The PWD21 reports much less (280.8 mm) precipitation than the average for the period (542.8 mm). The JWD records more accumulation than the TBR and remains closest to the mean of all instruments, but this should not be confused with being the most accurate. The PWS100 has a positive bias in July but this is counteracted with a negative bias at the end of September, resulting in a final accumulation close to the average. The RAL and Thies1 agree until November where the RAL develops a positive bias until the end of the period and records 618.8 mm in total. The Thies LPMs disagree; with Thies1 and Thies2 recording 591.7 and 744.8 mm respectively (26% difference).

Next the PRs are compared, first with a focus on the 5-minute time period (operational radar periodicity) before other time periods are considered. In Figure 3.6, every comparison with the PWD21 has a gradient of between $0.28 < \mathrm{m} < 0.42$ and the lowest coefficients of any instrument. The comparison reveals that the PWD21 is consistently measuring less rainfall than the other instruments. Similarly, the TBR records less precipitation than the PWS100, Thies LPMs, JWD and RAL but more precipitation than the PWD21, and has the second lowest r^2 values. Furthermore, the TBR versus the PWD21 has the lowest r^2 value of any case, indicating that the TBR and PWD21 are also inconsistent in their underestimations. The TBR has a resolution of 0.2 mm which in a 5-minute sampling period equates to a 2.4 mm h^{-1} minimum rainfall rate, hence the TBR data are aliased (insufficient sampling frequency for the signal frequency) at this time resolution. The intercept values for the TBR tests have large deviations from the origin which highlights the poor performance of the TBR in weak rainfall.

The remaining instruments (RAL, PWS100, JWD, Thies LPMs) have higher coefficient values of $0.89 < r^2 < 0.95$. The Thies2 generally has slightly higher coefficient values than the Thies1, which is also observed in Figure 3.5. The gradient and intercept of the RAL and JWD consistently show a slight negative bias compared with other instruments. The JWD–RAL comparison shows that both instruments must have similar observational characteristics because the r^2 of their comparison is high (0.947). Since the JWD is shown to underestimate drop counts during heavy PRs (Section 4.2), the RAL must also be underestimating PR during heavy precipitation events.

In general, Figure 3.7 shows an increase in agreement as the sample duration increases. This is to be expected due to the decrease in the influence of random error as the sample duration increases. Though Figure 3.7 demonstrates that using longer time periods results in higher r^2 values overall, the difference between 1 hour and 1 day for some instruments does not change or even slightly decreases. The RAL, JWD and PWS100 all have similar r^2 values throughout the time periods. At a 1-hour interval, the TBR has equivalent coefficients to the other instruments (excluding the PWD21). The PWD21 reaches a maximum r^2 value of 0.8 at a 1-day interval, but this is still far below the other instruments. The Thies2 has slightly higher albeit similar coefficients of determination as the RAL, JWD and PWS100 which is unexpected considering that the Thies LPMs use the same instrument design.

4.2. Drop Size Distribution

Overall, Figure 3.8 shows an increase in the steepness of the right tail of the DSD with increasing PR, as expected from Marshall and Palmer (1948). For more detailed analysis we split the results into three parts.

4.2.1. Small Drop Sizes (< 1.0 mm)

The PWS100 records fewer drops than the other instruments for small drops in all PRs and durations; multiple orders of magnitude less in the smallest sizes (< 0.8 mm). In case 15 the PWS100 barely detects the precipitation compared to the other three instruments. Case 15 was intended to represent the 1 hour, 0.05 mm h^{-1} PR scenario. However in Figure 3.3 it is noted that the actual PR in case 15 is slightly lower than the desired amount, due to the method of finding stable PR cases. The actual PR in case 15 is 0.02 mm h^{-1} and this slightly lower PR, combined with the PWS100 being unable to count small drop diameters, is likely the cause of the PWS100 barely detecting the precipitation. The Thies LPMs agree with the JWD for small drop-diameters up to PRs of 2 mm h^{-1}, where the Thies LPMs record more small droplets than the JWD. Above the 2 mm h^{-1} cases, the JWD count-values for small drops increasingly deviate from the Thies LPMs and more towards the PWS100 which is known to underestimate small drops across all PRs.

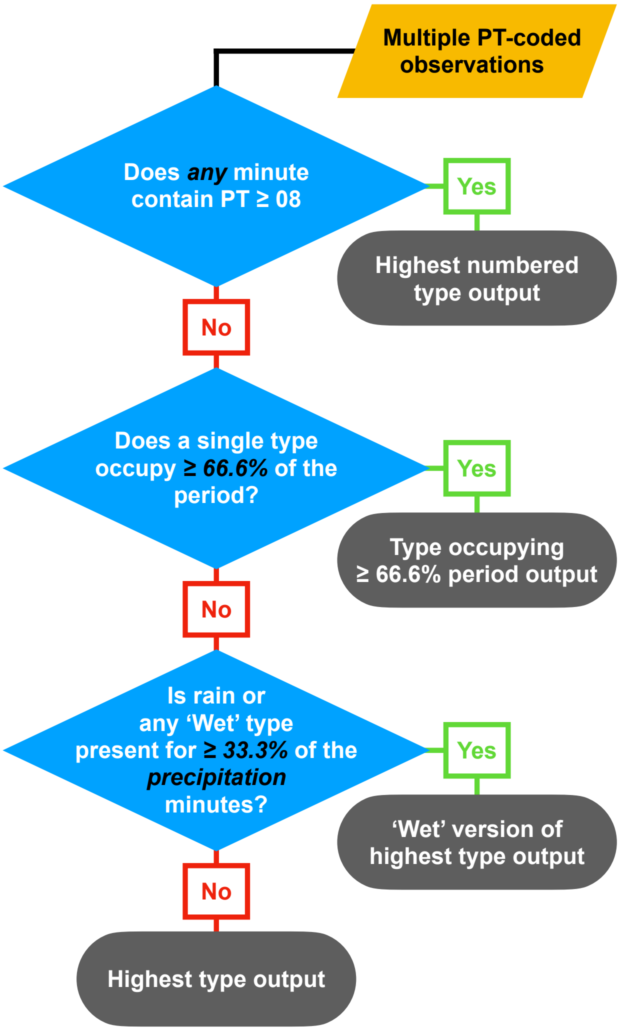


Figure 3.4.: A Boolean algorithm to take multiple precipitation type-data periods and merge them into longer periods. Any PT ≥ 8 in the input data results in that type as the output. This is because these events are rare, high-impact, and short duration, which may not meet the next criterion in the algorithm. Next, if any single PT occupies $\geq$ two-thirds of the input data, that is the output. Thirdly, the minutes with no precipitation are excluded. Of the remaining data, if $\geq$ one-third contain rain or any wet PT, then the highest wet PT code index that exists in the input data is used.

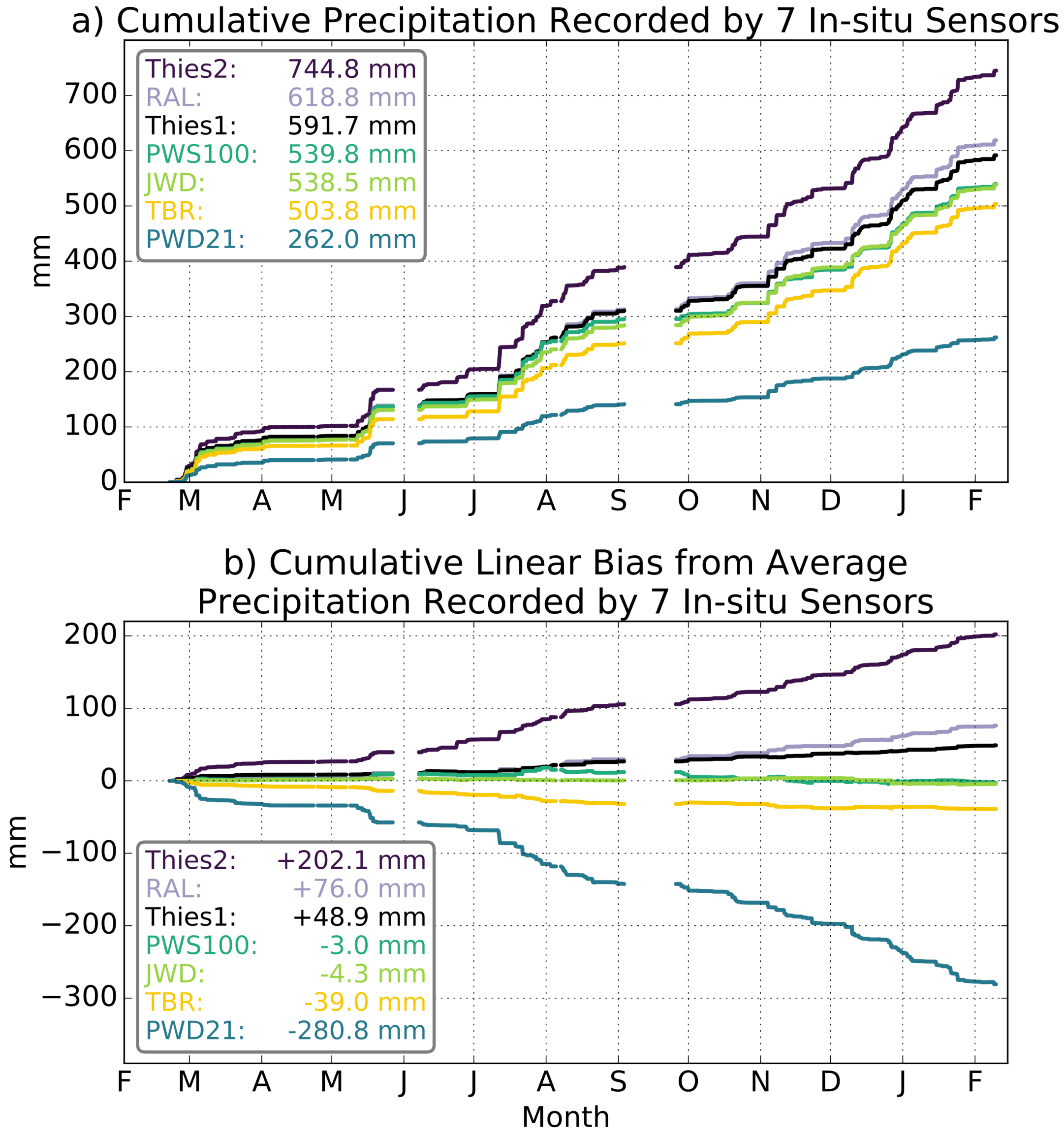
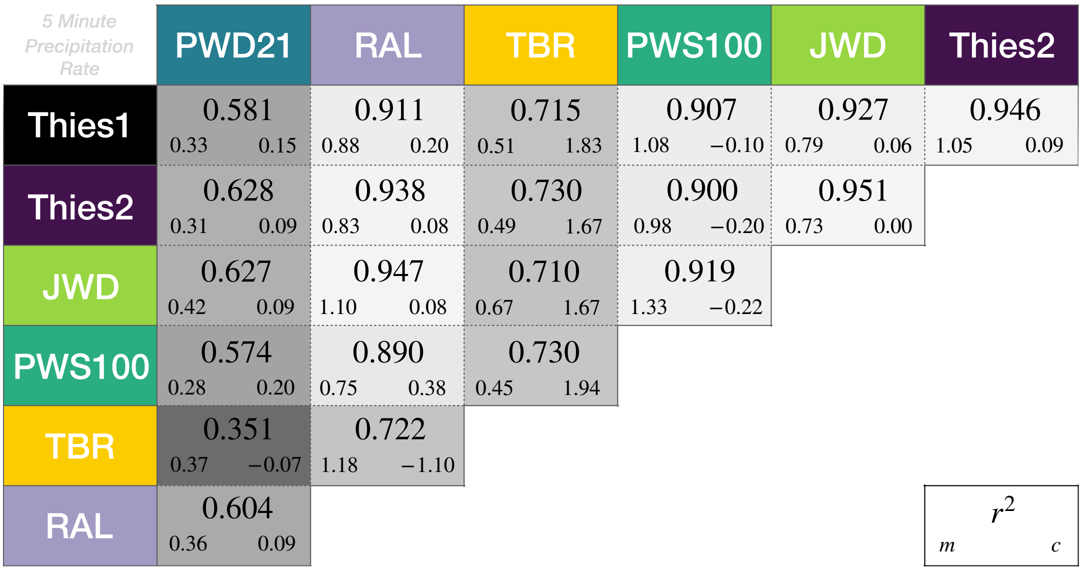


Figure 3.5.: Long term precipitation accumulation behaviour between the different instruments. 3.5a) shows the total precipitation accumulation, and 3.5b) shows the bias, using the mean of all sensors as the baseline. The total accumulation for each instrument is listed for each instrument in decreasing order in the legend of Figure 3.5a, and the total deviation from the mean of all instruments (542.8 mm) at the end of the period is listed in order in Figure 3.5b. The accumulation measured after 12 months ranges from 262.0 mm to 744.8 mm (284% difference), with a range of bias from -280.8 mm to +202.1 mm.



5 Minute Precipitation Rate	PWD21	RAL	TBR	PWS100	JWD	Thies2
Thies1	0.581 0.33 0.15	0.911 0.88 0.20	0.715 0.51 1.83	0.907 1.08 −0.10	0.927 0.79 0.06	0.946 1.05 0.09
Thies2	0.628 0.31 0.09	0.938 0.83 0.08	0.730 0.49 1.67	0.900 0.98 −0.20	0.951 0.73 0.00	
JWD	0.627 0.42 0.09	0.947 1.10 0.08	0.710 0.67 1.67	0.919 1.33 −0.22		
PWS100	0.574 0.28 0.20	0.890 0.75 0.38	0.730 0.45 1.94			
TBR	0.351 0.37 −0.07	0.722 1.18 −1.10				
RAL	0.604 0.36 0.09					r^2 m c

Figure 3.6.: Every instrument compared with another for 5 minute-averaged PR. Within each cell, the large upper value is the coefficient of determination r^2, whilst the lower left and lower right are the gradient m and y-axis intercept c of the linear least-squares regression line respectively. The cells are shaded based on the r^2 value. The instruments in the columns are the y-axis data and the instruments in the rows are the x-axis data, such that the uppermost left result of $r^2 = 0.581$ is taken from a scatterplot of Thies1 on the x-axis and PWD21 on the y-axis. Hence, the gradient of Thies1 vs. PWD21 ($m = 0.33$) indicates that the PWD21 records a third of the rainfall that the Thies1 recorded during the 12-month observation period.

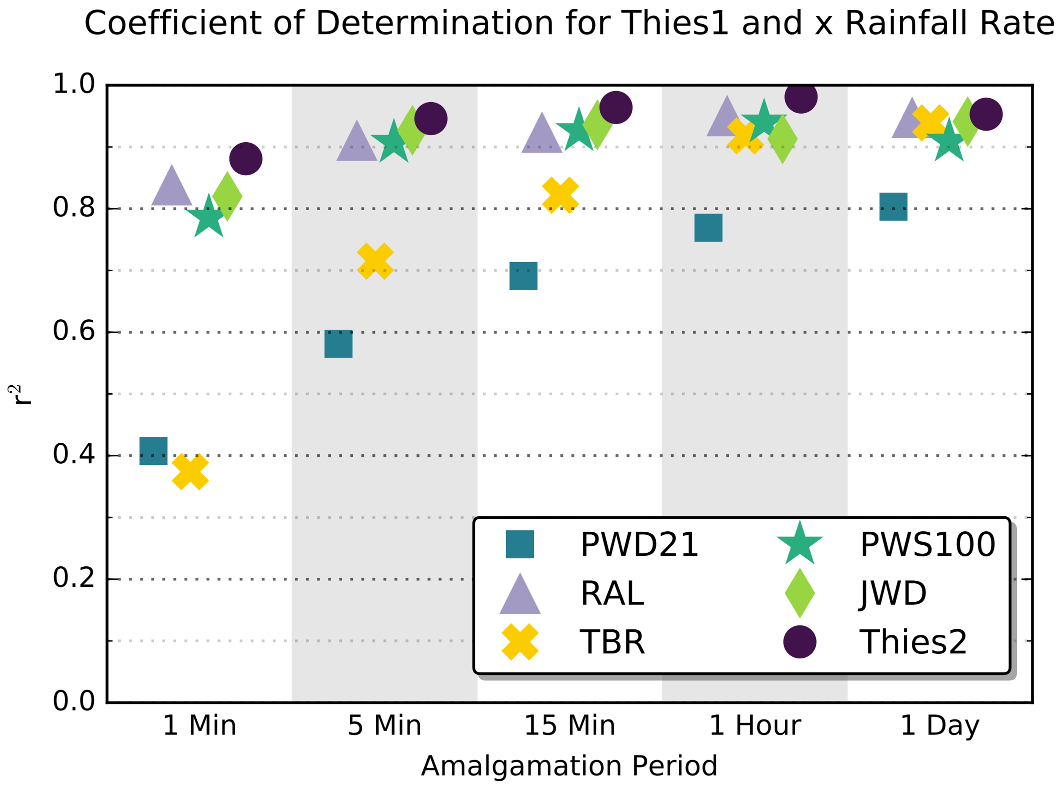


Figure 3.7.: Coefficient of determination (r^2) used as an indicator of instrument agreement between PRs measured by the labelled instrument and Thies1. Values are shown for multiple time durations to indicate the dependence of time interval on agreement. Data from the whole study period is binned into time durations labelled, and a coefficient is calculated using least-squares linear regression as explained in Section 3.1. Thies1 is used as the baseline instrument because Figure 3.6 showed it to be similar to the Thies2, RAL, PWS100 and JWD, and allows the Thies LPMs to be compared.

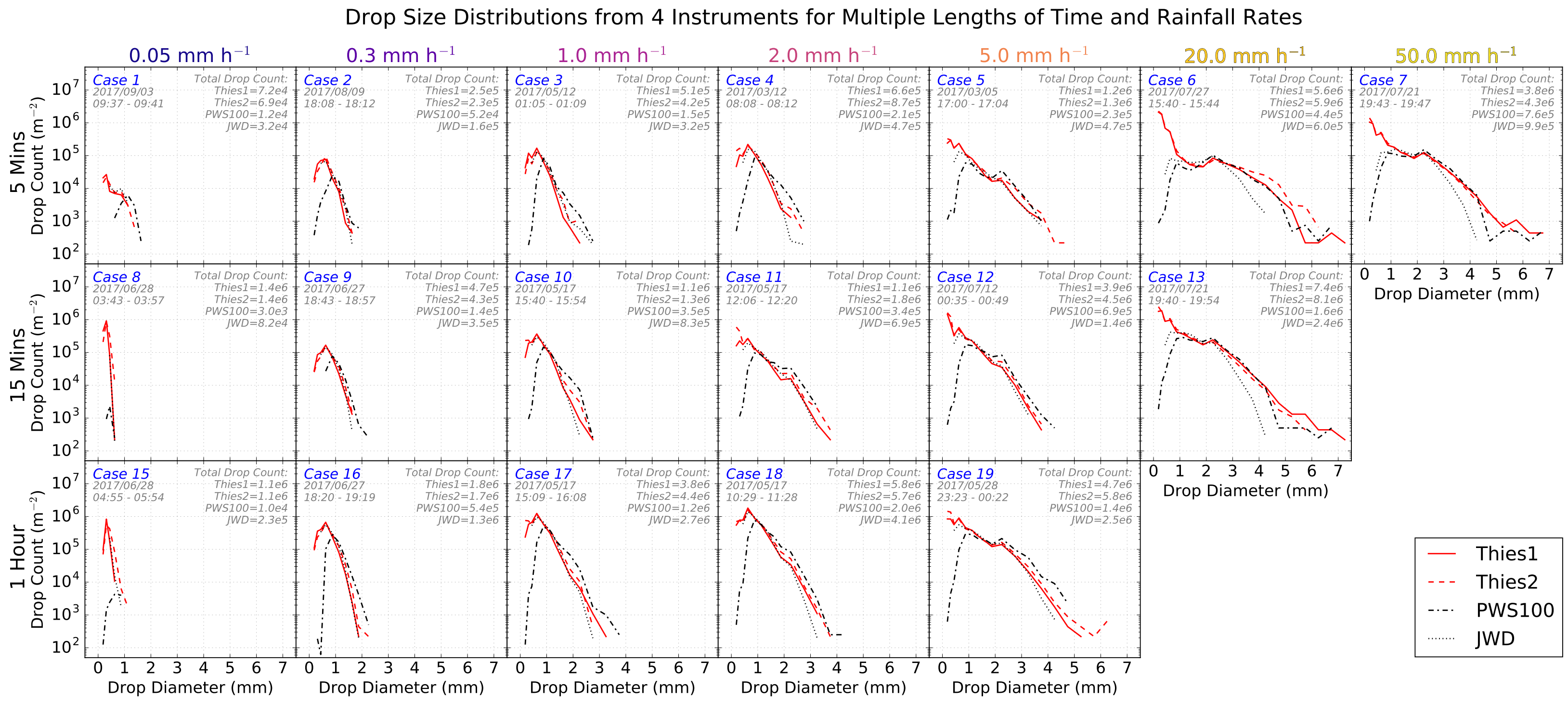


Figure 3.8.: Drop size distribution (DSD) of 4 instruments in 18 event periods. The event periods are chosen using the PRs in Figure 3.3. The rows denote the duration of the event period; 5 minutes, 15 minutes, and 1 hour. The columns denote the intensity of the rainfall during the period. The date and time of each case is displayed in the top left and the total drop count measured by each instrument in each case is displayed in standard form in the top right of each subplot.

4.2.2. Medium Drop Sizes (1.0–3.0 mm)

The spread of medium-sized drop counts measured by all instruments is within an order of magnitude, showing broad agreement. For the medium-intensity PRs (1.0, 2.0, 5.0 mm h^{-1}), the PWS100 typically records equal or more drops in the 1-3 mm diameter region of the distribution. The largest deviation is in case 17 where the PWS100 records 50,000 drops, Thies1 and JWD 10,000 drops, and Thies2 11,000 drops at 2 mm. The JWD agrees with the Thies LPMs for medium-sized drops in weak PR (0.3 mm h^{-1}), but in 3 of the cases (10, 17, 18) records the least drops in the 2.0–3.0 mm size bin (approaching 1 order of magnitude towards 3.0 mm drops) compared to the Thies LPMs and PWS100. Case 10 has the largest spread in medium-sized drop counts which is 1 order of magnitude at 2 mm.

4.2.3. Large Drop Sizes (> 3.0 mm)

The Thies LPMs and JWD are in agreement for large drops, except during the more intense rain rates (20, 50 mm h^{-1}), where the JWD records 1–1.5 orders of magnitude fewer large drops than the Thies LPMs and PWS100. The JWD also records the largest drops to be in the 4.0–4.5 mm size bin, and none in the 4.5–5.0 mm size bin in any of the cases. The largest difference between the Thies LPMs is at 4.25 mm in case 6 (5-minute, 20 mm h^{-1}) but only around 100 drops. The PWS100 is in agreement with the Thies LPMs in most cases for large drops. In 4 of the cases (12, 17, 18, 19) the large drop counts are marginally higher, but above 4.5 mm diameter drops, the counts from the PWS100 are less than the Thies LPMs. Out of the 18 cases, the Thies LPMs record the largest drop-diameters out of any instrument 5 times and equal largest drop-diameters 6 times, occurring more often in the higher PRs. The PWS100 records the largest drop-diameters in 7 of the cases, occurring more often in weak PRs.

4.3. Drop Velocity Distribution

Figure 3.9 shows the DVD results, in the same cases and layout as in Figure 3.8, but without the JWD which cannot measure drop velocity. Broadly the PWS100 counts fewer particles in total (the cases are identical to Figure 3.5) and has a DVD upper tail which ends at lower velocities in all cases compared to both of the Thies LPMs. The DVD lower tail and peak counts of drop velocities have less than an order of magnitude of spread between instruments, with the notable exception of cases 6, 8 and

15. Cases 8 and 15 are again affected by the PWS100 undercounting small drops. The two Thies LPMs agree to within half an order of magnitude (except in case 15, very weak precipitation) but Thies2 records more drops > 4 m s^{-1} than Thies1 in almost every case, and records the highest drop velocities in 8 cases (equal to Thies1 in 8 cases, less than Thies1 in 2 cases).

4.4. Precipitation Type

Figure 3.10 shows the agreement between instruments on coincident observations of PT at 5-minute intervals (operational radar periodicity). The most striking result is that the Thies LPMs report more PTs than the PWS100 and PWD21. The PWS100 never reports wet ice, and only observes 2 periods of ice and 1 period of hail. The PWD21 never reports drizzle and rain, ice, wet ice or hail. Along the diagonal boxes (bordered in white) the highest agreement between instruments occurs in the no precipitation class ($> 98\%$ across-the-board), the rain class ($> 80\%$) and the snow class ($> 30\%$).

The drizzle class has broad results; both of the Thies LPMs count more drizzle than rain. The PWS100 and PWD21 record ∼20 times fewer drizzle periods than the Thies LPMs, and yet in the lower right matrix (Figure 3.10f), both instruments rarely (6%) agree on the times of drizzle despite having similar total counts. The PWS100 frequently identifies rain when the other instruments identify drizzle (8-15%) or drizzle & rain (80–90%) classes.

The spuriosity of the Thies LPMs is evidenced by the first two rows and columns in Fig. 10a, where the none and drizzle classes all contain other hydrometeors. Wet snow is equally agreed upon as it is disagreed as rain by both Thies LPMs (∼40% each). The PWD21 has less spuriousness because it is unable to resolve many hydrometeor types—evident with the many zero count rows (grey boxes) in Figure 3.10c, 3.10e, and 3.10f. The PWS100 typically classifies Thies LPM drizzle as no precipitation (73-84%) or rain (8-16%), Thies LPM ice as no precipitation, and Thies LPM hail as snow, rain or none. The snow has some agreement compared to the other instruments (32-39%) but often the PWS100 classifies Thies LPM snow as no precipitation (52-64%). The PWS100 and PWD21 have large disagreements, as do the two LPMs with themselves and the other instruments, highlighting the difficulty in observing PT. The HSS from each comparison for each time period examined is summarised in Figure 3.11.

Unsurprisingly, Figure 3.11 shows the two identical Thies LPM instruments have the highest agreement. Second highest are the PWS100-PWD21 matrices. For the other

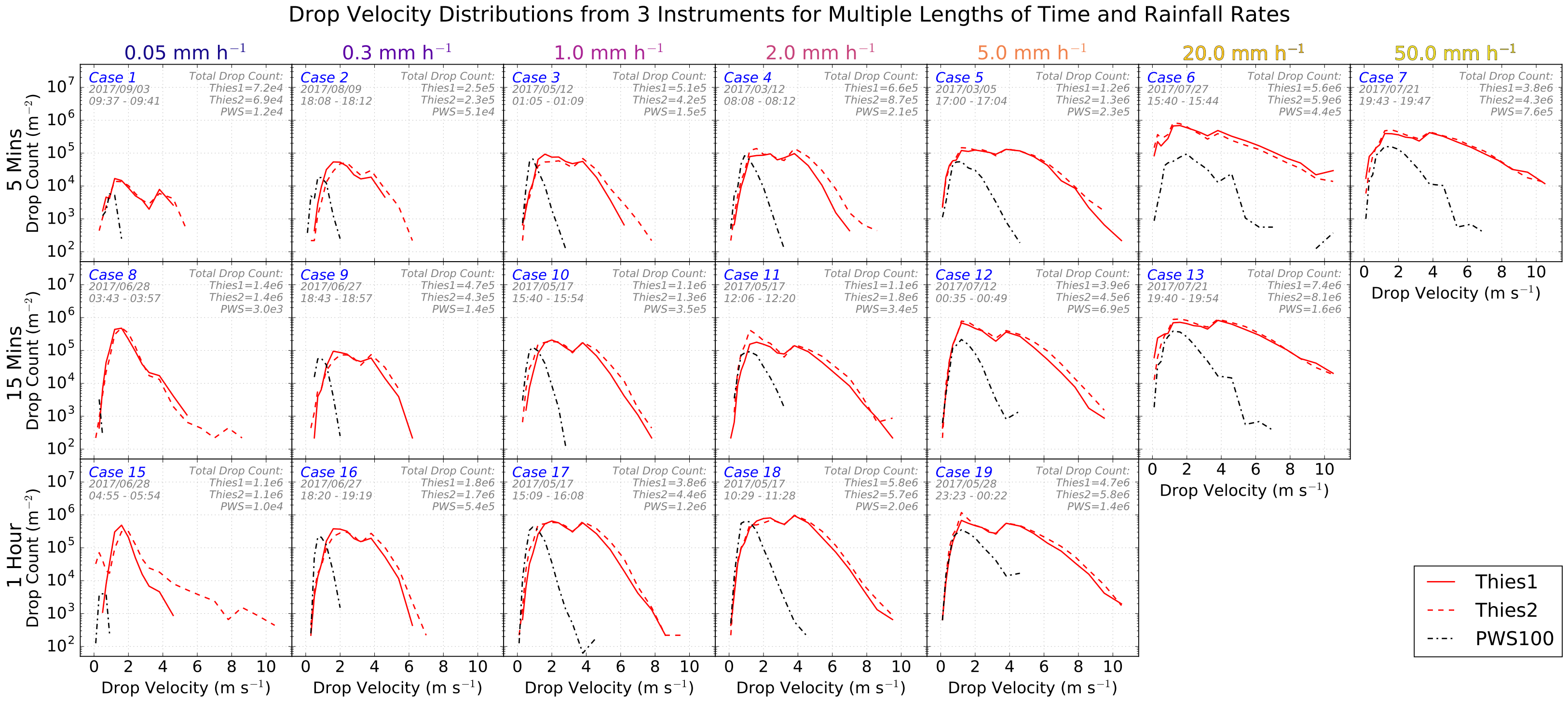


Figure 3.9.: Drop Velocity Distribution of 3 instruments in 18 study periods. The study periods are identical to those used in Figure 3.8. The rows denote the duration of the study period; 5 minutes, 15 minutes, and 60 minutes. The columns denote the intensity of the rainfall.

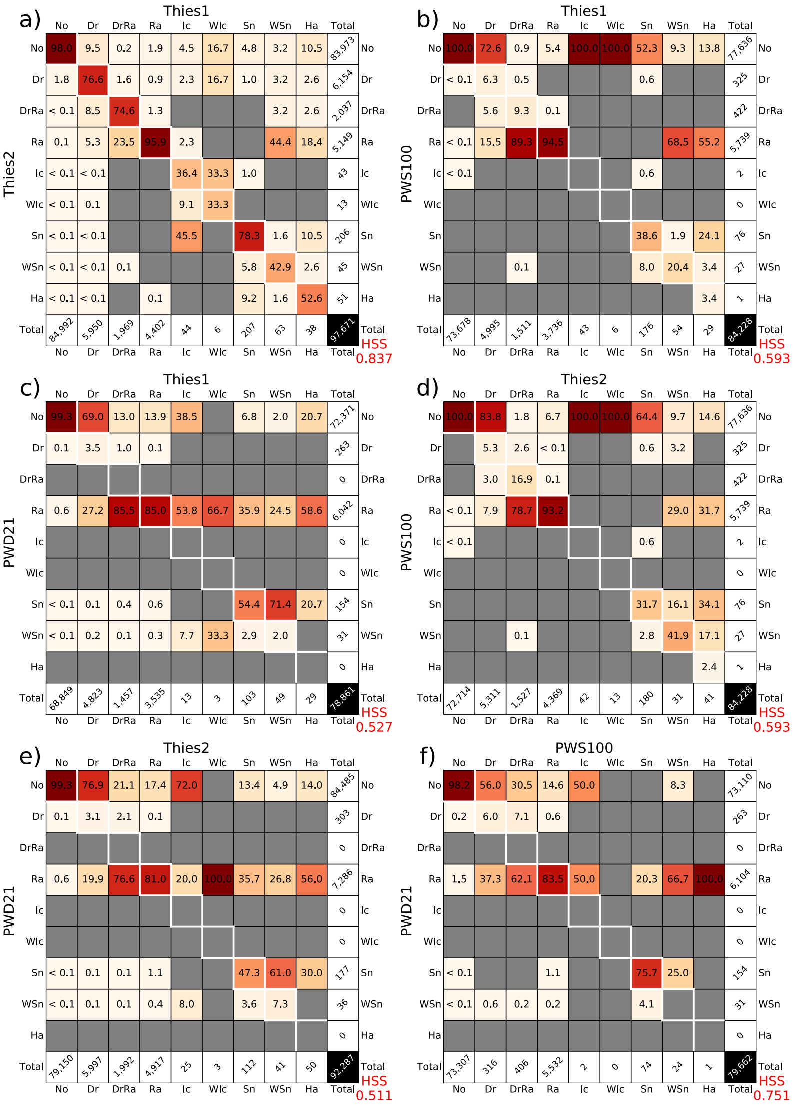


Figure 3.10.: Confusion matrices of hydrometeor type for 5-minute time intervals. Values shown inside the matrix are percentages normalized by the total number of observations of the type in that column, such that the instrument listed on the top of the matrix is considered truth in each plot. E.g. for the first matrix, for the ‘wet snow’ precipitation type (denoted WSn), 42.9% of the observations made by Thies2 agreed with Thies1. Looking further up in that same column, it shows that 44.4% of the Thies1 ‘wet snow’ events were classified as ‘rain’ by Thies2. The color intensity of each cell from white to dark red corresponds to the percentage values written inside each cell. Totals observed by an instrument are shown in the white boxes surrounding the matrix and the total number of observations in the black lower right box which differs for each matrix since both instruments in a matrix must report a PT (Section 3.3). The multi-class Heidke Skill Score (HSS; Heidke (1926)) of each confusion matrix is shown in red, in the lower right corner outside each matrix.

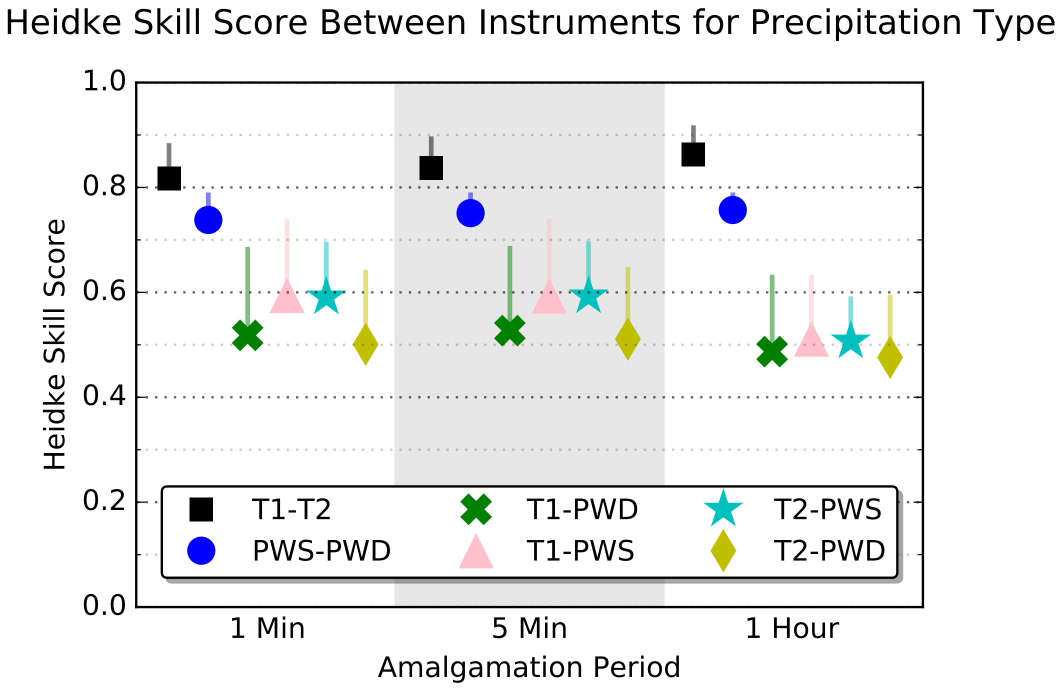


Figure 3.11.: Multi-class Heidke Skill Scores (HSS) between each instrument capable of PT observations. Scores are shown for 1, 5 and 60-minute intervals. The vertical lines indicate the improvement in score if a simpler hydrometeor class scheme is used (none, liquid, mixed-phase, solid). The PWS100 is shortened to PWS here, Thies1 to T1, Thies2 to T2 and the PWD21 to PWD for readability.

comparisons, the Thies-PWD21 agree less than the Thies-PWS100. The vertical lines on Figure 3.11 show the HSS if the confusion matrices were to be simplified into four classes: none, liquid, mixed-phase and solid, where mixed-phase are any wet-denoted type. Solid includes ice, snow and hail. HSS increases modestly across-the-board if a simpler hydrometeor class system is used. However this would not help in instances of hail and snow where the two classes are merged, nor does it indicate which type of mixed-phase precipitation is occurring.

5. Discussion

The performance of the instruments depends on the user. This section will consider applications in context with the results seen in the study to inform existing and future sensor deployments and design.

5.1. Precipitation Depth and Rate

The results confirm that the PWD21 is unable to measure over 20 mm h^{-1}—no precipitation is observed over this rate; convective events are significantly under-captured. Concurrently the sensitivity of weak PRs is poorly captured, evidenced by a positive intercept when compared with all other instruments at 5 minute sample duration (except the TBR which has time resolution limitations). The PWD21 should not be used as a PR or depth sensor where possible. Funnel gauges (TBR and RAL) are affected by air temperature and humidity. The funnels will take longer to wet in the summer because precipitation can more rapidly evaporate. There is evidence for this seasonality as Figure 3.5 shows that the largest increase in bias in the TBR occurs in June-August, and the RAL bias increases more than the other instruments from November onwards. The TBR can also lose liquid within a bucket through evaporation, which the RAL mitigated by employing a less exposed drop collection reservoir design. This may explain the larger summer differences in the TBR vs the RAL. Funnel gauges are also affected by wind (Sevruk 1996).

The Thies LPMs have a large difference in recorded precipitation over 12 months (591.7 and 744.8 mm for Thies1 and Thies2 respectively, or 25% more in Thies2). The Thies1 accumulated total is close to other instruments throughout the year and both Thies LPMs have similar r^2 values to other instruments in Figure 3.6, the largest difference being for the PWD21 with a difference of 0.047. However, the gradient m between the

Thies LPMs is 1.05 which suggests a 5% difference. Here, the limitation of the least squares regression is highlighted; examination of the scatterplot between the Thies LPMs (not shown) reveals several cases where the Thies2 reported 5-minute precipitation as much as half that of the Thies1, skewing the gradient of the best fit. The scatterplot shows that the overdiagnosis of PR (at 5-minute intervals) in Thies2 compared to Thies1 generally occurs across all PRs and is likely a systematic bias with the laser to diode occultation technique. Lanza and Vuerich (2009) noted an overestimation in precipitation from optical disdrometers compared with a reference gauge and Lanzinger (2006) quantified a 5-20% overestimation from optical disdrometers. Frasson et al. (2011) suggests that the Thies LPM is sensitive to precipitation particles outside of the manufacturer-quoted laser beam area, causing an overcount of particles for the specified area when compared with a TBR-gauge. Large differences between Thies LPMs were also noted in Frasson et al. (2011), suggesting that the build and calibration consistency by the manufacturer is poor.

A phenomenon observed in Figure 3.8 with the JWD is the underestimation of small and large drops in heavy PRs, also observed in Figure 3.6 because the gradient of the JWD versus RAL, PWS100 and Thies LPMs all show an underestimation by the JWD. The total PD over 12 months is near the mean of all instruments (-4.3 mm bias), but the DSD results indicate that the JWD underdetects PR, so the PD value should be lower than truth. The PWS100 has a similar 12 month PD (-3.0 mm from average of all instruments) which is likely to be an underestimate and r^2 values are around 0.9 (if the lower TBR and PWD21 are dismissed). These values are the lowest of the RAL, JWD and Thies LPMs and the PWS100 also has a lower gradient than the Thies1, JWD and RAL for 5-minute intervals seen in Figure 3.6, further supporting the conclusion that the PWS100 is underdiagnosing PR and PD.

5.2. Drop Size Distribution

The DSD comparisons are split up by PR and accumulation time. There are few differences between the time accumulations showing that 5 minutes is sufficient for a representative DSD shape. The Thies LPMs and PWS100 measure some drops larger than the JWD can measure (> 5.0 mm). It has been shown that the maximum stable diameter of a raindrop in stationary air before breakup occurs is ~6 mm (Villermaux and Bossa 2009; Marshall and Palmer 1948). This suggests that the largest drops measured by the Thies LPMs and PWS100 are realistic. Therefore the JWD appears to be limited and from the results in Figure 3.8 is also underdiagnosing the number of large droplets. This is important for heavy PRs which can cause flash flooding because larger drops contribute more volume of water to the rain rate than smaller drops. Capturing larger drops in disdrometer measurements is also important for comparison with, or calibration of, precipitation radars. Smith (2016) found that the largest 6% of drops in a simulated DSD contributed towards 85% of the value of the reflectivity factor, Z_h. The JWD should therefore not be used for applications where high PRs need to be well captured.

The JWD also underdiagnoses the small drops at high PRs ($>$ 2.0 mm h^{-1}), which may be due to the dead-time effect, not accounted for here, where large drops cause oscillations on the JWD plate which take a non-negligible time to dampen. In this time, motion from small drops is undetectable. However, the impact of the dead-time correction has been questioned in the literature (Ulbrich and Atlas 2007). The JWD agrees with the Thies LPM for lower PRs ($<$ 2.0 mm h^{-1}) so high sensitivity to drizzle and weak PRs makes the JWD suitable for those applications.

Alternatively, the Thies LPMs have a steeper DSD for small drops during high PRs ($>$ 5.0 mm h^{-1}) which is unrealistic; the Thies LPM housing has been suspected of splashing precipitation into the measuring beam (Pickering et al. 2019) which would increase the number of small drops seen in Figure 3.8. Sensitivity at weak PRs and drizzle is also high in the Thies LPMs, although this does occasionally result in some anomalous measurements of insects or debris. The PWS100 underdiagnosing small drops could be a critical failure for automated precipitation-detection applications. In the PWS100 manual it states “proportion of particles detected will fall off significantly below about 0.5 mm diameter”, not below 1.0 mm as seen here and should therefore not be used for high-sensitivity purposes. All instruments have suspect drop count values in the small drop diameter sizes for higher PRs ($>$ 5.0 mm h^{-1}), so it is not clear that any of the instruments tested capture the full DSD correctly for high PRs ($>$ 5.0 mm h^{-1}).

5.3. Drop Velocity Distribution

As with DSD, the difference in distribution shape is consistent for a given PR across multiple data collection time samples, indicating that a 5-minute sample is sufficiently representative. Pickering et al. (2019) note that there are often cases where the Thies LPM measures a portion of the smallest particles (0.125 mm) at the highest velocity bin ($>$ 10 m s^{-1}) which is incompatible with the empirically derived relationship between drop diameter and terminal fall velocity by Gunn and Kinzer (1949). The cause of these small diameter and high velocity drops is unknown but the behaviour exists in both Thies LPMs used in this study.

The Gunn-Kinzer curve is a widely-used relationship between drop size and terminal velocity. In the 5.0 mm h^{-1} cases (5, 12, 19), the maximum drop-diameters of around 4.0–4.5 mm for both the PWS100 and Thies LPMs should have a terminal velocity of 8.7-9.0 m s^{-1} according to the Gunn-Kinzer curve. In Figure 3.9 however, the PWS100 has a maximum velocity of 4.5 m s^{-1} and the Thies LPMs of 10.5 m s^{-1} which is higher than Gunn-Kinzer but closer than the PWS100. This trend continues across all DVDs. The Thies LPMs are close to the values expected from Gunn-Kinzer, but the PWS100 DVDs appear to be linearly stretched to approximately half the values expected.

Since the PWS100 under-detects small drops, the DVD will also be affected. Despite this the counts of low velocity drops are the same or higher than the Thies LPMs in 13 out of the 18 cases, which supports the observation that the Thies LPMs incorrectly measure a broad range of velocities for small drops, decreasing the true low velocity (< 1.0 m s^{-1}) count. Since the Thies LPMs use only DSD and DVD for PT, these anomalies in observation will lead to incorrect diagnoses of PT when particle sizes are small.

5.4. Precipitation Type

Both Thies LPMs have high drizzle counts, higher than rain which does not match the site climatology (Ventouras et al. 2006). This signals a systematic design error which may be due to the sensitivity of the instrument to any small particles in the atmosphere like insects or debris. Both the PWS100 and PWD21 counted fewer drizzle cases than the Thies LPMs. In Section 4.1 it was shown that the PWS100 underestimates small drops; drizzle is defined as drops < 0.5 mm in diameter, so the result is consistent with earlier findings. Similarly, the PWD21 does not output DSDs, so cannot explicitly distinguish drizzle. Ice and wet ice are the most disagreed classes simply because the climatology of the study location has rare occurrences of pristine ice crystals. Hail is also rare in the UK (Hand and Cappelluti 2011) and is therefore unreasonable to evaluate over a 12-month period.

The results for PT show weak agreement across the instruments, which is consistent with the literature for similar instruments. Bloemink and Lanzinger (2005) showed that for a simple PT classifier (none, liquid, mixed-phase, solid), there was little difference between the Thies LPM and Vaisala FD12P (akin to the PWD21 with a visiometer, thermometer and capacitive plate) when each were compared with a trained observer. They noted a particular weakness for the instruments to detect mixed-phase precipitation, also seen here.

There is little difference in the HSS for different time periods, which demonstrates that the amalgamation algorithm presented here is successful; if there were large increases or decreases in the HSS over time, then the algorithm is introducing a bias. The PT scheme and amalgamation algorithm should be applied to future instruments and to research analysis, so that there is consistency between instrument manufacturers in this field. The explicit PT scheme will also require instruments to be explicit, which the current PW code does not require since it is ambiguous.

5.5. Study Limitations

Several limitations exist with the current study which must be noted and considered. Firstly there is no single instrument which can be identified as the truth, and therefore the results are limited to comparisons between two instruments or an average of all instruments (referred to as a "composite working reference" in intercomparison studies). By contrast, Lanza and Vuerich (2009) use a composite working reference of 4 gauges (two TBRs and two weighing gauges) identified from laboratory comparisons (Lanza and Stagi 2009). The effect of wind on precipitation measurements has long been known (Heberden 1769; Jevons 1862; Koschmieder 1934) but here no considerations are made to isolate precipitation events by wind speed thresholds. This study was based on existing instruments on-site and their pre-installed locations which could not be changed due to the need to maintain consistent long-term measurements at CAO. The PWD21 and PWS100 instruments are on the roof of a small building, which is sub-optimal due to turbulence generated by the building—however, these instruments are raised above the building as shown in Figure 3.1 which reduces the turbulence within the instrument sampling volume. Finally, whilst 12 months of data covers seasonal changes, the study period cannot be considered a true climatological sample since 30 years of observations would be required. Note that other intercomparison studies also use a data collection period on the order of 1 year (Lanza and Vuerich 2009; Kochendorfer et al. 2017). The study is also conducted in a single location and therefore the results only apply to locations with similar precipitation climatologies as that of CAO (see Section 2.2) and up to PRs of 50 mm h^{-1}.

6. Summary

We examined six techniques for measuring precipitation to inform the future usage of their observations, sensor deployments and to provide a benchmark for new sensor development. Comparison techniques from the literature were modified and improved upon to further explore the consistency and reliability of the observations from sensors with different measurement techniques. Rainfall depth and rate were examined using linear regression. The regressions of each instrument were cross-compared to reveal consistent poor-agreement and other patterns. A novel technique for comparing drop size and velocity distributions was described and employed. By selecting cases of stable rainfall rate, the drop distributions are kept consistent for the comparisons. A new precipitation type look-up-table was created to convert and group the widely-used PW code format into more explicit classes which enables a standardised comparison with other instruments to be performed. An amalgamation algorithm was also introduced which merges multiple 1-minute precipitation type observations into a single code. The new PT codes and amalgamation algorithm were then used to compare 4 instruments over 3 timescales. Though the comparison of the instruments showed poor instrument agreement, the results demonstrate the successful application of the PT scheme and amalgamation algorithm.

For robust measurements, observing sites should employ multiple sensor designs. No single sensor in this study could satisfy all user applications. The Thies LPM makes reasonable observations for all variables, although the PT data is difficult to verify without a human observer and the PD between two Thies LPMs over 12 months had a 26% difference, signalling poor manufacturing calibration consistency. The PWS100 reported unrealistically low velocity measurements, fewer PTs than the Thies LPM and struggled to observe drizzle. The PWD21 performed poorly for every variable; PR and PD showed large negative biases of around 50% compared with the average of all instruments in the study, and PTs were narrowly reported (few classes). The JWD and PWS100 12-month PD were close to the mean PD but the DSD from the JWD underdiagnosed small and large drops, especially in higher PRs. The JWD and RAL had reasonable PR r^2 values with a fast response time. The RAL 12-month PD was 14% higher than the average for all instruments. The TBR has a slight negative PD bias compared with the average for all instruments over 12 months with 7.2% below the average. The low PD resolution of 0.2 mm makes the TBR unsuitable for response times less than 5 minutes for typical PRs observed at CAO.

A clear outcome of this study is that observations of PT, whilst useful and in grow-ing demand, are poor because there is only moderate agreement between instruments with mixed-phase precipitation disagreed upon the most. Other studies still use human observers as the most trusted PT which shows that there is a need for improved precipitation type sensors. Overall, the sensor design choice should change based on the user requirements, and this study serves as a reference for such a decision to be made.

References

Adolf Thies GmbH & Co. KG, 2011: Laser Precipitation Monitor - Instruction for Use. Tech. rep., Adolf Thies GmbH & Co. KG, Hauptstraße 76, 37083 Göttingen, Germany, 66 pp.

Agnew, J., 2014: Final report on the operation of a Campbell Scientific CS135 ceilometer at Chilbolton Observatory. **(March)**.

Al-Sakka, H., A. A. Boumahmoud, B. Fradon, S. J. Frasier, and P. Tabary, 2013: A new fuzzy logic hydrometeor classification scheme applied to the french X-, C-, and S-band polarimetric radars. *Journal of Applied Meteorology and Climatology*, **52 (10)**, 2328–2344, doi:10.1175/JAMC-D-12-0236.1.

Atlas, D., and F. H. Ludlam, 1961: Multi-wavelength radar reflectivity of hailstorms. *Quarterly Journal of the Royal Meteorological Society*, **87 (374)**, 523–534, doi:10.1002/qj.49708737407.

Bentley, W. A., 1904: Studies of raindrops and raindrop phenomena. *Monthly Weather Review*, **32 (10)**, 450–456, doi:10.1175/1520-0493(1904)32⟨450:SORARP⟩2.0.CO;2, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1904)32%3C450:SORARP%3E2.0.CO;2.

Bloemink, H., and E. Lanzinger, 2005: Precipitation type from the Thies disdrometer. *WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO-2005)*, 1–7, URL http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-82-TECO_2005/Papers/3(11)_Netherlands_4_Bloemink.pdf.

Campbell Scientific, 2012: PWS100 Present Weather Sensor User Guide. Tech. rep., 122 pp.

Chandrasekar, V., R. Meneghini, and I. Zawadzki, 2003: Global and Local Precipitation Measurements by Radar. *Radar and Atmospheric Science: A Collection of Essays in Honor of David Atlas*, R. M. Wakimoto, and R. Srivastava, Eds., American Meteorological Society, Boston, MA, 215–236, doi:10.1007/978-1-878220-36-3_9, URL https://doi.org/10.1007/978-1-878220-36-3_9.

Distromet Ltd., 2002: Disdrometer RD-80 Instruction Manual. Tech. rep., Distromet Ltd., Distromet AG, Islerenweg 2, 8126 Zumikon, Switzerland, 18 pp.

Distromet Ltd., 2012: Disdrometer RD-80 Operating Instructions. Tech. rep., Switzerland, 1–23 pp.

Eisenberg, D., S. Laustsen, and J. Stege, 2018: Wind turbine blade coating leading edge rain erosion model: Development and validation. *Wind Energy*, **21 (10)**, 942–951, doi:10.1002/we.2200.

Erkal, A., D. D'Ayala, and L. Sequeira, 2012: Assessment of wind-driven rain impact, related surface erosion and surface strength reduction of historic building materials. *Building and Environment*, **57**, 336–348, doi:10.1016/j.buildenv.2012.05.004.

Feingold, G., Z. Levin, and S. Tzivion, 1991: The evolution of raindrop spectra. Part III: downdraft generation in an axisymmetrical rainshaft model. 315–330 pp., doi: 10.1175/1520-0469(1991)048⟨0315:TEORSP⟩2.0.CO;2.

Frasson, R. P. d. M., L. K. da Cunha, and W. F. Krajewski, 2011: Assessment of the Thies optical disdrometer performance. *Atmospheric Research*, **101 (1-2)**, 237–255, doi:10.1016/j.atmosres.2011.02.014, URL http://dx.doi.org/10.1016/j.atmosres.2011.02.014.

Fujiyoshi, Y., T. Endoh, T. Yamada, K. Tsubokj, Y. Tachibana, and G. Wakahama, 1990: Determination of a Z-R Relationship for Snowfall Using a Radar and High Sensitivity Snow Gauges. *Journal of Applied Meteorology*, **29 (2)**, 147–152, doi: doi:10.1175/1520-0450.

Fulton, R. A., J. P. Breidenbach, D.-J. Seo, D. A. Miller, and T. O'Bannon, 1998: The WSR-88D Rainfall Algorithm. *Weather and Forecasting*, **13 (2)**, 377–395, URL http://journals.ametsoc.org/doi/abs/10.1175/1520-0434%281998%29013%3C0377%3ATWRA%3E2.0.CO%3B2.

Gunn, R., and G. D. Kinzer, 1949: The Terminal Velocity of Fall for Water Droplets in Stagnant Air. *Journal of Meteorology*, 243–248, doi:10.1175/1520-0469(1949)006⟨0243:TTVOFF⟩2.0.CO;2.

Hand, W. H., and G. Cappelluti, 2011: A global hail climatology using the UK Met Office convection diagnosis procedure (CDP) and model analyses. *Meteorological Applications*, **18 (4)**, 446–458, doi:10.1002/met.236.

Harimaya, T., 1978: Observation of Size Distribution of Graupel and Snow Flake. *Journal of the Faculty of Science, Hokkaido University, Ser. VII (Geophysics)*, **5 (3)**, 67–77.

Harrison, D., S. J. Driscoll, and M. Kitchen, 2000: Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorol. Appl.*, **6**, 135–144, doi:10.1017/S1350482700001468.

Heberden, W., 1769: XLVII. Of the different quantities of rain, which appear to fall, at different heights, over the same spot of ground. *Philosophical Transactions of the Royal Society of London*, **(59)**, 359–362.

Heidke, P., 1926: Berechnung Des Erfolges Und Der Güte Der Windstärkevorhersagen Im Sturmwarnungsdienst. *Geografiska Annaler*, **8 (4)**, 301–349, doi:10.1080/20014422.1926.11881138, URL https://doi.org/10.1080/20014422.1926.11881138.

Humphrey, M. D., J. D. Istok, J. Y. Lee, J. A. Hevesi, and A. L. Flint, 1997: A new method for automated dynamic calibration of tipping-bucket rain gauges. *Journal of Atmospheric and Oceanic Technology*, **14 (6)**, 1513–1519, doi:10.1175/1520-0426(1997)014⟨1513:ANMFAD⟩2.0.CO;2.

Jameson, A. R., and A. B. Kostinski, 2001: What is a raindrop size distribution? *Bulletin of the American Meteorological Society*, **82 (6)**, 1169–1177, doi:10.1175/1520-0477(2001)082⟨1169:WIARSD⟩2.3.CO;2.

Jevons, W. S., 1862: On the deficiency of rain in an elevated rain-gauge, as caused by wind. *Journal of the Franklin Institute*, **73 (5)**, 332, doi:10.1080/14786446108643180.

Joss, J., and A. Waldvogel, 1967: A spectrograph for raindrops with automatic interpretation. *Pure and Applied Geophysics*, **68 (3)**, 240–246.

Keegan, M. H., D. H. Nash, and M. M. Stack, 2013: On erosion issues associated with the leading edge of wind turbine blades. *Journal of Physics D: Applied Physics*, **46 (38)**, doi:10.1088/0022-3727/46/38/383001.

Kinnell, P. I., 1981: Rainfall intensity-kinetic energy relationships for soil loss prediction. *Soil Science Society of America Journal*, **45 (1)**, 153–155, doi:10.2136/sssaj1981.03615995004500010033x.

Kochendorfer, J., and Coauthors, 2017: Analysis of single-Alter-shielded and unshielded measurements of mixed and solid precipitation from WMO-SPICE. *Hydrology and Earth System Sciences*, **21 (7)**, 3525–3542, doi:10.5194/hess-21-3525-2017.

Koschmieder, H., 1934: Methods and Results of Definite Rain Measurements. *Monthly Weather Review*, **62 (1)**, 5–7, doi:10.1175/1520-0493(1934)62⟨5:marodr⟩2.0.co;2.

Langleben, M. P., 1954: The terminal velocity of snowflakes. *Quarterly Journal of the Royal Meteorological Society*, **80 (344)**, 174–181, doi:10.1002/qj.49708034404.

Lanza, L. G., and L. Stagi, 2009: High resolution performance of catching type rain gauges from the laboratory phase of the WMO Field Intercomparison of Rain Intensity Gauges. *Atmospheric Research*, **94 (4)**, 555–563, doi:10.1016/j.atmosres.2009.04.012, URL http://dx.doi.org/10.1016/j.atmosres.2009.04.012.

Lanza, L. G., and E. Vuerich, 2009: The WMO Field Intercomparison of Rain Intensity Gauges. *Atmospheric Research*, **94 (4)**, 534–543, doi:10.1016/j.atmosres.2009.06.012, URL http://dx.doi.org/10.1016/j.atmosres.2009.06.012.

Lanzinger, E., 2006: Rainfall amount and intensity measured by the Thies laser precipitation monitor. *TECO-2006, Geneva, Switzerland*, 1–9, doi:WMO-IOM-94,WMO-TDNo.135, URL http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-94-TECO2006/3(3)_Lanzinger_Germany.pdf.

Laws, J. O., and D. A. Parsons, 1943: The relation of raindrop-size to intensity. *Eos, Transactions American Geophysical Union*, **24 (2)**, 452–460.

Locatelli, J. D., and P. V. Hobbs, 1974: Fall speeds and masses of solid precipitation particles. *Journal of Geophysical Research*, **79 (15)**, 2185–2197, doi:10.1029/JC079i015p02185.

Löffler-Mang, M., and J. Joss, 2000: An optical disdrometer for measuring size and velocity of hydrometeors. *Journal of Atmospheric and Oceanic Technology*, **17 (2)**, 130–139, doi:10.1175/1520-0426(2000)017⟨0130:AODFMS⟩2.0.CO;2.

Lumb, F. E., 1963: Downward penetration of snow in relation to the intensity of precipitation. *Meteorology Magazine*, **92 (1086)**, 1–14.

Marshall, J. S., and W. M. Palmer, 1948: The distribution of raindrops with size. *Journal of Meteorology*, **5**, 165–166, doi:10.1002/qj.49707632704.

Norbury, J. R., and W. J. White, 1971: A rapid-response rain gauge. *Journal of Physics E: Scientific Instruments*, **4 (8)**, 601.

Oppenheim, I., and D. Shinar, 2012: A context-sensitive model of driving behaviour and its implications for in-vehicle safety systems. *Cognition, Technology & Work*, **14 (3)**, 261–281, doi:10.1007/s10111-011-0178-3, URL https://doi.org/10.1007/s10111-011-0178-3.

Park, H. S., A. V. Ryzhkov, D. S. Zrnić, and K.-E. Kim, 2009: The Hydrometeor Classification Algorithm for the Polarimetric WSR-88D: Description and Application to an MCS. *Weather and Forecasting*, **24 (3)**, 730–748, doi:10.1175/2008WAF2222205.1.

Pickering, B. S., R. R. Neely, and D. Harrison, 2019: The Disdrometer Verification Network (DiVeN): A UK network of laser precipitation instruments. *Atmospheric Measurement Techniques*, **12 (11)**, 5845–5861, doi:10.5194/amt-12-5845-2019.

Ralph, F. M., and Coauthors, 2005: Improving short-term (0-48 h) cool-season quantitative precipitation forecasting: Recommendations from a USWRP Workshop. *Bulletin of the American Meteorological Society*, **86 (11)**, 1619–1632, doi:10.1175/BAMS-86-11-1619.

Reeves, H. D., 2016: The uncertainty of precipitation-type observations and its effect on the validation of forecast precipitation type. *Weather and Forecasting*, **31 (6)**, 1961–1971, doi:10.1175/WAF-D-16-0068.1.

Rosewell, C. J., 1986: Rainfall kinetic energy in eastern Australia. 1695–1701 pp., doi:10.1175/1520-0450(1986)025⟨1695:RKEIEA⟩2.0.CO;2.

Rubel, F., and K. Brugger, 2009: 3-hourly quantitative precipitation estimation over Central and Northern Europe from rain gauge and radar data. *Atmospheric Research*, **94 (4)**, 544–554, doi:10.1016/j.atmosres.2009.05.005, URL http://dx.doi.org/10.1016/j.atmosres.2009.05.005.

Saltikoff, E., and L. Nevvonen, 2011: First experiences of the operational use of a dual-polarisation weather radar in Finland. *Meteorologische Zeitschrift*, **20 (3)**, 323–333.

Sevruk, B., 1996: Adjustment of tipping-bucket precipitation gauge measurements. *Atmospheric Research*, **42 (1-4)**, 237–246, doi:10.1016/0169-8095(95)00066-6.

Slot, H. M., E. R. Gelinck, C. Rentrop, and E. Van der Heide, 2015: Leading edge erosion of coated wind turbine blades: Review of coating life models. *Renewable Energy*, **80**, 837–848, doi:10.1016/j.renene.2015.02.036, URL http://dx.doi.org/10.1016/j.renene.2015.02.036.

Smith, P. L., 2016: Sampling Issues in Estimating Radar Variables from Disdrometer

Data. *Journal of Atmospheric and Oceanic Technology J. Atmos. Oceanic Technol*, **1**, 2305–2313, doi:10.1175/JTECH-D-16-0040.1.

Tang, W., and C. I. Davidson, 2004: Erosion of limestone building surfaces caused by wind-driven rain: 2. Numerical modeling. *Atmospheric Environment*, **38 (33)**, 5601–5609, doi:10.1016/j.atmosenv.2004.06.014.

Tapiador, F. J., and Coauthors, 2016: On the Optimal Measuring Area for Pointwise Rainfall Estimation: A Dedicated Experiment with Fourteen Laser Disdrometers. *Journal of Hydrometeorology*, 753–760, doi:10.1175/JHM-D-16-0127.1, URL http://journals.ametsoc.org/doi/10.1175/JHM-D-16-0127.1.

The Met Office, 2016: Southern england: climate. United Kingdom Meteorological Office, URL https://www.metoffice.gov.uk/climate/uk/regional-climates/so. URL https://www.metoffice.gov.uk/climate/uk/regional-climates/so.

Thurai, M., and V. N. Bringi, 2018: Application of the generalized gamma model to represent the full rain drop size distribution spectra. *Journal of Applied Meteorology and Climatology*, **57 (5)**, 1197–1210, doi:10.1175/JAMC-D-17-0235.1.

Tokay, A., W. A. Petersen, P. Gatlin, and M. Wingo, 2013: Comparison of raindrop size distribution measurements by collocated disdrometers. *Journal of Atmospheric and Oceanic Technology*, **30 (8)**, 1672–1690, doi:10.1175/JTECH-D-12-00163.1.

Ulbrich, C. W., 1983: Natural variations in the analytical form of the raindrop size distribution. **22 (10)**, 1764–1775, doi:10.1175/1520-0450(1983)022⟨1764:NVITAF⟩2.0.CO;2.

Ulbrich, C. W., and D. Atlas, 2007: Microphysics of raindrop size spectra: Tropical continental and maritime storms. *Journal of Applied Meteorology and Climatology*, **46 (11)**, 1777–1791, doi:10.1175/2007JAMC1649.1.

Vaisala Oyj, 2001: Present Weather Detector PWD21 User's Guide. Tech. rep., Vaisala Oyj, Vanha Nurmijärventie 21, 01670 Vantaa, Finland, 115 pp.

Ventouras, S., S. A. Callaghan, and C. L. Wrench, 2006: Long-term statistics of tropospheric attenuation from the Ka/U band ITALSAT satellite experiment in the United Kingdom. *Radio Science*, **41 (2)**, 1–19, doi:10.1029/2005RS003252.

Villermaux, E., and B. Bossa, 2009: Single-drop fragmentation determines size distribution of raindrops. *Nature Physics*, **5 (9)**, 697–702, doi:10.1038/nphys1340.

Wexler, R., 1955: *The melting layer*, Vol. 3. Harvard University.

Wiesner, J., 1895: Beiträge zur Kenntniss des tropischen Regens. *Sitzber. k. Akad. Wiss. Wein*, **Math.-natu (104)**, 1397–1434.

WMO, 1988: Manual on Codes. *WMO Publ. 306*, **1**, 203 pp.

WMO, 2017: WMO Guide to Meteorological Instruments and Methods of Observation (the CIMO Guide). Chapter 6. Masurement of Precipitation. Tech. rep., World Meteorological Organization.

Chapter 4.

Improving Observations of Precipitation Type at the Surface: A 5-year Verification of a Radar-derived Product from the United Kingdom Met Office. Part I: Rain, Mixed-phase and Snow

In revision to *AMS Journal of Hydrometeorology.*

Contributions from: Ryan R. Neely III, Steven Best, Darren Lyth, David Dufton and Maryna Lukach

Abstract

This study aims to evaluate the skill of a radar-based surface precipitation type (SPT) product with observations on the ground. Social and economic impacts can occur from SPT because it is not well forecast or observed. Observations from the United Kingdom Meteorological Office's weather radar network are combined with post-processed numerical weather prediction (NWP) freezing level heights in a Boolean logic algorithm to create a 1 km resolution Cartesian-gridded map of SPT. Here 5 years of discrete non-probabilistic outputs of rain, mixed phase, snow

(Part I) and hail (Part II) are compared against surface observations made by trained observers, automatic weather stations, disdrometers and public reports. The novel skill verification method developed as part of this study employs several tolerances of space and time from the SPT product, indicating the precision of the product for a desired accuracy. In general the results indicate that the tolerance verification method works well and produces reasonable statistical score ranges grounded in physical constraints. Using this method, we find that the mixed precipitation class is the least well diagnosed which is due to a negative bias in the input temperature height field, resulting in rain events frequently being classified as mixed. Snow is captured well by the product which is entirely reliant upon a post-processed NWP temperature field, although a single period of anomalously cold temperatures positively skewed snow scores with low-skill events. Furthermore, we conclude that more verification consistency is needed amongst studies to help identify successful approaches and thus improve SPT forecasts.

1. Introduction

The type of hydrometeors reaching the surface, known as the surface precipitation type (SPT), can severely impact human activities. In regions where solid precipitation types are common and expected occurrences, long-term adaptations are cost-effective, but where solid precipitation types are infrequent and uncommon (mid-latitudinal and certain mountainous regions) these adaptations are not cost-effective and (as in the case of the UK) events can significantly disrupt daily life (Kay 2016; Curtis et al. 2017). In the winter of 2009/10, the cost to the UK National Health Service from falls on snow and surface ice was £42 million (Beynon et al. 2011). Mitigative actions such as clearing roads, covering exposed crops, redirecting aircraft are cost-associated and require sufficient lead time and confidence (Cornford and Thornes 1996; Rasmussen et al. 2001; Handa et al. 2006; Clarke et al. 2009).

Real-time observations are often used by forecasters directly, or in nowcasting systems to issue precipitation type guidance, valid for timescales of 0-6 hours (Rasmussen et al. 2001; Schmid and Mathis 2004; Haiden et al. 2011). SPT is accurately reported by trained observers but their observations are infrequent, whereas automated ground instruments record continuously but with less accuracy (Bloemink and Lanzinger 2005; Landolt et al. 2019). The United Kingdom's Met Office operates a network of both station types across the UK but these do not provide complete spatial coverage at a high enough temporal resolution sufficient for animated, gridded map products which are essential for SPT nowcasting and public understanding. An ideal measurement system for SPT nowcasting is weather radar because it possesses a high spatiotemporal resolution. Additionally, the UK weather radar network has (at most) a 10-minute turnaround from measurement to dissemination (Harrison et al. 2000) so it is useful for real-time decision-makers.

This study aims to assess the skill of a UK radar-derived SPT product over a 5-year period. Since the product is deterministic and precipitation type is discrete non-probabilistic data, there are a limited number of statistical techniques suitable for performing a verification. Furthermore, snow and mixed-phase precipitation are an order of magnitude less frequent than rain (Kay 2016; Brown 2019). This discrepancy in the abundance of the classes can deceptively skew some statistical scores (Wilks 2011), further reducing the number of applicable verification techniques.

Added difficulty is introduced with the comparison of a radar-derived spatial product with point surface instruments, since the representative volumes differ. Weather radars measure distribution-weighted three-dimensional volumes of the atmosphere. The verification "truth" on the ground (often many hundreds of metres below the peak-weighting of the radar voxel) is a pinpoint measurement, typically a fraction of a cubic metre for automated instruments. Human reporters are capable of broader visual verification of the precipitation type but their observation volume is still an order of magnitude less than weather radars.

In this study, a new approach is applied to determine the skill-range of radar-based surface precipitation type products against several surface observation datasets, by varying the temporal and spatial tolerance of the product. Part I examines the ability for the Met Office SPT product to diagnose rain, mixed-phase and snow precipitation types, while Part II examines the skill of hail diagnosis, which uses a separate criterion in the Met Office product and requires separate verification data. Weaknesses and opportunities for improvement of the radar-based SPT product are presented in both parts. The verification techniques developed here are further useful for verification of NWP forecasts of precipitation type (or any discrete non-probabilistic variable) and thus facilitate more accurate diagnoses of precipitation type in atmospheric science.

The boundary between rain, mixed-phase and snow (R-M-S) is important because the presence of mixed-phase precipitation typically indicates that the hydrometeors are melting before they reach the ground and will therefore not accumulate. This is important for several industries - if wet precipitation meets a cold surface (or if it occurs with diurnal cooling), then ice is the primary risk. If the surface is warm (or if it is associated with diurnal heating) then the runoff water will drain away into rivers and lakes, potentially contributing to flood events.

The R-M-S boundaries in the UK (and similar geographies) are difficult to diagnose and forecast. Cases are often borderline since surface temperatures are non-extreme and fluctuate diurnally between −5 °C and +10 °C in winter (Parker et al. 1992; Brabson and Palutikof 2002), and many factors can influence the change of precipitation phase. The influence of the northern hemisphere mid-latitude jet stream and the enclosure of the North Atlantic warmed by the gulf stream create fluctuating synoptic patterns and coastal micrometeorology. Small changes in the vertical temperature structure of the atmosphere can also shift the R-M-S boundary by hundreds of kilometers horizontally.

1.1. Met Office SPT Product

To overcome the disparity between the radar-observed voxel and the surface precipitation type diagnosis, the Met Office created an SPT product which uses NWP output as input to a parameterised translational process below the lowest-usable radar beam. Since late 2013 the SPT product has been operational with the same spatiotemporal resolution as the Met Office precipitation rate product (1 km^2, 5-minute frequency). Figure 4.1 shows an example of the product at a single point in time. The product has 4 classes: hail, snow, mixed-phase and rain. Note that the term ‘mixed-phase’ refers to the mixture of snow and rain and does not include partially melted graupel or hail. These types are determined with a Boolean logic decision tree described in Table 4.1. The algorithm inputs are radar-derived surface precipitation rate (Harrison et al. 2000), 0 °C wet-bulb isotherm altitude (above local surface) derived from the Met Office Unified Model run in a Euro4 configuration, and radar reflectivity.

The criterion from Waldvogel et al. (1979) is employed for hail diagnosis:

$$\begin{gathered} Z \geq 45 \text{ dBZ} \\ \& \\ h \geq (FZL + 1.4 \text{ km}), \end{gathered} \tag{4.1}$$

where Z is radar reflectivity factor at a radar bin, h is the height of the radar bin, and FZL is the 0 °C wet-bulb isotherm altitude in km. Lumb’s critical rate is used for the mixed-phase diagnosis and is defined as:

$$R_c = 0.2909e^{\left(0.004\frac{FZL}{f(v)}\right)} \tag{4.2}$$

Table 4.1.: The Boolean logic algorithm steps used for the Met Office surface precipitation type product. Note that the term ‘mixed-phase’ refers to the mixture of snow and rain and does not include partially melted graupel or hail.

Precipitation Type	Criterion
Hail	if a radar reflectivity of ≥ 45 dBZ occurs ≥ 1.4 km above the 0 °C isotherm height (Waldvogel et al. 1979)
Snow	if the NWP model freezing level height (0 °C wet-bulb isotherm) is negative (i.e. below the surface)
Mixed-phase	if the surface rain rate is higher than Lumb’s critical rate (Lumb 1963)
Rain	if none of the previous criteria are satisfied

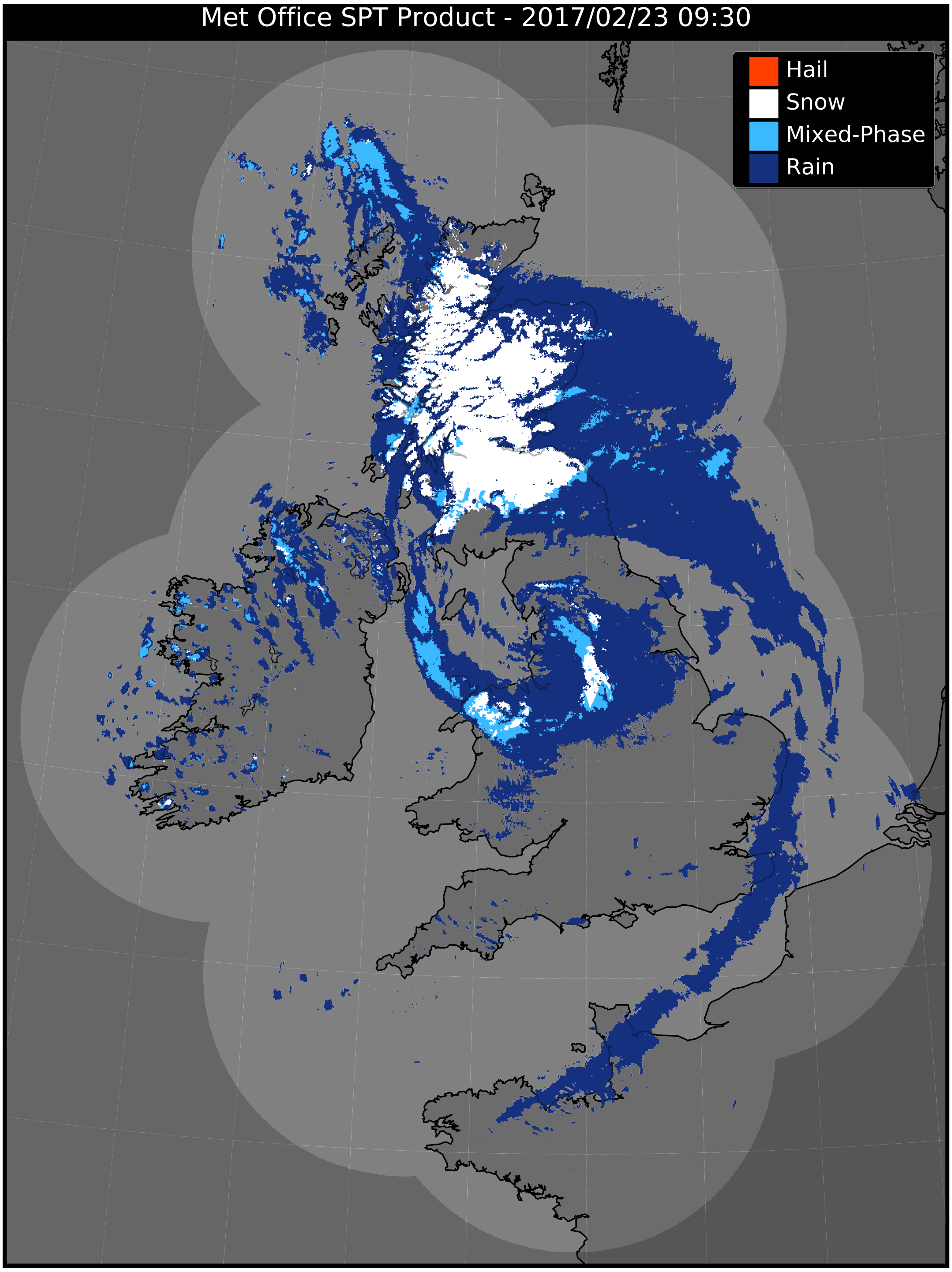


Figure 4.1.: An example of the Met Office SPT product, during named winter storm Doris at 0930 UTC 23rd February 2017. An animated video of the whole day is supplied in the supplementary material.

where R_c is the critical rate in mm h^{-1}, FLZ is the 0 °C wet-bulb isotherm height above the local surface in metres, and $f(v)$ is a function of wind speed but is set equal to 1 in the Met Office implementation and is therefore neglected. The notion is that for a given 0 °C wet-bulb isotherm height, precipitation will be observed at the ground as still containing a proportion of solid hydrometeors if the critical rate is met, due to evaporative cooling (Lumb 1963).

This process is applied initially to each pixel from all 18 radars (15 Met Office, 2 Met Éireann and 1 Channel Islands Meteorological Department). All data are then composited onto a Cartesian 1 km^2 grid using the modal-value of all contributing pixels since a single location in the UK is typically observed by many radar sites simultaneously.

1.2. Verification Data

Data which are used to evaluate the performance of the SPT product are described here. The known capabilities and limitations of the ground instruments are critical to aid the discussion of the results. Table 4.2 summarises each dataset and Figure 4.2 shows the locations of all surface stations as well as the locations of all radar sites which contribute to the SPT product.

1.2.1. Automatic SYNOP

The Met Office operates a network of surface weather stations called SYNOP stations which report observations once every hour. At the automatic stations, precipitation type is reported using the World Meteorological Organization (WMO) “Present Weather” (PW) code from table 4680 (WMO 1988, 2017). The PW code is determined using an arbiter which combines multiple measurements: a Vaisala FD12P present weather sensor, a precipitation detector, a visiometer, a ceilometer and an air temperature thermometer (Green 2010). Known limitations of the arbiter are insensitivities to weak precipitation rates, poor detection of ‘sleet’ (UK nomenclature for mixed precipitation), no quantitative uncertainty and difficulties calibrating or tracing errors since the arbiter “has many assumptions” (Lyth and Molyneux 2006; Lyth 2008). 172 automatic SYNOP station locations were available for inclusion during this study.

Table 4.2.: Summary of the 3 ground verification datasets used in this study. Includes the different measurement techniques, the format of the data when received, the frequency of data available, the number of locations available, and the availability over the duration of this study period.

	Automatic	Manual	DiVeN
Measurement Technique	A Vaisala FD12P present weather sensor, precipitation detector, visiometer, ceilometer and air temperature thermometer combined into an arbiter.	Trained meteorological observer.	A laser disdrometer measures particle diameter and fall velocity and uses empirical relationships to determine precipitation type.
Format	PW Code (WMO Table 4680), 83 codes reported.	PW Code (WMO Table 4677), 91 codes reported.	PW Code (WMO Table 4680), 21 codes reported.
Frequency	Hourly	Hourly	5-minute
Locations	172	38	14
Availability	2014-2018 (5 years)	2014-2018 (5 years)	2017-2018 (18–23 months, depending on the install date)

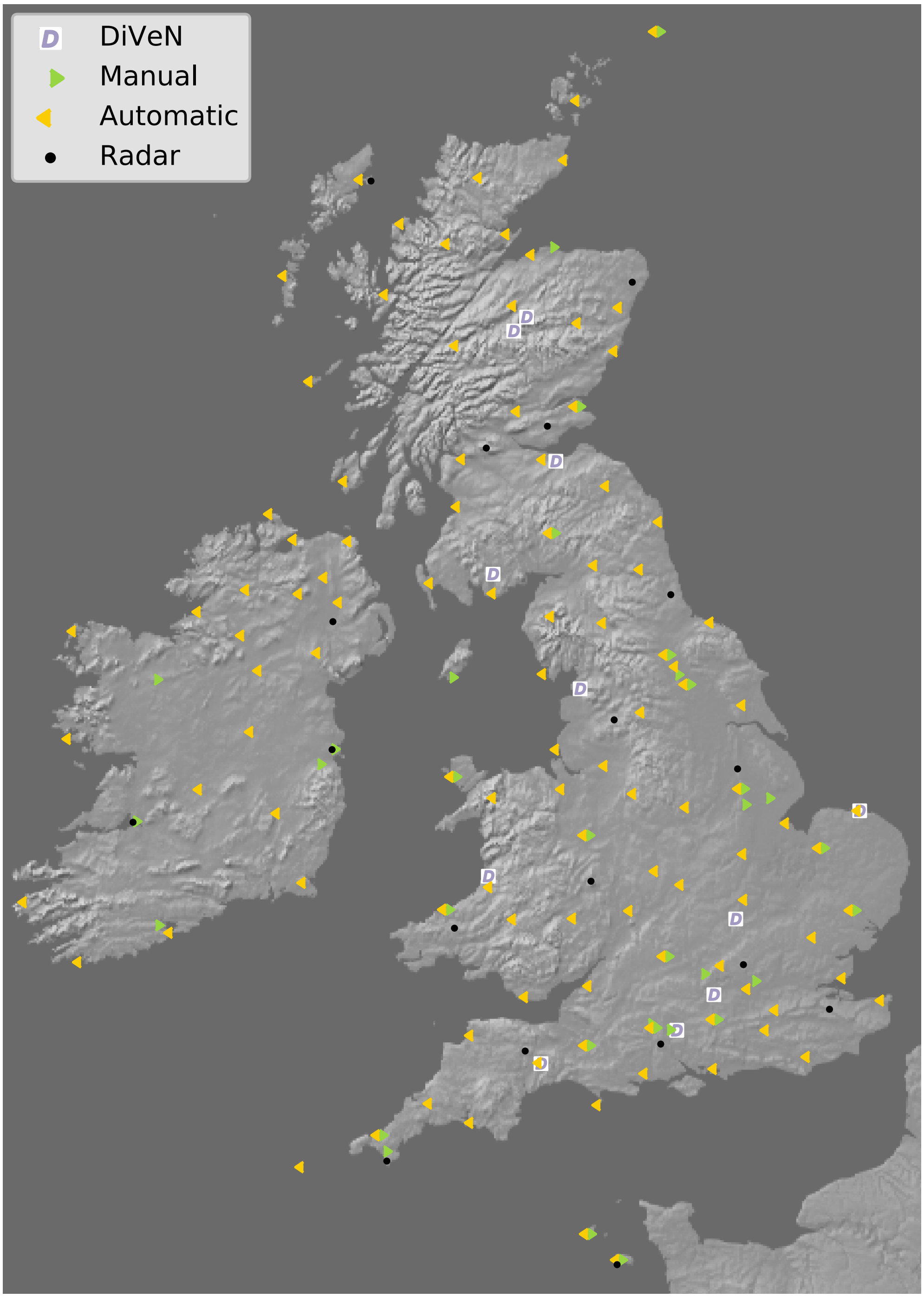


Figure 4.2.: A map of the UK showing all surface station sites (automatic, manual and DiVeN) used in the verification in this study, as well as the locations of all radar sites used in the UK Met Office SPT product. Some stations are a hybrid (denoted with adjacent yellow left-pointing and green right-pointing triangles), where the observations are mostly automatic but are sometimes overridden with manual observations if an observer is present and disagrees with the automated diagnosis.

1.2.2. Manual SYNOP

Met Office manual SYNOP stations are those where a qualified employee has physically observed meteorological conditions every hour. WMO Table 4677 is used to record PW observations (WMO 1988, 2017). Manual reports are considered to be the highest quality standard of PW observation and observers are well trained with handbooks to minimise inconsistencies between sites. The range of PW codes available cover more obscure weather conditions and many do not refer to precipitation at all. The main limitation of the manual stations is that there are few locations; 38 manual SYNOP station locations were available for inclusion during this study.

1.2.3. DiVeN Disdrometers

With the support of the Met Office and the National Centre for Atmospheric Science (NCAS), the Disdrometer Verification Network (DiVeN) was installed in the UK in early 2017 (Pickering et al. 2019). The Thies laser disdrometers (Adolf Thies GmbH & Co. KG 2011) measure the diameter and fall velocity of hydrometeors and use empirical relationships (such as those developed by Gunn and Kinzer (1949); Locatelli and Hobbs (1974)) to estimate WMO table 4680 PW codes (WMO 1988, 2017). Prior studies have shown that the Thies laser disdrometers have good ability to distinguish between solid and liquid precipitation types but less skill in the mixed-phase or during light precipitation (Bloemink and Lanzinger 2005; Lyth 2008; Pickering et al. 2019, 2020). Hail detection from the Thies laser disdrometer is possible but is less well studied, so the instruments will not be used here for verification of the SPT hail class. Data are openly available (NERC et al. 2019) from February 2017 (18–23 months depending on the site install date) at a 5-minute frequency and 14 locations exist.

2. Study Period Characteristics

In this study, the Met Office SPT product is veified over a 5 year period of 2014 to 2018 inclusive (60 months total). Before verifying the product an overview of the data characteristics throughout the study period is provided here.

2.1. Frequency Maps

SPT-product classes from the 5-year study period are summed in time to create total radar-diagnosed frequencies of precipitation, and then each precipitation type as a percentage of total precipitation observed. High-resolution zoomable PDF maps are provided in the supplementary material. Figure 4.3 shows the percentage of the 5-year period where a pixel prescribed precipitation of any kind. The spatial distribution of precipitation frequency in Figure 4.3 shows higher precipitation frequency in the north and western areas, and over higher terrain. The radar network covers the whole of the UK (except the Shetland Islands) but some artefacts are visible. Note that the western and southern edges of the product are constrained by the extent of the UKPP 0 °C wet-bulb isotherm field. The furthest extent of the radar network detects precipitation less frequently because the beam is less sensitive with range and may overshoot precipitation.

In a similar fashion, azimuths which experience long-term partial or total beam blockage (by terrain, buildings or trees) exhibit radial streaks of decreased percentages. The edge of some radar maximum-range boundaries are visible notably in northern Scotland and this is due to dual-polarisation upgrade downtime at individual sites (see supplementary material). The patches of decreased precipitation frequency are likely due to the removal of ground or sea clutter (reflective human or natural structures) which also removed some weaker precipitation events. Annual and monthly plots (see supplementary material) show that the Channel Islands (most southern radar) sea clutter has been almost entirely eradicated by the dual-polarisation upgrade - a well-documented ability of the technology (Hubbert et al. 2009; Dufton and Collier 2015).

For the precipitation classes, the total occurrences are normalised against occurrences of any precipitation type e.g. for each pixel, the total number of snow reports as a percentage of the total number of precipitation reports from Figure 4.3. Since rain is overwhelmingly common in the UK (greater than 90% in most areas), the rain frequency map is dominated by the signals shown in Figure 4.3 and is therefore not shown here (see supplementary material). Maps for mixed-phase and snow are shown in Figures 4.4 and 4.5.

Orography is clearly resolved in the SPT product, which can be attributed to the 0 °C wet-bulb local height for the mixed and snow classes. The mixed-phase class is also influenced by the enhancement of precipitation rate over orography applied by the Met Office (Harrison et al. 2000) due to Lumb's critical rate. The highest snow frequency is

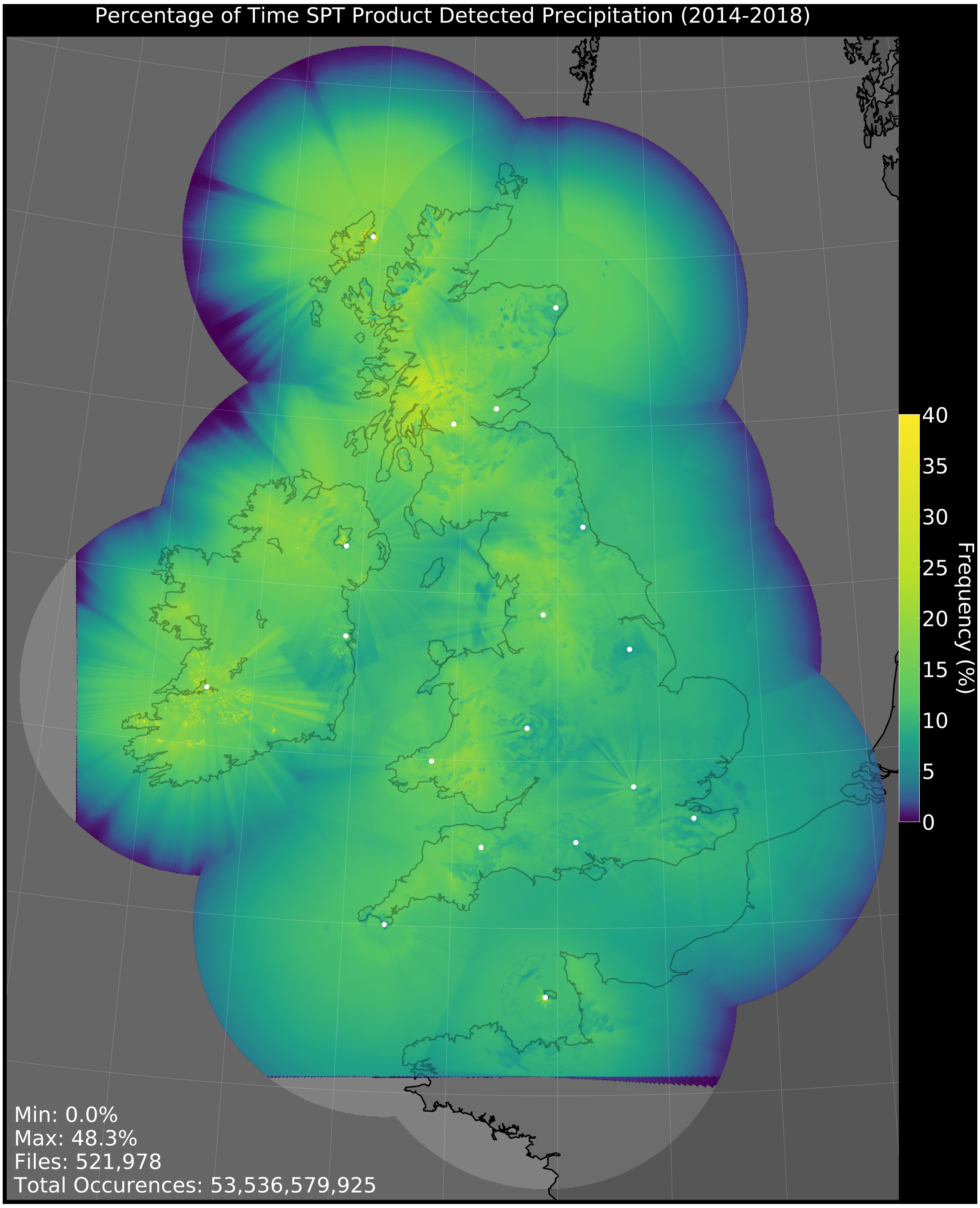


Figure 4.3.: Percentage of time that precipitation of any class is detected by the Met Office radar network from the start of 2014 to the end of 2018 (5 years). The Met Office, Met Éireann and the Channel Islands Meteorological Department radar locations are marked as white dots.

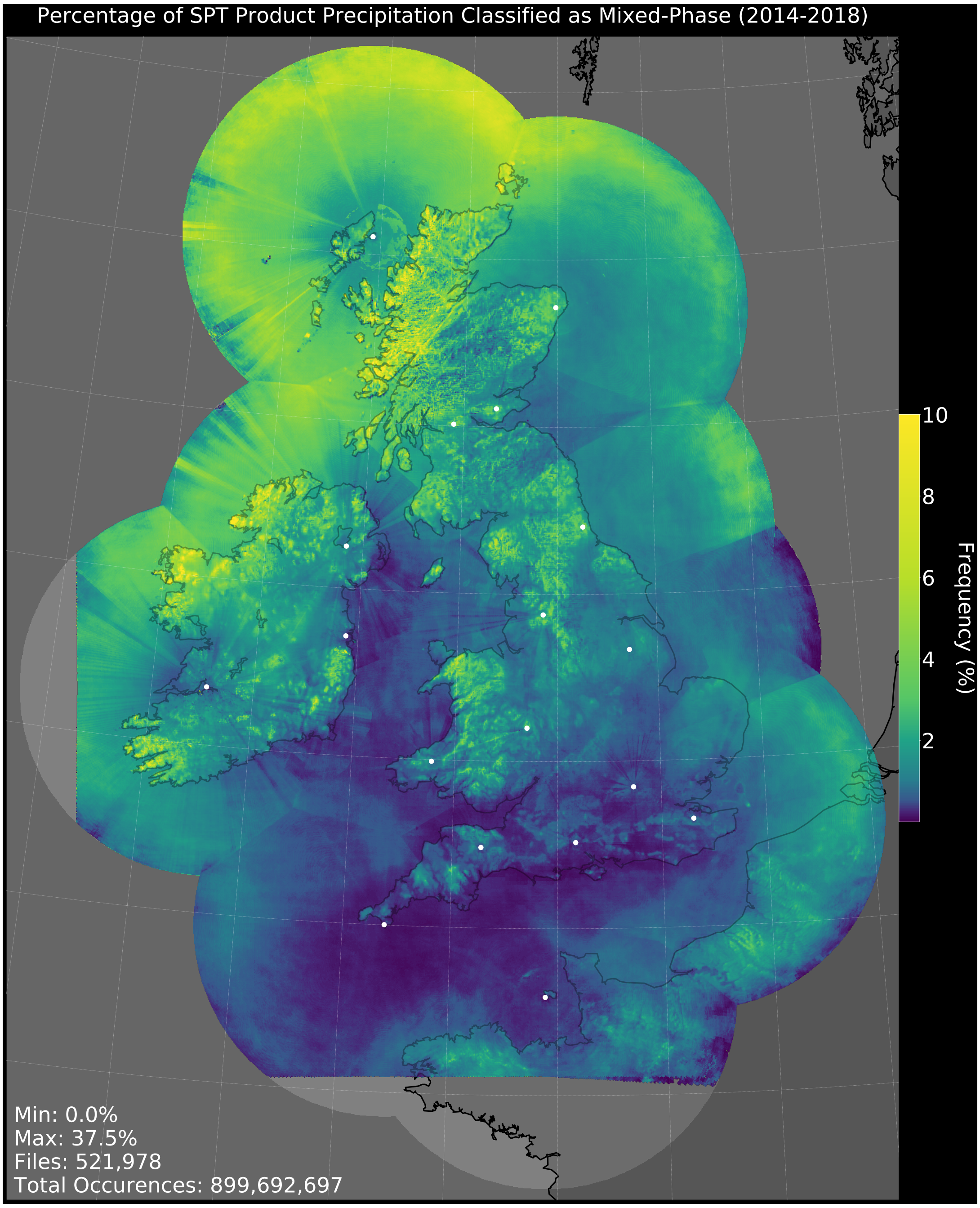


Figure 4.4.: Percentage of precipitation detected by the Met Office radar network which the SPT product diagnosed as the precipitation type mixed-phase, between 2014 and 2018 inclusive. The Met Office, Met Éireann and the Channel Islands Meteorological Department radar locations are marked as white dots. The scale is set from 0 to 10%.

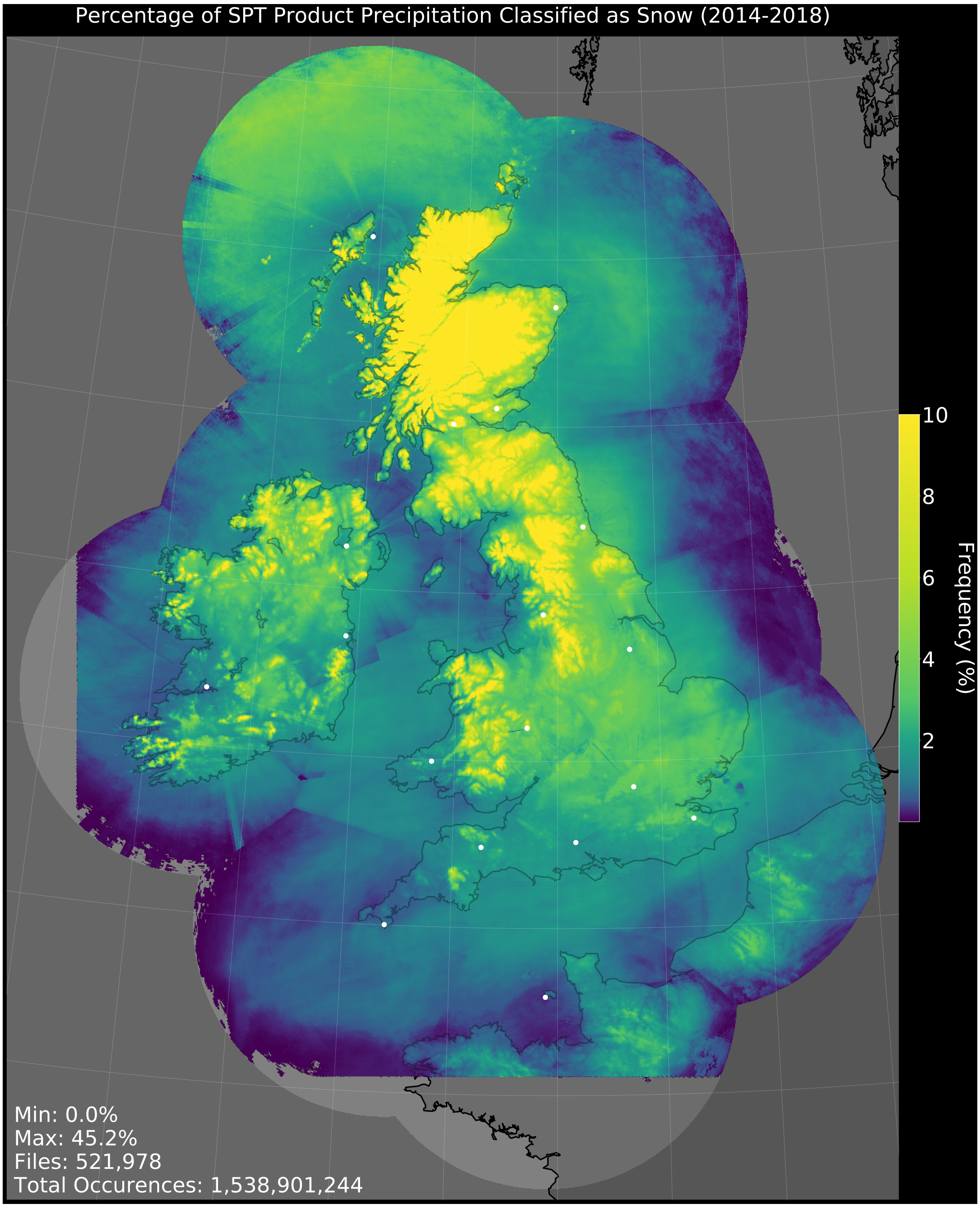


Figure 4.5.: Percentage of precipitation detected by the Met Office radar network diagnosed as snow by the SPT product, between 2014 and 2018 inclusive. The Met Office, Met Éireann and the Channel Islands Meteorological Department radar locations are marked as white dots. The scale is set from 0% to 10% to highlight features. The maximum percentage is 45.2%, which occurs over the Scottish Grampians.

over the Scottish mountains where 45.2% of the precipitation detected receives a snow classification. Between 2014-2018, every km^2 pixel of UK-land is diagnosed as experiencing snow at least once. Lowland areas of England typically experience ∼0.5-1.0% of precipitation as mixed-phase and ∼3-4% of precipitation as snow. The mixed-phase class occurs more frequently over the western-facing coasts of Scotland and the Republic of Ireland, which experience heavier precipitation more often due to exposure to westerly-dominated synoptic weather and thus meet Lumb's critical rate more frequently.

In Figures 4.4 and 4.5, offshore wind farms are visible east of London and the Thames Estuary. Wind turbines are reflective so the precipitation rate will be falsely higher and thus Lumb's critical rate will be met more often. Mixed-phase frequency also decreases in both plots where a reflectivity correction is made for known wind farms; for snow, this means the minimum reflectivity for precipitation diagnosis is met less often. These plots show that the correction is too strong and that the polygon is not large enough since a halo-effect is seen around these locations, even after the dual-polarisation upgrade. A feathered-edge polygon would give improved results.

The Ingham radar (Lincolnshire) has fewer mixed-phase precipitation events at maximum range from the radar, caused by lower reflectivity such that Lumb's critical rate is met less frequently. Borders between preferred radars during the compositing process are visible but mainly over the ocean (with the exception of East Anglia). Banding occurs in the mixed and snow plots particularly around the edge of the network; the insensitivity to weaker precipitation at long ranges (because the radar is less sensitive generally and the beam is at a high altitude) means that the percentage of events detected that are heavy (and are therefore more likely to meet Lumb's critical ratio) is higher.

In general, long-term frequency plots are useful for exposing artefacts, events and trends within the radar and SPT product data. The sensitivity of the SPT product to changes in reflectivity and radar scan geometry are well highlighted here. A limitation of using this method to find radar artefacts is that many years of observations are needed if seasonal changes are to be observed.

2.2. Verification Data Statistics

The SYNOP (automatic and manual) reports are hourly and cover the full 5-year study period. DiVeN began in February 2017 and therefore contributes 18–23 months of data (depending on the site install date), but every 5 minutes. The automatic stations contributed a total of 330,369 precipitating PW code reports, of which 321,111 (97.20%) were rain, 2,408 (0.73%) were mixed, and 6,850 (2.07%) were snow. Manual sites are less common and contributed 75,647 precipitation reports, consisting of 73,609 (97.31%) rain, 716 (0.95%) mixed-phase, and 1,322 (1.75%) snow. DiVeN disdrometer instruments contributed 148,441 precipitation reports, of which 135,083 (91.00%) were rain, 2,787 (1.88%) were mixed-phase and 10,571 (7.12%) were snow. DiVeN sites observe higher frequencies of mixed and snow cases because several of the sites are at high elevation (5 sites > 250 m a.m.s.l. out of 14 total). The Met Office SYNOP sites are more commonly at lower elevations on flat terrain (∼10% > 250 m a.m.s.l.).

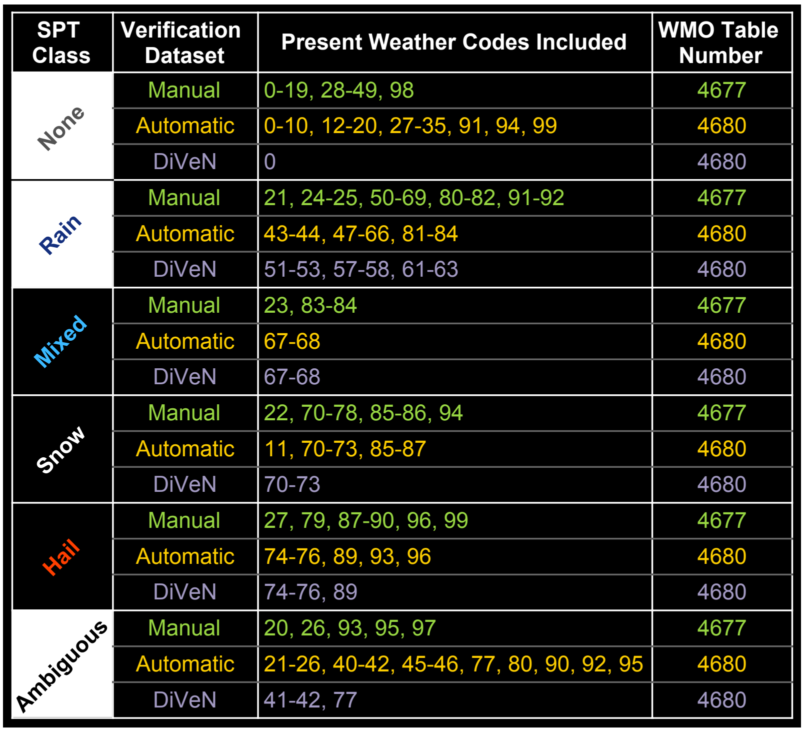
3. Methodology

The aim of this study is to evaluate the skill of the Met Office SPT product over a 5-year period. To achieve this, several ground-based datasets are used to increase the volume of data available and to have multiple perspectives since all ground-based data have their own artefacts and biases. The sections below outline the steps taken to evaluate the skill of the SPT product.

3.1. Data Handling and Quality Control

A limitation of the ground-based data is that all are coded using the PW system; many codes contain multiple precipitation types or are ambiguous (i.e. multiple conditions are described). To facilitate comparison to the SPT product, the WMO Table 4680 and 4677 codes are translated into the Met Office SPT product classes (none, rain, mixed-phase, snow, hail) or 'ambiguous' as shown in Figure 4.6. The number of ambiguous (containing more than one SPT product class) reports were as follows: manual 16,961 (1.8%), automatic 489,481 (7.7%), DiVeN 12,888 (0.5%).

In this study, an event constitutes one surface observation paired with a co-located SPT product diagnosis. There are 9,894,007 events in total available to this study from combined automatic, manual, DiVeN sites. The purpose of this study is to examine the SPT-classification skill of the product, not whether the radar correctly detects precipitation. Therefore, events that contain no precipitation (from either or both data sources), events that are erroneous (SPT data missing, codes outside of the PW coding scheme) or are ambiguous, are removed (562,590 events remain). The SPT product



SPT Class	Verification Dataset	Present Weather Codes Included	WMO Table Number
None	Manual	0-19, 28-49, 98	4677
	Automatic	0-10, 12-20, 27-35, 91, 94, 99	4680
	DiVeN	0	4680
Rain	Manual	21, 24-25, 50-69, 80-82, 91-92	4677
	Automatic	43-44, 47-66, 81-84	4680
	DiVeN	51-53, 57-58, 61-63	4680
Mixed	Manual	23, 83-84	4677
	Automatic	67-68	4680
	DiVeN	67-68	4680
Snow	Manual	22, 70-78, 85-86, 94	4677
	Automatic	11, 70-73, 85-87	4680
	DiVeN	70-73	4680
Hail	Manual	27, 79, 87-90, 96, 99	4677
	Automatic	74-76, 89, 93, 96	4680
	DiVeN	74-76, 89	4680
Ambiguous	Manual	20, 26, 93, 95, 97	4677
	Automatic	21-26, 40-42, 45-46, 77, 80, 90, 92, 95	4680
	DiVeN	41-42, 77	4680

Figure 4.6.: Conversion Look up Table (LuT) for converting ground observations from WMO Present Weather code into the SPT product classes for this study. Also shown are the ranges of PW codes supported by each instrument and the specific table used, since autonomous and human observations use different WMO tables. Many of the codes available in the WMO tables are ambiguous (contain multiple SPT product classes) and are shown in the last row. All supported PW codes from each surface dataset are assigned an "SPT class" in the table. Note that the term 'mixed-phase' refers to the mixture of snow and rain and does not include partially melted graupel or hail.

should also be functioning nominally in the wider vicinity; if the SPT product has any erroneous flags in the 5 × 5 km ± 15-minute SPT pixel region around the ground report location, then the event-pair is discarded (555,993 events remain). Additionally, events where either of the event-pair report hail are removed. After filtering, 554,457 events remain from which the analysis is performed.

Ground-based observations are paired with the next available SPT file because output files are labelled with the end time of a 5-minute period. Note that the Met Office operates a 10-minute radar scan strategy with 3 elevation descents containing both high and low elevation angles.

Table 4.3.: The structure of the 3 × 3 confusion matrix applied in this study.

		Surface: Rain	Surface: Mixed	Surface: Snow	
SPT	Rain	r	s	t	y_1
	Mixed	u	v	w	y_2
	Snow	x	y	z	y_3
		x_1	x_2	x_3	Total, n

Table 4.4.: The layout of the 2 × 2 contingency table used in this study.

		Surface: Yes	Surface: No	
SPT	Yes	Hit, a	False Alarm, b	y_1
	No	Miss, c	Correct Null, d	y_2
		x_1	x_2	Total, n

3.2. Confusion Matrices and Contingency Table Metrics

Discrete non-probabilistic datasets are typically veified by confusion matrices where events are allocated a position in the matrix based on the ground-truth dataset (the class-designated column) and the dataset under examination (the class designated row). Table 4.3 shows the confusion matrix that will be employed in this analysis. The top-left to bottom-right diagonal entries are therefore instances where the dataset under examination is in agreement with the truth and a ‘hit’ occurs. The remaining entries reveal where the scrutinised dataset (the SPT product) is misdiagnosing.

Furthermore, the confusion matrix ($n \times n$) is reformulated into dichotomous (yes/no) contingency tables (2×2, shown in Table 4.4) for each of the SPT product precipitation classes (Wilks 2011). Three metrics are then applied to each table: frequency bias (B), probability of detection (POD), false alarm ratio (FAR):

$$B = \frac{(a+b)}{(a+c)} \tag{4.3}$$

$$POD = \frac{a}{(a+c)} \tag{4.4}$$

$$FAR = \frac{b}{(a+b)} \tag{4.5}$$

where a = hit, b = false alarm and c = miss. Bias shows whether the class is being under or over-diagnosed by the SPT product which can range from 0 (under-diagnosis) to ∞ (over-diagnosis); 1 is the perfect score. POD is the chance of a correct diagnosis when the precipitation type does occur and thus ranges from 0 (the event is never detected) to 1 (the event is always detected). FAR is the chance of a false diagnosis when the event is diagnosed and ranges from 0 (no false alarms) to 1 (all diagnoses are false alarms).

3.3. Heidke Skill Score and Bootstrapping

An overall score is sought for the SPT product, before narrowing in to identify the strengths and weaknesses of the product on a per-precipitation-class basis. Generally, a skill score (SS) takes the form:

$$SS = \frac{V - V_{ref}}{V_{perf} - V_{ref}} \tag{4.6}$$

Where V is the verification metric, V_{ref} is the verification metric for a reference diagnosis, and V_{perf} is the verification metric for a perfect diagnosis. Several scores exist and each come with strengths and limitations. Since the SPT data is discrete non-probabilistic (rain, mixed-phase or snow) as opposed to dichotomous (yes or no), two appropriate higher-dimension generalised skill scores are considered: the Heidke Skill Score (HSS) and the Peirce Skill Score (PSS). The n-dimension Heidke Skill Score (HSS) is defined following the structure of Eq. 6 as:

$$HSS = \frac{\sum_{i=1}^{I} p\left(y_i, x_i\right) - \sum_{i=1}^{I} p\left(y_i\right) p\left(x_i\right)}{1 - \sum_{i=1}^{I} p\left(y_i\right) p\left(x_i\right)} \tag{4.7}$$

where $\sum_{i=1}^{I} p\left(y_i, x_i\right)$ is the proportion correct (the normalised sum of all diagonal confusion matrix terms),$\sum_{i=1}^{I} p\left(y_i\right) p\left(x_i\right)$ is the random proportion correct (the product of diagnosed and observed normalised probabilities summed over each class), 1 is the perfect score, I is the length of the confusion matrix, y_i is the ith row and x_i is the ith column (Doolittle 1888; Heidke 1926). The HSS indicates the fractional improvement in diagnosis over the probability of a correct diagnosis by chance, which would score zero. The highest score (V_{perf}) is 1, and the lowest possible score is $-\infty$; negative values therefore indicate that a random guess would have been more skillful. For a dichotomous 2 × 2 contingency table the HSS collapses to:

$$HSS = \frac{2(a \times d - b \times c)}{(a + c)(c + d) + (a + b)(b + d)} \tag{4.8}$$

where $d =$ correct nulls. Applying the HSS to both the higher-dimension classifier (all classes simultaneously) and the individual classes allow the contributions from each precipitation type to be quantified.

The PSS is a modification on the HSS where the denominator V_{ref} term is the unbiased random proportion $\sum_{j=1}^{J} \left[p\left(x_j\right)\right]^2$, defined by the climatology of the observation dataset. If the climatology of the verification region differs substantially, or if seasonal changes occur during a verification period, the score must be re-calculated for each subset of the events (Wilks 2011). This adds computational expense and obscures the analysis as the subsets of events have no rigorous boundaries for climatology or seasonality. Therefore this study uses the HSS as an overall SPT product metric, which is applied to each ground-based dataset (automatic, manual, DiVeN) separately.

To show the stability of the overall skill score, a bootstrapping technique is employed (Efron and Tibshirani 1994; Chernick 2011). A similar approach for SPT verification is taken by Wandishin et al. (2005); Elmore et al. (2015). Events are extracted at random with replacement (an event can be extracted multiple times) to form a new subset of data. Bootstrapping is repeated 100 times to create many new subsets of randomised events which give an indication of the sensitivity of the HSS to rare events.

The spread of HSS for the subset of data produced by bootstrapping is heavily dependent on the number of random samples taken in each bootstrap and must, therefore, be chosen with physical justification. The more data that is ingested, the less variability the HSS exhibits with a random subset. The full 5-year dataset will have a narrow spread when bootstrapped, whereas a single event could have any HSS value and therefore the maximum possible spread. This study aims to show the realistic range of HSS values possible with a single month and a single year of the SPT product. Two bootstrap sample sizes are chosen to represent the number of events typically reported (after the quality control procedures described in Section 3.1) in one month and in one year, from each ground observation dataset (automatic, manual and DiVeN).

3.4. Tolerance

Due to the disparity of the lowest usable radar beam height and the surface, precipitation observed by radar is often not vertically co-located with the surface. Sandford (2015) showed that the uncertainty in radar drift estimates can vary from 1 km below the melting layer to 10 km at the extreme distance of the maximum range of a radar. Hail has a terminal fall velocity of > 10 m s^{-1} (Matson and Huggins 1980;

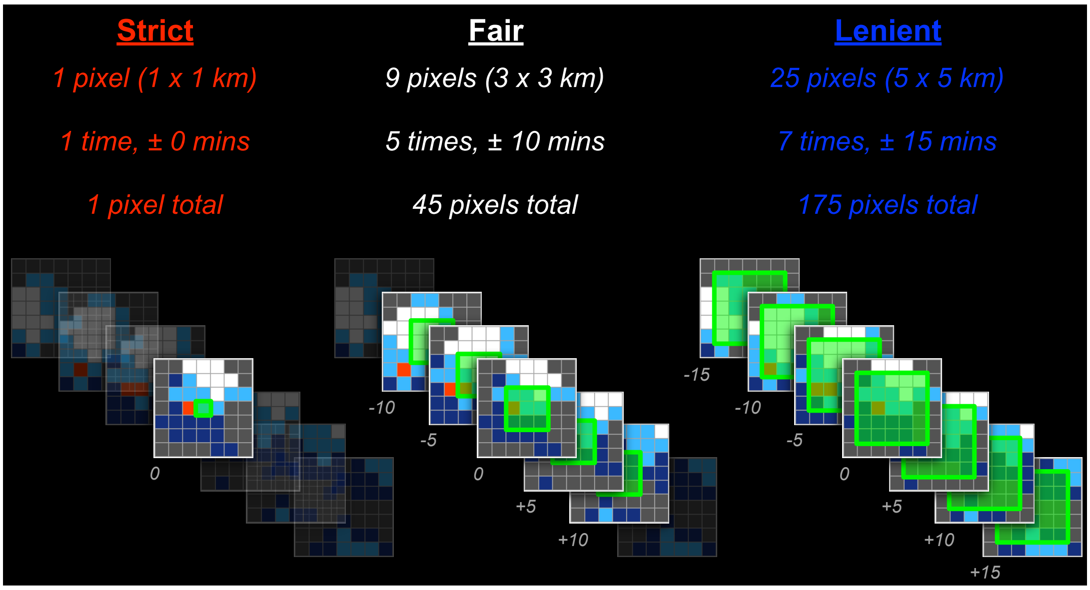


Figure 4.7.: An example of a timeseries of the SPT product stacked, representing time (5-minute frequency). The green-outlined area is the sample used for verification in three tolerances. The strict tolerance uses only the pixel co-located with the ground report. The fair tolerance uses a 3 × 3 km region around the ground report and ± 10 minute product outputs for a total of 45 pixels. The lenient tolerance uses a 5 × 5 km region around the ground report and ± 15 minute product outputs for a total of 175 pixels. If any of the green-shaded pixels are in agreement with the ground observation, then the SPT product is correct and a 'hit' is recorded.

Bohm 1989) while snow aggregates fall with a velocity of less than 2 m s^{-1} (Langleben 1954; Zikmunda 1972; Locatelli and Hobbs 1974), so the descent time varies between precipitation types. Furthermore, the horizontal wind advects precipitation as it falls and, therefore, the amount of horizontal displacement during descent will differ between precipitation types.

There are several factors determining the trajectory of hydrometeors as they fall to the Earth's surface, which makes verification difficult. Here a general solution is applied which increases the spatial and temporal tolerance for the SPT product to inform how the product skill is impacted. This informs a user of what spatiotemporal specificity corresponds with a desired accuracy. Three tolerances of the SPT product are used; strict: only the 1 × 1 km area and 5-minute period co-located with the surface report; fair: a 3 × 3 km area and ± 10 minutes around the surface observation will be considered; lenient: a 5 × 5 km area and ± 15 minutes around the surface observation will be considered. Figure 4.7 shows the three tolerances diagrammatically.

If any of the SPT product pixels in the fair or lenient tolerances agrees with the surface, then it is considered a 'hit'. The lenient tolerance is approximately the maximum reasonable displacement (∼2.5 km radius) and fall time (15 minutes) a hydrometeor could experience from the lowest usable beam height given the Met Office radar network coverage. To apply this verification technique to other products, the choice in tolerance may differ. There must exist a physical meaning to the minimum (strict) and maximum (lenient) possible extent of the gridded product under examination, which is dependent upon the specific variable being examined and also the measurement technique.

4. Results

4.1. Heidke Skill Score and Bootstrapping

First, the higher-dimension generalised HSS is examined to give an overall value to the SPT product, before examining each precipitation class. Note that only the SPT product pixel which directly encapsulates the location and time of the ground-based observation is used here (i.e. strict tolerance). While the hit and correct null quadrants are simple, the higher-dimension thresholds for false alarm or miss criterion from multiple SPT pixels would be subjective.

Figure 4.8 shows the higher-dimension HSS for all classes of the SPT product. Overall the SPT product has absolute HSS values (using the full dataset without bootstrapping, indicated by black dots on Figure 4.8) from 0.48 for automatic, 0.60 for manual, and 0.73 for DiVeN. If all surface-based observations are combined, the HSS of the SPT product is 0.61.

The spread of HSSs represents the possible scores if a random month or random year of data were considered. HSS distributions are markedly different between yearly and monthly bootstrap representations, with a much narrower spread for the yearly than monthly. Between verification datasets there are also differences. The manual station verification has the largest spread with a standard deviation (2σ) of 0.147 monthly and 0.038 yearly. Automatic stations give the second largest spread but the lowest overall score, with a standard deviation (2σ) of 0.058 monthly and 0.018 yearly (approximately half compared to manual sites). The DiVeN dataset has the highest scores and a standard deviation (2σ) of 0.024 monthly and 0.008 yearly (approximately half compared to automatic sites). Ultimately the differences in HSS spread tell us more about the ground-based dataset than the SPT product, but taking into consideration all three

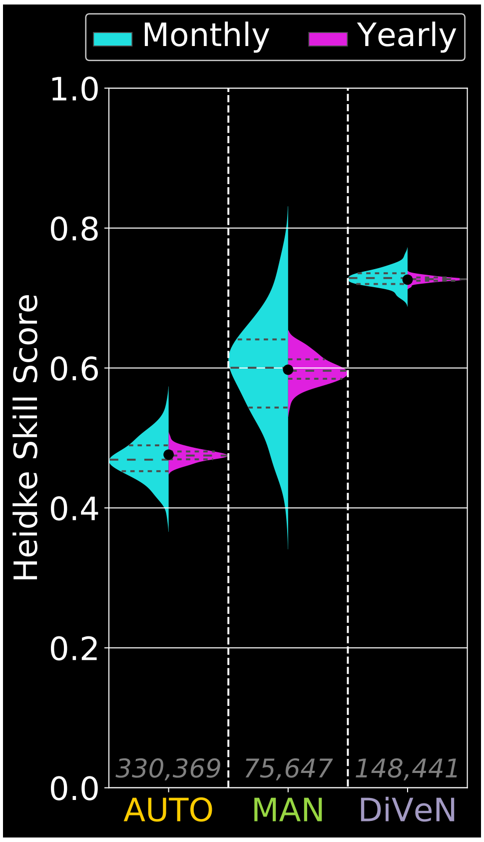


Figure 4.8.: Higher-dimension HSS (rain, mixed-phase and snow simultaneously) with probability distributions produced by a bootstrapping technique. Note that each distribution is scaled to fit half the width of the column for ease of viewing. Each ground dataset is shown (automatic, manual, DiVeN) and each has monthly and yearly representative distributions. The black dot indicates the HSS for the full dataset.

ground-based datasets gives a broader picture of the variability of the skill of the SPT product on different timescales, from approximately 0.4 to 0.8.

The HSS is recalculated with adjustments to some of the SPT product classes. Including the hail class of the SPT product in the calculation makes little difference because the HSS gives proportional weighting to rare events, and the ground-based datasets rarely report hail; automatic stations never report hail. If the mixed precipitation class is removed, the score (for all ground-based datasets) improves significantly from 0.61 to 0.77. This is unhelpful as the SPT product would in this scenario have an ‘unknown’ class for these events. If all mixed-phase diagnoses are re-classified as rain the HSS increases to 0.73 and if all mixed-phase diagnoses are re-classified as snow then the HSS decreases slightly to 0.59. This indicates that mixed diagnoses are more likely to be rain than either mixed-phase or snow.

4.2. Confusion Matrices

Confusion matrices are useful for showing where each class is being misdiagnosed. Figure 4.9 shows the results for the rain, mixed-phase and snow classes for each of the three ground observation sets available. Note again that the tolerance approach cannot be applied (see previous section), so the values shown are using only the encapsulating SPT product pixel area and time.

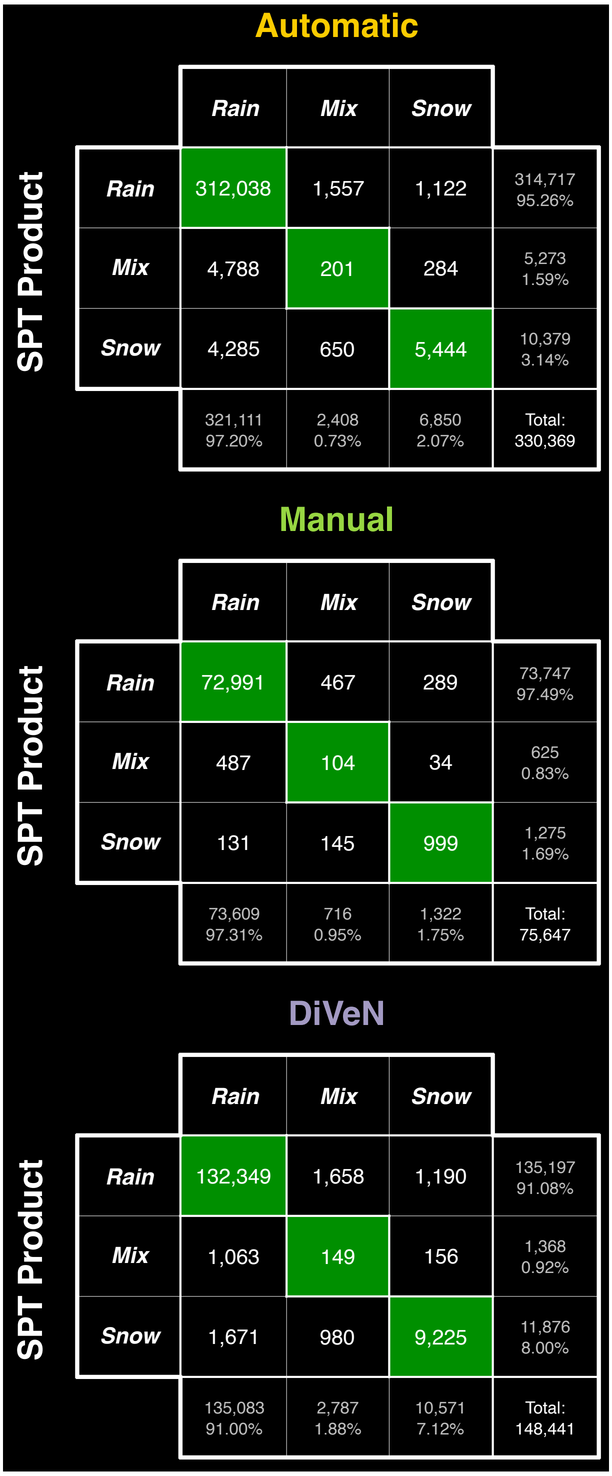
Firstly for the overall frequency of diagnoses, the rain type is underdiagnosed by the SPT product for automatic stations (-1.94%) but is close to the observed occurrences by manual (+0.18%) and DiVeN (+0.08%) sites. For mixed precipitation, the SPT product diagnoses this class twice as often compared with automatic sites, around the same compared with manual sites, and half as often compared with DiVeN sites. Finally, snow is diagnosed 50% more by the SPT product compared with automatic stations, around the same for manual stations, and 12% more for DiVeN sites.

Next, the rows of the confusion matrices are examined so that for a given SPT product diagnosis, the true observed precipitation type can be discussed. For example, given that the rain class is diagnosed, it is correct most often, but there are some miss events where the ground station observed mixed-phase or snow and in all ground datasets the mixed-phase class is the missed truth more often. The mixed class is poorly diagnosed, and rain is the observed ground event 23.8, 4.7, and 7.1 times more often (automatic, manual and DiVeN). Finally, the snow diagnosis is correct 52.5%, 78.4%, and 77.7% of the time (automatic, manual, DiVeN). The miss events differ between ground datasets. For automatic, the majority of miss events are rain (41.3% of all snow diagnoses), with 6.3% miss events being mixed. For manual, miss events are more evenly split over rain (10.3%) and mixed (11.4%). For DiVeN, rain is the missed event for 14.1% of the snow diagnoses and mixed is the missed event for 8.3% of the snow diagnoses.

4.3. Contingency Table Metrics with Tolerance

Next, skill scores are examined for each precipitation class where a contingency table has been produced from the confusion matrices. Three realistic tolerances based on the maximum horizontal displacement during descent from the lowest-usable radar beam have been applied to the SPT product as described in Section 3.4. All of the results are composed into Figure 4.10.

The hierarchy of the next section is as follows: each verification metric is discussed indi-



Automatic

SPT Product	Rain	Mix	Snow	
Rain	312,038	1,557	1,122	314,717 95.26%
Mix	4,788	201	284	5,273 1.59%
Snow	4,285	650	5,444	10,379 3.14%
	321,111 97.20%	2,408 0.73%	6,850 2.07%	Total: 330,369

Manual

SPT Product	Rain	Mix	Snow	
Rain	72,991	467	289	73,747 97.49%
Mix	487	104	34	625 0.83%
Snow	131	145	999	1,275 1.69%
	73,609 97.31%	716 0.95%	1,322 1.75%	Total: 75,647

DiVeN

SPT Product	Rain	Mix	Snow	
Rain	132,349	1,658	1,190	135,197 91.08%
Mix	1,063	149	156	1,368 0.92%
Snow	1,671	980	9,225	11,876 8.00%
	135,083 91.00%	2,787 1.88%	10,571 7.12%	Total: 148,441

Figure 4.9.: Confusion matrices of SPT product against ground observations, for each ground observation type. a) Automatic SYNOP, b) Manual SYNOP c) Disdrometer Verification Network (DiVeN).

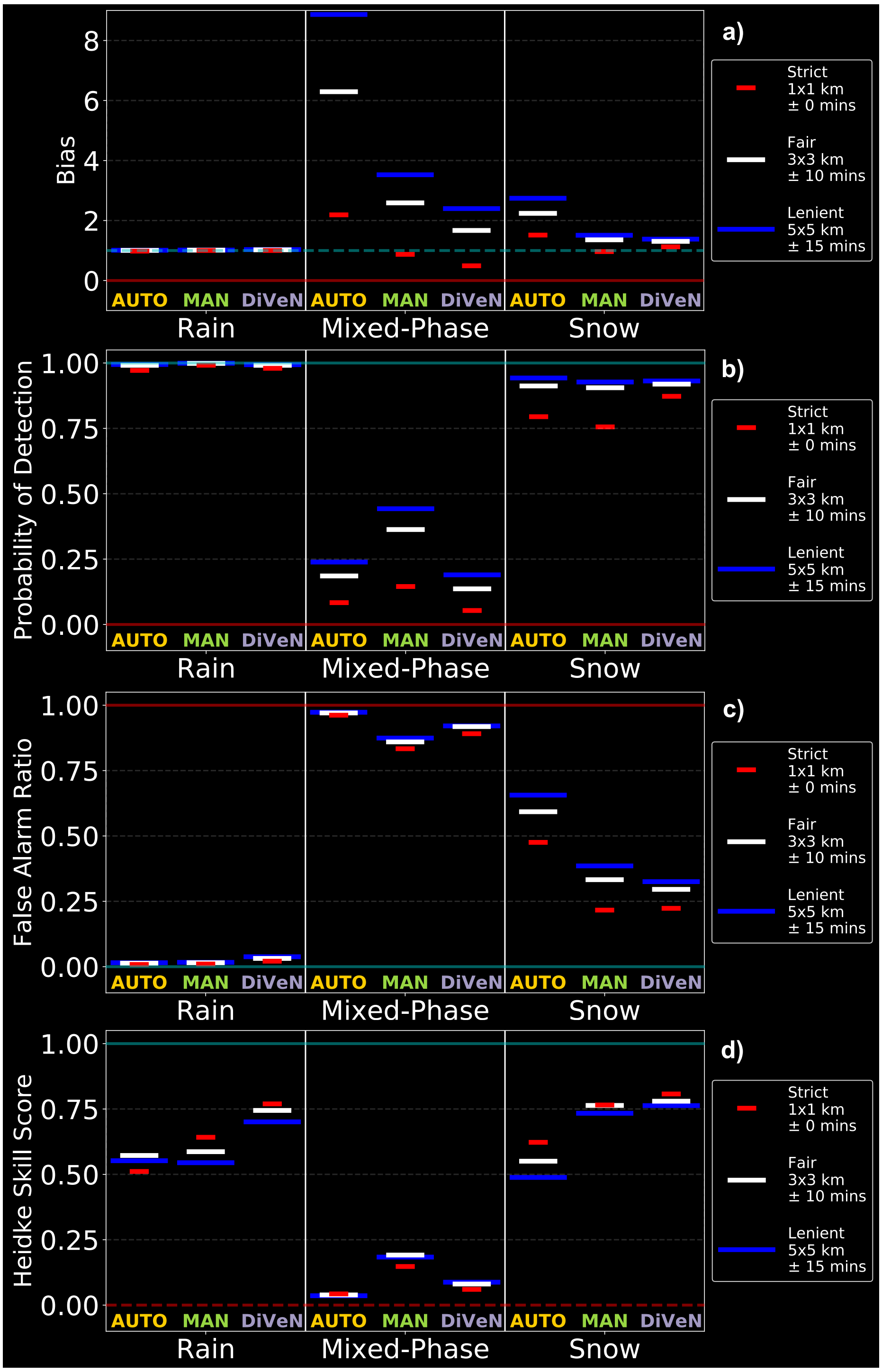


Figure 4.10.: Skill scores for each precipitation class and ground dataset. a) Bias, b) Probability of Detection (POD), c) False Alarm Ratio (FAR), d) Heidke Skill Score (HSS). Cyan horizontal lines indicate a perfect score, and red horizontal lines indicate a 'no skill' score. Solid cyan or red lines are fixed value limits, dashed are surpassable (bias and HSS).

vidually, going through the precipitation types (as some scores have interdependencies between the precipitation classes) and commenting on differences between the ground datasets and tolerances throughout.

4.3.1. Bias

The frequency bias indicates the scale to which precipitation classes are being under or over-diagnosed. Generally speaking, the mixed-phase and snow classes are overdiagnosed at the expense of rain. The high frequency of rain events makes the bias close to 1 but a slight underdiagnosis is occurring. Bias changes with increased tolerance are also small. The mixed-phase has the largest positive biases of any class, with the highest being 8.87 (automatic, lenient tolerance), whereas some are close to an ideal bias (0.87, manual, strict tolerance). The strict DiVeN result shows an underdiagnosis of mixed-phase (0.49) but increased tolerance shows an overdiagnosis (1.67 and 2.40). For snow, biases are overall smaller than the mixed-phase class but still show a positive tendency. With strict tolerance, biases against the manual and DiVeN data are 0.96 and 1.12, whereas bias against automatic is 1.52.

4.3.2. Probability of Detection

The POD tells us the probability of the SPT product being correct given that the precipitation class is occurring. Again the rain class is weighted by the frequency of occurrence (91-97% of precipitation) in the study period and has values close to a perfect score of 1. The lowest rain class POD score is in the automatic dataset (0.97, strict) due to underdiagnosis. For the mixed-phase class, POD is low, ranging from 0.08, 0.15 and 0.05 to 0.24, 0.44, and 0.19 (automatic, manual, DiVeN). The snow class has POD values similar to rain, with lenient/fair tolerances consistently 0.91-0.94 for all ground datasets. The strict tolerance varies: 0.79 (automatic), 0.76 (manual), 0.87 (DiVeN). Given that an SPT is occurring, increasing tolerance makes a correct diagnosis more likely.

4.3.3. False Alarm Ratio

The FAR indicates the probability of a false alarm when the SPT product diagnoses a precipitation type. The rain FAR is consistently low due to its high occurrence frequency. The DiVeN dataset gives a slightly higher rain FAR of 0.04 (lenient tolerance) which is indicative of the lower occurrence frequency from DiVeN (91% vs. 97% of precipitation for the other datasets). The mixed-phase class has high FAR (from 0.83 to 0.97) for all verification datasets consistent with a positive bias. The snow class has different FAR depending on the verification dataset: against manual and DiVeN, FAR values are around 0.22-0.39 but against automatic, FAR values are 0.48-0.66. Increasing the SPT product tolerance increases the chance of a false alarm.

4.3.4. Heidke Skill Score

The HSS indicates the fractional improvement of the SPT product diagnoses over random diagnoses, where a value of 0 is no skill and a value of 1 is a perfect diagnosis every time. The decimal value can be described as a percentage improvement over random chance. The HSS values for rain take into account the high frequency of occurrence and range between 0.51-0.64 for automatic and strict, but are higher (0.70-0.77) for DiVeN (lower rain occurrence frequency). The HSS values are not correlated with increasing or decreasing tolerance as is the case with the other verification metrics; this is explained in the discussion (Section 5.4). The weaknesses in the mixed-phase class are highlighted by the HSS, with low values across the ground datasets and tolerances. Automatic observations give the lowest scores (∼0.04), DiVeN the middle scores (0.06-0.09), and manual the highest scores (0.15-0.19), but all indicate poor skill. Snow has skill on-par or better than the rain class, with values ranging between 0.73-0.81 for manual and DiVeN datasets, while the automatic dataset gives scores slightly lower with a wider range from 0.49-0.62.

5. Discussion

5.1. Rain

Since rain is the dominant class with $> 90\%$ frequency, most skill scores for this precipitation type are skewed. The bias appears close to 1 but is underdiagnosed, POD is deceptively high and, similarly, FAR is deceptively low. The HSS takes the frequency into account and shows a 50-65% improvement over random chance diagnoses which is caused by the mixed-phase class diagnosing rain events. A fairer verification should not include low-skill rain cases; product users would not look for snow during heatwaves, for example. Events could be limited to the Met Office snow warnings, or periods of

0 °C wet-bulb isotherm below 500 m a.m.s.l., criteria that operational forecasters use (Sabrina Lee, MeteoGroup, personal communication, 2019). Alternatively, the occasions when the SPT product is opened could be recorded to build up cases targeted to user activity. The number of events would be reduced but SPT frequencies would be more equitable and the verification more applicable to certain product users, dependent on the criteria used.

5.2. Mixed-phase

POD for the mixed-phase class ranges from 0.08-0.24. Combined with a positive bias tendency up to 8.87, this indicates that the mixed-phase class has very little skill. This is reinforced by FAR values ranging from 0.83-0.97 and HSS scores between 0.04-0.19. Typically overdiagnosis increases the POD, but the mixed-phase class in the Met Office SPT product is the most overdiagnosed and still has the lowest POD of any class.

The HSS re-classification results (Section 4.1) and the confusion matrices in Figure 4.9 show that the mixed-phase class diagnoses are more often rain than mixed-phase or snow. Combining all verification datasets, 87.2% of mixed-phase class diagnoses are rain, 6.2% are correct and 6.5% are snow. The height of the mixed-phase to rain boundary being too low would be consistent with these results. Assuming Lumb's critical rate to be correct, this bias would be attributable to either a negative bias in the local 0 °C wet-bulb isotherm height, a positive bias in precipitation rate diagnosed by the radar, or both. Figures 4.4 and 4.5 showed the sensitivity of the SPT product to precipitation rate, as 'corrected' artefacts in precipitation rate still show a signal in the mixed-phase frequency map.

Lumb's critical rate uses the work of Langleben (1954), setting the boundary between rain and mixed-phase at 90% of the precipitation as liquid, based on the behaviour of the velocity of the particle. Lumb (1963) also assumed spherical aggregates and a saturated atmospheric column. Note that the data used in the derivation of Lumb's critical rate only covered 1-4 mm h^{-1} precipitation rates. These assumptions and limitations of Lumb's critical rate should be revisited and examined with modern measurement techniques to ensure that the SPT product is valid under all atmospheric conditions.

Finally, the effect of topographic representativity must be discussed. The method of calculating the local 0 °C wet-bulb isotherm height results in a topographic resolution of 1 km^2. For the majority of the United Kingdom this is an acceptable approach. Where deviations of topography altitude are large such as in mountainous regions, if the station providing verification data is situated in a valley or on a peak in the terrain, then the verification will have systematic errors, since the SPT product is calculating precipitation type for the average topographic altitude within 1 km^2. To combat this, a higher resolution topography could be used with the existing framework, for higher-resolution product output. Topographic representativity will also affect the snow diagnosis since the local height of the 0 °C wet-bulb isotherm is the only criterion, meaning a perfect diagnosis at 1 km^2 resolution is not possible.

5.3. Snow

Overall the snow class has similar HSS to rain diagnoses, but is overdiagnosed and, thus, has a higher FAR than rain. Since the diagnosis is entirely dependent on the height of the UKPP 0 °C wet-bulb isotherm being below the ground (i.e. surface temperatures below zero), the results suggest that the height is negatively biased. This conclusion would also agree with the results of the mixed-phase precipitation class.

For the snow class the skill of the Euro4 temperature field is essentially being veified, which itself has many influencing factors. The only other source for misclassification is the previously mentioned 1 km^2 resolution of the local terrain input data. The SPT product might be seen as an attractive candidate for verifying NWP model SPT forecasts against. However, be aware that this would be a closed-loop verification for the snow class since its diagnosis is entirely reliant upon the model.

DiVeN data give higher verification metric values (73-81% improvement over random chance). The sites contain more snow events (5 sites > 250 m a.m.s.l.) which are often observed when the 0 °C wet-bulb isotherm height is several hundred metres below the surface. Borderline cases are less common in DiVeN compared with the other data. Similar to rain cases being low-skill in summer, low-skill winter events make a difference to the snow verification results. In late February and early March 2018, the exceptional snowfall associated with the "Beast from the East" (Galvin et al. 2019; Greening and Hodgson 2019) brought many low-skill snow cases into the verification dataset. If 2018 data is removed then scores using all datasets are reduced dramatically. The SPT product has diminished value in these scenarios since it is clear to users that all precipitation will reach the surface as snow.

5.4. Tolerance Method

The tolerance method used in this study demonstrates the sensitivity of the product's skill when adjusting the spatiotemporal inclusion, which a user typically considers when viewing a graphical map. Given the spatiotemporal range (from 1 pixel at one time, to 175 pixels over 30 minutes) the range of values provided by this method is often quite narrow, and is therefore informative to users. A wide range of score results would add negligible value to a single verification score result with no indication of spread. The tolerance method is therefore applicable to future verification of precipitation type diagnosis from any spatial-coverage product using single-point reference datasets.

When viewing a contingency table, the sum of all events remains constant between strict, fair and lenient tolerances but events can only move vertically in a contingency table between tolerances. If more events move from 'miss' to 'hit' compared with the number moving from 'correct null' to 'false alarm', then the HSS improves, and vice versa. The initial distribution of events differs significantly between precipitation class and ground dataset, hence the HSS values sometimes increase and sometimes decrease (notably rain against automatic observations) between SPT product tolerances in the results of this study.

If a user desires a higher POD then a larger domain should be considered from the SPT product. If a lower FAR is desirable then a smaller domain should be taken around a desired location, which will depend on the specific user and their application of the Met Office SPT product. The results are more complex for the HSS values. If a user wants a higher skill score, then the spatiotemporal sample should be different for each precipitation class and always dependent on which ground dataset is most trusted. For the mixed-phase class, the HSS values reveal that a larger sample increases the skill of the diagnosis (except when considering the automatic data (explained above) which has the lowest HSS of any ground dataset). Generally for rain and snow, using the specific pixel encapsulating a location area and time increases product skill. Note that this does not take into account the skill of detecting or not detecting precipitation accurately since all events that feed into the verification have precipitation in both sources.

5.5. Comparison to Other Verification Work

Comparing other SPT products in the literature is difficult since there are many variables affecting the verification. Table 4.5 shows a sample of literature verifying many SPT products based on NWP and various observational inputs. In addition to the different inputs to each algorithm, the location, time period, method and verification scores also differ, influencing the verification results of each study. Here it was noted that even the inclusion of a fifth year onto four existing years dramatically changed the true climatology and therefore the overall results.

Different statistical approaches are applied in different studies. Chen et al. (2016) and Gascón et al. (2018) use critical success index (CSI) as a verification metric, but CSI cannot be applied to the higher-dimension confusion matrices. Wandishin et al. (2005) use Brier Skill Score (BSS) but this is only applicable for probabilistic data. Elmore et al. (2015) use the PSS for contingency and the Gerrity skill score (GSS; Gerrity Jr. (1992)) for higher-order which emphasises weighted ranking to each class based on climatological rarity. The PSS is as justifiable as the HSS as a skill metric and both have higher-order applicability and give similar score values (Wilks 2011). However, using the same score for contingency tables and confusion matrices (as was done here) demonstrates the contributions from each class to the overall score.

6. Summary and Further Work

Reliable observations of precipitation type are needed both to verify and improve forecast microphysics, and also to operationally force NWP models with more accurate initial conditions through data assimilation. The Met Office surface precipitation type (SPT) product was examined with 3 datasets of ground-based observations over 5 years (2014-2018). The product uses Boolean logic to diagnose hail, snow, mixed-phase and rain using two empirical relationships based on radar reflectivity, radar precipitation rate and the 0 °C wet-bulb isotherm from an NWP model. In this paper, which is the first part of two papers, snow, mixed-phase and rain were veified. An overall product score was obtained using the higher-order Heidke skill score (HSS) and a bootstrapping technique to infer the monthly and yearly sensitivity to the overall product score. Statistical metrics applied to individual precipitation classes from contingency tables were bias (B), probability of detection (POD), false alarm ratio (FAR), and the HSS. A novel tolerance method was introduced which shows the realistic spatiotemporal spread

Table 4.5.: Comparative literature of verifications on surface precipitation type products, based on various input data, verification data, study location, time period and methods. This study is also shown, along with results from all studies for comparison. M-P stands for mixed-phase, T_{dry} and T_{dew} are dry-bulb and dewpoint temperature, PSS is the Peirce skill score, and GSS is the Gerrity skill score (Gerrity Jr. 1992). Continues in Table 4.6.

Study	Product	Method	Results
This work (Pickering et al. 2020)	Radar + NWP Observation Rain, M-P, Snow	UK, 5 years Automatic, Manual, DiVeN Varied spatiotemporal tolerance	Bias: Rain ~ 1, M-P $> > 1$, Snow > 1 POD: Rain ~ 0.98, M-P 0.05-0.44, Snow 0.76-0.94 FAR: Rain ~ 0.02, M-P 0.83-0.97, Snow 0.22-0.66 HSS: Rain 0.51-0.77, M-P 0.04-0.19, Snow 0.49-0.81 Overall HSS: 0.61 (0.4-0.8 w/ bootstrapping)
Gascón et al. (2018)	NWP Ensemble 0-24h forecast Rain, Snow, Others	Europe, 4 months (winter) Manual reports Strict tolerance	POD: Rain 0.52, M-P 0.07, Snow 0.55 - Probability of any precipitation must exceed 50% - Highest probabilistic type chosen
Reeves (2016)	Sounding Observation Rain, Snow, Others	U.S. East of Rockies, 27 months Automatic + Crowdsourced Auto: strict tolerance Crowd: closest (max 35 km ± 1h)	POD: Rain 0.68-0.92, Snow 0.79-0.90 - SPT must be consistent for > 4h - Lower rain frequency (62%)
Schmid and Mathis (2004)	$T_{dry} + T_{dew}$ Observation Rain, Snow	Switzerland, 2 months (winter) Automatic reports Strict tolerance	Bias: Rain 0.92, Snow 1.10 POD: Rain 0.81, Snow 0.99 - Higher snow frequency (63%)

Table 4.6.: Table 4.5 continued.

Study	Product	Method	Results
Chen et al. (2016)	Radar + NWP Observation Rain, Snow	8 U.S. cities, 4 months (winter) Crowdsourced reports Strict tolerance	POD: Rain 0.93-0.98, Snow 0.69-0.82 FAR: Rain 0.08-0.29, Snow 0.02-0.15
Wandishin et al. (2005)	NWP Ensemble 0-48h forecast Rain, Snow, Others	U.S., 3 months (winter) Manual reports Strict tolerance	POD: Rain 0.45-0.95, Snow 0.70-0.95 - Probability thresholds - Low-skill rain ($T_{dry} > 5$ °C) removed
Elmore et al. (2015)	NWP 0-24h forecast Rain, Snow, Others	U.S., 8 months (2 winters) Crowdsourced reports Strict tolerance	Bias: Rain 1.10-1.60, Snow 0.80-1.05 PSS: Rain 0.58-0.76, Snow 0.58-0.74 Overall GSS: 0.48 (3h), 0.34-0.45 (18h)
Ikeda et al. (2013)	NWP 1-8h forecast Rain, M-P, Snow	Eastern U.S., 2 months (winter) Automatic reports 18x18 km ± 6-minute tolerance Fractional confusion matrices	POD: Rain 0.90, M-P 0.66, Snow 0.86 - ‘No precipitation’ class included

of scores taking into consideration the fall time and the horizontal displacement precipitation may experience between the lowest-usable radar beam from the UK Met Office radar network and the ground.

The results show that the 0 °C wet-bulb isotherm from the UKPP (interpolation from the Euro4 NWP model) is too low, causing an overdiagnosis of snow ($B > 1$) leading to FAR values of 0.22-0.48 (strict tolerance). The 0 °C wet-bulb isotherm height also controls the height at which mixed-phase precipitation is fully melted into rain, and may contribute to the significant overdiagnosis of mixed-phase ($B >> 1$) with FAR values of 0.83-0.97 and POD values of 0.05-0.44 (all verification datasets and tolerances). Due to the overdiagnosis of snow and mixed-phase, by elimination the rain class is underdiagnosed. Rain has a bias of just under 1 which is skewed by the high frequency of the rain class, 91-97% between verification datasets. The HSS takes into account high frequency of occurrence, and this gives values of 0.51-0.77, which are similar to snow where HSS values are 0.49-0.81 (all verification datasets and tolerances). The mixed-phase has low HSS values of 0.04-0.19.

Overall the higher-dimension HSS value for all datasets combined is 0.61, which improves to 0.73 if all mixed-phase diagnoses are re-labelled as rain. Between verification datasets, the higher-dimension HSS are 0.48 ± 0.058 (automatic), 0.60 ± 0.147 (manual) and 0.73 ± 0.024 (DiVeN), where the uncertainty is representative of a 2σ confidence interval produced through bootstrapping.

Ground-based observations should capture the climatology of the location or target audience of the users of the product. Thus, the representativity of the data used to evaluate the product at a certain location is important. The automatic and manned SYNOP stations run by the Met Office may not capture the most extreme climatologies of the UK due to their siting requirements for optimal measurement standard. Similarly, the Disdrometer Verification Network likely does not capture the UK climatology since many instruments are located at high elevations.

Improvements to the Met Office radar-based SPT product are ongoing based on the results of this study. The Euro4 model has been marked for depreciation at the end of 2021 and there has been a freeze on scientific upgrades for several years. The implementation of a newer, higher resolution NWP model temperature field, particularly a model with improved microphysics schemes, should improve the snow class diagnosis in a future Met Office SPT product. Note that to evaluate the improvement in future SPT products, the current SPT product can be statistically implemented as a baseline. Currently, V_{ref} in Equation 4.6 is set here as the random proportion correct but this can be changed to be the proportion correct from this ‘baseline’ SPT product instead. Thus the score can then be used to show the percentage improvement over the current SPT product. The methods employed here may be easily implemented for verifying a range of observation-based or simulation-based classifiers, however the most important aspect of verification is consistency of score choice between studies to enable comparisons and to identify successful SPT diagnosis techniques.

In Part II, the hail precipitation class which uses a separate empirically-derived criterion from Waldvogel et al. (1979) is veified. Since the 3 ground-based datasets used here rarely detected true hail cases (automatic SYNOP stations are incapable of detecting hail and DiVeN instruments mistake graupel for hail), crowdsourced reports are used as verification data. Strengths and limitations are discussed, with recommendations to improve hail detection in the UK. Further, precipitation classes which are not included in the current radar-based Met Office SPT product are discussed for future inclusion.

References

Adolf Thies GmbH & Co. KG, 2011: Laser Precipitation Monitor - Instruction for Use. Tech. rep., Adolf Thies GmbH & Co. KG, Hauptstraße 76, 37083 Göttingen, Germany, 66 pp.

Beynon, C., S. Wyke, I. Jarman, M. Robinson, J. Mason, K. Murphy, M. A. Bellis, and C. Perkins, 2011: The cost of emergency hospital admissions for falls on snow and ice in England during winter 2009/10: A cross sectional analysis. *Environmental Health: A Global Access Science Source*, **10 (1)**, 1–7, doi:10.1186/1476-069X-10-60.

Bloemink, H., and E. Lanzinger, 2005: Precipitation type from the Thies disdrometer. *WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO-2005)*, 1–7, URL http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-82-TECO_2005/Papers/3(11)_Netherlands_4_Bloemink.pdf.

Bohm, H. P., 1989: A general equation for the terminal fall speed of solid hydrometeors. 2419–2427 pp., doi:10.1175/1520-0469(1989)046⟨2419:AGEFTT⟩2.0.CO;2.

Brabson, B. B., and J. P. Palutikof, 2002: The evolution of extreme temperatures in the Central England temperature record. *Geophysical Research Letters*, **29 (24)**, 1–4, doi:10.1029/2002GL015964.

Brown, I., 2019: Snow cover duration and extent for Great Britain in a changing climate: Altitudinal variations and synoptic-scale influences. *International Journal of Climatology*, **39 (12)**, 4611–4626, doi:10.1002/joc.6090.

Chen, S., J. J. Gourley, Y. Hong, Q. Cao, N. Carr, P. E. Kirstetter, J. Zhang, and Z. Flamig, 2016: Using Citizen Science Reports to Evaluate Estimates of Surface Precipitation Type. *Bulletin of the American Meteorological Society*, **97 (2)**, 187–193, doi:10.1175/BAMS-D-13-00247.1.

Chernick, M. R., 2011: *Bootstrap methods: A guide for practitioners and researchers*, Vol. 619. John Wiley & Sons.

Clarke, J. P. B., S. Solak, Y. H. Chang, L. Ren, and A. E. Vela, 2009: Air traffic flow management in the presence of uncertainty. *Proceedings of the 8th USA/Europe Air Traffic Management Research and Development Seminar, ATM 2009*, 571–580.

Cornford, D., and J. E. Thornes, 1996: A Comparison Between Spatial Winter Indices and Expenditure on Winter Road Maintenance. *International Journal of Climatology*, **16 (3)**, 339–357, doi:10.1002/(sici)1097-0088(199603)16:3⟨339::aid-joc40⟩3.3.co;2-l.

Curtis, S., A. Fair, J. Wistow, D. V. Val, and K. Oven, 2017: Impact of extreme weather events and climate change for health and social care systems. *Environmental Health: A Global Access Science Source*, **16 (Suppl 1)**, doi:10.1186/s12940-017-0324-3.

Doolittle, M. H., 1888: Association ratios. *Bull. Philos. Soc. Washington*, **7**, 122–127.

Dufton, D. R. L., and C. G. Collier, 2015: Fuzzy logic filtering of radar reflectivity to remove non-meteorological echoes using dual polarization radar moments. *Atmospheric Measurement Techniques*, **8 (10)**, 3985–4000, doi:10.5194/amt-8-3985-2015.

Efron, B., and R. J. Tibshirani, 1994: *An introduction to the bootstrap.* CRC Press.

Elmore, K. L., H. M. Grams, D. Apps, and H. D. Reeves, 2015: Verifying forecast precipitation type with mPING. *Weather and Forecasting*, **30 (3)**, 656–667, doi: 10.1175/WAF-D-14-00068.1.

Galvin, J., M. Kendon, and M. McCarthy, 2019: Snow cover and low temperatures in February and March 2018. *Weather*, **74 (3)**, 104–110, doi:10.1002/wea.3469.

Gascón, E., T. Hewson, and T. Haiden, 2018: Improving Predictions of Precipitation Type at the Surface: Description and Verification of Two New Products from the ECMWF Ensemble. *Weather and Forecasting*, **33 (1)**, 89–108, doi:10.1175/WAF-D-17-0114.1, URL http://journals.ametsoc.org/doi/10.1175/WAF-D-17-0114.1.

Gerrity Jr., J. P., 1992: A Note on Gandin and Murphy's Equitable Skill Score. *Monthly Weather Review*, **120 (11)**, 2709–2712, doi:10.1175/1520-0493(1992)120⟨2709:ANOGAM⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0493(1992)120%3C2709:ANOGAM%3E2.0.CO.

Green, A., 2010: From Observations to Forecasts – Part 7 . A new meteorological monitoring system for the United Kingdom's Met Office. *Weather*, **65 (10)**, 272–277.

Greening, K., and A. Hodgson, 2019: Atmospheric analysis of the cold late February and early March 2018 over the UK. *Weather*, **74 (3)**, 79–85, doi:10.1002/wea.3467.

Gunn, R., and G. D. Kinzer, 1949: The Terminal Velocity of Fall for Water Droplets in Stagnant Air. *Journal of Meteorology*, 243–248, doi:10.1175/1520-0469(1949)006⟨0243:TTVOFF⟩2.0.CO;2.

Haiden, T., A. Kann, C. Wittmann, G. Pistotnik, B. Bica, and C. Gruber, 2011: The integrated nowcasting through comprehensive analysis (INCA) system and its validation over the Eastern Alpine region. *Weather and Forecasting*, **26 (2)**, 166–183, doi:10.1175/2010WAF2222451.1.

Handa, H., L. Chapman, and X. Yao, 2006: Robust route optimization for gritting/salting trucks: A CERCIA experience. *IEEE Computational Intelligence Magazine*, **1 (1)**, 6–9, doi:10.1109/MCI.2006.1597056.

Harrison, D., S. J. Driscoll, and M. Kitchen, 2000: Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorol. Appl.*, **6**, 135–144, doi:10.1017/S1350482700001468.

Heidke, P., 1926: Berechnung Des Erfolges Und Der Güte Der Windstärkevorhersagen Im Sturmwarnungsdienst. *Geografiska Annaler*, **8 (4)**, 301–349, doi:10.1080/20014422.1926.11881138, URL https://doi.org/10.1080/20014422.1926.11881138.

Hubbert, J. C., M. Dixon, S. M. Ellis, and G. Meymaris, 2009: Weather radar ground clutter. Part I: Identification, modeling, and simulation. *Journal of Atmospheric and Oceanic Technology*, **26 (7)**, 1165–1180, doi:10.1175/2009JTECHA1159.1.

Ikeda, K., M. Steiner, J. Pinto, and C. Alexander, 2013: Evaluation of cold-season precipitation forecasts generated by the hourly updating high-resolution rapid refresh model. *Weather and Forecasting*, **28 (4)**, 921–939, doi:10.1175/WAF-D-12-00085.1.

Kay, A. L., 2016: A review of snow in Britain: The historical picture and future projections. *Progress in Physical Geography*, **40 (5)**, 676–698, doi:10.1177/0309133316650617.

Landolt, S. D., J. S. Lave, D. Jacobson, A. Gaydos, S. Divito, and D. Porter, 2019: The impacts of automation on present weather–type observing capabilities across the conterminous United States. *Journal of Applied Meteorology and Climatology*, **58 (12)**, 2699–2715, doi:10.1175/JAMC-D-19-0170.1.

Langleben, M. P., 1954: The terminal velocity of snowflakes. *Quarterly Journal of the Royal Meteorological Society*, **80 (344)**, 174–181, doi:10.1002/qj.49708034404.

Locatelli, J. D., and P. V. Hobbs, 1974: Fall speeds and masses of solid precipitation particles. *Journal of Geophysical Research*, **79 (15)**, 2185–2197, doi:10.1029/JC079i015p02185.

Lumb, F. E., 1963: Downward penetration of snow in relation to the intensity of precipitation. *Meteorology Magazine*, **92 (1086)**, 1–14.

Lyth, D., 2008: Results from UK Met Office investigations into new technology present weather sensors. *WMO Technical Conference on Instruments and Methods of Observation (TECO-2008), St. Petersburg, Russian Federation*, 27–29.

Lyth, D., and M. Molyneux, 2006: Results of Using Present Weather Instruments in the United Kingdom. Tech. rep., The Met Office, 46 pp.

Matson, R. J., and A. W. Huggins, 1980: The Direct Measurement of the Sizes, Shapes and Kinematics of Falling Hailstones. 1107–1125 pp., doi:10.1175/1520-0469(1980)037⟨1107:tdmots⟩2.0.co;2.

NERC, UK Met Office, B. S. Pickering, R. R. Neely III, and D. Harrison, 2019: The Disdrometer Verification Network (DiVeN): particle diameter and fall velocity measurements from a network of Thies Laser Precipitation Monitors around the UK (2017-2019). URL https://catalogue.ceda.ac.uk/uuid/602f11d9a2034dae9d0a7356f9aeaf45, doi:10.5285/602f11d9a2034dae9d0a7356f9aeaf45.

Parker, D. E., T. P. Legg, and C. K. Folland, 1992: A new daily central England temperature series, 1772–1991. *International Journal of Climatology*, **12 (4)**, 317–342, doi:10.1002/joc.3370120402.

Pickering, B. S., R. R. Neely, and D. Harrison, 2019: The Disdrometer Verification Network (DiVeN): A UK network of laser precipitation instruments. *Atmospheric Measurement Techniques*, **12 (11)**, 5845–5861, doi:10.5194/amt-12-5845-2019.

Pickering, B. S., R. R. Neely III, J. Jeffery, D. Dufton, and M. Lukach, 2020: Intercomparison of multiple precipitation sensor designs for precipitation rate and depth, drop size and velocity distribution, and precipitation type. *Journal of Hydrometeorology*, **In review**.

Rasmussen, R., and Coauthors, 2001: Weather Support to Deicing Decision Making (WSDDM): A winter weather nowcasting system. *Bulletin of the American Meteorological Society*, **82 (4)**, 579–595, doi:10.1175/1520-0477(2001)082⟨0579:WSTDDM⟩

2.3.CO;2.

Reeves, H. D., 2016: The uncertainty of precipitation-type observations and its effect on the validation of forecast precipitation type. *Weather and Forecasting*, **31 (6)**, 1961–1971, doi:10.1175/WAF-D-16-0068.1.

Sandford, C., 2015: Correcting for wind drift in high resolution radar rainfall products : a feasibility study. *Journal of Hydrology*, **531**, 284–295, doi:10.1016/j.jhydrol.2015.03.023, URL http://dx.doi.org/10.1016/j.jhydrol.2015.03.023.

Schmid, W., and A. Mathis, 2004: Validation of methods to detect winter precipitation and retrieve precipitation type. *Annalen der Meteorologie*, **(January)**, 1–8, URL http://www.iac.ethz.ch/staff/wueest/sirwec/conferences/bingen2004/topic_iv/sirwec-iv-3-schmid.pdf.

Waldvogel, A., B. Federer, and P. Grimm, 1979: Criteria for detection of hail cells. *Journal of Applied Meteorology*, **18 (12)**, 1521–1525, doi:10.1175/1520-0450(1979)018⟨1521:CFTDOH⟩2.0.CO;2.

Wandishin, M. S., M. E. Baldwin, S. L. Mullen, and J. V. Cortinas, 2005: Short-range ensemble forecasts of precipitation type. *Weather and Forecasting*, **20 (4)**, 609–626, doi:10.1175/WAF871.1.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*, Vol. 100. Third edit ed., Academic press, 306 pp.

WMO, 1988: Manual on Codes. *WMO Publ. 306*, **1**, 203 pp.

WMO, 2017: WMO Guide to Meteorological Instruments and Methods of Observation (the CIMO Guide). Chapter 6. Masurement of Precipitation. Tech. rep., World Meteorological Organization.

Zikmunda, J., 1972: Fall Velocities of Spatial Crystals and Aggregates. *Journal of The Atmospheric Sciences*, **29**, 1511–1515, doi:10.1175/1520-0469(1972)029⟨1511:FVOSCA⟩2.0.CO;2.

Chapter 5.

Improving Observations of Precipitation Type at the Surface: A 5-year Verification of a Radar-derived Product from the United Kingdom Met Office. Part II: Hail.

In revision to *AMS Journal of Hydrometeorology.*

Contributions from: Ryan R. Neely III, Steven Best, Maryna Lukach and David Dufton

Abstract

In Part II of this study, the hail precipitation class from the United Kingdom Met Office surface precipitation type (SPT) product is verified. The criterion used for hail detection is that a radar reflectivity factor $\geq$ 45 dBZ must exist at an altitude of greater than 1.4 km above the 0 °C isotherm. Since hail is uncommon in the UK, the verification dataset used in Part I of this verification for rain, mixed-phase and snow, is not suitable for verifying hail. Instead, a crowdsourced set of reports from the European Severe Weather Database are employed. The tolerance method described and applied in Part I is applied to the hail class verification over 5 years from 2014 through to 2018, where 111 reports of hail occurred in the total radar network domain and 32 hail reports occurred in the UK itself. Hail events from the SPT product are also mapped which shows the broad distribution and frequency of hail in the UK and also reveals artefacts in the SPT product. Verification results show that the Met Office SPT product has a higher probability of detection for hail-producing storms in mainland Europe than in the UK due to the exact method of applying the criterion and the scan strategy of individual radars. Further, there is a slight northeasterly spatial bias in hail diagnoses by the SPT product. Improvements to the ground verification datasets for hail are needed and would permit more extensive verification of hail diagnosis and forecast products in the future.

1. Introduction

Pickering et al. (2020, hereafter Part I) introduced the United Kingdom Met Office's radar-based surface precipitation type (SPT) product and verified the rain, mixed-phase and snow diagnoses over 5 years with a novel tolerance technique. In this paper, the same technique is applied to verify the SPT product for hail diagnoses over the same time period. Strengths and limitations are discussed, with recommendations to improve hail detection in the UK.

Hailstones begin with an embryo, which are most commonly pristine ice crystals, aggregates or frozen supercooled drops (Knight and Knight 1970a; Federer et al. 1986; Rasmussen and Heymsfield 1987; Dennis and Kumjian 2017; Allen et al. 2020). The embryos are then grown by supercooled water instantly freezing into an opaque layer on the surface, also known as riming. When particles grown by this process are small, they are often referred to as graupel; with continued growth, the larger particles are referred to as hail. Defining criteria between graupel and hail have included their size but also their shape and density which affects their diameter-velocity relationship (Locatelli and Hobbs 1974). AMS (2017) defines hail as ice particles larger than 5 mm in diameter and is the most widely used definition.

Large hail is grown in the updraft region of convective systems where supercooled liquid water is lofted and collides with larger ice particles which have greater mass and therefore faster fall velocities (Nelson 1983; Allen et al. 2020). The supercooled liquid water will either immediately freeze onto the larger ice particle (dry growth, riming) or remain as a liquid until the whole liquid-coated ice particle sheds enough energy to the surrounding atmosphere to freeze (wet growth, accretion). When an ice particle is small, the impacting supercooled water freezes instantly onto its surface and the temperature of the particle is slightly increased from the latent heat of freezing. The larger a hailstone gets, the more supercooled liquid water is deposited onto the surface and the more latent heat is generated which leads to the wet growth stage (Pruppacher and Klett 1997). Several atmospheric conditions determine the growth type, which often changes several times during a hailstone growth cycle and leads to the distinctive internally-layered opaque-translucent structure. During wet growth the liquid coating is distorted by aerodynamic drag, causing icicle lobes to form which gives larger hail its spiked appearance (Knight and Knight 1970b).

Hail can have varying degrees of impact on society based upon the kinetic energy transferred upon impact (Hohl et al. 2002). The kinetic energy of a hailstone is overwhelmingly determined by its size (mass). However, the transfer of energy upon impact is, to a lesser extent, determined by the mechanical properties of the hailstone (elasticity, tensile strength, elongation, hardness and fatigue limit), which may vary between storms but are less understood due to measurement difficulties (Heymsfield et al. 2014; Giammanco et al. 2017). In 2016, U.S. property (which includes vehicles but excludes agriculture) losses due to convective storms were similar to hurricanes at 11.23 billion USD, where 60% of property losses are attributed to hail by insurance risk modelling (Guntur and Tippett 2017). A single severe hail event in a dense urban area can result in nearly 1 billion USD of property and vehicular losses alone (Brown et al. 2015), the cost of which is burdened upon insurance companies and their policyholders, or directly upon uninsured owners. In Europe, hail can be the costliest natural hazard (Schemm et al. 2016) with single events known to cost on the order of 1 billion EUR (Punge et al. 2014; Kaspar et al. 2009). In the UK, hail is less common (Punge and Kunz 2016; Webb et al. 2009) but damaging events do occur; Webb et al. (2001) and Clark et al. (2018) discuss several historical hailfalls that caused significant damage, and more recently a storm in Sheffield produced large spiked hail with icicle lobes, shown in Figure 5.1.

Hail is difficult to forecast with numerical weather prediction (NWP) models because it precipitates in narrow swaths on the order of 1-10 km as storms propagate (Frisby 1961; Webb et al. 2001). Due to this, forecasts mostly exist as probabilities of hail-producing convective activity occurring within a radius of a given location. Certain properties of the predicted atmosphere can be used as diagnostics to estimate the likely size of hail produced if convective activity occurs, such as surface temperature, CAPE, deep-layer shear, storm-relative helicity, mid-level lapse rate (Rasmussen and Straka 1998; Manzato 2005; Johnson and Sugden 2014); however there are many drawbacks and these methods lack specificity. Observation-based hail warnings use the real-time observed properties of a storm to nowcast which areas will be impacted. Radar-based products such as the Met Office SPT product are suitable for nowcasting because

Figure 5.1.: An image of large hail taken during a storm in Sheffield, UK on the 26th June 2020, courtesy of Twitter user @ImEmmaB. The hail has a spiked appearance indicating wet growth occurred to produce icicle lobes. The coins in the image used for comparison indicate that the largest hailstones during this event had a maximum diameter approximately greater than 5 cm.

of their broad coverage and high spatiotemporal resolution—if the reliability of the product is sufficient. Furthermore, NWP improvements rely upon observations with higher diagnosis skill to be used as verification data. Verifying an NWP model against a poorer-skill observation will not lead to improvements. An added layer of complexity to consider is that climate change may lead to the occurrence of hail becoming more common in places such as the UK, along with increases in maximum hailstone sizes and therefore greater property damage and losses. However, there is low certainty and much disagreement in the literature over climatological changes for hail in Europe (Eccel et al. 2012; Mohr et al. 2015; Sanderson et al. 2015; Púčik et al. 2017). Reliable and long-term observations will help in understanding the effect of climate change on hail, which radar-based products are suitable for. Long term climatologies also have value in the construction industry, to inform material choices based on risk. For example, wind turbines are particularly sensitive to hail (Keegan et al. 2013), so their placement or protective tolerance (which is associated with a cost) could be decided based on long term radar observations for hail, if the observations are reliable. However, the reliability of the Met Office product has not yet been verified, which is the purpose of this study.

Due to the rare nature of hail events in the UK, the methods used in Part I must be adapted. Further, the verification datasets used in Part I are unreliable for hail diagnosis, so a crowdsourced dataset is utilised to bring greater certainty to the verification. The following sections describe the hail diagnosis algorithm used in the Met Office SPT product, and the verification dataset used in this verification and how it differs from those used in Part I.

1.1. Met Office Radar Product

The radar-based hail detection product is part of the broader precipitation type product detailed in Part I. Briefly, the SPT product has been running since late 2013, covers the United Kingdom and has a spatiotemporal resolution of 1 km^2 and 5 min. Hail is the first class in a Boolean logic decision tree, using the criterion of Waldvogel et al. (1979) as follows:

$$\begin{gathered} Z \geq 45 \text{ dBZ} \\ \& \\ h \geq (FZL + 1.4 \text{ km}), \end{gathered} \tag{5.1}$$

where Z is the radar reflectivity factor on a polar grid, in units of dBZ. FZL is the 0 °C wet-bulb isotherm altitude which is taken from the UKPP, a post-processed (regridded) version of the Euro4 NWP model based on the Met Office Unified Model architecture, updated hourly. Figure 5.2 gives an example of the SPT product during a severe hail event on the 1st July 2015, which is described in detail by Clark et al. (2018). Three individual storms produced hailstones reaching the surface widely in excess of 25 mm with the highest up to 60 mm in diameter at several locations.

Waldvogel et al. (1979) originally set out to examine how many hail embryo seeding rockets were being wasted in the Soviet hail suppression work by Sulakvelidze (1969). The criterion used for selection of hail seeding was complex and Waldvogel et al. (1979) used a sample of 195 cells (chosen if 35 dBZ or greater existed at the -5 °C isotherm) to show that a) 32% of rockets were being wasted on non-hail producing cells and that b) the same hit rate could be achieved using the simplified criterion of Equation 5.1 with fewer false alarms. Note that graupel events were specifically avoided in the study by excluding cold low synoptic characteristics.

Equation 5.1 is applied to every polar voxel observed by individual radars in the precipitation radar network, which consists of 15 radars managed by the Met Office, 2 managed by Met Éireann, and 1 managed by the Channel Islands Meteorological De-

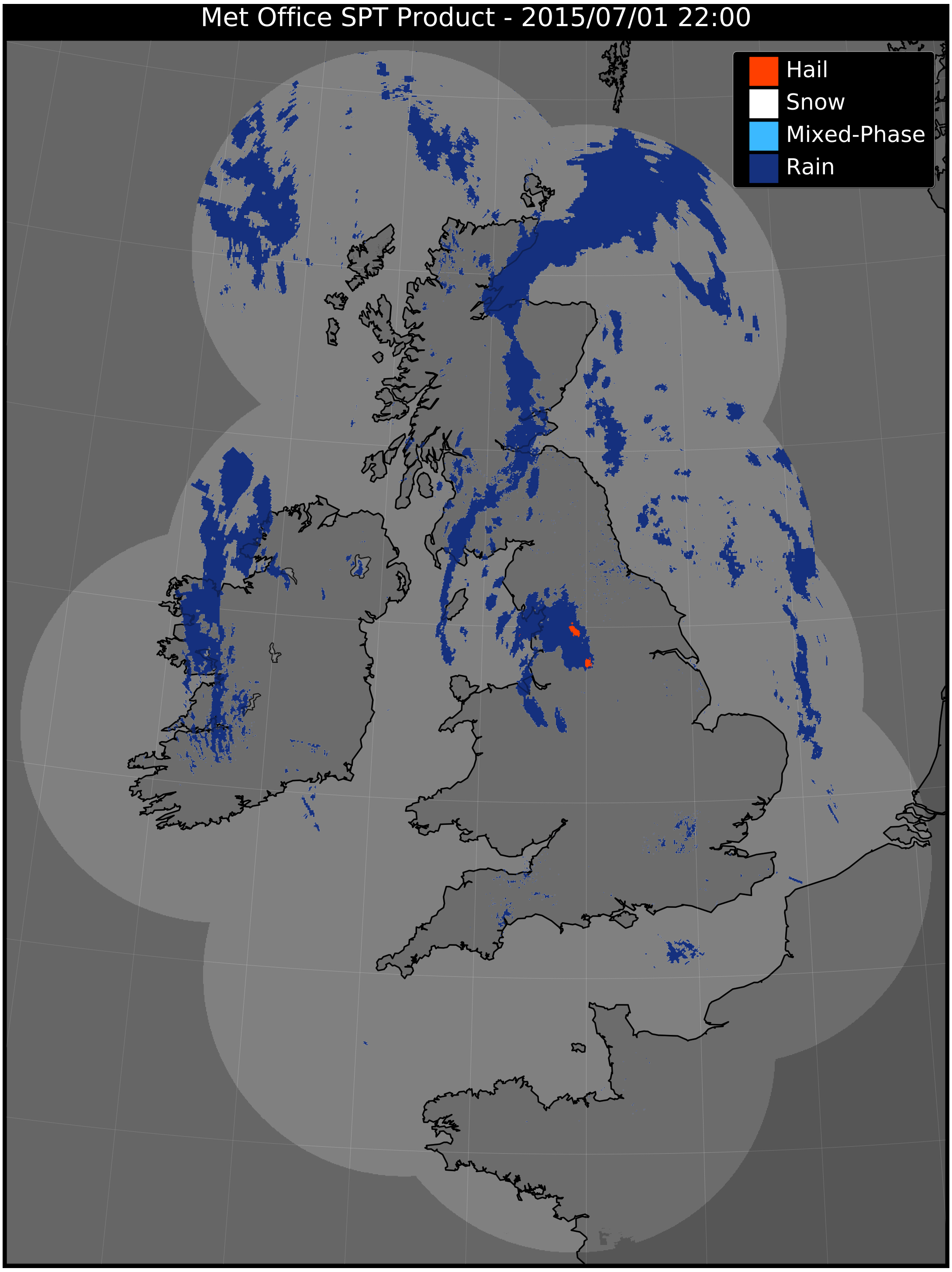


Figure 5.2.: An example of the Met Office SPT product, during two disruptive hail-producing supercells at 2200 UTC 1st July 2015 as discussed in Clark et al. (2018). An animated video of the entire day is supplied in the supplementary material.

partment. These individual radars have significant coverage area overlap in the network. The final composited product is a Cartesian 1 km^2 grid over the UK and Ireland. For each pixel, any radar which observes that location is checked for hail diagnoses. If the statistical mode (the most common) diagnosis out of the available radars is the hail class (versus any other precipitation type in the SPT product, or none), then the final gridded product pixel is labelled as hail. For example, if five radars observe a chosen 1 km^2 location, three out of the five must be diagnosing hail there for the pixel in the final gridded product to receive the hail label.

1.2. Verification Data

The verification datasets used in Part I [(automatic, manual and Disdrometer Verification Network (DiVeN)] all use the Present Weather (PW) code system (WMO 1988, 2017) to record observations, yet there is no entry for graupel meaning that hail is often used to describe both events, with no delineation. Pickering et al. (2019) demonstrated graupel being reported as hail by the DiVeN instruments with a first-hand case study. Furthermore, hail is rarely captured by traditional fixed-point observing networks due to sparse coverage and infrequent event occurrence. Crowdsourced reports such as the U.S. storm reports (Carey et al. 2003), mPING (Elmore et al. 2014), or home automatic weather stations (Clark et al. 2018) are more successful at capturing sporadic events, although in the case of manually submitted reports, the lack of a report co-located with a diagnosis does not mean the absence of the event in reality. Therefore the choice of statistical metrics can be somewhat limited. The technique used in this study gives added insight where these verification limitations exist.

1.2.1. ESWD Reports

The verification dataset used in this study is a crowdsourced dataset from the European Severe Weather Database (ESWD), created by the European Severe Storms Laboratory (ESSL), which does delineate between hail and graupel (Dotzek et al. 2009). While graupel is a relatively unknown term to the general public who will typically label all dense ice precipitation of any size as hail, the ESWD only records hail and excludes graupel events. The maximum hail diameter is recorded, and a human checks the feasibility of reports to ensure that all reports are likely to be hail as per the > 5 mm diameter definition (AMS 2017). The definition of a hail report for the ESWD in Groenemeijer and Liang (2019) is: “Hailstones that have a diameter (in the longest direction) of at least 2.0 centimetres, or hailstones that form a layer of 2.0 cm thickness or more on flat parts of the earth's surface." This gives the verification confidence that all verification data points are certain hail occurrences. Table 5.1 summarises the dataset characteristics and includes the datasets from Part I for comparison.

The ESWD reports also contain quality control designations (considered analogous to uncertainty) which enables filtering to increase the overall quality of the dataset. The designations are: "as received" (QC0), "plausibility check passed" (QC0+), "report confirmed by a reliable source" (QC1), and "scientific case study" (QC2), described in Groenemeijer and Liang (2019). The limitation of crowdsourced datasets like the ESWD reports is that spatial biases inevitably exist which are correlated with population density and community awareness of the reporting system. For example, the ESWD does not capture the frequency of hail events known to exist around the Mediterranean from Michaelides et al. (2018). Further, Punge and Kunz (2016) showed that there is a bias in the ESWD towards central Europe where more people are reporting hail events to the database. Since this study does not directly compare the spatial distribution of the ESWD hail reports against the Met Office SPT product, these biases should not hamper the verification. The ESWD reports have been used successfully in numerous studies (Kaltenböck et al. 2009; Gatzen et al. 2011; Hov et al. 2013; Antonescu et al. 2016; Groenemeijer et al. 2017) and reports are available for the entire 5-year study period (2014-2018).

2. Study Period Characteristics

In this study, the Met Office SPT product is verified for hail over a 5 year period of 2014 to 2018 inclusive (60 months total). This section of the paper describes the characteristics of the study period through a frequency map and statistics derived from the SPT product and the ESWD verification dataset.

2.1. Frequency Map

Following Part I, the SPT-product diagnoses of hail over the 5-year study period are summed in time to create total radar-diagnosed frequencies, shown in Figure 5.3. High-resolution zoomable PDF maps for different timescales are also provided in the supplementary material. In addition, Figure 5.4 shows a map of the UK and the Republic of Ireland, with locations, regions and radar sites labelled to aid with the interpretation

Table 5.1.: Summary of the ESWD verification dataset used in this study, compared with the datasets used for rain, mixed-phase and snow class verification in Part I. Includes the measurement technique, the format of the data when received, the frequency of data available, the number of locations available, and the availability over the duration of this study period.

	ESWD	Automatic	Manual	DiVeN
Measurement Technique	Reports submitted by the general public.	A Vaisala FD12P present weather sensor, precipitation detector, visiometer, ceilometer and air temperature thermometer combined into an arbiter.	Trained meteorological observer.	A laser disdrometer measures particle diameter and fall velocity and uses empirical relationships to determine precipitation type.
Format	Confirmation of hail, often maximum hail diameter (cm), with quality assessment.	PW Code (WMO Table 4680), 83 codes reported.	PW Code (WMO Table 4677), 91 codes reported.	PW Code (WMO Table 4680), 21 codes reported.
Frequency	Ad-hoc	Hourly	Hourly	5-minute
Locations	N/A	172	38	14
Availability	2014-2018 (5 years)	2014-2018 (5 years)	2014-2018 (5 years)	2017-2018 (18 months)

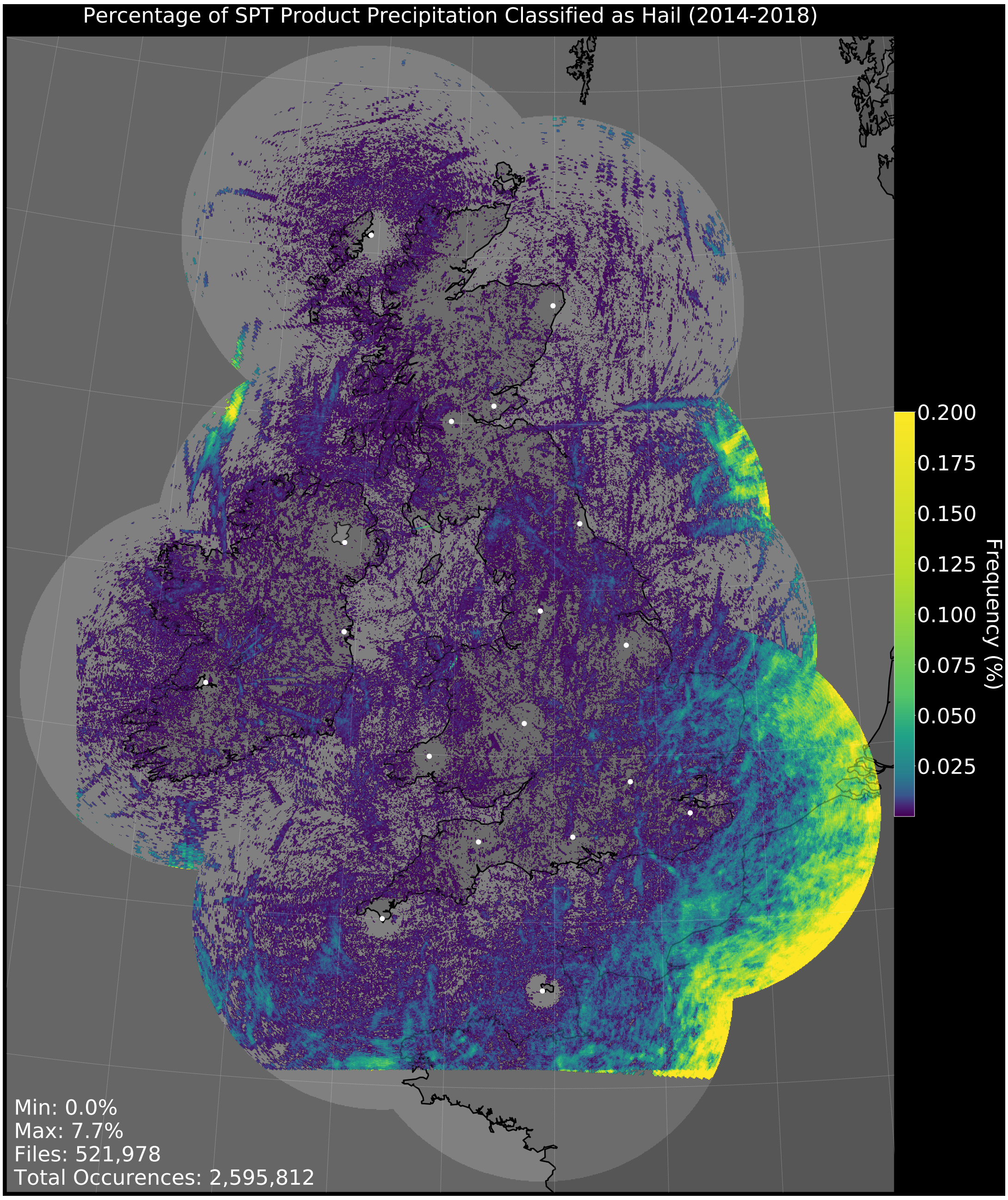


Figure 5.3.: Percentage of precipitation which the Met Office SPT product diagnosed as hail from the start of 2014 to the end of 2018 (5 years). The Met Office, Met Éireann and the Channel Islands Meteorological Department radar locations are marked as white dots. The scale is set from 0% to 0.2% to show additional features. The maximum of 7.7% is a bug in the product on the southern boundary. The highest percentages on the product domain fringes over mainland Europe are around 2-3%.

of the following section.

The hail class frequency map highlights the disparity in hail event occurrence in the UK. Regions which are diagnosed hail more frequently are, in rank order: continental Europe (even when factoring in long-distance effects), East Anglia (possibly emphasised by radar distance), the English Channel and the Pennines. These frequency-rankings are in broad agreement with climatological studies of hail occurrences over Western Europe (Punge and Kunz 2016; Webb et al. 2009), where continental Europe experiences higher frequencies of hail than the UK. However, the lack of spatial coverage in the verification reference data does not allow for informed specific comments to be made on the exact frequency values suggested by the SPT product for the occurrence of hail.

Radar scan geometry is important for the hail class. The hail criterion has an altitude requirement so the edge of the radar network is more likely to observe hail; the only level of the atmosphere being observed is high and there are no low-altitude beams to 'disagree' (given that the mode of several overlapping radars is output, as described in Section 1.1) with a diagnosis simply because the altitude is insufficient. However, the Ingham (Lincolnshire) radar diagnoses significantly less hail than adjacent radars at long range, signalling a low reflectivity bias at long-range concurrent with the frequency map interpretations from Part I.

Close to radar site locations, the detection of hail appears to be less frequent. The Channel Islands (furthest south) radar is the most prominent example of this behaviour, which results in the Channel Islands having almost no hail diagnoses over the 5-year period. This close range 'hail hole' effect is due to a combination of the criterion used for hail detection and also the method of compositing individual, overlapping radars into a single end product. For the Channel Islands, coverage is only provided by one radar, so the effect is most notable. The scan strategy has a maximum elevation angle of 4° and thus the beams are not high enough at close range to be greater than 1.4 km above the 0 °C wet-bulb isotherm. Even if 45 dBZ or greater is observed, the height criterion will not be satisfied and thus no hail can be diagnosed. For other radars in the UK and Ireland, the close range area to one radar site is also observed by one or multiple other radar sites and therefore, the 'hail hole' is less pronounced. Coverage from distant radars (if not blocked by geographical features) will consist of higher elevation beams which more frequently observe the atmospheric region required for the Waldvogel et al. (1979) criterion. However, for a single 1 km^2 Cartesian pixel location in the final SPT product, the majority of individual radars observing that location on their polar grid must diagnose hail. For example, the atmosphere above

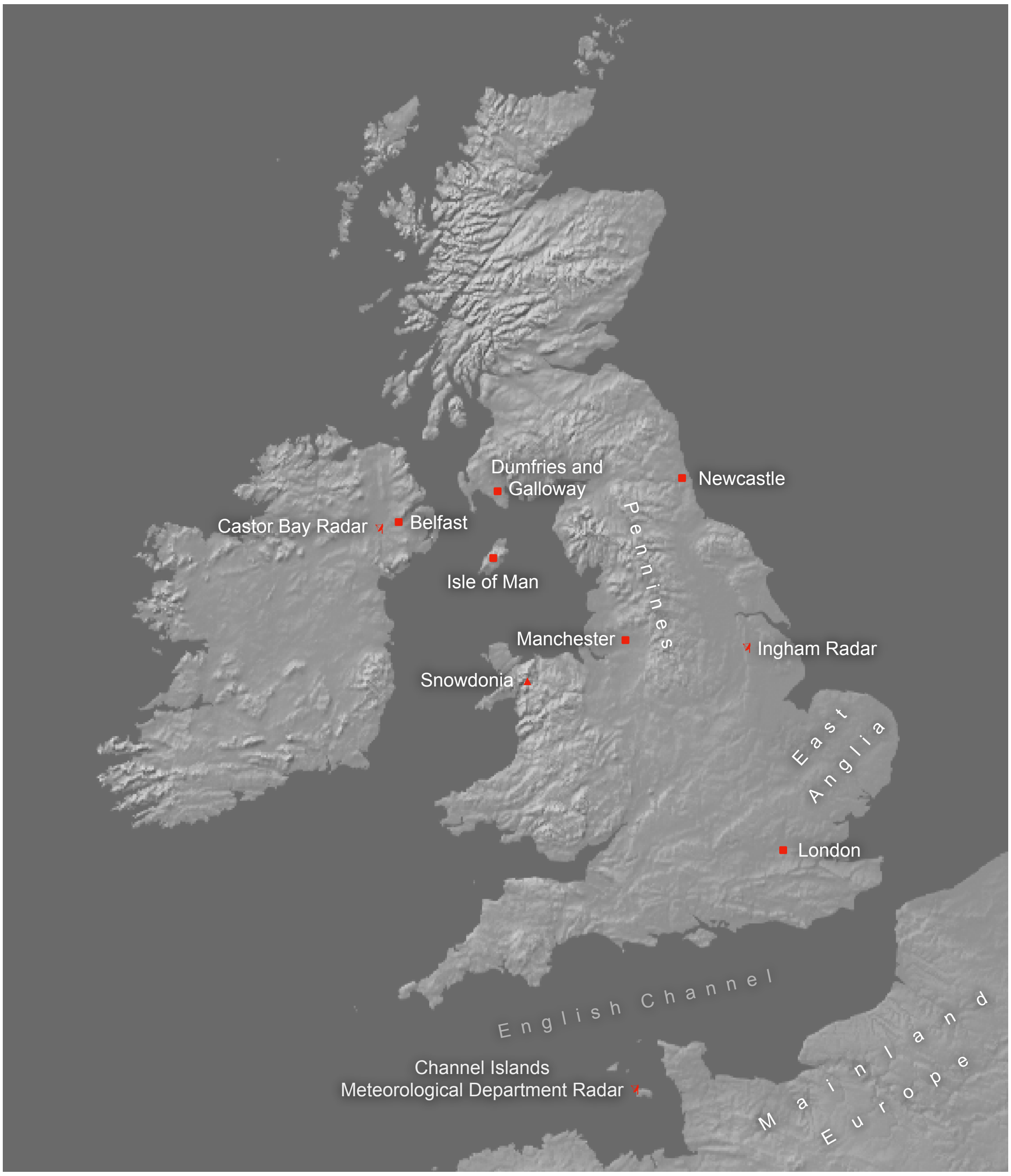


Figure 5.4.: A map of the UK and the Republic of Ireland showing locations discussed in Section 2.1.

London is simultaneously observed by 5 radar sites; in order for hail to be diagnosed for a given 1 km^2 pixel in London, 3 out of the 5 radars must diagnose hail in that 1 km^2 Cartesian pixel. Two of the radars are close to London, and will exhibit the same 'hail hole' individually as the Channel Islands radar does, making hail diagnosis in the final composited product less likely but not as unlikely as the Channel Islands 'hail hole'. Sub-seasonal versions of this plot (see supplement) show that the border of the hail hole changes over the year. The height of the 0 °C wet-bulb isotherm in an average month is lower in the winter than in the summer, so the radar beams reach the altitude needed to be 1.4 km above the 0 °C level at closer ranges to the radar compared to the summer months.

Just north of the Isle of Man in Dumfries and Galloway, a partial beam blockage correction on the Castor Bay radar (located at latitude, longitude: 54.50, -6.34, near Lurgan, Belfast, Northern Ireland) may be the cause of a high streak of hail detections which would be caused by a high reflectivity bias. Snowdonia in North Wales also exhibits an increased hail frequency, possibly due to elevated terrain being observed with high reflectivity.

Narrow swaths of individual storm-cell motion can be seen in Figure 5.3 (not to be confused with the ray-interference of the Met Éireann radar in Shannon, furthest west of the network). The three storm tracks of the long-lived hail-producing cells on 1st July 2015, documented by Clark et al. (2018), are also visible on Figure 5.3. These start near Manchester and propogate to the NNW towards Newcastle. The streak of hail on the edge of the network NW of the Republic of Ireland and Northern Ireland is observed by 2 radars and associated with a single event from 1800 UTC 19 July 2016 to 0100 UTC 20 July 2016 where consistently high precipitation rates of > 32 mm h^{-1} existed for over 6 hours. This event was associated with a slow-moving shortwave trough leading ahead of a cold front pushing in from the west replacing southerly flow in the region, which triggered a long line of convection with substantial lightning activity.

In general, long-term frequency plots are useful for exposing artifacts, events and trends within the radar and SPT product data. The sensitivity of the SPT product to changes in radar scan geometry are well highlighted. The limitation of using this method to find radar artifacts is that many years of observations are needed, particularly for hail, and usually more than one year if seasonal changes are to be observed.

2.2. Verification Data Statistics: ESWD

The ESWD database contains 111 reports of hail within the Met Office radar domain occurring on 35 days out of the 5-year study period (7 hail days per year on average). Of these, 32 were located in the UK and 79 were in mainland Europe, where the frequency of hail is much higher (Punge et al. 2014). Maximum hail diameter seen in each event (85/111 reported this variable) varies from 15 to 90 mm with an average of 37 mm. The months of June, July and August (JJA) contain the majority of the hail reports (92, 82.9%) and also the largest hail diameters compared to the rest of the year. For comparison, the SPT product diagnoses 58.0% of hail occurrences in JJA across the whole domain (so including over the ocean, which is unobserved by the ESWD). 14 of the 32 ESWD hail reports occurring in the UK (43.8%) occurred on 1 July 2015.

3. Methodology

The aim of this study is to verify the skill of the Met Office SPT product for hail over a 5-year period. To achieve this, traditional ground-based verification instruments do not provide frequent observations to conduct a verification. Further, graupel is often reported as hail due to the absence of a graupel code in the PW system. Graupel is a hydrometeor which forms in a similar process as hail, with supercooled liquid drops freezing as rime onto an ice crystal or aggregate, but does not exceed 5 mm in diameter (AMS 2017). The product criterion (Waldvogel et al. 1979) was designed for hail, not graupel. The sections below outline the steps taken to verify the skill of the SPT product.

3.1. Quality Control and Data Handling

The ESWD reports contain a quality control rating, designated by staff at the ESSL. Out of the 111 reports within the radar domain and study period, 93 (83.8%) were rated QC1: “report confirmed by reliable source” and 18 (16.2%) were rated QC0+: “plausibility check passed” which are both deemed acceptable for this study. Therefore, all events in the database are used and no removal is required. No events were QC0: “as received” nor were any rated a QC2: “scientific case study” despite Clark et al. (2018) and the 14 ESWD reports associated with those storms.

Table 5.2.: The layout of the 2 × 2 contingency table used in this study. Compared with Part I, there is no ground verification report of ‘no hail’ and therefore several of the contingency table elements are N/A.

		Surface		
		Yes	N/A	
SPT	Yes	Hit, a	N/A	N/A
	No	Miss, c	N/A	N/A
		o_1	N/A	N/A

3.2. Statistical Metrics

Since the hail reports are only confirmatory (i.e. there are zero reports of “no hail”), only the ‘hit’ and ‘miss’ quadrants of a contingency table (described in Part I) are satisfied (see Table 5.2).

Therefore the statistical metrics used in Part I to verify the precipitation classes, such as the Heidke skill score (HSS; Heidke (1926)) are not applicable in this study. With only the ‘yes’ hail verification data, the metrics that can be used here are probability of detection (POD or hit rate) and probability of miss (POM or miss rate):

$$POD = \frac{a}{(a+c)} \tag{5.2}$$

$$POM = \frac{c}{(a+c)} \tag{5.3}$$

where a = hit and c = miss. POD is the chance of a correct diagnosis when hail does occur and thus ranges from 0 (the event is never detected) to 1 (the event is always detected). POM is the chance of no diagnosis when hail does occur and thus ranges from 0 (the event is never missed) to 1 (the event is always missed). The values are complementary such that POD + POM = 1. Since hail has such a high impact on society, the SPT product is also simplified into a dichotomous (yes / no) hail output, such that cases where the radar saw no precipitation are included as ‘miss’ events. This is in contrast to Part I where the no-precipitation cases were excluded. This makes the results of the hail verification incomparable to the results of the rain, mixed-phase and snow categories and demonstrates that without improvements to ground-based hail observations, all precipitation types cannot be verified equitably.

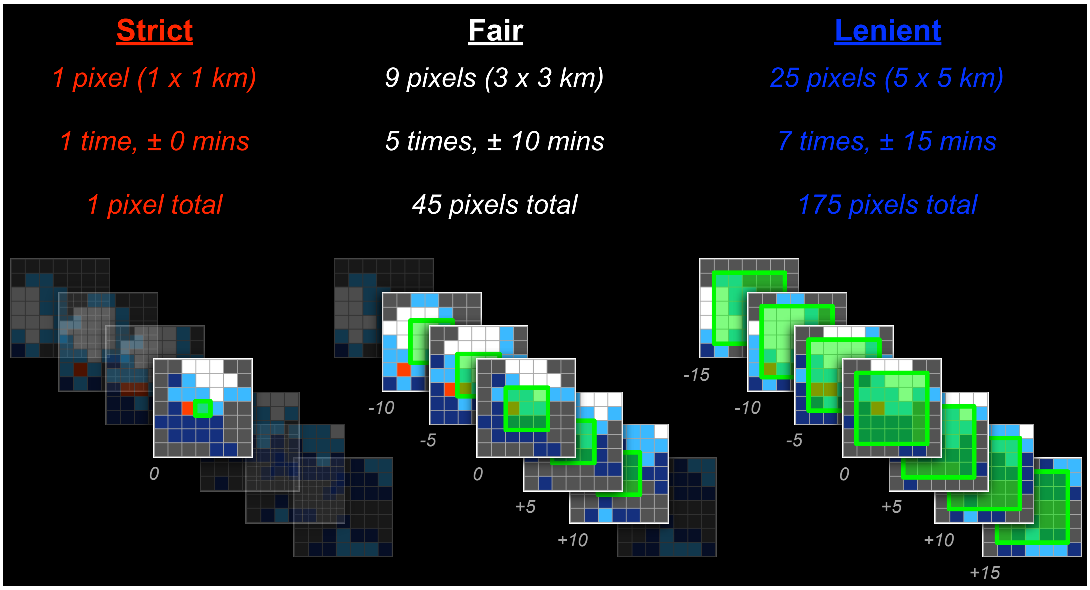


Figure 5.5.: An example of a timeseries of the SPT product stacked representing time (5-minute frequency). The green-shaded area is the sample used for verification in three tolerances. The strict tolerance uses only the pixel co-located with the ground report. The fair tolerance uses a 3×3 region around the ground report and ± 10 minute product outputs for a total of 45 pixels. The lenient tolerance uses a 5 × 5 km region around the ground report and ± 15 minute product outputs for a total of 175 pixels. If any of the green-shaded pixels are in agreement with the ground observation, then the SPT product is correct and a ‘hit’ is recorded.

3.3. Tolerance

Part I introduced the idea of using a tolerance method on the SPT product to incorporate the disparity of the lowest-usable radar beam height and the horizontal advection that may occur as the hydrometeors fall to the surface. The same method used for the rain, mixed-phase and snow verification of varying the spatial and temporal tolerance of the radar between strict (only the pixel containing the report), fair (3×3 km & ± 10 min), and lenient (5×5 km & ± 15 min) is applied here. If the radar does not diagnose hail in the 5 × 5 km area and within ± 15 min of the ESWD report, it is labelled as a miss. Figure 5.5 shows diagrammatically the tolerance method applied in this analysis.

4. Results

4.1. Statistical Metrics

The tolerance method is applied to enhance the informedness of the hail class results, where the only outcome is hit or miss since all of the ESWD reports are confirmatory. The strict tolerance, where only the pixel encapsulating the hail report is examined, has an overall hit rate of 30.6%, the fair tolerance (3×3 km ± 10 min) 65.8%, and the lenient tolerance (5×5 km ± 15 min) 75.7%. Note that the scores are additive such that all the events captured by the strict tolerance will also be captured by the fair tolerance, and so on. Complete misses where the radar did not report any hail diagnoses within 5×5 km and ± 15 min occurred in 24.3% (27) of the ESWD events.

A map is shown in Figure 5.6 which denotes which tolerance included a hail diagnosis by the SPT product, or a miss. For example, the blue triangles over Northern England mean that the strict and fair (3×3 km, ± 10 min) tolerance did not contain any hail diagnoses, but that the lenient (5×5 km, ± 15 min) did. Only 3 events in the UK were detected by the SPT product at the exact time and location of the ESWD report (i.e. strict tolerance), and they all occurred in East Anglia. The results visually suggest a discrepancy between hail events in the UK compared with hail reports in mainland Europe. The analysis is therefore further broken down into hail events in the UK (of which there are 32) and in continental Europe (79) to check for any discrepancy. The UK results are: strict (9.4%), fair (46.9%), lenient (62.5%) and miss (37.5%) whereas the continental Europe results are: strict (39.2%), fair (73.4%), lenient (81.0%) and miss (19.0%). The POD is higher and the POM is lower for the ESWD reports in continental Europe compared with those in the UK. For each tolerance, the improvement of continental European hit scores are: strict (+29.8%), fair (+26.5%), lenient (+18.5%) and miss (-18.5%).

4.2. Spatiotemporal Variability

To provide further insight into the performance of the SPT algorithm for diagnosing hail, the SPT product around the 111 ESWD-reported events (in a Lagrangian perspective) is visualised in Figure 5.7. The radar pixels around the report location in space and time are tallied up for hail diagnosis across all of the 111 ESWD reports. For example, if the radar product were perfect then the pixel directly over the ESWD report should contain 111 hail classifications (out of a total of 111). The same can be

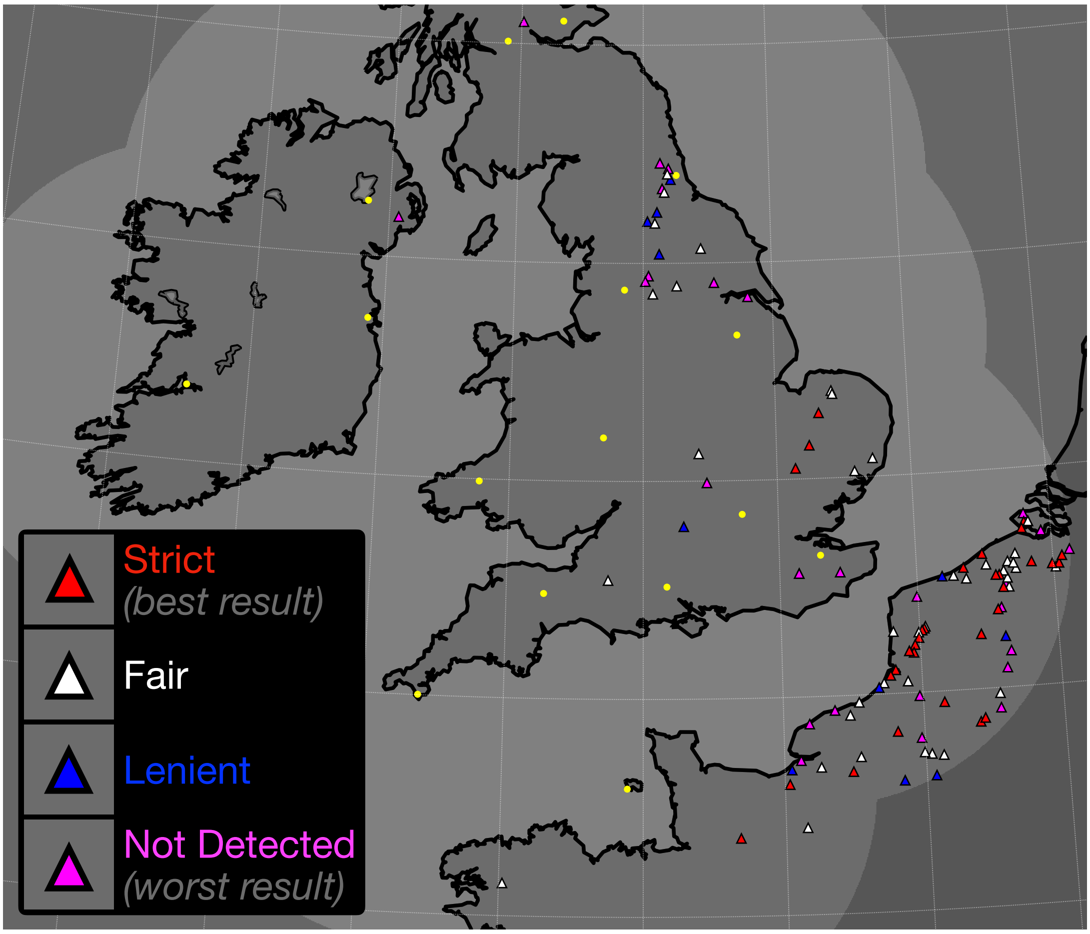


Figure 5.6.: A map of all of the ESWD hail reports within the UKMO-domain from 2014 to 2018 inclusive. The colours indicate which tolerance (if any) included a hail diagnosis by the SPT product, starting with the strict tolerance and increasing the spatiotemporal domain through fair and lenient tolerance. If the SPT product did not diagnose a hail pixel in the lenient tolerance then it is a complete miss event. The yellow circles represent the locations of the radar sites.

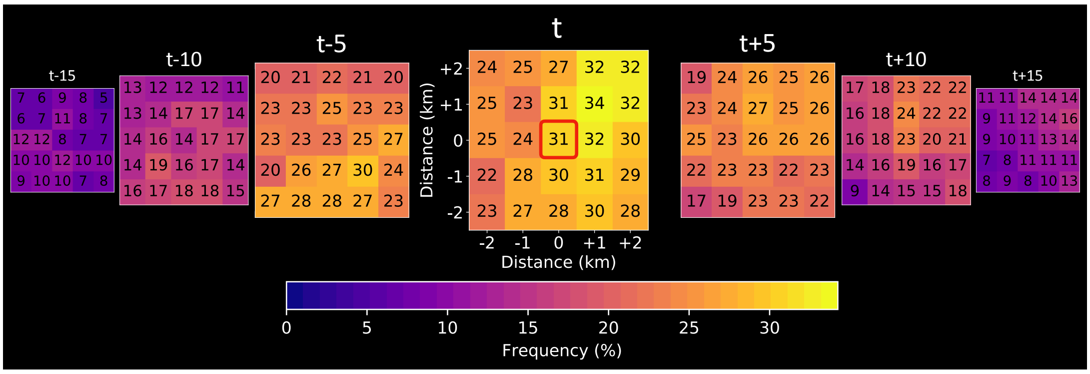


Figure 5.7.: Spatiotemporal map of hail hits surrounding an ESWD report at the ground. For each report of hail, the surrounding radar pixels in space and time from a Lagrangian perspective are captured. Shown is a sum of all the reports such that the total number of hits possible for each pixel is the number of the ESWD hail reports, or 111, meaning every hail report was associated with a hail detection from the radar at that time and space away from the hail report.

done for the pixel 1 km north, 5 min prior to the hail report, and so on, to produce a spatiotemporal spread map (Figure 5.7).

Every pixel in a 5 × 5 km region and ± 15 min around the report contains at least one radar-reported hail diagnosis (shown in Figure 5.7), possibly in-part due to the true size and motion of the hail-producing region of a convective feature. The temporal variation shows the general movement of hail-producing cells in the Met Office radar domain. At t-15 min to t-5 min, the hail pixels are concentrated to the south and southwest, whereas in the t+5 to t+15 minute period, hail pixels are more frequent to the northeast. The spatial variation indicates that the radar product has a slight northeasterly bias from the location of a hail report from the ESWD. The most frequently hit pixel is not directly over the report (30.6%) but 1 km NE of the report (34.2%). At first the results could be interpreted that a time delay is the cause of the spatial shift. However, note that if the product was shifted back in time by 5 min, then the time t+5 grid would become time t, which has lower hits across all pixels, so the timing of the product is at peak accuracy.

4.3. Maximum Hail Diameter

Since Figure 5.6 shows that the Met Office SPT product has higher POD over continental Europe, it was hypothesised that the hail sizes being larger on the continent than over the UK (Punge and Kunz 2016), may be the cause of the increase in skill. Most

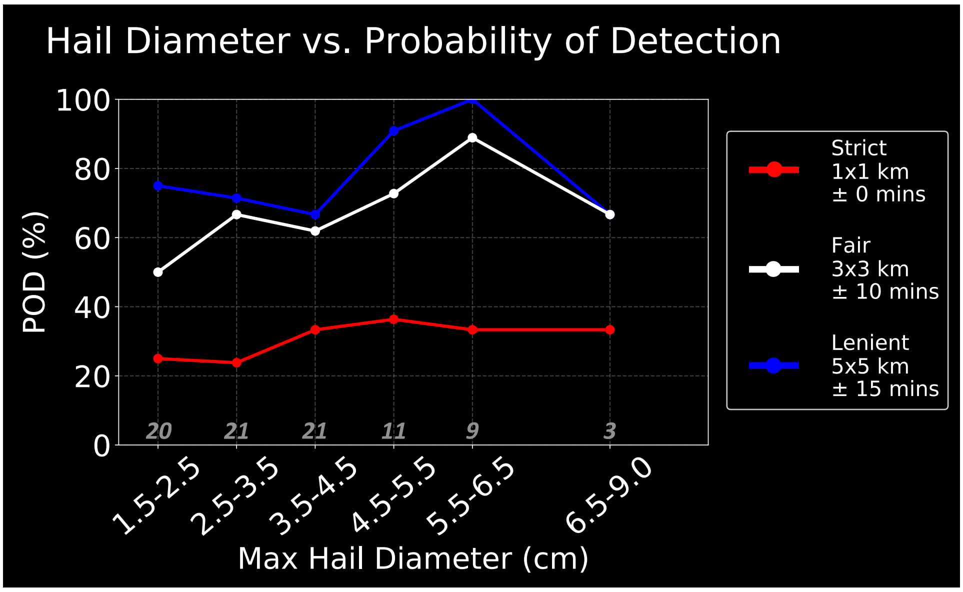


Figure 5.8.: Probability of detection (POD) for binned groups of hail reports of increasing maximum diameter, for each tolerance method. The grey values inside the lower region of the plot indicate the number of events in each bin. Bin-widths are all 1 cm with the exception of the final diameter bin which is larger (3.5 cm) to capture the remaining 3 events.

of the ESWD reports (85 out of 111) include an estimate of the maximum diameter of the hail being observed, so to check for a relationship between hail diameter and skill of the SPT product, POD values for each binned size range are calculated for each tolerance, and are displayed in Figure 5.8.

The strict tolerance has lower POD in smaller diameter hail events (approximately 25%) but the POD increases to 35% in larger hail diameter events. The lenient and fair tolerances also broadly increase with increasing hail diameter, with lower diameter hail having a POD around 50-75%, increasing up to 90-100% for 5.5-6.5 cm diameter hail (although there are only 9 events in this bin). The largest diameter hail bin has only 3 events and is therefore more likely to be affected by anomalies in the data due to a very small sample size. Also note that the radar reflectivity factor will not increase linearly with size for particles greater than ∼1–2 cm diameter since Mie scattering dominates over Rayleigh at this size for C-band (5.6 GHz) radars (Fabry 2015). The number of complete miss events is the area between the lenient line and 100%, which is around 25% for smaller diameter hail (1.5–2.5 cm) and drops to 0% for the 5.5-6.5 diameter hail bin.

5. Discussion

5.1. Hail

As shown, the diagnosis of hail by the SPT product is variable, and hail is a rare event in the UK with only 32 events reported to the ESWD in 5 years, 14 of which are on a single day. With a lenient tolerance the UK POD is 62.5% which is reasonable; for the strict tolerance the POD is 9.4%, which would suggest poor skill. However, there are large representativity issues with this product. Hail is an atmospheric process which can occur in very narrow swaths, the product resolution is only 1 km^2, and the measurement of high reflectivity aloft may not be co-located directly above the location where the hail reaches the surface. For hail, the strict tolerance shows that the SPT product has low skill at diagnosing hail with an accuracy of less than 1 km^2.

The main outcome is that there is lower skill over the UK compared with mainland Europe. One hypothesis is that the hail diameters are larger on the mainland but Section 4.3 shows that hail diameter does not correlate with sufficiently large changes in POD to explain the POD differences between the UK and mainland Europe. This result appears to be a combination of radar coverage overlap, scan geometry and the criterion used for hail detection. Radars close together increase the amount of overlap meaning one ground location may have several radars observing over it. Since the SPT algorithm for hail uses the modal output from overlapping radars, for hail to be detected the majority of radars covering a single point need to agree in the diagnosis of hail. Note that the Waldvogel et al. (1979) criterion relies on altitude; a reflectivity threshold of 45 dBZ or greater must be met ≥ 1.4 km above the 0 °C wet-bulb isotherm. The scan strategy of the Met Office radars is such that the radar beam does not observe higher than 4° elevation angle. These factors translate into a horizontal radial distance close to the radar (approximately 20 km) where the radar does not observe at the required altitude. Conversely, areas in mainland Europe are almost entirely being observed by a single radar; if it detects hail then there is no agreement needed between other radars. In addition, the increased distance from the radar in mainland Europe means that all radar elevation angles are at higher altitudes, making it more likely that the hail criterion will be met. This increased likelihood of hail detection extends around the entire edge of the UK radar domain, visible in the frequency maps in Figure 5.3.

The northeasterly spatiotemporal bias seen in Figure 5.7 may be related to several sources. Human observers are more likely to report the hail event with a time delay rather than before the event occurred. Also, by the time the hail has reached the ground from the altitude where the radar criterion is met, the storm may have proceeded NE and therefore the radar associates a ground report with the new location of the storm at a later time. Microphysically, the bias may be due to the tilted updraft of hail-producing storms. Dennis and Kumjian (2017) showed that an increase in east-west shear should produce larger hailstones, which would also tilt the updraft region meaning that the radar is observing the hail at a slightly eastward bias compared to where the hail reaches the ground. Dennis and Kumjian (2017) also note that north-south shear reduces the supply of hail embryos into the updraft. Allen et al. (2020) show with hail trajectory data from Dennis and Kumjian (2017) that the hail does not reach the surface directly below the area of highest radar reflectivity factor (Z) but instead reaches the ground to the southwest of the peak Z, which is consistent with the bias seen here in the Met Office SPT hail class.

The use of the criterion from Waldvogel et al. (1979) is somewhat questionable from the onset. Waldvogel et al. (1979) used a 3 cm, 10 GHz, X-band radar which begins to exhibit Mie scattering at smaller hydrometeor sizes than C-band does. The Met Office (and partners) radar network will begin to have decreasing and then non-linearly changing reflectivity for hailstones greater than ∼1–2 cm diameter. The testing for Waldvogel et al. (1979) also took place in Switzerland and used only 71 convective cells (33 hail and 38 rain) to derive the criterion used here. The POD achieved was 100% but this was at the expense of false alarms, which were high; 28 rain events met the hail criterion of greater than 45 dBZ $>$ 1.4 km above the 0 °C isotherm, representing a 39% false alarm ratio (FAR, see Part I). Finally, Waldvogel et al. (1979) discussed the geometric limitations of the criterion since it was used on range-height indicators (RHIs) and therefore all elevation angles were observed. For this reason, they state that "a real-time application of this method is therefore excluded at the present time", saying that this criterion should not be used on plan-position indicator (PPI) data such as the Met Office (and partners) radar network scan strategy. In fact, Waldvogel et al. (1979) suggest a different criterion for hail on PPI data but it uses 6 variables (hail probability from a complex table of parameters, maximum reflectivity Z and its height, height of the 45 dBZ reflectivity contour, height of the cloud top and height of the 0 °C isotherm) and is therefore more computationally expensive for an operational product which has to be published to users a set time after the initial observation.

The major limitation of this study is that it is not possible to say when the SPT product falsely diagnosed hail that did not occur. For example, some of the hail occurrences on the frequency map in Figure 5.3 appear to be artifacts. The POD could be low at the expense of a high FAR since (as described in Part I) these values are interdependent.

Reducing FAR may be desirable such that costly mitigative action is not taken too frequently. However, the frequency map in Figure 5.3 indicates that hail seems to occur significantly more frequently than the literature would suggest (Punge and Kunz 2016) and almost in every location in the UK over the 5-year study period. Anecdotally and from the ESWD reports, this seems unlikely to be true. The limitations of the ground verification datasets that exist at this current time do not allow the false alarm ratio to be verified, and the question therefore remains and will remain unanswered until ground-based datasets improve. Without improvements to ground-based hail observations, all precipitation types cannot be verified equitably.

5.2. Comparison to Other Products

Several studies test the Waldvogel et al. (1979) criterion or similar. Holleman et al. (2000) compared 5 radar-based hail detection criterion on two C-band radars in the Netherlands during summer 1999 against ground data from 19 synoptic sites, 321 volunteers, insurance filings and media reports. The Waldvogel et al. (1979) criterion achieved the highest critical skill index (CSI) of 0.44 with a POD of ∼65%. Tweaking the height above the 0 °C isotherm to 1.75 km instead of the original 1.4 km criterion achieved a slightly higher critical success index of 0.46 but lower POD of ∼60%. Skripniková and Řezáčová (2014) tested seven radar-based hail detection algorithms on well documented recent hail events from Czechia and southwest Germany from 2002 to 2011. For the Waldvogel et al. (1979) criterion, POD (FAR) varied from 83% (80%) down to 25% (45%) as Skripniková and Řezáčová (2014) adjusted the altitude threshold from 4 km to 8.5 km. Further tweaks to the Waldvogel et al. (1979) criterion produced a maximum CSI of 0.32 for 56 dBZ at 3.5 km above the 0 °C isotherm, where the POD reached ∼50%. Donavon and Jungbluth (2007) found a strong relationship between the 50-dBZ echo height and the height of the melting level, similar to Waldvogel et al. (1979), for hail of 19-25 mm in the contiguous United States. A criterion developed with linear regression of past observations was verified over three convective seasons. The criterion achieved a POD of 90%, a FAR of 22%, and a critical success index (CSI) of 0.72. Donavon and Jungbluth (2007) noted that performance of the criterion was surpassed by other indicators in strong-shear cases such as supercells, which typically produce the largest hailstone diameters.

Several other studies find conflicting results with other criterion. Blair et al. (2011) studied 568 giant-hail reports (greater than 102 mm in diameter) over 15 years (1995-2009) in the contiguous United States, compared with WSR-88D radar observations and reanalysis data. The verification determined that vertically integrated liquid, maximum storm reflectivity and reflectivity in the hail-growth region showed “little to no skill in discriminating between giant hail and smaller hail sizes.” Ortega (2018) used a database where residents were cold-called after a suspected hail-producing cell passed over a location, to verify 9 algorithms in the United States. The method of data collection enabled a dichotomous surface verification dataset unlike the storm reports method such as the ESWD used in the present study. The greatest Heidke skill score (HSS; Heidke (1926)) value of 0.42 was reached for a reflectivity threshold of 55 dBZ on the lowest-usable radar beam at a given location when used to detect the presence of hail of any diameter. Allen et al. (2020) comment on the empirical nature of all hail-detection studies: “empirical relationships developed in one part of the world may not be applicable globally or even regionally depending on the sample used in development.” Allen et al. (2020) also comment on the scan strategy limitations of hail detection methods, which is the case in this study with the Met Office SPT product.

For a fair (3 × 3 km, ± 10 min) tolerance which encapsulates the northeasterly spatiotemporal bias, the Met Office SPT product for UK-only events have a POD of 46.9%, which is lower than the aforementioned implementations of the Waldvogel et al. (1979) in other studies. The compositing method used, demanding the majority of radars to agree upon hail diagnosis, is likely to be the cause of the comparative underperformance. However, the false alarm ratio is unknown in the present study. Where the POD is high in other studies, the FAR is also high which is an undesirable product trait. Future verification work for hail detection in the UK should prioritise the ability for ground verification datasets to capture false diagnoses in order to balance FAR and POD with more complex skill metrics such as CSI, HSS and more.

5.3. Tolerance Method

The tolerance method used in Part I of this study is modified here but still retains a sense of realistic range to the presented POD values. It also demonstrates the sensitivity of the product’s skill when adjusting the spatiotemporal inclusion, which a user typically considers when viewing a graphical map. Adjacent pixels to a location would not be dismissed by a user viewing the product. More realistically, a user would associate a particular convective cell with a hail detection and note it as a hail-capable feature, rather than expecting kilometre-level accuracy. The broad spread of the POD scores for hail shows that choosing a single tolerance is not desirable since it may mislead the conclusion of the verification. In this instance, the spatiotemporal bias would not have been detected with a strict tolerance verification. As in Part I, if a user desires a higher POD then a larger sample should be considered from the SPT product. Vice versa if a lower FAR is desirable then a smaller sample should be taken around a desired location, which will depend on the specific SPT product user.

6. Summary and Further Work

In Part I of this study, the rain, mixed-phase and snow classes of the Met Office SPT product were verified against synoptic (automatic and manual; Green (2010)) and DiVeN (Pickering et al. 2019) datasets. The second part of this study verified the hail class against a crowdsourced database of reports from the EWSD. In both instances, a novel tolerance method is applied which shows the sensitivity of the verification to a larger area and time period based on radar scan geometry and atmospheric behaviours from the literature. Part I showed that the SPT product had some skill in the rain and snow classes, but the mixed-phase class was overdiagnosed and had low skill. Here in Part II, the skill of the hail class has been shown to be moderate with lenient tolerance, although large uncertainty remains over the false diagnoses due to limited ground verification data, nor can the results of Part I and Part II be directly compared.

The SPT product has 4 classes; by comparison, dual-polarisation radar hydrometeor classification algorithms or HCAs (Liu and Chandrasekar 2000; Straka et al. 2000; Park et al. 2009; Al-Sakka et al. 2013) often have many more. With increasing classes, more specificity is required by the algorithm to achieve the same level of skill. Adding more classes without the justified ability to do so will degrade the skill of a discrete, non-probabilistic product. Therefore a balance is required between what classes are beneficial to users of a product and what is reasonably achievable. The Met Office SPT product does not contain a criterion for freezing rain, so misclassifies a freezing level above the surface as either mixed-phase or rain. If two freezing levels exist from the Euro4 model then the highest one is used. This atmospheric setup is rare for the UK, but it did occur during the study period; on 1 March 2018 affecting southern and southwestern parts of England and Wales.

As was discussed in Section 1, graupel are small ice crystals rimed with supercooled water and typically form the embryos for hail. Graupel are not part of the SPT class but have useful properties for the signal of electrification of a storm (Carey and Rutledge 1996; Deierling et al. 2008). The WMO PW code system also does not contain a graupel entry, only hail, so the verification of graupel versus hail would be difficult with the datasets which exist today. DiVeN instruments or similar instruments which measure the diameter and fall velocity distributions of hydrometeors have the raw data to enable graupel to be distinguished from hail, but there are very few sites so UK coverage is poor (Pickering et al. 2019) and the verification of these instruments does not cover the graupel/hail delineation (Pickering et al. 2020). In general, a standardised table of precipitation types is needed across the atmospheric sciences, which reduces the ambiguity supplied in the PW code system and is also widely adopted by both instrument manufacturers and national weather services globally.

In dual-polarisation HCAs, the graupel class can exceed the 45 dBZ criterion used in Waldvogel et al. (1979). The UK Met Office has an operational HCA on the newly upgraded dual-polarisation radar network (Best and Harrison 2018), based on a modified version of Al-Sakka et al. (2013). An added graupel class is one of the modifications, supervised by a backscattering microphysical model described in Marzano et al. (2007) where graupel exists up to a reflectivity of 60 dBZ. However, because there is no altitude dependency on HCAs, many of the issues found in the SPT product in this study can be overcome. Note that the HCA classifies precipitation at the radar beam-height, and that a translational product is still required to deal with below-beam effects as precipitation falls to the surface, which is of particular importance for the rain, mixed-phase and snow classes. A newer version of the SPT product based on the outcomes of this verification and the dual-polarisation HCA is in operation at the Met Office. Future work will apply the same verification techniques used here onto the new product, in order to quantify the improvements.

References

Al-Sakka, H., A. A. Boumahmoud, B. Fradon, S. J. Frasier, and P. Tabary, 2013: A new fuzzy logic hydrometeor classification scheme applied to the french X-, C-, and S-band polarimetric radars. *Journal of Applied Meteorology and Climatology*, **52 (10)**, 2328–2344, doi:10.1175/JAMC-D-12-0236.1.

Allen, J. T., I. M. Giammanco, M. R. Kumjian, H. Jurgen Punge, Q. Zhang, P. Groenemeijer, M. Kunz, and K. Ortega, 2020: Understanding Hail in the Earth System. *Reviews of Geophysics*, **58 (1)**, 1–49, doi:10.1029/2019RG000665.

AMS, 2017: Hail. Glossary of meteorology. The American Meteorological Society.

Antonescu, B., D. M. Schultz, F. Lomas, and T. Kühne, 2016: Tornadoes in Europe: Synthesis of the observational datasets. *Monthly Weather Review*, **144 (7)**, 2445–2480, doi:10.1175/MWR-D-15-0298.1.

Best, S., and D. Harrison, 2018: A Hydrometeor Classification Algorithm for the UK C-Band Weather Radar Network. *10th European Conference on Radar in Meteorology & Hydrology*, ERAD, Observations R&D, Met Office, Fitzroy Road, Exeter, EX1 3PB, UK, 1–9.

Blair, S. F., D. R. Deroche, J. M. Boustead, J. W. Leighton, B. L. Barjenbruch, and W. P. Gargan, 2011: A Radar-Based Assessment of the Detectability of Giant Hail. *Electronic J. Severe Storms Meteor.*, **6 (7)**, 1–30.

Brown, T. M., W. H. Pogorzelski, and I. M. Giammanco, 2015: Evaluating hail damage using property insurance claims data. *Weather, Climate, and Society*, **7 (3)**, 197–210, doi:10.1175/WCAS-D-15-0011.1.

Carey, L. D., and S. A. Rutledge, 1996: A multiparameter radar case study of the microphysical and kinematic evolution of a lightning producing storm. *Meteorology and Atmospheric Physics*, **59 (1-2)**, 33–64, doi:10.1007/BF01032000.

Carey, L. D., S. A. Rutledge, and W. A. Petersen, 2003: The Relationship between Severe Storm Reports and Cloud-to-Ground Lightning Polarity in the Contiguous United States from 1989 to 1998. *Monthly Weather Review*, **131 (7)**, 1211–1228.

Clark, M. R., J. D. C. Webb, and P. J. Kirk, 2018: Fine-scale analysis of a severe hailstorm using crowd-sourced and conventional observations. *Meteorol. Appl.*,

492 (April), 472–492, doi:10.1002/met.1715.

Deierling, W., W. A. Petersen, J. Latham, S. Ellis, and H. J. Christian, 2008: The relationship between lightning activity and ice fluxes in thunderstorms. *Journal of Geophysical Research Atmospheres*, **113 (15)**, 1–20, doi:10.1029/2007JD009700.

Dennis, E. J., and M. R. Kumjian, 2017: The impact of vertical wind shear on hail growth in simulated supercells. *Journal of the Atmospheric Sciences*, **74 (3)**, 641–663, doi:10.1175/JAS-D-16-0066.1.

Donavon, R. A., and K. A. Jungbluth, 2007: Evaluation of a technique for radar identification of large hail across the Upper Midwest and Central Plains of the United States. *Weather and Forecasting*, **22 (2)**, 244–254, doi:10.1175/WAF1008.1.

Dotzek, N., P. Groenemeijer, B. Feuerstein, and A. M. Holzer, 2009: Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD. *Atmospheric Research*, **93 (1-3)**, 575–586, doi:10.1016/j.atmosres.2008.10.020, URL http://dx.doi.org/10.1016/j.atmosres.2008.10.020.

Eccel, E., P. Cau, K. Riemann-Campe, and F. Biasioli, 2012: Quantitative hail monitoring in an alpine area: 35-year climatology and links with atmospheric variables. *International Journal of Climatology*, **32 (4)**, 503–517, doi:10.1002/joc.2291.

Elmore, K. L., Z. L. Flamig, V. Lakshmanan, B. T. Kaney, V. Farmer, H. D. Reeves, and L. P. Rothfusz, 2014: MPING: Crowd-sourcing weather reports for research. *Bulletin of the American Meteorological Society*, **95 (9)**, 1335–1342, doi:10.1175/BAMS-D-13-00014.1.

Fabry, F., 2015: *Radar Meteorology: Principles and Practice*. Cambridge University Press, 254 pp., doi:10.1017/CBO9781107707405.

Federer, B., A. Waldvogel, W. Schmid, H. H. Schiesser, F. Hampel, M. Schweingruber, W. Stahel, and J. Bader, 1986: Main Results of Grossversuch IV. *Journal of Climate & Applied Meteorology*, **25 (July)**.

Frisby, E. M., 1961: Relationship of Ground Hail Damage Patterns to Features of the Synoptic Map in the Upper Great Plains of the United States. *Journal of Applied Meteorology*, **1 (3)**, 348–352.

Gatzen, C., T. Púčik, and D. Ryva, 2011: Two cold-season derechoes in Europe. *Atmospheric Research*, **100 (4)**, 740–748, doi:10.1016/j.atmosres.2010.11.015.

Giammanco, I. M., B. R. Maiden, H. E. Estes, and T. M. Brown-Giammanco, 2017: Using 3D laser scanning technology to create digital models of hailstones. *Bulletin of the American Meteorological Society*, **98 (7)**, 1341–1347, doi:10.1175/BAMS-D-15-00314.1.

Green, A., 2010: From Observations to Forecasts – Part 7 . A new meteorological monitoring system for the United Kingdom's Met Office. *Weather*, **65 (10)**, 272–277.

Groenemeijer, P., and Z. Liang, 2019: ESWD data format specification. Tech. rep., European Severe Storms Laboratory e.V., Münchener Straße 20, 82234 Wessling, Germany, 1–65 pp.

Groenemeijer, P., and Coauthors, 2017: Severe convective storms in Europe: Ten years of research and education at the European Severe Storms Laboratory. *Bulletin of the American Meteorological Society*, **98 (12)**, 2641–2651, doi:10.1175/BAMS-D-16-0067.1.

Guntur, P., and M. Tippett, 2017: Impact of ENSO on U.S. Tornado and Hail frequencies. Tech. Rep. March, Willis Re, Minneapolis, MN, USA. doi:10.1175/JAMC-D-16-0249.1.

Heidke, P., 1926: Berechnung Des Erfolges Und Der Güte Der Windstärkevorhersagen Im Sturmwarnungsdienst. *Geografiska Annaler*, **8 (4)**, 301–349, doi:10.1080/20014422.1926.11881138, URL https://doi.org/10.1080/20014422.1926.11881138.

Heymsfield, A. J., I. M. Giammanco, and R. Wright, 2014: Terminal velocities and kinetic energies of natural hailstones. *Geophysical Research Letters*, **41 (23)**, 8666–8672, doi:10.1002/2014GL062324.

Hohl, R., H. H. Schiesser, and D. Aller, 2002: Hailfall: The relationship between radar-derived hail kinetic energy and hail damage to buildings. *Atmospheric Research*, **63 (3-4)**, 177–207, doi:10.1016/S0169-8095(02)00059-5.

Holleman, I., H. R. A. Wessels, J. R. A. Onvlee, and S. J. M. Barlag, 2000: Development of a hail-detection-product. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, **25 (10-12)**, 1293–1297.

Hov, Ø., and Coauthors, 2013: Extreme Weather Events in Europe: preparing for climate change adaptation October. Tech. Rep. October, Norwegian Meteorological

Institute, Oslo, 140 pp.

Johnson, A., and K. Sugden, 2014: Evaluation of Sounding-Derived Thermodynamic and Wind-Related Parameters Associated with Large Hail Events. *E-Journal of Severe Storms Meteorology*, **9 (5)**, 1–42, URL http://www.ejssm.org/ojs/index.php/ejssm/article/viewArticle/137.

Kaltenböck, R., G. Diendorfer, and N. Dotzek, 2009: Evaluation of thunderstorm indices from ECMWF analyses, lightning data and severe storm reports. *Atmospheric Research*, **93 (1-3)**, 381–396, doi:10.1016/j.atmosres.2008.11.005, URL http://dx.doi.org/10.1016/j.atmosres.2008.11.005.

Kaspar, M., M. Müller, V. Kakos, D. Rezacova, and Z. Sokol, 2009: Severe storm in Bavaria, the Czech Republic and Poland on 12-13 July 1984: A statistic- and model-based analysis. *Atmospheric Research*, **93 (1-3)**, 99–110, doi:10.1016/j.atmosres.2008.10.004, URL http://dx.doi.org/10.1016/j.atmosres.2008.10.004.

Keegan, M. H., D. H. Nash, and M. M. Stack, 2013: On erosion issues associated with the leading edge of wind turbine blades. *Journal of Physics D: Applied Physics*, **46 (38)**, doi:10.1088/0022-3727/46/38/383001.

Knight, C. A., and N. C. Knight, 1970a: Hailstone Embryos. *Journal of the Atmospheric Sciences*, **27 (4)**, 659–666, doi:10.1175/1520-0469(1970)027⟨0659:HE⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0469(1970)027%3C0659:HE%3E2.0.CO;2.

Knight, C. A., and N. C. Knight, 1970b: Lobe structures of hailstones. *Journal of the Atmospheric Sciences*, **27 (4)**, 667–671.

Liu, H., and V. Chandrasekar, 2000: Classification of hydrometeors based on polarimetric radar measurements: Development of fuzzy logic and neuro-fuzzy systems, and in situ verification. *Journal of Atmospheric and Oceanic Technology*, **17 (2)**, 140–164, doi:10.1175/1520-0426(2000)017⟨0140:COHBOP⟩2.0.CO;2.

Locatelli, J. D., and P. V. Hobbs, 1974: Fall speeds and masses of solid precipitation particles. *Journal of Geophysical Research*, **79 (15)**, 2185–2197, doi:10.1029/JC079i015p02185.

Manzato, A., 2005: The use of sounding-derived indices for a neural network short-term thunderstorm forecast. *Weather and Forecasting*, **20 (6)**, 896–917, doi:10.1175/WAF898.1.

Marzano, F. S., D. Scaranari, and G. Vulpiani, 2007: Supervised Fuzzy-Logic Classification of Hydrometeors Using C-Band Weather Radars. *IEEE Transactions on Geoscience and Remote Sensing*, **45 (11)**, 3784–3799.

Michaelides, S., and Coauthors, 2018: Reviews and perspectives of high impact atmospheric processes in the Mediterranean. *Atmospheric Research*, **208 (July 2017)**, 4–44, doi:10.1016/j.atmosres.2017.11.022, URL https://doi.org/10.1016/j.atmosres.2017.11.022.

Mohr, S., M. Kunz, and B. Geyer, 2015: Hail potential in Europe based on a regional climate model hindcast. *Geophysical Research Letters*, **42 (24)**, 10 904–10 912, doi:10.1002/2015GL067118.

Nelson, S. P., 1983: The Influence of Storm Flow Structure on Hail Growth. *Journal of the Atmospheric Sciences*, **40 (8)**, 1965–1983, doi:10.1175/1520-0469(1983)040⟨1965:TIOSFS⟩2.0.CO;2, URL https://doi.org/10.1175/1520-0469(1983)040%3C1965:TIOSFS%3E2.0.CO;2.

Ortega, K. L., 2018: Evaluating multi-radar, multi-sensor products for surface hailfall diagnosis. Evaluating Multi-Radar, Multi-Sensor Products for Surface Hail-Fall Diagnosis. *Electronic J. Severe Storms Meteor*, **13 (1)**, 1–36, URL http://www.ejssm.org/ojs/index.php/ejssm/article/viewFile/163/113.

Park, H. S., A. V. Ryzhkov, D. S. Zrnić, and K.-E. Kim, 2009: The Hydrometeor Classification Algorithm for the Polarimetric WSR-88D: Description and Application to an MCS. *Weather and Forecasting*, **24 (3)**, 730–748, doi:10.1175/2008WAF2222205.1.

Pickering, B. S., R. R. Neely, and D. Harrison, 2019: The Disdrometer Verification Network (DiVeN): A UK network of laser precipitation instruments. *Atmospheric Measurement Techniques*, **12 (11)**, 5845–5861, doi:10.5194/amt-12-5845-2019.

Pickering, B. S., R. R. Neely III, J. Jeffery, D. Dufton, and M. Lukach, 2020: Intercomparison of multiple precipitation sensor designs for precipitation rate and depth, drop size and velocity distribution, and precipitation type. *Journal of Hydrometeorology*, **In review**.

Pruppacher, H. R., and J. D. Klett, 1997: *Microphysics of Clouds and Precipitation*. 2nd ed., Reidel, Dordrecht.

Púčik, T., and Coauthors, 2017: Future changes in European severe convection environments in a regional climate model ensemble. *Journal of Climate*, **30 (17)**, 6771–6794, doi:10.1175/JCLI-D-16-0777.1.

Punge, H. J., K. M. Bedka, M. Kunz, and A. Werner, 2014: A new physically based stochastic event catalog for hail in Europe. *Natural Hazards*, **73 (3)**, 1625–1645, doi:10.1007/s11069-014-1161-0.

Punge, H. J., and M. Kunz, 2016: Hail observations and hailstorm characteristics in Europe: A review. *Atmospheric Research*, **176-177**, 159–184, doi:10.1016/j.atmosres.2016.02.012, URL http://dx.doi.org/10.1016/j.atmosres.2016.02.012.

Rasmussen, E. N., and J. M. Straka, 1998: Variations in supercell morphology. Part I: Observations of the role of upper-level storm-relative flow. *Monthly Weather Review*, **126 (9)**, 2406–2421, doi:10.1175/1520-0493(1998)126⟨2406:VISMPI⟩2.0.CO;2.

Rasmussen, R. M., and A. J. Heymsfield, 1987: Melting and Shedding of Graupel and Hail. Part III: Investigation of the Role of Shed Drops as Hail Embryos in the 1 August CCOPE Severe Storm. *Journal of the Atmospheric Sciences*, **44 (19)**, 2783–2803, doi:10.1175/1520-0469(1987)044⟨2783:MASOGA⟩2.0.CO;2.

Sanderson, M. G., W. H. Hand, P. Groenemeijer, P. M. Boorman, J. D. Webb, and L. J. Mccoll, 2015: Projected changes in hailstorms during the 21st century over the UK. *International Journal of Climatology*, **35 (1)**, 15–24, doi:10.1002/joc.3958.

Schemm, S., L. Nisi, A. Martinov, D. Leuenberger, and O. Martius, 2016: On the link between cold fronts and hail in Switzerland. *Atmospheric Science Letters*, **17 (5)**, 315–325, doi:10.1002/asl.660.

Skripniková, K., and D. Řezáčová, 2014: Radar-based hail detection. *Atmospheric Research*, **144**, 175–185, doi:10.1016/j.atmosres.2013.06.002.

Straka, J. M., D. S. Zrnić, and A. V. Ryzhkov, 2000: Bulk Hydrometeor Classification and Quantification Using Polarimetric Radar Data: Synthesis of Relations. *Journal of Applied Meteorology*, **39 (8)**, 1341–1372, doi:10.1175/1520-0450(2000)039⟨1341:BHCAQU⟩2.0.CO;2, URL http://dx.doi.org/10.1175/1520-0450(2000)039%3C1341:BHCAQU%3E2.0.CO%5Cn2.

Sulakvelidze, G. K., 1969: Rainstorms and hail. *Israel Program for Scientific Translations*, 310.

Waldvogel, A., B. Federer, and P. Grimm, 1979: Criteria for detection of hail cells.

Journal of Applied Meteorology, **18 (12)**, 1521–1525, doi:10.1175/1520-0450(1979)018⟨1521:CFTDOH⟩2.0.CO;2.

Webb, J. D., D. M. Elsom, and G. T. Meaden, 2009: Severe hailstorms in Britain and Ireland, a climatological survey and hazard assessment. *Atmospheric Research*, **93 (1-3)**, 587–606, doi:10.1016/j.atmosres.2008.10.034, URL http://dx.doi.org/10.1016/j.atmosres.2008.10.034.

Webb, J. D., D. M. Elsom, and D. J. Reynolds, 2001: Climatology of severe hailstorms in Great Britain. *Atmospheric Research*, **56 (1-4)**, 291–308, doi:10.1016/S0169-8095(00)00081-8.

WMO, 1988: Manual on Codes. *WMO Publ. 306*, **1**, 203 pp.

WMO, 2017: WMO Guide to Meteorological Instruments and Methods of Observation (the CIMO Guide). Chapter 6. Masurement of Precipitation. Tech. rep., World Meteorological Organization.

Chapter 6.

Conclusions

1. Summary of the Achieved Research

The type of precipitation reaching the surface (SPT) has a significant impact on society and can often be more impactful than the amount of precipitation reaching the surface. In the United Kingdom (UK), uncommon precipitation types are often the cause of significant disruption to daily activities (Thornes 1992; Agnew and Thornes 1995; Webb et al. 2001; Clark et al. 2018). Despite the high impact of rare surface precipitation types (SPTs), they are both poorly observed (Schmid and Mathis 2004; Chen et al. 2016, and more) and are poorly forecasted (Elmore et al. 2015; Gascón et al. 2018, and more), in part due to the complex and unstable microphysical processes which determine the SPT, as explained in Chapter 1. To improve the skill of SPT data, a framework of suitable assessment techniques must be established. The assessment of SPT data is challenging for many reasons. Firstly, the frequency of SPTs is highly skewed in places like the UK, where rain is by far the most common type but mixed-phase, snow, freezing rain, ice pellets, graupel and hail all occur orders of magnitude less frequently while being more impactful to society (Carrière et al. 2000; Fairman et al. 2015). Capturing a statistically significant and climatologically representative sample of SPT data is difficult due to the rarity of some SPTs. Secondly, ambiguous reporting standards such as the descriptive present weather (PW) code tables (WMO 1988, 2017) are used widely for recording both manual and automatic SPT observations at a single point location. Slight differences exist in the PW code descriptions between the manual stations and the automatic stations, and many of the PW codes mention several SPTs or are listed alongside other atmospheric phenomena which dilutes the specificity of the reference dataset. Finally, the verification of spatial SPT products against point SPT observations, such as those derived from radar observations or from numerical weather prediction (NWP), requires compensation of representativity errors which exist between the two datasets. All of these issues can result in obfuscated assessments which is evident from the wide variety of verification metrics used in the literature. The inconsistency of applied assessment techniques obscures the comparison of different SPT diagnosis techniques, thereby suppressing the advancement of the most accurate methods into widespread operational use.

A framework of assessment techniques was developed throughout this thesis, with a focus on three areas. First, new automated instruments were deployed in the UK which had the potential to provide more SPT observations and with higher accuracy. Second, these instruments were evaluated against other precipitation-measuring instruments with novel techniques to determine their accuracy, including the development of a new SPT classification standard. Third, the new instruments were used alongside existing observational datasets to verify a radar-based spatial SPT product in the UK over 5 years with a tolerance approach to address representativity errors. The methodologies applied successfully throughout the thesis can be used to more effectively assess the skill of SPT data in meteorology.

The following sections summarise the research that has been completed, connecting the research to the objectives from Section 5 of Chapter 1. The limitations of the research and then recommendations for future research are presented, ending with some final remarks.

1.1. Point SPT Instrument Deployment

The end goal of the thesis was to perform a verification of a spatial SPT product against surface point SPT observations (objective 3). To begin with, a new dataset was needed to supplement the existing surface point SPT observations that were being collected by the Met Office. In particular, the existing automated stations were known for measuring the mixed-phase SPT class with poor skill (Lyth and Molyneux 2006). While some manual stations were available (16) which had trained observers and reported SPTs with high confidence (daytime only and through PW codes), the majority of UK surface SPT observations were automatic (160). Furthermore, all of the Met Office surface station SPT reports were recorded hourly, while the spatial SPT product diagnosed SPT at 5-minute intervals, meaning that verification could only take place for every twelfth opportunity, reducing the verification sample size. In Chapter 2 a new reference point dataset was collected which had the potential to identify SPT with greater confidence, in more extreme environments and at higher temporal resolution than those currently available from the UK Met Office surface network (Green 2010).

Chapter 2 describes the low-cost installation of a network of Thies LPMs around the UK called the Disdrometer Verification Network (DiVeN) (Pickering et al. 2019), contributing to the objective (1a) of this thesis to increase the available surface point SPT reference data. 14 instruments were taken on loan from the UK Met Office and were installed in 13 locations (the Chilbolton site had two instruments placed two metres apart). The DiVeN locations were chosen to represent the breadth of known radar sampling geographies and precipitation climatologies in the UK (Harrison et al. 2000; Fairman et al. 2015), with consideration to the intended aim of performing verification on a radar-based spatial SPT dataset. Some DiVeN sites were close to Met Office radar dishes and in flat terrain so that the lowest usable radar beam was at low altitude and at high polar resolution, thereby providing high quality data. Alternatively, some DiVeN sites were far from the radar sites in mountainous areas so that the lowest usable radar beam was at high altitude and at coarse polar resolution, thereby providing low quality data. The choice of DiVeN site locations attempted to reduce any bias resulting from only performing the verification in high quality radar locations, a criticism of the existing Met Office surface stations which were located in non-remote and non-extreme terrain locations, typically well-observed by radars (Green 2010).

The datalogging system was engineered in-house using commercial off-the-shelf (COTS) components and the data were published to a website in near-real time (see Appendix A.1), satisfying objective 1b of this thesis. Per-minute plots (see Appendix A.2) appeared with at most a 7-minute delay from the initial observation, and daily plots (see Appendix A.3) appeared at 01 UTC. Both types of plot were used primarily to maintain the network and spot artefacts in the data but they also found wider use. For example, the Cairngorm Ski Centre used daily plots to determine when the snow level had surpassed the instrument and to determine conditions at the site mid-point before leaving the base station (Jim Cornfoot, personal communication, 2018). The website was available to Met Office forecasters to use in daily operations to determine whether the type of information would have value beyond the DiVeN project. The feedback was that too much information was presented at high temporal frequency which slowed the loading times of the website, and that spatial data would be preferred to point observations (Dawn Harrison, personal communication, 2017). This user feedback further motivated the use of these instruments to perform verification and to improve a spatial product rather than being useful standalone. While the instruments and the datalogging systems were robust and performed well, the uploading of data to the server was sporadic from many sites, due to the need for cellular data renewal and some power losses. An extended version of the upload performance plot (Figure 2.5) from Chapter 2 is shown in Appendix A.4.

In addition to the real-time website, to increase the impact of the research, DiVeN data were uploaded to an online repository: the Centre for Environmental Data Analysis (CEDA) Archive (NERC et al. 2019, see Appendix A.5). The data exist on the CEDA Archive in a citable, CF-compliant netCDF format which makes it an accessible standard to the environmental research community. One of the included variables is the new SPT classification standard which was proposed in Chapter 3, to encourage its wider adoption. The outcome is that DiVeN data have been used on numerous published (e.g. Lin et al. 2020) and in-preparation research projects from the University of Leeds, two commercial studies on wind turbine blade erosion by Vestas and WeatherQuest, and is also referenced in a Danish study on wind turbine blade erosion (Tilg et al. 2020). In the first 11 months the CEDA Archive entry for DiVeN has been accessed by 19 users from 6 countries (as of 1 October 2020). By these metrics, making the observational dataset available to the wider research community (objective 1c) was successfully achieved.

Note that the case studies in Chapter 2, as well as two other occurrences of graupel being diagnosed as snow (noted by trained observers viewing the live website), led to the DiVeN dataset being unusable for observations of hail in Chapter 5. In total there were nearly 2,000 minutes of hail PW code reports across 13 sites which is significantly higher than the expected climatology shown by Punge and Kunz (2016), reaffirming that graupel and hail are not distinguished by the PW code standard. While the research described in Chapter 2 validated that the DiVeN instruments were functional and verified that SPT reports had some merit, the case study results were qualitative and used only a few samples. Hence, an evaluation of the DiVeN Thies LPM instruments against existing precipitation instruments was required, in order to determine their overall performance, which was conducted in Chapter 3.

1.2. Point SPT Instrument Evaluation

Before the new DiVeN Thies LPMs could be used as a reference dataset for verification, the quality of the measurements needed to be assessed. Two DiVeN instruments were co-located 2 metres apart at the Chilbolton Atmospheric Observatory (CAO) for the purposes of a consistency comparison between the two instruments, as well as an evaluation against the other precipitation sensors located at CAO. Chapter 3 describes the precipitation sensor evaluation study, which utilised a 12-month period starting from the initial Thies LPM installations in February 2017. Precipitation depth and rate, the drop size and drop velocity distributions, and SPT were compared at several timescales. Although this thesis is focused on measurements of SPT, the analysis of other variables, in particular the ability to accurately capture the drop size and velocity distributions, has implications for the diagnosis of SPT. Six measurement techniques were included: tipping-bucket (TBR), drop-counting (RAL), acoustic (JWD), optical (Thies), single-angle visiometer with capacitor (PWD21) and dual-angle visiometer (PWS100). Not all of the measurement techniques were able to record every variable under examination and only the PWD21, PWS100 and the two Thies LPMs were able to directly infer SPT.

Firstly, the Thies LPMs were compared with the other instruments for precipitation rate and depth, drop size and drop velocity distribution (objective 2a). There were large differences in the measured precipitation depth and rate between the instruments (25%) with both Thies LPMs recording more precipitation than the average of all instruments. However, these results do not affect the Thies LPM for SPT diagnosis as the drop size distribution (DSD) results show that the high precipitation totals are due to a high drop count rather than the drop volumes inferred from diameter measurements, which is in agreement with the literature (Lanza and Vuerich 2009; Frasson et al. 2011). Long-term qualitative examination of the diameter and velocity output combined (made possible with the live website) revealed that the velocity of the small drops varied significantly and often had unrealistically high values when compared with the literature (Gunn and Kinzer 1949). The anomaly appeared to be associated with high wind speeds, although this was not examined directly by the research in Chapter 3. The interpretation of the anomaly is that SPT diagnosis will be highly uncertain for low precipitation rates where small hydrometeors tend to be dominant, especially in high winds. This is unlikely to affect the verification in Chapter 4, since both the radar-based pixel and the surface point must detect precipitation to be included and radars are generally insensitive to very low precipitation rates. For the larger drop sizes in the evaluation in Chapter 3, the velocities and the velocity distribution were consistent with the literature, whereas the only other velocity measuring instrument in the study (the PWS100) measured a much slower velocity distribution compared to the literature. However, the PWS100 was located on the roof of a small building which may cause turbulence and reduce fall velocities; the results are inconclusive. These results partially answer how the Thies LPM compares to existing instruments (objective 2a), but a more rigorous reference instrument for particle size and velocity would have been more conclusive. For example, the World Meteorological Organization (WMO) intercomparison study (liquid precipitation only) of Lanza and Vuerich (2009), used a composite working reference of four laboratory calibrated gauges whereas the research in Chapter 3 used an average of all instruments as the reference value for precipitation depth and rate.

Another objective of this thesis was to assess the impact that timescale had on the measurements (2c), which the results show to be more significant to certain instruments and variables. The most notable was the TBR for precipitation rate which is due to the small resolution of a single increment of rainfall (0.2 mm) and is a widely known issue. At time periods of 15 minutes or longer, the TBR had similar coefficients of determination with the other instruments in the study, a result which could be used to inform future studies, along with the extensive body of research on TBR gauges which exists in the literature. After 1-hour periods the coefficients of determination did not significantly increase for any of the instruments, likely related to the typical duration of a precipitation event being less than one hour. For DSDs and drop velocity distribution (DVDs), it was found that a 5-minute sample was sufficient to represent the distributions at all precipitation rates PR $\geq$ 0.3 mm h^{-1}; a longer sample duration did not change the shape of the distributions nor did it change the differences between each instrument. These results support the use of the Thies LPM for diagnosis of SPT on 5-minute timescales since it relies on the diameter-velocity relationship for the diagnosis.

Comparing SPT measurements (objective 2b) required a new approach because the existing standard for point SPT data (the PW code) increases the uncertainty in the measurements. All of the instruments recorded SPT in the PW code format, in which some codes are ambiguous and reference multiple precipitation types as well as different intensities which are not of concern to the objective of SPT assessment. A new standard set of SPT classes were created which is based on the microphysical processes leading to different SPTs (Chapter 1, Section 2) and also utilised many of the categories used in dual-polarisation radar hydrometeor classification algorithms (Liu and Chandrasekar 2000; Park et al. 2009; Al-Sakka et al. 2013; Best and Harrison 2018), making the standard compatible for future assessments. A conversion table between PW codes and the new precipitation type (PT) index was also presented—for the unambiguous PW codes. The ambiguous PW codes containing many different SPTs or conditionals (‘and’/‘or’) could not be translated and were therefore excluded from the evaluation dataset. Since PW codes do not distinguish between graupel and hail (WMO 1988, 2017), graupel cannot be translated or appear in the confusion matrix results of the instrument comparison (Figure 3.10 of Chapter 3), although it is part of the proposed SPT classification standard (Table 3.3 of Chapter 3). PW codes were recorded across different time intervals, covering 1, 5, 10, 15 and 60 minutes depending on the instrument, however all instruments recorded the 1-minute interval so this was chosen as the baseline. The method of producing multi-minute amalgamations of PW code also differed between the instruments and often led to unintuitive SPT outputs not suitable for high impact SPT events. A new Boolean algorithm was designed to amalgamate multiple 1-minute SPT observations together. When applied in Chapter 3, the amalgamation algorithm performed well and did not introduce a time-dependent bias. The development of an SPT classification standard and an amalgamation algorithm contribute towards a more suitable method for evaluating SPT observations (objective 2b).

The instruments were then compared across the 12-month period for their SPT diagnoses. The PWD21 and PWS100 never reported several of the SPT classes which the Thies LPMs reported, which resulted in the agreement between the PWD21 and PWS100 being higher than each of the instruments compared with any of the Thies LPMs. The mixed-phase (denoted by the class ‘wet snow’) was the least agreed upon by the instruments. The number of cases was limited and without a trained meteorological observer as the reference ‘truth’ the results are inconclusive. The results could therefore not satisfactorily answer whether the Thies LPMs were comparable to existing sensors for SPT observations (objective 2d). However, verification studies in the literature have compared the Thies instrument to a trained observer albeit for short periods of time, which provides some evidence of the skill of the Thies LPM for SPT observations. The prior study results do not clarify whether the Thies LPM used in DiVeN will produce more skilful diagnoses of mixed-phase precipitation compared with the automated arbiter of instruments used at Met Office surface stations. Therefore, in Chapter 4 it was decided to present all three of the ground reference datasets separately, since the skill of each was not definitively quantified by the earlier research.

1.3. Spatial SPT Verification—Rain, Mixed-phase and Snow

Chapter 4 discusses the verification of a spatial SPT product from the Met Office. The output SPT classes are rain, mixed-phase, snow and hail. Hail is diagnosed separately and also requires separate verification data and is therefore separated into Chapter 5. The delineation of rain, mixed-phase and snow is calculated using Lumb’s critical rate which accounts for evaporative cooling effects with an assumed lapse rate (Lumb 1963).

The wider issues to address in Chapters 4 and 5 were the choice of verification metrics and how to overcome the representativity issue of a spatial SPT product being verified against point SPT observations (objective 3a).

The choice of verification metrics (Bias, probability of detection (POD), false alarm ratio (FAR), Heidke Skill Score (HSS)) was motivated by the information provided by each metric which could inform both the users of the product and the development of a future product. Bias indicates the tendency of the product to under- or over-diagnose a specific SPT class and shows which direction borders between classes need to be adjusted. The POD indicates what chance the product has of detecting a certain SPT event, whereas the FAR indicates the probability that a diagnosed SPT event is false. The HSS was chosen because it indicates the percentage of improvement between random chance and a perfect diagnosis (matching the reference dataset), so the HSS considers the climatological rarity of the SPT class. Importantly, the HSS can be applied to both a single SPT class on a contingency table and also a multi-dimensional confusion matrix of the whole product, which allows for the comparison of SPT classes, or the contribution of a specific SPT to the overall product skill score value. Chapter 4 discusses both of these methods for the Met Office radar-based spatial SPT product.

A novel method was developed to perform verification of spatial products against point reference data, addressing the representativity errors between the datasets (objective 3a). Using realistic values of the maximum horizontal advection and fall velocity from the lowest-usable radar beam, a range of spatiotemporal tolerances of the spatial SPT data were verified against the point reference data. Three tolerances were used: 1 × 1 km ± 0 mins, 3 × 3 km ± 10 mins, and 5 × 5 km ± 15 mins. The result is a verification score range which can be considered representative of the realistic best and worst case verification result. The score range is more informative and understandable to an end user than a single verification score which can be improved or deteriorated by the verification parameters. The developed method was applied in Chapters 4 and 5 and demonstrated the benefit added to the interpretation of the product skill compared to a single tolerance approach, thus addressing objective 3a.

Using only the strict tolerance on the spatial SPT data, the overall HSS for all SPT classes combined was 0.61, which can be interpreted as a 61% improvement over a random chance diagnosis, where 100% would be an exact match with the reference dataset. When each reference dataset was considered individually, the overall HSS of the spatial SPT product was 0.48 for automatic, 0.60 for manual, and 0.73 respectively. The verification was further regimented with a 2 × 2 contingency table for each precipitation class, where any correct diagnosis in the wider tolerance was considered a hit. These results showed that the mixed-phase is overdiagnosed with the lowest HSS for any precipitation class. Together these results quantify the skill of an existing spatial SPT product (objective 3b) and delineates the skill between the SPT classes (objective 3c) and between the reference datasets (objective 3e).

How the skill of the spatial SPT product varies at different locations (objective 3d) was not possible to verify quantitatively because the verification scores were highly dependent upon the climatology of the location. For example, in mountainous regions there are more ‘obvious’ snow cases, so the spatial SPT product has more hit events and thus has a higher skill score value, which is not realistic or useful to an end user. This meant that the effect of high or low quality radar data was not discernible. A similar outcome occurred for seasonal changes in skill, where in summer months the skill of the snow precipitation class was undefinable, since no events occurred or were diagnosed. For a qualitative validation of the product, the normalised frequency of each precipitation type was plotted on a map of the UK for the 5-year study period. Many artefacts were found with radar beam blockage, wind turbines, ground and sea clutter, and more complex artefacts caused by seasonal changes in the meteorological features when observed by radar. These results highlight where the skill is qualitatively very poor due to artefacts and allows the Met Office to overcome these artefacts in the future for a more accurate product.

The last objective of the thesis (3f) was to suggest changes to the product which could improve its skill. Chapter 4 showed that the product skill, defined as the multi-class HSS, could be improved by adjusting the mixed-phase class. If the mixed-phase class is removed, the overall HSS improved from 0.61 to 0.77 but this is unhelpful since the new class would be ‘undefined’. If mixed-phase diagnoses are changed to another class, the overall HSS improves from 0.61 to 0.73 for rain re-classification, and deteriorates from 0.61 to 0.59 for snow re-classification. This is consistent with the other results indicating that the mixed-phase class has very little skill, and is over-diagnosed at the expense of rain. Also, since a bias was identified in the 0 °C wet-bulb height from NWP when compared to the radar-diagnosed freezing level height, a change is expected to improve the skill of the product. However, this test was not performed due to the resources required to re-run the product algorithm over the large dataset. Therefore, the research on how to improve spatial SPT products (objective 3f) has only been minimally achieved. Improvements to the spatial SPT products is where future research should be focused now that a verification framework and a baseline skill value have been established in the research conducted here.

1.4. Spatial SPT Verification—Hail

The final results chapter of this thesis was very limited by the reference data available due to the rarity of true hail events in the UK, defined as ice particles larger than 0.5 mm in diameter (AMS 2012). The PW code system used by manual, automatic and DiVeN instruments do not differentiate between graupel and hail and therefore cannot be used. Automatic stations never report a hail PW code between 2014–2018. Crowdsourced hail reports from the European Severe Weather Database (ESWD) (Dotzek et al. 2009) were relied upon as the trusted source of ground point reference data. These data are confirmatory such that they are only a 'yes' observation of a hail event, which rules out the method of using a dichotomous 2 × 2 contingency table. The tolerance method is again applied, but to the limited 'yes only' ground point reference data. In addition, spatial SPT product diagnoses of 'no precipitation' were considered a 'no hail' event, which differs from the rain, mixed-phase and snow verification. There, 'no precipitation' cases are instead removed from the events set in order to ignore the skill of the product to detect precipitation at all. Since hail is a high-impact event due to the damage it can cause, if the radar product diagnoses 'no precipitation' when hail occurs, it was determined that such events should be included in the verification.

The results in Chapter 5 showed that over the entire domain, hail was correctly diagnosed 30.6%, 65.8% and 75.7% for strict, fair and lenient tolerances respectively. These POD values are low, but the FAR values are unknown and thus may be intentionally low for the end user, as noted in the product description from the Met Office. When the ESWD reports were split into UK events and mainland Europe events (the UK Met Office radar products cover part of mainland Europe), a difference was discovered. POD values for strict, fair and lenient tolerance were +29.8%, +26.5% and +18.5% higher over mainland Europe compared with the UK. The size of the observed hail had a weak relationship to the spatial SPT product skill but overall the sample sizes were very small, limiting the statistical rigour of the results. The interpretation of the difference in results between the UK and mainland Europe was that the geometry of the radar scans makes a hail detection more likely the further away from a radar it occurs. This is due to the radar beam height increasing with range and the hail criterion having an altitude dependence. Artefacts were noted in the 5-year normalised frequency maps in Chapter 5 (Figure 5.3) where isolated radar sites had a ring of no hail diagnoses within a short range from the radar, concurrent with the radar geometry being unsuitable for a high-skill spatial SPT hail product. Even the source of the algorithm (Waldvogel et al. 1979) highlighted radar geometry issues and suggested that the criterion only be used on range height indicator scans rather than volumetric scans.

The results in Chapter 5 confirm the poor performance and frequent artefacts that were expected by users of the single-polarisation radar-based spatial SPT product for hail. The results can be used as a baseline for new products that were developed by the Met Office in recent years. However, more reference data is needed for a more conclusive result. While it is difficult to quantify how large a reference dataset should be, one major improvement would be the ability to report detections of hail from the spatial product as false alarms, which would require the continuous operation of a hail detection sensor network. Some methods for increasing the amount and the quality of reference data for hail are discussed in Section 4.1. Dual-polarisation radar technology allows for hail detection without an altitude criterion and will also be discussed, in Section 4.3.

2. Wider Impact of the Research

The overarching theme of this thesis research is that verification of SPT is inherently difficult and that the existing infrastructure does not permit high confidence SPT observations. Without confident and rigorous SPT observations, improvements in the skill of forecasts or nowcasts of SPT will be unverifiable. DiVeN showed that low-cost real-time precipitation type observations are possible at high temporal frequency using a high resolution diameter–velocity distribution. The Thies LPM provides more information about the precipitation characteristics and could produce less ambiguous SPT data than those provided by the PW code. The Thies LPM instruments in the DiVeN network that were originally loaned from the Met Office have now been donated to the National Centre for Atmospheric Science (NCAS) Atmospheric Measurement and Observation Facility (AMOF) so that their scientific research impact can continue. Furthermore, the Met Office have made plans to change their data collection system for surface stations which will allow more complex DSD and DVD data from disdrometers, with the eventual goal of deploying disdrometers to their ground station network. By having a dense network of disdrometers, NWP models could be verified through DSDs and DVDs directly using continuous statistics rather than having to infer the discrete SPT. The PW code system is also under examination by the Met Office for a potential replacement which can meet all future user requirements (Darren Lyth, personal communication, 2020).

The evaluation of different measurement techniques at CAO in Chapter 3 showed the disparity in precipitation measurements and should motivate research designing more accurate instruments. Over 12 months there were significant differences in total measured precipitation depth. In addition, the technique of using consistent precipitation rates for DSD and DVD comparisons allows for greater interrogation of the instruments compared with the techniques used in the literature, and should be applied to all precipitation distribution comparisons in future research. The SPT comparisons, while inconclusive due to the lack of a trusted reference, used a novel approach with a new SPT classification standard and an algorithm to amalgamate observations over time. Both the standard and the amalgamation algorithm should be implemented as the new consistent standard in SPT-measuring instruments, such as the planned UK Met Office surface station renewal. Furthermore, consideration should be given to the activity of verification when surface networks are designed, as it was for DiVeN.

The results of Chapter 4 which show a positive bias in the diagnosis of mixed-phase precipitation, led to the discovery of a negative bias in the height of the 0 °C isotherm from the NWP model when compared with radar-diagnosed freezing level, shown in Appendix A.6. A correction has been applied to a new Met Office spatial SPT product which also uses dual-polarisation hail diagnosis. Hydrometeor classification algorithms on dual-polarisation radars do not rely on a reflectivity–altitude relationship like the Waldvogel et al. (1979) criterion did. Therefore, the geometry of the radar scan will not influence the ability for the radar to detect hail, and the artefacts of the product evaluated in Chapter 5 of this thesis such as the rings around radar sites where hail could not be detected, should not exist in the new product. Future research includes the verification of the new spatial SPT product against the existing spatial SPT product examined here. Overall this thesis research provides an improved SPT observing standard and a framework on which to perform the verification of SPT data. The application of the new framework to existing ground point SPT measurements and to a spatial SPT product will motivate new approaches to surface precipitation type measurements, and enable them to be quantitatively assessed.

3. Limitations of the Research

The main limitations of the research performed in Chapter 2 were the number of instruments (14), the budget, and the available time which all restricted the data collection abilities of the Disdrometer Verification Network. However, the choice of site locations did cover a wide array of geographies and precipitation climatologies, and the high temporal resolution of the DiVeN sites enabled a verification “event” in Chapter 4 to occur with every spatial SPT product output (5 minutes). Conversely, the automatic and manual station datasets had a combined 121 sites (8.6 times more than DiVeN) but reported every hour (12 times less frequently). The DiVeN data were also hypothesised to be of higher accuracy than the automatic station reports for mixed-phase and hail due to the more direct measurement technique of diameter–velocity distributions over the Met Office arbiter of temperature, forward-scattering and rain gauge. This hypothesis could ultimately not be resolved by the research conducted due to the lack of cases in Chapter 2 and the absence of a true reference and the limited number of events in Chapter 2 where the experimental design was not fit for purpose. Ideally the true reference should be a human observer for multiple borderline cases of mixed-phase SPT, beyond the resource capabilities of this thesis. The opportunistic human observer validation in Chapter 2 only covered two transitions from rain to snow and did not do so continuously. Bloemink and Lanzinger (2005) compared a Thies LPM and a FD12P sensor against a human observer and provided some evidence of the accuracy of the Thies LPM in different SPTs but the study only considered a six-week period of precipitation with 10-minute observations (341 rain, 48 mixed-phase, 247 snow). The lack of a temperature sensor connected to the Thies LPM (an optional extra which improves the PW code diagnosis skill according to the manufacturer) meant that the verification of a spatial SPT product in Chapter 4 could not be isolated to the borderline cases, which would be a more stringent method since the ‘obvious’ cases of rain and snow unfairly raise the skill value.

The low-cost implementation of DiVeN combined with the remote-only access to most of the sites meant that the instruments were offline for a large percentage of the study period resulting in a reduction to the overall number of verification cases. A critical bug in September 2017 was fixed within 1 month but resulted in the loss of several weeks of data. The Feshie (Druim nam Bo) site was located at 882 m a.m.s.l. and was frequently covered in rime ice as shown in Figure 2.6 in Chapter 2, which hampered the ability to record precipitation at the site in the winter months. While heated Thies LPMs are available, the renewable energy generated at the Feshie site was insufficient to support the required power demands. More examples of site-specific challenges are given in Chapter 2. Overall the impact to the research was minimal due to the large number of total cases, but more cases of the rare SPTs would have increased the rigour of the verification.

The PW codes were relied on throughout the research, which are a limitation when the raw drop size and drop velocity measurements are provided by the Thies LPM. Deriving SPT from the raw DSD and DVD values would have been possible from the literature on the relationships of the 2D distribution of diameter and velocity to SPT (Gunn and Kinzer 1949; Langleben 1954; Locatelli and Hobbs 1974; Bohm 1989). However, with some artefacts and spurious particles noted during the early measurements, and from contact with the instrument manufacturer, it was determined that the internal processing of the Thies LPM considered artefacts and also relied on the literature (although the exact methodology was not supplied since it is considered the intellectual property of Thies). In addition, for the research in Chapters 3 and 4, all other instruments and manual observers reported PW code, so it was a set standard by which to perform the verification (with translation to the new SPT classifier and the spatial SPT classifications). More research is needed to implement the new SPT standard proposed in Chapter 3 into future instruments and spatial product updates to replace the ambiguous PW code for the SPT variable.

In Chapter 3, no absolute truth reference could be obtained for any of the precipitation variables considered. For this reason the research is an evaluation (comparing multiple methods of reaching an end result) as opposed to a verification (comparing a new method with an existing method considered to be the truth), but a verification is more valuable and is needed for the ground point SPT instruments (see Section 4.1 of Chapter 1). As mentioned in Chapter 3, the WMO intercomparison studies use a composite working reference of four laboratory-tested gauges for rain rate, which could be considered a verification (Lanza and Vuerich 2009). While a set of trusted instruments could have been implemented at the start of the study period at CAO, this was not considered at the time nor would it have likely been possible due to the available resources. In addition the most trusted instrument differs depending on the precipitation rate. Some of the instruments were installed sub-optimally, such as the PWD21 and PWS100 being on the roof of a small building, but these could not be altered because of the need to maintain a consistent long term data record at CAO. Ultimately, the most trusted reference for SPT measurements is a human using vision to make a manual observation (without the constraint of the PW code system). Finally, the evaluation study in Chapter 3 was limited to a single location in Southern England and only a 12-month period. Therefore, the variety of SPTs and the number of rare SPTs captured by the study was low. The applicability of the results is limited to locations with similar climatologies but furthermore, a 12-month period is also not a sufficiently long period to be labelled as a representative climatological sample.

In Chapter 4 the main limitation was that the accuracy of the in-situ reference point ground data was not well quantified. The research in Chapters 2 and 3 showed some deficiencies in the DiVeN dataset and the automated station sensors are also suspected of performing poorly in cases of mixed-phase precipitation (Lyth and Molyneux 2006). Due to these uncertainties, all the reference datasets were presented separately in Chapter 4 to show the disagreements. Part of the disagreements can be attributed to the locations of the sensors in their network. For example, the DiVeN instruments have more locations at high altitude (36% > 250 m a.m.s.l.) compared with the Met Office sites (~10% > 250 m a.m.s.l.) which means there are more 'obvious' snow cases at these sites so the skill appears to be higher. Chapter 4 demonstrated that the verification scores were highly sensitive to the meteorology of the study period. Removing the year 2018 from the verification sample, the overall HSS scores were reduced. There was a period in March 2018 where temperatures were below zero for many days and snow was the dominant precipitation type. While snow is usually a borderline event in the UK and an SPT product requires high skill to capture it accurately, March 2018 contained many 'obvious' snow events. The value of the spatial SPT product to end users in an obvious SPT scenario is unclear; in summer, the value to end users of the snow and mixed-phase classes approaches zero.

Generally the results of Chapters 4 and 5 are only applicable to the UK and only cover a 5-year study period which may not capture the true climatology of extremely rare events like hail. The overwhelming number of rain cases compared with any other type of precipitation makes the verification challenging because scalar metrics become skewed. The future research in Section 4.2 suggests a method to make all SPT classes equitably represented which requires additional ground data not present in this research. In both Chapter 4 and Chapter 5, the SPT product is examined independently at each time step. No consideration is given to the behaviour of the product for a single meteorological feature (convective cell, frontal system) over time. Whether the features behave realistically compared to the expected atmospheric dynamics within convective and frontal features, is unknown. For example, the verification of an NWP model for SPT performed by Ikeda et al. (2013) noted that larger synoptic-scale features were more accurately predicted than convective-scale features, a delineation which was not explored here.

In Chapter 5 the reference dataset of ESWD crowdsourced and quality controlled hail reports was the major limiting factor due to the statistically insignificant number of cases (111). Since the PW code system does not definitively separate hail and graupel, they could not be used, and the automatic surface stations never report the hail PW codes. The crowdsourced dataset from ESWD has high certainty because of the quality control processes that are performed on submitted reports. However, the database is not widely known by the UK public and therefore only 32 reports were submitted within

5 years. The critical success measure of a crowdsourced dataset is its popularity which will result in a larger sample size. The spatial SPT product frequency map of hail shows around half of all land pixels in the UK being diagnosed hail within the 5-year study period, which is high compared to the literature (Punge and Kunz 2016). However, the false alarm ratio of the radar-based product is not known because of the absence of a 'no hail' category in the reference dataset. The lack of data meant that the same verification methodology from the rain, mixed-phase and snow verification could not be repeated, nor could the results be compared, despite the SPT categorisations being codependent (i.e. a hail diagnosis is an event removed from the other precipitation classes). The fact that consistent methodologies are not applicable to all SPT classes demonstrates that improvements are needed to the ground point reference datasets. An ideal goal for the ground point reference datasets is for all precipitation types to be verified with the same, equitable statistical framework.

Finally, the major limitation of the spatial SPT product itself is that a feedback loop exists if the Met Office radar-based product is used to verify NWP forecasts because the 0 °C wet-bulb isotherm height is taken from an NWP model. Therefore, the verification of NWP using the spatial SPT product will simply be verifying a combination of the location of the storm and the precipitation rate (or reflectivity for hail), which is possible without the SPT product. The ground data is a more independent reference by which to conduct verification of NWP forecasts of SPT but these are not spatial. Note that the verification methods currently used in the literature for NWP forecasts against surface point SPT observations (see Section 4.6 in Chapter 1) have the same issues as the verification of spatial and point SPT data. The verification framework developed in this thesis, particularly the tolerance method in Chapter 4, is also suitable for the verification of NWP SPT forecasts against ground point reference SPT data, and should be employed in the future for this task.

4. Future Research and Recommendations

The research in Chapters 2, 3, 4 and 5 have demonstrated the real-time capture of precipitation type data and have advanced our ability to perform both evaluation of point SPT data and statistical verification of spatial SPT products. Despite the advances, there are some limitations which have been described in the previous section. There are several fields of future research and research that can be built upon and motivated by the research laid out in this thesis. The following sections describe these potential avenues in three categories: point measurements, statistical techniques, and spatial products.

4.1. Point SPT Measurements

Having a trusted set of point reference observations is the necessary foundation on which improvements to forecasting can be built upon to realise socioeconomic value. Some skill is lost between the observations and the forecast simply due to the stochastic nature of the atmosphere and the inconsistent geometry of the datasets. While it may be possible to make an NWP model more skilful than the reference dataset is, it could not be proven through verification. However, the more accurate the initial point observations are, the more skill and understanding can be passed onto the NWP model and improve decision making and informedness to the end users. For these reasons, the following paragraphs describe future research that would improve the quality and the amount of available reference data for SPT.

The deployment of the Disdrometer Verification Network as described in Chapter 2 demonstrated that there is added value to operational data on precipitation characteristics more than just the typically observed rate or the accumulated depth of the precipitation. In addition, the research in Chapter 2 showed that the cost to achieve that value can be minimised, motivating widespread disdrometer use. Deutscher Wetterdienst (DWD) has operated disdrometers in Germany for several years. The UK Met Office plan to introduce disdrometers into their network in the next few years to supplement the existing arbiter for SPT. This is needed and will improve the reliability of any reference dataset by which to perform verification of spatial SPT diagnosis products and NWP forecasts.

More research is needed to produce observations using the new SPT standard from the raw diameter–velocity data that the Thies LPM disdrometer collects. This will require both laboratory studies and more fieldwork comparing the SPT observations of the Thies LPM against human observations. The scale of the work required is outside of the scope of this thesis and will require thousands of SPT events to be captured in order to provide a statistically rigorous sample. The PW code system should be replaced with the SPT classifier developed in Chapter 3. The Met Office is aware of the shortcomings of the PW code system and has plans to replace it in the future, which will be aided by the deployment of disdrometers at automatic surface station sites. Again this will improve all future verification work by providing a more accurate reference dataset.

While crowdsourced data was utilised in Chapter 5 for hail verification, the number of datapoints in the 5-year study period was statistically small. Other crowdsourced datasets exist but must be thoroughly quality controlled to avoid spurious reports, in particular by delineating graupel from hail. One method would be to trawl social media for images of hail that were uploaded with location information. The limitation is that such an endeavour would traditionally be labour intensive and time consuming, and that estimating the size of hail from images is difficult. Other data collection methods have demonstrated success for the hail SPT. A novel dataset was collected between 2006–2015 where a team of volunteers watched potential hail-producing storms on radar and then call businesses and homes with publicly listed phone numbers to ask them about what was occurring or what had occurred (Ortega et al. 2009; Ortega 2018), resulting in a total dataset of 54,299 reports. While this method is effective, it requires an active group of volunteers, is labour intensive and may be considered intrusive.

An automated approach to collecting hail sizes is described in Soderholm et al. (2020). An unmanned aerial system (UAS) is used to build up a composite image of an area where hailstones are lying on the ground. Numerical image processing techniques are then used to estimate the size of the individual hailstones through photogrammetry, which produces a hail size distribution. Although piloting a UAS is labour intensive, the photogrammetry technique could be applied to static webcams, cameras in vehicles, and social media images to reduce the physical labour required and thus increase the amount of data available. Autonomous UAS would be a path forward to survey suspected hailstorms but regulatory permission is currently difficult to obtain. Calibration is also a difficult task when the cameras will differ in their distance to the ground and viewing angle, and melting means that the measured hail sizes are smaller than when they impacted the ground. Machine learning with convolutional neural network approaches (also referred to as ‘computer vision’) may be a way to account for camera differences and to recognise the size of known reference objects in the image to increase the accuracy of the hail size distribution. Being able to capture these images immediately after or even during a hail event will minimise the effect of melting. Computer vision could also be applied more generally to the discrimination of all SPTs and even estimate their quantity, as some work in the literature has shown (Garg and Nayar 2007; Ma et al. 2016).

The research conducted in this thesis can be used to motivate the need for higher temporal frequency and spatially dense SPT data, which are needed to support the radar-based products, in the same way that observations currently are with TBRs used for radar quantitative precipitation estimation (QPE) calibration. More accurate and more dense SPT data would improve our understanding of the transitional processes, allow for finer calibration, and would enable the work of Lumb (1963) to be revisited, as was suggested by the conclusions of Chapter 4. While hail is not observable with the current automated Met Office surface stations, there is some evidence that hail will become an increasingly likely and increasingly damaging event in the UK (Eccel et al. 2012; Mohr et al. 2015; Sanderson et al. 2015; Púčik et al. 2017), so capturing hail events accurately may become more important in the future. The ESWD is a good resource of trusted hail events and should be maintained, but more work needs to be done to publicise the database so that more hail reports are submitted by members of the public.

4.2. Statistical Verification Techniques

While the study period used in this thesis to verify the spatial SPT product was 5 years, there were only 5,911 mixed-phase events and 18,743 snow events recorded by all ground point reference instruments. Hail was notably infrequent with only 32 events in the UK from the ESWD dataset. The study should be repeated with a longer study period of 10 or more years, in combination with another hail reference dataset source. With a more than 10 year study-period, all precipitation types would be widely detected by the spatial SPT product, allowing for smoother frequency maps, particularly for the hail SPT.

One issue faced in Chapter 4 was the unequal number of SPT events by class. Rain made up between 91–97% of the reference point observations, which skewed some of the scalar verification metrics. In addition, the product was being verified for snow and mixed-phase events during the summer, when users would not find any value in the spatial SPT product because the SPT was ‘obvious’ as rain. In future research, the set of events to be verified should be truncated based on when users (forecasters, decision makers) actually use the spatial SPT product. For example, some forecasters would only look at such a product for mixed-phase and snow SPT in the UK when the freezing level was below 1 km (Sabrina Lee, personal communication, 2019). Alternatively, a temperature range could be chosen. Snow may be considered ‘obvious’ at surface temperatures less than -2 °C, whereas rain may be considered ‘obvious’ at surface temperatures greater than 4 °C. A temperature criterion for verification could not be applied consistently for all reference datasets in this research because the DiVeN instruments did not have co-located temperature information. In a future verification, the verification sample set could be limited to days where a winter precipitation warning had been issued, or days when thunderstorm warnings had been issued (for the hail class). However, limiting the verification sample would remove the majority of SPT occurrences, and introduces a bias based on the current skill of weather warning issuance from the Met Office.

There are a wide variety of verification metrics, some of which are described in Chapter 1. The choice to use one metric over another is difficult to make and is often subjective. In this research, the choices were bias, probability of detection, false alarm ratio, and Heidke skill score based on the questions that were asked about the spatial SPT product in the thesis objectives, but the needs of the end user should also influence the choice. For example, the hail SPT class may prioritise the reduction of the false alarm ratio over the probability of detection, so that users do not become complacent after many false alarms. The contingency table and confusion matrix in the verification could be weighted by how impactful or costly one precipitation type is over another, to reduce the overall skill score if the most impactful SPTs have poor skill. The weighting could be different for each category of product user and therefore the skill values would also differ between users. The usefulness of verification is in how the results reveal which action should be taken to improve a product by some predetermined standard. In practical terms the results would become more complex with this suggested approach and thus would be harder to interpret, perhaps making the required action unclear.

Finally, the spatial SPT product verified in this thesis is deterministic which gives no indication of the certainty of the diagnosis and is less informative to an end user. A probabilistic version of the spatial SPT product should be developed in the future, as will be explained in the next section. From a verification perspective, the toolset of applicable statistical methods for verification of a probabilistic product would be much greater if the product were probabilistic. In such a product, clear definitions should exist for the mixed-phase class, where confusion could occur between the percentage probability of occurrence versus the percentage of precipitation that was liquid or solid.

Furthermore, a validation is required with the end users of the spatial SPT product. The level of specificity (with regards to Figure 1.24 in Section 4.2 of Chapter 1) needed by the users is not evident. Are the rain, mixed-phase, snow and hail SPT classifications sufficient? Would the delineation of drizzle and rain, or would the addition of graupel, ice pellets and freezing rain SPT classes be valuable to an end user? A validation checks whether a product is fit for purpose (see Section 4.1 in Chapter 1), which has not been performed for the spatial SPT product but should be before future development of a new product occurs.

4.3. Spatial SPT Products

The research going forward should be concentrated on objective 3f of this thesis—to iterate and improve upon spatial SPT observation products. Work is ongoing at the UK Met Office to develop a new radar-based spatial SPT product based on the verification research performed in this thesis. Two major changes have been made. Firstly, a correction has been applied to the height of the 0 °C wet-bulb isotherm which was shown in Chapter 4 to be too low, causing an over-diagnosis of mixed-phase when rain occurred, and an over-diagnosis of snow when mixed-phase occurred. In addition, the Euro4 NWP model which supplies the 0 °C wet-bulb isotherm will likely be replaced by a newer NWP model and thus take advantage of increased resolution and improved microphysics schemes to further improve the freezing level height estimation. Secondly, hail is now detectable at any altitude and distance from the radar (within the radar beam) with a dual-polarisation hydrometeor classification algorithm. The hit rate is expected to increase and the circles of no hail detection for isolated radars, which was noted in Chapter 5, should not exist in the new product. The verification framework developed in this thesis should be reapplied to the new product in order to quantify the improvements.

In addition, based on the conclusions of Chapter 4, Lumb's critical rate (Lumb 1963) should be re-examined. The current derived formula only considers cases with precipitation rates between 1–4 mm h^{-1} and uses 90% precipitation melted (so 10% remains solid) as the boundary between liquid and mixed-phase, which does not follow the definition used in the new SPT classification standard. The ability for the ground point SPT instruments to detect the 10% remaining solid mass and record a mixed-phase PW code when Lumb's critical rate would diagnose rain, was not examined in this research. Also, Lumb's critical rate does not consider time; a saturated atmospheric column is always assumed due to the precipitation evaporatively cooling the air (Lumb 1963) but this process is not instantaneous. When precipitation begins, the atmospheric column may not be saturated and an SPT of rain may take some time to transition to mixed-phase and snow through evaporative cooling. The time taken for the column to saturate will be strongly dependent on the precipitation rate such that any time-dependence will be more important for stratiform (typically weaker) precipitation events, as explained by Lumb (1963).

For rain, mixed-phase and snow, the current spatial SPT product essentially tries to estimate the surface temperature through the NWP 0 °C wet-bulb isotherm height and an empirical relationship for evaporative cooling. Future research should attempt to instead utilise the real-time observations of temperature from the ground since there already exists an operational capability to produce high-resolution surface temperature maps in near-real time (Green 2010). For locations without surface temperature data, maps could use interpolation and could also consider terrain changes where surface stations do not exist. The average distance from one UK Met Office surface station to the next nearest station is 17.6 km. With crowdsourced home weather stations that distance has been shown to decrease to 2.1 km (Coney et al. 2020) but significant quality control is required because the data is low-quality and is associated with many artefacts. Using the suggested surface temperature approach would make the spatial product completely independent from NWP and allow the product to be a reference to NWP verification with higher statistical rigour. However, some rarer SPTs such as freezing rain and ice pellets would require a full vertical temperature profile of the atmosphere in order to be diagnosed. For these SPTs, vertical temperature profiles from NWP may be crucial and irreplaceable inputs.

A limitation of the current spatial SPT product itself is that it gives no indication of hail size which is often a critical variable in order to determine whether the event will be a minor spectacle or cause costly damage. Similarly, the SPT product does not indicate the intensity of each precipitation type. Rainfall and snowfall require different relationships for radar QPE because their electromagnetic scattering behaviour differs. Future research should utilise the SPT information to aid the quantification of the amount of precipitation, which could lead to more accurate values (Berne and Krajewski 2013). Combining the amount of precipitation with an awareness of time, the spatial SPT product could be used to estimate the depth of snow on the ground, which would be a useful input to hydrology models. With memory, a product would take into account that rain was falling over snow and would therefore accelerate melting, which is known to exacerbate flood events (Muchan et al. 2015).

Finally, there is no indication of uncertainty in the spatial SPT product. Probabilistic values give additional information to decision makers on the confidence of the product and can increase informedness when compared to a deterministic product. There are many ways to produce a probabilistic output, such as with fuzzy logic algorithms or with neural networks. The recommended research to revisit Lumb's critical rate should attempt to produce a probabilistic function instead of a discrete relationship, as well as using knowledge of the recent precipitation rate to estimate how saturated the atmospheric column is. Alternatively, surface-based humidity values or wet-bulb temperatures could indicate whether a change in SPT was likely to occur, for a given precipitation rate.

5. Final Remarks

This thesis has demonstrated firstly that low-cost point observations of SPT have value. Secondly, the classification of SPTs has been constrained compared to the existing SPT reporting standards. Finally, spatial SPT products have been verified with point SPT data using novel tolerance techniques which address representativity issues and can be applied to other spatial products in the future. As NWP attempts to resolve small-scale features with increasing resolution, there is more research to be done on SPT. Observations of SPT need to be improved with increased specificity in order for the verification of NWP to add value. To address this, novel SPT data sources are being developed which will require equitable assessment. The research conducted in this thesis demonstrates a verification framework which should be used to quantify the skill of, and to motivate improvements to, SPT products. In addition, a baseline skill value has been set for future products to improve upon. With further research these improvements may lead to more accurate and informed NWP model predictions of SPT and increase the preparedness of society for disruptive SPT events.

References

Agnew, M. D., and J. E. Thornes, 1995: The weather sensitivity of the UK food retail and distribution industry. *Meteorological Applications*, **2 (2)**, 137–147, doi:10.1002/met.5060020207.

Al-Sakka, H., A. A. Boumahmoud, B. Fradon, S. J. Frasier, and P. Tabary, 2013: A new fuzzy logic hydrometeor classification scheme applied to the french X-, C-, and S-band polarimetric radars. *Journal of Applied Meteorology and Climatology*, **52 (10)**, 2328–2344, doi:10.1175/JAMC-D-12-0236.1.

AMS, 2012: Hail. Glossary of Meteorology. URL https://glossary.ametsoc.org/wiki/Hail.

Berne, A., and W. F. Krajewski, 2013: Radar for hydrology: Unfulfilled promise or unrecognized potential? *Advances in Water Resources*, **51**, 357–366, doi:10.1016/j.advwatres.2012.05.005, URL http://dx.doi.org/10.1016/j.advwatres.2012.05.005.

Best, S., and D. Harrison, 2018: A Hydrometeor Classification Algorithm for the UK C-Band Weather Radar Network. *10th European Conference on Radar in Meteorology & Hydrology*, ERAD, Observations R&D, Met Office, Fitzroy Road, Exeter, EX1 3PB, UK, 1–9.

Bloemink, H., and E. Lanzinger, 2005: Precipitation type from the Thies disdrometer. *WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO-2005)*, 1–7, URL http://www.wmo.int/pages/prog/www/IMOP/publications/IOM-82-TECO_2005/Papers/3(11)_Netherlands_4_Bloemink.pdf.

Bohm, H. P., 1989: A general equation for the terminal fall speed of solid hydrometeors. 2419–2427 pp., doi:10.1175/1520-0469(1989)046⟨2419:AGEFTT⟩2.0.CO;2.

Carrière, J. M., C. Lainard, C. Le Bot, and F. Robart, 2000: A climatological study of surface freezing precipitation in Europe. *Meteorological Applications*, **7 (3)**, 229–238, doi:10.1017/S1350482700001560.

Chen, S., J. J. Gourley, Y. Hong, Q. Cao, N. Carr, P. E. Kirstetter, J. Zhang, and Z. Flamig, 2016: Using Citizen Science Reports to Evaluate Estimates of Surface Precipitation Type. *Bulletin of the American Meteorological Society*, **97 (2)**, 187–193, doi:10.1175/BAMS-D-13-00247.1.

Clark, M. R., J. D. C. Webb, and P. J. Kirk, 2018: Fine-scale analysis of a severe hailstorm using crowd-sourced and conventional observations. *Meteorol. Appl.*, **492 (April)**, 472–492, doi:10.1002/met.1715.

Coney, J., B. S. Pickering, D. Dufton, and M. Lukach, 2020: Goldmine or Bust? Crowd-sourced Meteorological Data for Atmospheric Science. Ph.D. thesis, University of Leeds, 68 pp.

Dotzek, N., P. Groenemeijer, B. Feuerstein, and A. M. Holzer, 2009: Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD. *Atmospheric Research*, **93 (1-3)**, 575–586, doi:10.1016/j.atmosres.2008.10.020, URL http://dx.doi.org/10.1016/j.atmosres.2008.10.020.

Eccel, E., P. Cau, K. Riemann-Campe, and F. Biasioli, 2012: Quantitative hail monitoring in an alpine area: 35-year climatology and links with atmospheric variables. *International Journal of Climatology*, **32 (4)**, 503–517, doi:10.1002/joc.2291.

Elmore, K. L., H. M. Grams, D. Apps, and H. D. Reeves, 2015: Verifying forecast precipitation type with mPING. *Weather and Forecasting*, **30 (3)**, 656–667, doi:10.1175/WAF-D-14-00068.1.

Fairman, J. G., D. M. Schultz, D. J. Kirshbaum, S. L. Gray, and A. I. Barrett, 2015: A radar-based rainfall climatology of Great Britain and Ireland. *Weather*, **70 (5)**, 153–158, doi:10.1002/wea.2486.

Frasson, R. P. d. M., L. K. da Cunha, and W. F. Krajewski, 2011: Assessment of the Thies optical disdrometer performance. *Atmospheric Research*, **101 (1-2)**, 237–255, doi:10.1016/j.atmosres.2011.02.014, URL http://dx.doi.org/10.1016/j.atmosres.2011.02.014.

Garg, K., and S. K. Nayar, 2007: Vision and rain. *International Journal of Computer Vision*, **75 (1)**, 3–27, doi:10.1007/s11263-006-0028-6.

Gascón, E., T. Hewson, and T. Haiden, 2018: Improving Predictions of Precipitation Type at the Surface: Description and Verification of Two New Products from the ECMWF Ensemble. *Weather and Forecasting*, **33 (1)**, 89–108, doi:10.1175/WAF-D-17-0114.1, URL http://journals.ametsoc.org/doi/10.1175/WAF-D-17-0114.1.

Green, A., 2010: From Observations to Forecasts – Part 7 . A new meteorological monitoring system for the United Kingdom's Met Office. *Weather*, **65 (10)**, 272–277.

Gunn, R., and G. D. Kinzer, 1949: The Terminal Velocity of Fall for Water Droplets in Stagnant Air. *Journal of Meteorology*, 243–248, doi:10.1175/1520-0469(1949)006⟨0243:TTVOFF⟩2.0.CO;2.

Harrison, D., S. J. Driscoll, and M. Kitchen, 2000: Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorol. Appl.*, **6**, 135–144, doi:10.1017/S1350482700001468.

Ikeda, K., M. Steiner, J. Pinto, and C. Alexander, 2013: Evaluation of cold-season precipitation forecasts generated by the hourly updating high-resolution rapid refresh model. *Weather and Forecasting*, **28 (4)**, 921–939, doi:10.1175/WAF-D-12-00085.1.

Langleben, M. P., 1954: The terminal velocity of snowflakes. *Quarterly Journal of the Royal Meteorological Society*, **80 (344)**, 174–181, doi:10.1002/qj.49708034404.

Lanza, L. G., and E. Vuerich, 2009: The WMO Field Intercomparison of Rain Intensity Gauges. *Atmospheric Research*, **94 (4)**, 534–543, doi:10.1016/j.atmosres.2009.06.012, URL http://dx.doi.org/10.1016/j.atmosres.2009.06.012.

Lin, D., B. Pickering, and R. R. Neely, 2020: Relating the radar bright band and its strength to surface rainfall rate using an automated approach. *Journal of Hydrometeorology*, **21 (2)**, 335–353, doi:10.1175/JHM-D-19-0085.1.

Liu, H., and V. Chandrasekar, 2000: Classification of hydrometeors based on polarimetric radar measurements: Development of fuzzy logic and neuro-fuzzy systems, and in situ verification. *Journal of Atmospheric and Oceanic Technology*, **17 (2)**, 140–164, doi:10.1175/1520-0426(2000)017⟨0140:COHBOP⟩2.0.CO;2.

Locatelli, J. D., and P. V. Hobbs, 1974: Fall speeds and masses of solid precipitation particles. *Journal of Geophysical Research*, **79 (15)**, 2185–2197, doi:10.1029/JC079i015p02185.

Lumb, F. E., 1963: Downward penetration of snow in relation to the intensity of precipitation. *Meteorology Magazine*, **92 (1086)**, 1–14.

Lyth, D., and M. Molyneux, 2006: Results of Using Present Weather Instruments in the United Kingdom. Tech. rep., The Met Office, 46 pp.

Ma, Y., P. Ding, Q. Li, W. Lu, J. Yang, and W. Yao, 2016: A Vision-Based Precipitation Sensor for Detection and Classification of Hydrometeors. *IEEE Sensors Journal*, **16 (11)**, 4546–4554, doi:10.1109/JSEN.2016.2542282.

Mohr, S., M. Kunz, and B. Geyer, 2015: Hail potential in Europe based on a regional climate model hindcast. *Geophysical Research Letters*, **42 (24)**, 10 904–10 912, doi:10.1002/2015GL067118.

Muchan, K., M. Lewis, J. Hannaford, and S. Parry, 2015: The winter storms of 2013/2014 in the UK: hydrological responses and impacts. *Weather*, **70 (2)**, 55–61.

NERC, UK Met Office, B. S. Pickering, R. R. Neely III, and D. Harrison, 2019: The Disdrometer Verification Network (DiVeN): particle diameter and fall velocity measurements from a network of Thies Laser Precipitation Monitors around the UK (2017-2019). URL https://catalogue.ceda.ac.uk/uuid/602f11d9a2034dae9d0a7356f9aeaf45, doi:10.5285/602f11d9a2034dae9d0a7356f9aeaf45.

Ortega, K. L., 2018: Evaluating multi-radar, multi-sensor products for surface hailfall diagnosis. Evaluating Multi-Radar, Multi-Sensor Products for Surface Hail-Fall Diagnosis. *Electronic J. Severe Storms Meteor*, **13 (1)**, 1–36, URL http://www.ejssm.org/ojs/index.php/ejssm/article/viewFile/163/113.

Ortega, K. L., T. M. Smith, K. L. Manross, K. A. Scharfenberg, W. Arthur, A. G. Kolodziej, and J. J. Gourley, 2009: The severe hazards analysis and verification experiment. *Bulletin of the American Meteorological Society*, **90 (10)**, 1519–1530, doi:10.1175/2009BAMS2815.1.

Park, H. S., A. V. Ryzhkov, D. S. Zrnić, and K.-E. Kim, 2009: The Hydrometeor Classification Algorithm for the Polarimetric WSR-88D: Description and Application to an MCS. *Weather and Forecasting*, **24 (3)**, 730–748, doi:10.1175/2008WAF2222205.1.

Pickering, B. S., R. R. Neely, and D. Harrison, 2019: The Disdrometer Verification Network (DiVeN): A UK network of laser precipitation instruments. *Atmospheric Measurement Techniques*, **12 (11)**, 5845–5861, doi:10.5194/amt-12-5845-2019.

Púčik, T., and Coauthors, 2017: Future changes in European severe convection environments in a regional climate model ensemble. *Journal of Climate*, **30 (17)**, 6771–6794, doi:10.1175/JCLI-D-16-0777.1.

Punge, H. J., and M. Kunz, 2016: Hail observations and hailstorm characteristics in

Europe: A review. *Atmospheric Research*, **176-177**, 159–184, doi:10.1016/j.atmosres.2016.02.012, URL http://dx.doi.org/10.1016/j.atmosres.2016.02.012.

Sanderson, M. G., W. H. Hand, P. Groenemeijer, P. M. Boorman, J. D. Webb, and L. J. Mccoll, 2015: Projected changes in hailstorms during the 21st century over the UK. *International Journal of Climatology*, **35 (1)**, 15–24, doi:10.1002/joc.3958.

Schmid, W., and A. Mathis, 2004: Validation of methods to detect winter precipitation and retrieve precipitation type. *Annalen der Meteorologie*, **(January)**, 1–8, URL http://www.iac.ethz.ch/staff/wueest/sirwec/conferences/bingen2004/topic_iv/sirwec-iv-3-schmid.pdf.

Soderholm, J. S., M. R. Kumjian, N. McCarthy, P. Maldonado, and M. Wang, 2020: Quantifying hail size distributions from the sky - Application of drone aerial photogrammetry. *Atmospheric Measurement Techniques*, **13 (2)**, 747–754, doi:10.5194/amt-13-747-2020.

Thornes, E., 1992: The impact of weather and climate on transport in the UK. *Process in Physical Geography*, **16 (2)**, 187–208.

Tilg, A.-M., F. Vejen, C. B. Hasager, and M. Nielsen, 2020: Rainfall Kinetic Energy in Denmark: Relationship with Drop Size, Wind Speed, and Rain Rate. *Journal of Hydrometeorology*, **21 (7)**, 1621–1637, doi:10.1175/jhm-d-19-0251.1.

Waldvogel, A., B. Federer, and P. Grimm, 1979: Criteria for detection of hail cells. *Journal of Applied Meteorology*, **18 (12)**, 1521–1525, doi:10.1175/1520-0450(1979)018⟨1521:CFTDOH⟩2.0.CO;2.

Webb, J. D., D. M. Elsom, and D. J. Reynolds, 2001: Climatology of severe hailstorms in Great Britain. *Atmospheric Research*, **56 (1-4)**, 291–308, doi:10.1016/S0169-8095(00)00081-8.

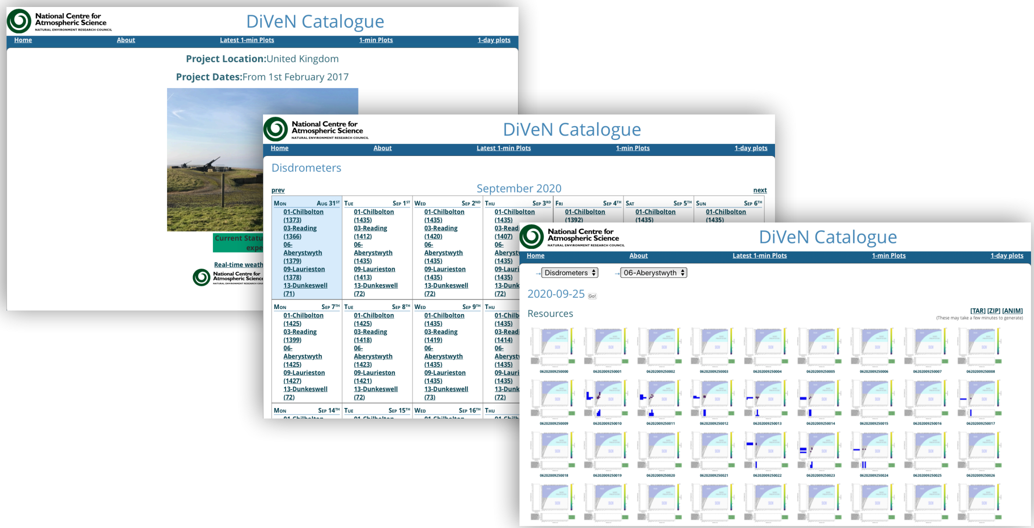
WMO, 1988: Manual on Codes. *WMO Publ. 306*, **1**, 203 pp.

WMO, 2017: WMO Guide to Meteorological Instruments and Methods of Observation (the CIMO Guide). Chapter 6. Masurement of Precipitation. Tech. rep., World Meteorological Organization.

Appendix A.

Supplementary Figures

The following section is a selection of supplementary figures (A.1–A.6) that are not critical to the interpretation of the thesis but that add additional context to the wider work accomplished.



DiVeN Catalogue

Project Location: United Kingdom

Project Dates: From 1st February 2017

Disdrometers

September 2020

Mon Aug 31st	Tue Sep 1st	Wed Sep 2nd	Thu Sep 3rd	Fri Sep 4th	Sat Sep 5th	Sun Sep 6th
01-Chilbolton (1373) 03-Reading (1366) 06-Aberystwyth (1379) 09-Laurieston (1378) 13-Dunkeswell (71)	01-Chilbolton (1435) 03-Reading (1412) 06-Aberystwyth (1435) 09-Laurieston (1413) 13-Dunkeswell (72)	01-Chilbolton (1435) 03-Reading (1420) 06-Aberystwyth (1435) 09-Laurieston (1435) 13-Dunkeswell (72)	01-Chilbolton (1435) 03-Readi (1407) 06-Aberyst (1435) 09-Lauri (1435) 13-Dunk (72)	01-Chilbolton (1392)	01-Chilbolton (1435)	01-Chilbolton (1435)
Mon Sep 7th	**Tue Sep 8th**	**Wed Sep 9th**	**Thu**			
01-Chilbolton (1425) 03-Reading (1399) 06-Aberystwyth (1425) 09-Laurieston (1427) 13-Dunkeswell (72)	01-Chilbolton (1425) 03-Reading (1418) 06-Aberystwyth (1423) 09-Laurieston (1421) 13-Dunkeswell (72)	01-Chilbolton (1435) 03-Reading (1419) 06-Aberystwyth (1435) 09-Laurieston (1435) 13-Dunkeswell (73)	01-Chilb (1435) 03-Readi (1414) 06-Aberyst (1435) 09-Lauri (1435) 13-Dunk (72)			
Mon Sep 14th	**Tue Sep 15th**	**Wed Sep 16th**	**Thu**			

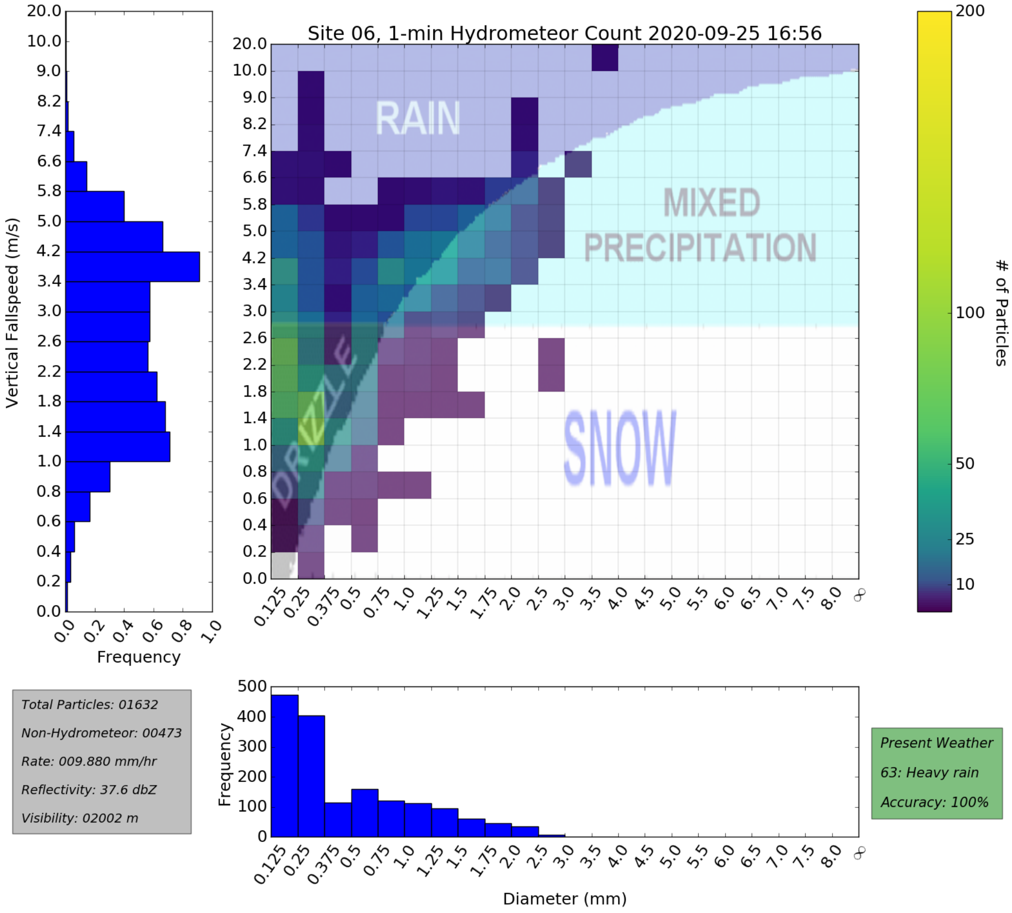
Disdrometers: 06-Aberystwyth

2020-09-25

Resources

[TAR] [ZIP] [ANIM]
(These may take a few minutes to generate)

Figure A.1.: The design of the publicly-accessible Disdrometer Verification Network (DiVeN) website containing real-time (within 2–7 minutes) precipitation data from the Thies LPM instruments.



Total Particles: 01632

Non-Hydrometeor: 00473

Rate: 009.880 mm/hr

Reflectivity: 37.6 dbZ

Visibility: 02002 m

Present Weather

63: Heavy rain

Accuracy: 100%

Figure A.2.: An example of the 1-minute plot available to users on the DiVeN website. Animated versions were available for each minute of the day to show changes in the 2D precipitation diameter-velocity distribution over time.

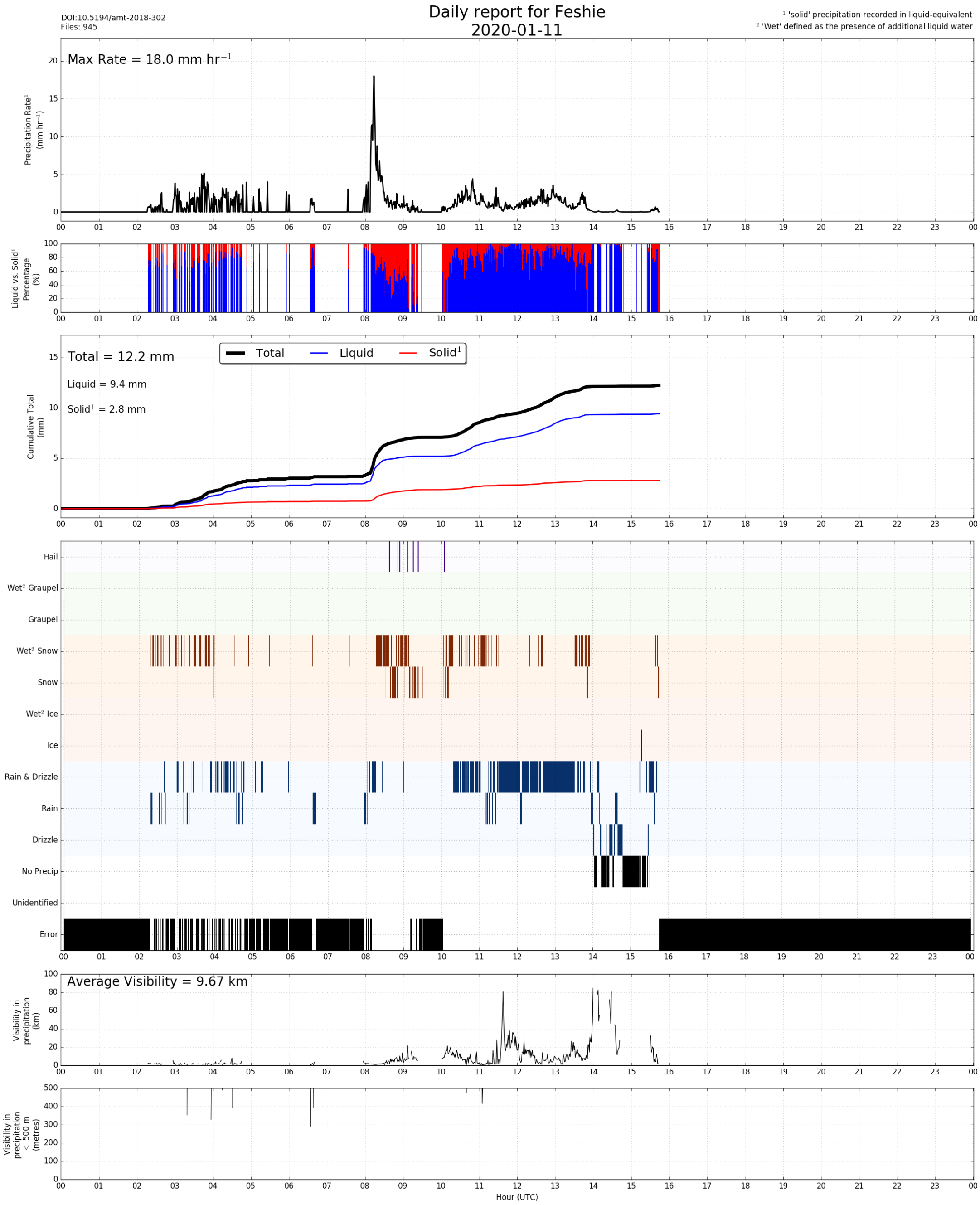


Figure A.3.: An example of a daily summary plot from the DiVeN website. The plot is showing data from 11 January 2020 at the Feshie (Druim nam Bo) site which is in the Scottish Cairngorms and is typically covered in rime ice in the winter (as shown in Figure 2.6 in Chapter 2). Subplots from top to bottom are: precipitation rate per minute (liquid equivalent, mm h^{-1}); liquid versus solid fraction (%); accumulation of liquid, solid and any precipitation since midnight (mm); precipitation type from present weather (PW) codes using the classifier and LuT described in Chapter 3; visibility due to precipitation (km).

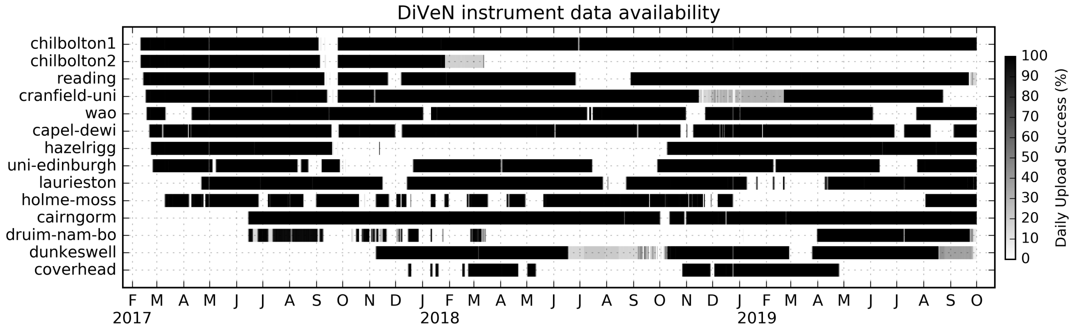
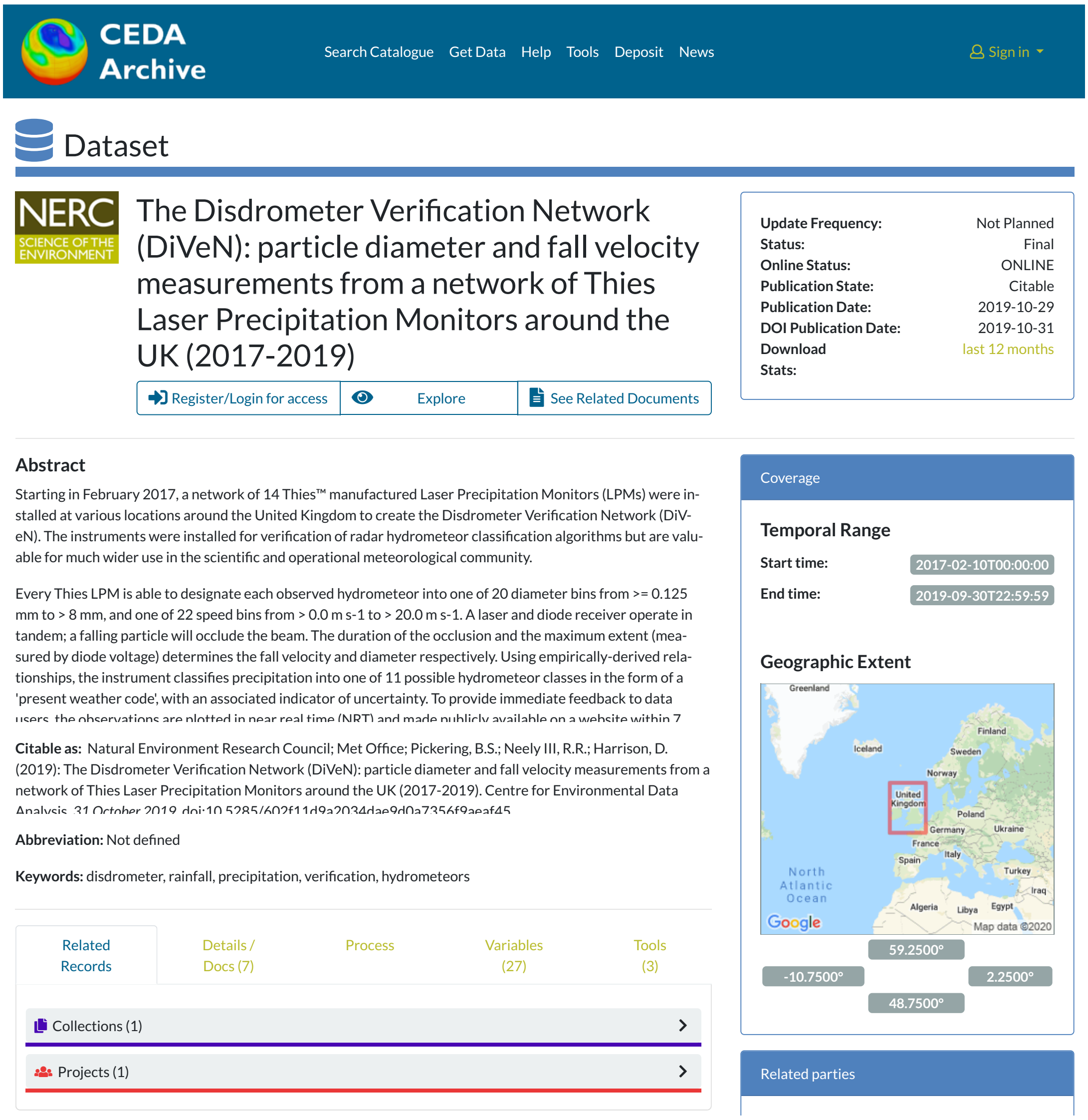


Figure A.4.: Extended version of Figure 2.5 in Chapter 2 showing the longer term availability of the DiVeN instruments, with site names changed to match those with datasets already labelled on the Centre for Environmental Data Analysis (CEDA) archive. Some sites were very successful in uploading data almost for the entire duration of the project, whereas some sites were very poor, which was most commonly due to loss of internet rather than the instrument itself not functioning. Note that the study period of Chapter 4 ends on 31st December, 2018. Also note that although the recurring internet fees were not paid from 2020 onwards, the instruments are still functioning and storing data locally. When the instruments are collected to be ingested by the National Centre for Atmospheric Science for redeployment, any data that was recorded can be converted to netCDF and uploaded to CEDA with minimal effort due to the coding approach taken. Depending on travel restrictions imposed by the novel coronavirus COVID-19, there could be almost 5 years of data collected in total, an invaluable resource for future precipitation studies in the United Kingdom (UK).



Dataset

The Disdrometer Verification Network (DiVeN): particle diameter and fall velocity measurements from a network of Thies Laser Precipitation Monitors around the UK (2017-2019)

Update Frequency: Not Planned
Status: Final
Online Status: ONLINE
Publication State: Citable
Publication Date: 2019-10-29
DOI Publication Date: 2019-10-31
Download Stats: last 12 months

Abstract

Starting in February 2017, a network of 14 Thies™ manufactured Laser Precipitation Monitors (LPMs) were installed at various locations around the United Kingdom to create the Disdrometer Verification Network (DiVeN). The instruments were installed for verification of radar hydrometeor classification algorithms but are valuable for much wider use in the scientific and operational meteorological community.

Every Thies LPM is able to designate each observed hydrometeor into one of 20 diameter bins from >= 0.125 mm to > 8 mm, and one of 22 speed bins from > 0.0 m s-1 to > 20.0 m s-1. A laser and diode receiver operate in tandem; a falling particle will occlude the beam. The duration of the occlusion and the maximum extent (measured by diode voltage) determines the fall velocity and diameter respectively. Using empirically-derived relationships, the instrument classifies precipitation into one of 11 possible hydrometeor classes in the form of a 'present weather code', with an associated indicator of uncertainty. To provide immediate feedback to data users, the observations are plotted in near real time (NRT) and made publicly available on a website within 7

Citable as: Natural Environment Research Council; Met Office; Pickering, B.S.; Neely III, R.R.; Harrison, D. (2019): The Disdrometer Verification Network (DiVeN): particle diameter and fall velocity measurements from a network of Thies Laser Precipitation Monitors around the UK (2017-2019). Centre for Environmental Data Analysis, *31 October 2019*. doi:10.5285/602f11d9a2034dae9d0a7356f9aeaf45

Abbreviation: Not defined

Keywords: disdrometer, rainfall, precipitation, verification, hydrometeors

Related Records | Details / Docs (7) | Process | Variables (27) | Tools (3)

Collections (1)

Projects (1)

Coverage

Temporal Range

Start time: 2017-02-10T00:00:00
End time: 2019-09-30T22:59:59

Geographic Extent

59.2500°
-10.7500° 2.2500°
48.7500°

Related parties

Figure A.5.: The CEDA Archive containing the DiVeN dataset.

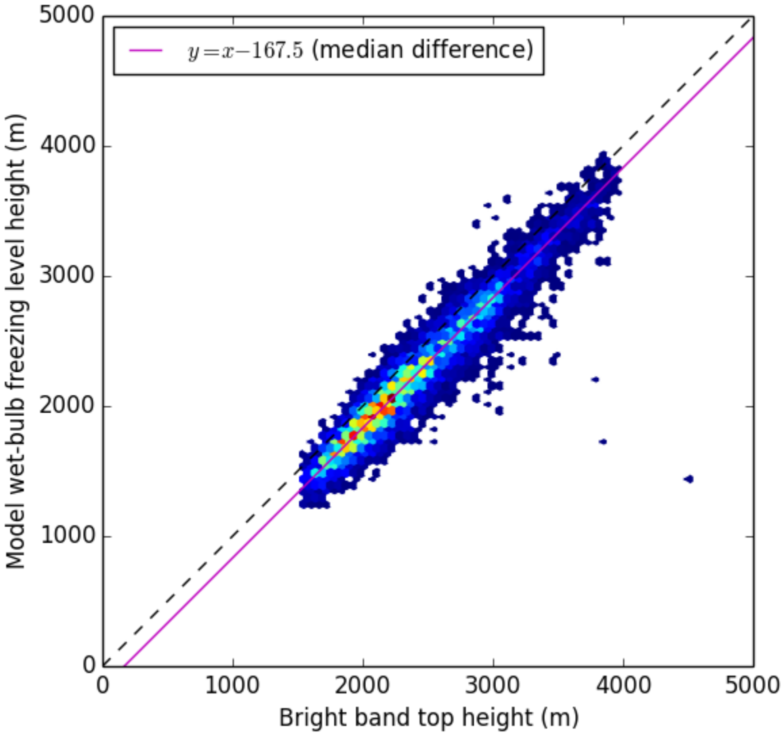


Figure A.6.: The bias between the 0 °C wet-bulb isotherm height from the UK post-processed (UKPP) model and the bright band top height from the melting layer detections of vertically-pointing radar scans. The data were collected between January 2019 and March 2020. This examination performed by colleagues at the UK Met Office was investigated as a result of the verification results of this thesis work in Chapter 4. The offset in median height difference of negative 167.5 m indicates that the UKPP-diagnosed freezing level is too low, which is in agreement with the verification results of Chapter 4 of this thesis.