

**Automating Agriculture: Using UAS and machine learning to monitor weed populations.**

By:

James Peter Thomas Lambert

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

The University of Sheffield

Faculty of Science

Department of Animal and Plant Sciences

Submission Date: 14/12/2018

Supervisors: Robert Freckleton, Dylan Childs

**Acknowledgements**

I would like to thank my supervisors Robert Freckleton & Dylan Childs for giving me the opportunity and putting their faith in me, ‘a guy with a drone’. Their supervision and guidance has given me skills I never thought I would possess, for which I will always be grateful. Helen Hicks deserves a special thank you for showing me the ways of black-grass over the course of this thesis and for putting up with my driving and music choices during the field seasons. To all the other field technicians who have accompanied me over the years, from undergraduates to the wonderful crew from Rothamsted, thank you.

I am grateful to the Jeremy Grantham and team for having the vision and commitment to set up the Grantham Centre for Sustainable Futures and for funding multidisciplinary work. This has allowed me to go and achieve things beyond what’s normally expected of a PhD student, from public engagement, competitions, internships to TV and other media appearances.

To everyone I have met over the course of my time here in the special city of Sheffield, thank you. Whether directly, indirectly, intentional or unintentional you have made my time here the best of my university experience. To the pub club crew thank you. To all my traveling companions; the Brat pack, Beth and Saz, Thank you. Special thanks to all my housemates over the years, plenty of good house meals and parties.

I am unsure how to put into words how thankful I am for my family and the support they have given me over the entirety of my life and schooling career. To Mum and Dad, you are my greatest source of inspiration and your work ethic continually spurs me on. To my sister Hattie, you are amazing, thank you for letting me come along with you on some wonderful travels, from Run Disney to cherry blossom (Sakura) in Japan and Hong Kong. You have already accomplished so much, and I can’t wait to see what you do next. To my lovely grandparents, I hope I am making you proud and that I have been able to communicate my research in such a way that you can then explain it too. To Orlando, thank you and woof!

All models are wrong, but some are useful.

George Box - 1976

**Abstract**

Agriculture currently faces many challenges including a changing climate, the need to produce more nutritionally healthier foods, concerns about environmental impacts and a growing population to provide for. This means that new technologies and methodologies need to be applied to increase the efficiency of agriculture.

*Alopecurus myosuroides* presents a significant challenge to crop production owing to its competitive effects on yields of arable crops in the UK and beyond. Apart from effects on yields, weeds such as this are expensive and time-consuming to control. Precisely knowing where the weed is in a field potentially allows farmers to tailor management practices to the specific site, generating better control of the weed populations year to year.

Here I present an investigation into the use of Unmanned Aerial Systems (UAS) for mapping populations of *A. myosuroides* across multiple locations and years, whilst assessing the use of alternative machine learning techniques to model the data generated from the UAS. I undertook 3 seasons of field walking and UAS flights. This generated the largest known ground-truthed labelled dataset of remotely sensed images of an arable weed.

Using the first year’s data I developed a novel methodology for the combination of ground-based sampling in conjunction with aerial mapping to showcase the applicability of UAS for mapping *A. myosuroides* using linear models and random forests (published in Weed Research).

In the latter two field seasons I advanced the data collection methodology to create a more standardised process, and implemented a Convolution Neural Network, resulting in an improved ability to classify *A. myosuroides* in new previously unseen plots (published in Pest Management Sciences).

Finally, I assessed the use of alternative vegetation indices for the classification of *A. myosuroides* and compared the best performing CNN models to skilled human observers. This showed that Green Normalized Difference Vegetation Index (GNDVI) was the optimum index and that the models performed better than the observers across all vegetation indices.

Overall, this work established:

* A methodology that can be expanded for the use of UAS to map plant populations.
* The suitability of CNN for mapping *A. myosuroides.*
* How data engineering can be used to increase the performance of CNN.
* The need for further analysis into the factors driving the limited transferability of models.

**Declaration of original authorship**

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

James Lambert

**Publications**

Included in this thesis is one published paper and one paper currently under review for publication, each appropriately formatted for their respective journal. I am the lead author for each, with my supervisory team on both papers and postdoctoral researcher in charge of field operations (Dr. Helen Hicks) included in the authors list for chapter 2.

Chapter 2: Lambert, J.P.T., Hicks, H.L., Childs, D.Z. and Freckleton, R.P., 2018. Evaluating the potential of Unmanned Aerial Systems for mapping weeds at field scales: a case study with Alopecurus myosuroides. *Weed research*, *58*(1), pp.35-45.

Chapter 3: Lambert, J. P., Childs, D. Z. and Freckleton, R. P. (2019), Testing the ability of unmanned aerial systems and machine learning to map weeds at subfield scales: a test with the weed Alopecurus myosuroides (Huds). Pest. Manag. Sci.. doi:10.1002/ps.5444.

**Contents**

Acknowledgements. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2

Abstract. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

Declaration of original authorship. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .6

Publications. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

1 Introduction 12

1.1 Crop production and weeds. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

1.2 Remote sensing. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14

1.3 Statistical techniques for big data. . . . . . . . . . . . . . . . . . . . . . . . . 17

1.4 Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .19

1.5 References. . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . . . . .20

2 Evaluating the potential of Unmanned Aerial Systems (UAS) for mapping weeds at field scales: a case study with Alopecurus myosuroides. 25

2.1 Summary. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

2.2 Introduction. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27

2.3 Materials and Methods. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

2.3.1 Study system. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

2.3.2 Weed population monitoring. . . . . . . . . . . . . . . . . . . . . . . 31

2.3.3 Collection of UAS images. . . . . . . . . . . . . . . . . . . . . . . . . .32

2.3.4 Data processing: image stitching. . . . . . . . . . . . . . . . . . . . .33

2.3.5 Data Analysis. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .34

2.3.6 Analysing correlations between weed surveys & imagery. 34

2.3.7 Testing predictive performance of images. . . . . . . . . . . . . .34

2.3.8 Field-to-field predictions. . . . . . . . . . . . . . . . . . . . . . . . . . .35

2.4 Results. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37

2.4.1 Explanatory power of UAS imagery. . . . . . . . . . . . . . . . . .37

2.4.2 Predictive ability. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39

2.4.3 Field-to-field predictions. . . . . . . . . . . . . . . . . . . . . . . . . . .39

2.5 Discussion. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40

2.5.1 Choice of spectral frequency. . . . . . . . . . . . . . . . . . . . . . . .41

2.5.2 Choice of analysis. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41

2.5.3 Automation. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42

2.5.4 Limitations and future work. . . . . . . . . . . . . . . . . . . . . . . . 44

2.6 Acknowledgments. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .46

2.7 References. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .46

3 Testing the ability of Unmanned Aerial Systems and machine learning to map weeds at subfield scales: a test with the weed Alopecurus myosuroides (Huds). 51

3.1 Abstract. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 51

3.2 Introduction. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53

3.3 Materials & Methodology. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56

3.3.1 Description of dataset. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56

3.3.2 UAS platform. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57

3.3.3 Modelling approach and metrics. . . . . . . . . . . . . . . . . . . . . 59

3.3.4.1 Model refinement: data balancing. . . . . . . . . . . . . . . . . . 61

3.3.4.2 Model refinement: data cleaning. . . . . . . . . . . . . . . . . . .62

3.3.5.1 Model transferability: Cross validation. . . . . . . . . . . . . .63

3.3.5.2 Modelling workflow – baseline model. . . . . . . . . . . . . . 64

3.4 Results. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64

3.4.1 Baseline Model. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64

3.4.2 Data Balancing. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65

3.4.3 Data Cleaning. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 66

3.4.4 Analysis artefact data. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 70

3.4.5 Out of Sample predictions - LOFO-CV. . . . . . . . . . . . . . . . 73

3.4.6 Temporal Effects. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74

3.5 Discussion. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74

3.5.1 Concluding remarks. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77

3.6 Acknowledgments. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78

3.7 Appendix. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79

3.8 References. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80

4 Are Humans better than Convolutional Neural Networks across multiple Vegetation indexes for the identification of Black-grass in UAS imagery? 85

4.1 Introduction. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 85

4.2 Materials & Methodology. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 88

4.2.1 Description of dataset. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 88

4.2.2 Vegetation Index creation. . . . . . . . . . . . . . . . . . . . . . . . . .89

4.2.3 Skilled human observer approach. . . . . . . . . . . . . . . . . . . 89

4.2.4 Statistical methods. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .90

4.2.5 CNN modelling approach. . . . . . . . . . . . . . . . . . . . . . . . . . 92

4.2.6 Data Engineering. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .92

4.2.7 Reasons for misclassification. . . . . . . . . . . . . . . . . . . . . . .93

4.3 Results. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .93

4.3.1 Trained Observers. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .93

4.3.2 CNN model. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .94

4.3.3 Error analysis. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 95

4.4 Discussion. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .98

4.4.1 Implications of findings. . . . . . . . . . . . . . . . . . . . . . . . . . . 98

4.4.2 Conclusions. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 100

4.5 Appendix. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 100

4.6 References. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 101

5 Discussion 105

5.1 Summary. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .105

5.2 Single field vs Multi field. . . . . . . . . . . . . . . . . . . . . . . . . . . . .106

5.3 Usage of Vegetation indexes. . . . . . . . . . . . . . . . . . . . . . . . . . .108

5.4 Late season vs Early season. . . . . . . . . . . . . . . . . . . . . . . . . . . 109

5.5 Statistical approaches. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 110

5.6 Conclusion. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .111

5.7 References. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .111

**Chapter 1**

**1. Introduction**

**1.1** **Crop production and weeds**

The most important and widely produced food source for human consumption are cereal crops (Gennari *et al*., 2015). In the UK, as of 2017, the total croppable area was 6.1 million hectares, of which cereal crops were cultivated on 3.2 million hectares. Wheat and barley were the predominant cereal crops at 1.8 and 1.2 million hectares respectively (National Statistics, 2017). In 2017 total wheat imports were 28% higher than 2016 at almost 1.9 million tonnes, the highest annual quantity of imports since the harvest failures of 2013 (Agriculture and Horticulture Development Board, 2018, September 27). This highlights the national gap between demand and crop production for food security (Enghiad *et al*., 2017)

Historical increases in crop production have come from advances in agricultural technologies including irrigation, the use of pesticides and synthetic nitrogen fertilisers (Grassini *et al*., 2013). The green revolution brought about widespread increases in crop yields through improved crop varieties (Conway, 1997). There is however a growing gap between the biological potential yields of crops and on-farm yields, with a national average for wheat production of 8 tonnes per ha, whilst the highest recorded yield is 16 tonnes per ha (van Ittersum *et al*., 2013). This area of uncertainty adds to the pressures of needing to produce more crops with less inputs, more sustainably (Lichtfouse *et al*., 2009).

Weeds are plants growing out of place, (Radosevich *et al*., 2007) and they are the most significant factor in limiting crop yields worldwide (Ziska *et al*., 2011). They represent significant potential yield losses of up to 34%: pathogens and animal pests (the two next biggest threats) when combined, generate the same loss potential, with various environmental variables accounting for the remaining 36% (Cousens, 1985). The most direct ways for weeds to affect crops are via competition, increased production costs or reduced crop quality (Zimdahl, 2018). Baker (1956) identified the key characteristics that would make the ‘ideal’ weed (Baker, 1965). These included rapid vegetative growth and the ability to germinate, grow and reproduce under a wide range of environmental conditions following periods of dormancy (Booth *et al*., 2003).

The most significant weed threatening UK arable crop production is *Alopecurus myosuroides* Huds. (black-grass). It is also found extensively throughout Europe and shows many traits of the “ideal” weed (Twining and Clarke, 2009). Data from the Black-Grass Research Initiative (BGRI) shows that the weed was found in 88% of all 20 x 2m plots surveyed, with significantly higher numbers in the south than the north of the UK (Hicks *et al*., 2018). This mirrors the pattern of arable crop cultivation in the UK. However, climate change predictions indicate an extension of the northern range of black-grass (Asseng *et al*., 2013, Stratonovitch *et al*., 2012) with (Comont *et* al., 2019) showing frequent occurrences in locations in which the weed has been historically absent. The growth cycle of black-grass coincides with that of most common arable crops resulting in competition and consequent yield loss (5%-50%) (Colbach *et al*., 2006). Prior to harvest of the arable crops, black-grass seeds “shatter” resulting in seed shed and dispersal (Beckie *et al*., 2000).

Concurrently, herbicide resistance is driving increased weed abundance at a national scale (Hicks *et al*., 2018). Therefore, poor control of black-grass populations can lead to rapid increases in population sizes over time. Weeds are known to be non-uniformly distributed within fields at all observable scales (Wilson and Brain, 1991, Nordmeyer, 2006). Consequently, mapping the distributions of weeds in fields is a significant area of interest for crop production and food security.

**1.2 Remote Sensing**

Remote sensing is the detection and recording of emitted or reflected electromagnetic radiation with sensors (Turner *et al*., 2003). Over the past fifty years, our ability to classify and estimate changes in vegetation cover has increased. This is due to the increased availability of imagery acquired by sensors on‐board aerial and satellite based platforms (Wulder and Franklin, 2006).

There are multiple properties of a sensor, each utilise the term resolution in a different context, ‘spatial’, ‘spectral’ and ‘temporal’ resolutions (Lillesand *et al*., 2014). Each type of resolution varies from sensor to sensor, and different applications of remote sensing depend on different combinations of these resolutions. Spatial resolution describes the smallest visible ground unit discernible on an image/pixel e.g. 2 meters (Rocchini, 2007). Spectral resolution is the ability to differentiate between spectral features along the Electromagnetic Spectrum (EMS). The spectral resolution needed to differentiate between land and water is low, but in contrast is high when attempting to differentiate between vegetation types (Martin and Aber, 1997). Temporal resolution (revisit time) is as a measure of how frequently images are taken of the area in question (Tarhan, 2007). Application of remote sensing to any problem is therefore a trade-off between spatio-temporal-spectral resolution and cost.

No matter the trade-off of the resolutions, all the sensors generate spectral data. The use of spectral data and associated techniques have created new tools for researchers from a wide range of disciplines to monitor and analyse areas of interest, both on earth and beyond (Des Marais *et al*., 2002).One key aspect of RS is that it creates an unbiased record for comparison and historical analysis, this allows trends, models and future projections to be used in decision making processes (Mustafa., *et al* 2011). The breadth of variety of sensors, sensor types and subsequent applications of them means that not all can be fully explored in this text, however a few notable alternative examples will be highlighted below.

The desired application of the remotely sensed data, i.e. the monitoring of rainforest canopy, determines the selection of sensor type. In the highlighted example of canopy monitoring, a number of differing sensor technologies could be used, utilising various parts of the EMS. One such sensor is Light Detection and Ranging (LiDAR), it is an active sensor in that it emits a portion of the EMS onto the target area and records the reflected values. LiDAR and other such active sensors are less sensitive to prevailing climactic conditions, such as cloud cover and variable lighting due to the active sensor technology. Differences in the time from the emittance of the EMS signal to it being received again by the sensor can be used to generate topographic data, this can be used to study the growth in height of the canopy over time.

Hand held remote sensing equipment has been used for the hyperspectral mapping of Arctic tundra (Davidson., *et al* 2016). Hand held sensors are often used when spatial and spectral resolution are a priority i.e. sampling a small area across many discrete spectral bands (Buchhorn., *et al* 2013).

Different species of plants and different parts of those species each reflect different intensities of the EMS which makes up their spectral signature (Durgante., *et al* 2013). These signatures can be characterised by manual inspection and classification, allowing for inferences to be made about entire tree canopies, without the need for ground based assessments. Remote sensing can be combined with ground truthing via Geographic Information Systems (GIS), and is providing new methods for advanced ecological management by integrating known reference points from ground based observations to then be applied over time and larger ecological scales (Mather and Koch, 2011). However, ground truthing of ecological data is typically challenging to collect at scales relevant to remote sensing applications (Michener and Jones, 2012).

The application of GIS and remote sensing in the context of agriculture has enabled and extended many concepts in the field of Precision Agriculture (PA) and placed it as one of the top ten revolutions in agriculture in the past fifty years (Crookston, 2006). PA has been defined by Mulla (2013) as doing the right management practice at the right place and right time. Precision management practices may involve the application of herbicides, fertilisers and seeds. A review of PA has shown that it increases farm profitability by reducing inputs and increasing outputs (Lambert and Lowenberg-De Boer, 2000, Daberkow and McBride, 2003). The reduction in inputs has also lead to a reduced impact on the environment (McBratney *et al*., 2005). The non-invasive nature and scalability of remote sensing make it a potentially powerful method for mapping weeds. However as with any remote sensing application, this requires certain criteria to be met. For example, the weeds and crop need to be suitably spectrally different and that the sensor has appropriate spatial and spectral resolution (Lamb and Brown, 2001). However, the cost and availability of high-resolution satellite imagery often limits the application in PA (Wu *et al*., 2007). Remote sensing has been used on farms for the estimation of crop yields (Benedetti and Rossini, 1993), drought estimation (Unganai and Kogan, 1998) and to monitor soil erosion amongst other applications (Dabral *et al*., 2008).

At its core PA harnesses new technologies into the methodologies of farm practitioners. One of the largest areas of increased interest has come from the use of Unmanned Aerial Vehicles/Systems (UAV/S) or drones (Zhang and Kovacs, 2012) for remote sensing. The high spatial resolution, ability to attach specific spectral sensors and low operational costs make them a perfect technology for PA and ecological monitoring (Hardin and Hardin, 2010).

UAS have been used in a range of studies, with many encouraging results. Notable studies on wheat and barley crops have demonstrated their potential for in season monitoring, demonstrating a correlation between Vegetation Indices (VI) and commonly used agronomic metrics (Hunt *et al*., 2010, Rasmussen *et al*., 2013). The review by Zhang and Kovacs (2012) offers many examples of UAS in agriculture. Some key examples are for site-specific management of farms (Huang *et al*., 2008), vineyards (Primicerio *et al*., 2012), grass species identification (Hardin and Jackson, 2005) and disease sampling (Schmale Iii *et al*., 2008). Ultimately UAS are another tool to generate data with. This proliferation of data generating tools has led to the era of “big data” in agriculture (Wolfert *et al*., 2017).

**1.3** **Statistical techniques for big data**

IBM estimates that the world produces 2.5x1018 bytes (2.5 million terabytes) of data per day and that the rate of data creation is only predicted to rise (Jacobson, 2013). This data is being produced by a large variety of processes including self-driving cars (Bojarski *et al*., 2016), medical reports (Obermeyer and Emanuel, 2016), online recommendations (Jordan and Mitchell, 2015) and many more. Most of these processes rely on recognizing patterns in observable data in order to understand the data, or make predictions on unseen data (Reddi, 2017). However, many patterns will be commonplace, and others will be contingent on accidental coincidences in the dataset. Additionally, data collection is often imperfect with many missing values. Therefore, algorithms that are used to identify the patterns in these data need to be robust enough to cope with the imperfect data (Witten *et al*., 2016).

The algorithms and processes involved in recognising patterns as outlined above, fall under the umbrella category of Artificial Intelligence and its subfield, Machine Learning (ML). ML can be defined as a set of computational methods using available past information to improve performance or to make accurate predictions (Mohri *et al*., 2012). ML can be applied to a variety of problem classes such as classification, regression and clustering (Bost *et al*., 2015, Robert, 2014, Jordan and Mitchell, 2015). There are a variety of learning methods that can be applied to these problem classes such as Supervised learning, where both the inputs and outputs are known (Kotsiantis *et al*., 2007), and Unsupervised learning when only the inputs are known (Srivastava *et al*., 2015). For image based supervised classification problems, Convolutional Neural Networks (CNN) offer state of the art performance and are now the dominant approach for almost all recognition and detection tasks (Szegedy *et al*., 2015, Taigman *et al*., 2014, LeCun *et al*., 2015).

Conventionally, in order to extract information from remotely sensed images, manual features such as texture, colour or shape have been used as the inputs to models like Logistic Regression (LR) or Random Forests (RF) (Camps-Valls *et al*., 2014, Benediktsson *et al*., 2003). However, when facing the big data of remotely sensed images, these manual features do not capture the true spatial variability of spectral signatures (Zhang *et al*., 2016, Camps-Valls and Bruzzone, 2005). When applying CNN to images, abstract feature maps (non-human defined features) are created by the CNN model and then tuned over the course of refining the model automatically (Zeiler and Fergus, 2014). There have been applications of CNN on remote sensing images to detect manmade objects and classify land use (Montavon *et al*., 2017, Chen *et al*., 2014).An inductive bias is a set of assumptions that are used to predict outputs given new inputs (Mitchell, 1980). The key to choosing an effective inductive bias is having domain knowledge (Sewell, 2017). So, in order to successfully apply a machine learning algorithm to weed mapping, we need to understand the agroecological domain.

**1.4** **Objectives**

My main objective is to use imagery generated from UAS to model density states of *A. myosuroides* across multiple field sites in the UK over multiple years*.*

In chapter 2, I present an exploration of the practicality of using UAS and a modified RGB sensor to assess the presence or absence of weeds within fields and if the imagery of one field can be used to estimate the densities of weeds in another field. I use linear models and random forests to assess these questions respectively.

In chapter 3, I expand both the capabilities of our UAS data collection and statistical methodologies. I also extend the data collection over multiple years. Here I assess the CNN against the full five density states that are generated from the ground truthing of the BGRI data collection.

In chapter 4, I examine the use of different Vegetation Indices that can be generated from this methodology and compared how the CNN models perform against skilled observers for the classification of the presence or absence of black-grass in an image.

**1.5 References**

AGRICULTURE AND HORTICULTURE DEVELOPMENT BOARD. 2018, September 27. *Supply & Demand* [Online]. Available: https://cereals.ahdb.org.uk/media/1440740/201718EndSeasonBalanceSheets.pdf [Accessed].

ASSENG, S., EWERT, F., ROSENZWEIG, C., JONES, J., HATFIELD, J., RUANE, A., BOOTE, K. J., THORBURN, P. J., RÖTTER, R. P. & CAMMARANO, D. 2013. Uncertainty in simulating wheat yields under climate change. *Nature Climate Change,* 3**,** 827.

BAKER, H. G. 1965. Characteristics and modes of origin of weeds. *Characteristics and modes of origin of weeds.***,** 147-72.

BECKIE, H. J., HEAP, I. M., SMEDA, R. J. & HALL, L. M. 2000. Screening for herbicide resistance in weeds. *Weed technology,* 14**,** 428-445.

BENEDETTI, R. & ROSSINI, P. 1993. On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Romagna. *Remote Sensing of Environment,* 45**,** 311-326.

BENEDIKTSSON, J. A., PESARESI, M. & AMASON, K. 2003. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing,* 41**,** 1940-1949.

BOJARSKI, M., DEL TESTA, D., DWORAKOWSKI, D., FIRNER, B., FLEPP, B., GOYAL, P., JACKEL, L. D., MONFORT, M., MULLER, U. & ZHANG, J. 2016. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.

BOOTH, B. D., MURPHY, S. D. & SWANTON, C. J. 2003. *Weed ecology in natural and agricultural systems*, CABI Pub.

BOST, R., POPA, R. A., TU, S. & GOLDWASSER, S. Machine learning classification over encrypted data. NDSS, 2015.

BUCHHORN, M., Walker, D., Heim, B., Raynolds, M., Epstein, H. and Schwieder, M., 2013. Ground-based hyperspectral characterization of Alaska tundra vegetation along environmental gradients. Remote Sensing, 5(8), pp.3971-4005.

CAMPS-VALLS, G. & BRUZZONE, L. 2005. Kernel-based methods for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing,* 43**,** 1351-1362.

CAMPS-VALLS, G., TUIA, D., BRUZZONE, L. & BENEDIKTSSON, J. A. 2014. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine,* 31**,** 45-54.

CHEN, Y., LIN, Z., ZHAO, X., WANG, G. & GU, Y. 2014. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected topics in applied earth observations and remote sensing,* 7**,** 2094-2107.

COLBACH, N., DÜRR, C., ROGER-ESTRADE, J., CHAUVEL, B. & CANEILL, J. 2006. AlomySys: Modelling black-grass (Alopecurus myosuroides Huds.) germination and emergence, in interaction with seed characteristics, tillage and soil climate: I. Construction. *European Journal of Agronomy,* 24**,** 95-112.

CONWAY, G. 1997. The Double Green Revolution. Penguin Books, London, UK.

COMONT, D., Hicks, H., Crook, L., Hull, R., Cocciantelli, E., Hadfield, J., Childs, D., Freckleton, R. and Neve, P., 2019. Evolutionary epidemiology predicts the emergence of glyphosate resistance in a major agricultural weed. *New Phytologist*.

COUSENS, R. 1985. A simple model relating yield loss to weed density. *Annals of applied biology,* 107**,** 239-252.

CROOKSTON, R. K. 2006. A top 10 list of developments and issues impacting crop management and ecology during the past 50 years. *Crop science,* 46**,** 2253-2262.

DABERKOW, S. G. & MCBRIDE, W. D. 2003. Farm and operator characteristics affecting the awareness and adoption of precision agriculture technologies in the US. *Precision agriculture,* 4**,** 163-177.

DABRAL, P., BAITHURI, N. & PANDEY, A. 2008. Soil erosion assessment in a hilly catchment of North Eastern India using USLE, GIS and remote sensing. *Water Resources Management,* 22**,** 1783-1798.

DAVIDSON, S., Santos, M., Sloan, V., Watts, J., Phoenix, G., Oechel, W. and Zona, D., 2016. Mapping Arctic tundra vegetation communities using field spectroscopy and multispectral satellite data in North Alaska, USA. Remote Sensing, 8(12), p.978.

DES MARAIS, D.J., Harwit, M.O., Jucks, K.W., Kasting, J.F., Lin, D.N., Lunine, J.I., Schneider, J., Seager, S., Traub, W.A. and Woolf, N.J., 2002. Remote sensing of planetary properties and biosignatures on extrasolar terrestrial planets. Astrobiology, 2(2), pp.153-181.

DURGANTE, F.M., Higuchi, N., Almeida, A. and Vicentini, A., 2013. Species spectral signature: discriminating closely related plant species in the Amazon with near-infrared leaf-spectroscopy. Forest Ecology and Management, 291, pp.240-248.

ENGHIAD, A., UFER, D., COUNTRYMAN, A. M. & THILMANY, D. D. 2017. An Overview of Global Wheat Market Fundamentals in an Era of Climate Concerns. *International Journal of Agronomy,* 2017.

GENNARI, P., HEYMAN, A. & KAINU, M. 2015. FAO Statistical Pocketbook. World food and agriculture. *Food and Agriculture Organization of the United Nations, Rome, Italy*.

GRASSINI, P., ESKRIDGE, K. M. & CASSMAN, K. G. 2013. Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nature communications,* 4**,** 2918.

HARDIN, P. J. & HARDIN, T. J. 2010. Small‐scale remotely piloted vehicles in environmental research. *Geography Compass,* 4**,** 1297-1311.

HARDIN, P. J. & JACKSON, M. W. 2005. An unmanned aerial vehicle for rangeland photography. *Rangeland Ecology & Management,* 58**,** 439-442.

HICKS, H. L., COMONT, D., COUTTS, S. R., CROOK, L., HULL, R., NORRIS, K., NEVE, P., CHILDS, D. Z. & FRECKLETON, R. P. 2018. The factors driving evolved herbicide resistance at a national scale. *Nature ecology & evolution,* 2**,** 529.

HUANG, Y., LAN, Y., HOFFMANN, W. & FRITZ, B. Development of an unmanned aerial vehicle-based remote sensing system for site-specific management in precision agriculture. Proceedings of 9th International Conference on Precision Agriculture, Denver, Colorado, USA, 2008. 20-23.

HUNT, E. R., HIVELY, W. D., FUJIKAWA, S. J., LINDEN, D. S., DAUGHTRY, C. S. & MCCARTY, G. W. 2010. Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring. *Remote Sensing,* 2**,** 290-305.

JACOBSON, R. 2013. 2.5 quintillion bytes of data created every day. how does cpg and retail manage it. IBM.

JORDAN, M. I. & MITCHELL, T. M. 2015. Machine learning: Trends, perspectives, and prospects. *Science,* 349**,** 255-260.

KOTSIANTIS, S. B., ZAHARAKIS, I. & PINTELAS, P. 2007. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering,* 160**,** 3-24.

LAMB, D. & BROWN, R. B. 2001. Pa—precision agriculture: Remote-sensing and mapping of weeds in crops. *Journal of Agricultural Engineering Research,* 78**,** 117-125.

LAMBERT, D. & LOWENBERG-DE BOER, J. 2000. *Precision agriculture profitability review*, Purdue Univ.

LECUN, Y., BENGIO, Y. & HINTON, G. 2015. Deep learning. *nature,* 521**,** 436.

LICHTFOUSE, E., NAVARRETE, M., DEBAEKE, P., SOUCHÈRE, V., ALBEROLA, C. & MÉNASSIEU, J. 2009. Agronomy for sustainable agriculture: a review. *Sustainable agriculture.* Springer.

LILLESAND, T., KIEFER, R. W. & CHIPMAN, J. 2014. *Remote sensing and image interpretation*, John Wiley & Sons.

MARTIN, M. E. & ABER, J. D. 1997. High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. *Ecological applications,* 7**,** 431-443.

MATHER, P. M. & KOCH, M. 2011. *Computer processing of remotely-sensed images: an introduction*, John Wiley & Sons.

MCBRATNEY, A., WHELAN, B., ANCEV, T. & BOUMA, J. 2005. Future directions of precision agriculture. *Precision agriculture,* 6**,** 7-23.

MICHENER, W. K. & JONES, M. B. 2012. Ecoinformatics: supporting ecology as a data-intensive science. *Trends in ecology & evolution,* 27**,** 85-93.

MITCHELL, T. M. 1980. *The need for biases in learning generalizations*, Department of Computer Science, Laboratory for Computer Science Research ….

MOHRI, M., ROSTAMIZADEH, A. & TALWALKAR, A. 2012. *Foundations of machine learning*, MIT press.

MONTAVON, G., SAMEK, W. & MÜLLER, K.-R. 2017. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*.

MULLA, D. J. 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems engineering,* 114**,** 358-371.

MUSTAFA, A.A., Singh, M., Sahoo, R.N., Ahmed, N., Khanna, M., Sarangi, A. and Mishra, A.K., 2011. Land suitability analysis for different crops: a multi criteria decision making approach using remote sensing and GIS. Researcher, 3(12), pp.61-84.

NORDMEYER, H. 2006. Patchy weed distribution and site-specific weed control in winter cereals. *Precision Agriculture,* 7**,** 219-231.

OBERMEYER, Z. & EMANUEL, E. J. 2016. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine,* 375**,** 1216.

PRIMICERIO, J., DI GENNARO, S. F., FIORILLO, E., GENESIO, L., LUGATO, E., MATESE, A. & VACCARI, F. P. 2012. A flexible unmanned aerial vehicle for precision agriculture. *Precision Agriculture,* 13**,** 517-523.

RADOSEVICH, S. R., HOLT, J. S. & GHERSA, C. M. 2007. *Ecology of weeds and invasive plants: relationship to agriculture and natural resource management*, John Wiley & Sons.

RASMUSSEN, J., NIELSEN, J., GARCIA‐RUIZ, F., CHRISTENSEN, S. & STREIBIG, J. C. 2013. Potential uses of small unmanned aircraft systems (UAS) in weed research. *Weed Research,* 53**,** 242-248.

REDDI, S. J. 2017. *New Optimization Methods for Modern Machine Learning.* Stanford University.

ROBERT, C. 2014. Machine learning, a probabilistic perspective. Taylor & Francis.

ROCCHINI, D. 2007. Effects of spatial and spectral resolution in estimating ecosystem α-diversity by satellite imagery. *Remote sensing of Environment,* 111**,** 423-434.

SCHMALE III, D. G., DINGUS, B. R. & REINHOLTZ, C. 2008. Development and application of an autonomous unmanned aerial vehicle for precise aerobiological sampling above agricultural fields. *Journal of Field Robotics,* 25**,** 133-147.

SEWELL, M. V. 2017. *Application of Machine Learning to Financial Time Series Analysis.* UCL (University College London).

SRIVASTAVA, N., MANSIMOV, E. & SALAKHUDINOV, R. Unsupervised learning of video representations using lstms. International conference on machine learning, 2015. 843-852.

STRATONOVITCH, P., STORKEY, J. & SEMENOV, M. A. 2012. A process‐based approach to modelling impacts of climate change on the damage niche of an agricultural weed. *Global Change Biology,* 18**,** 2071-2080.

SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D., ERHAN, D., VANHOUCKE, V. & RABINOVICH, A. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 1-9.

TAIGMAN, Y., YANG, M., RANZATO, M. A. & WOLF, L. Deepface: Closing the gap to human-level performance in face verification. Proceedings of the IEEE conference on computer vision and pattern recognition, 2014. 1701-1708.

TARHAN, C. I. D. 2007. Detection of environmental and urban change using remote sensing and GIS.

TURNER, W., SPECTOR, S., GARDINER, N., FLADELAND, M., STERLING, E. & STEININGER, M. 2003. Remote sensing for biodiversity science and conservation. *Trends in ecology & evolution,* 18**,** 306-314.

TWINING, S. & CLARKE, J. 2009. Future of UK winter oilseed rape production. *Crop Protection Association Agricultural Industries Confederation: ADAS,* 14.

UNGANAI, L. S. & KOGAN, F. N. 1998. Drought monitoring and corn yield estimation in Southern Africa from AVHRR data. *Remote sensing of environment,* 63**,** 219-232.

VAN ITTERSUM, M. K., CASSMAN, K. G., GRASSINI, P., WOLF, J., TITTONELL, P. & HOCHMAN, Z. 2013. Yield gap analysis with local to global relevance—a review. *Field Crops Research,* 143**,** 4-17.

WILSON, B. & BRAIN, P. 1991. Long‐term stability of distribution of Alopecurus myosuroides Huds. within cereal fields. *Weed Research,* 31**,** 367-373.

WITTEN, I. H., FRANK, E., HALL, M. A. & PAL, C. J. 2016. *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann.

WOLFERT, S., GE, L., VERDOUW, C. & BOGAARDT, M.-J. 2017. Big data in smart farming–a review. *Agricultural Systems,* 153**,** 69-80.

WU, J., WANG, D., ROSEN, C. J. & BAUER, M. E. 2007. Comparison of petiole nitrate concentrations, SPAD chlorophyll readings, and QuickBird satellite imagery in detecting nitrogen status of potato canopies. *Field Crops Research,* 101**,** 96-103.

WULDER, M. A. & FRANKLIN, S. E. 2006. *Understanding forest disturbance and spatial pattern: remote sensing and GIS approaches*, CRC Press.

ZEILER, M. D. & FERGUS, R. Visualizing and understanding convolutional networks. European conference on computer vision, 2014. Springer, 818-833.

ZHANG, C. & KOVACS, J. M. 2012. The application of small unmanned aerial systems for precision agriculture: a review. *Precision agriculture,* 13**,** 693-712.

ZHANG, L., ZHANG, L. & DU, B. 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine,* 4**,** 22-40.

ZIMDAHL, R. L. 2018. *Fundamentals of weed science*, Academic Press.

ZISKA, L. H., DUKES, J. & DUKES, J. S. 2011. *Weed biology and climate change*, John Wiley & Sons.

National Statistics, 2017. *Agriculture in the United Kingdom 2017*. Produced by: Department for Environment, Food and Rural Affairs Department of Agriculture, Environment and Rural Affairs (Northern Ireland) Welsh Assembly Government, The Department for Rural Affairs and Heritage The Scottish Government, Rural and Environment Science and Analytical Services

**Chapter 2**

**Evaluating the potential of Unmanned Aerial Systems (UAS) for mapping weeds at field scales: a case study with *Alopecurus myosuroides***

JPT LAMBERT, HL HICKS, DZ CHILDS & RP FRECKLETON

*Department of Animal & Plant Science, University of Sheffield, Sheffield S10 2TN, UK*

Received 10 May 2017

Revised version accepted 25 August 2017

Subject Editor: José Gonzalez-Andujar, CSIC, Spain

**Running head**: Mapping weeds at field scales using UAS

*Correspondence:* James PT Lambert*,* Dept. of Animal & Plant Science,University of Sheffield, Western Bank, Sheffield, S10 2TN, UK. Tel.: (+44) 114 222 0017; E-mail: jptlambert1@sheffield.ac.uk

**2.1 Summary**

Mapping weed densities within crops has conventionally been achieved either by detailed ecological monitoring or by field walking, both of which are time-consuming and expensive. Recent advances have resulted in increased interest in using Unmanned Aerial Systems (UAS) to map fields, aiming to reduce labour costs and increase the spatial extent of coverage. However, adoption of this technology ideally requires that mapping can be undertaken automatically and without the need for extensive ground-truthing. This approach has not been validated at large-scale using UAS-derived imagery in combination with extensive ground-truth data. We tested the capability of UAS for mapping a grass weed, *Alopecurus myosuroides*, in wheat crops. We addressed two questions: (i) can imagery accurately measure densities of weeds within fields and (ii) can aerial imagery of a field be used to estimate the densities of weeds based on statistical models developed in other locations? We recorded aerial imagery from 26 fields using a UAS. Images were generated using both RGB and Rmod (Rmod 670–750 nm) spectral bands. Ground-truth data on weed densities were collected simultaneously with the aerial imagery. We combined these data to produce statistical models that (i) correlated ground-truth weed densities with image intensity and (ii) forecast weed densities in other fields. We show that weed densities correlated with image intensity, particularly Rmod image data. However, results were mixed in terms of out of sample prediction from field-to-field. We highlight the difficulties with transferring models and we discuss the challenges for automated weed mapping using UAS technology.

**Keywords**: precision agriculture, site-specific weed management, wheat, modelling, distribution, integrated weed management, black-grass, drones*.*

**2.2 Introduction**

One of the problems with managing weed populations is that weeds are non-uniformly distributed at almost every spatial scale at which we study them. Weeds are undoubtedly patchily distributed within fields (Wilson & Brain, 1991; Nordmeyer, 2006). At the higher scales of fields, farms and landscapes, there can also be considerable variations in weed abundance (Thornton *et al*., 1990; Gabriel *et al*., 2005). Indeed, even at the national scale, some regions contain more weeds than others (Marshall, 2009). Such variations reflect the combined imprint of environment and management history (Fried *et al*., 2008). This non-uniform distribution of weeds has long been recognised and, for over a century, attempts have been made to understand the factors that contribute to variation in weed distributions; for example, as long ago as 1913, Brenchley (1913) attempted to understand how soil and management contribute to variation in the occurrence of weed species in the UK.

Understanding the distributions of weeds requires that they are monitored. Monitoring of weed populations has typically focused on small scale detailed studies. For example, a literature search focusing on weed surveys prior to 2008, showed that 84% of all previous field plots were smaller than 1 m2 (Queenborough *et al*., 2011). Moreover, monitoring effort is typically limited in terms of the number of observers available, so that most sampling protocols sacrifice spatial scale for intensity. This means that effective sampling areas may be relatively small; for instance, in one of our previous demographic studies, we estimate that the sampled area was only 3% of the total experimental area (total experimental area = 36 x 48 = 1728 m2; monitored area = 48 m2; (Lintell Smith *et al*., 1999)). Limited sampling of this sort presents many issues, because systems can vary dramatically both spatially and temporally (Craufurd & Wheeler, 2009).

Large-scale mapping has been undertaken to build up pictures of weed distributions at regional and landscape scales (Lawrence *et al*., 2006; Barnett *et al*., 2007; Cuneo *et al*., 2009). These analyses are usually based on coarse estimates of weed abundance. In the coarsest form, there are atlas measurements of occurrence at a scale as large as 10 x10 km (Preston *et al*. 2002). Even at this scale, data are useful in analysing large-scale geographic drivers of occurrence such as climate (Storkey *et al*., 2014). Field-scale estimates of occurrence (presence/absence) or prevalence (density) have also been used to build up large-scale pictures of the abundance of weeds (Joseph *et al*., 2006). Such data are extremely valuable in generating insights into the factors that drive weed abundance (Henckel *et al*. 2015; Westbury *et al*. 2008)

Mapping weed densities is thus a trade-off between precision and extent; fine-scale ecological monitoring generates detailed data on small scales, while large-scale surveys generate coarse data across large extents. To bridge this gap, Queenborough *et al*. (2011) developed Density Structured monitoring approaches for arable weeds. This approach generates field-scale maps of weed distributions. Within-field mapping is relatively coarse (a 5-point scale for assigning density states within large plots of size 20 x 20 m), but the approach is readily up-scaled to hundreds of fields during a field season for a small team (e.g. 2 or 3 observers). Based on readily available resources (i.e. field walking / monitoring in small teams) this represents a compromise approach that generates large numbers of within-field maps at among-farm and regional scales. Data from such monitoring can be used to parametrise predictive ecological models (Freckleton *et al*., 2011) and henceforth be useful in solving a key problem that many models face, lack of comprehensive data (Tredennick *et al*. 2017).

Measuring weeds in an agricultural setting is undoubtedly useful for the monitoring and management of farm systems (Huang & Asner, 2009), but arguably limited by the trade-off between precision and extent. However, recent technological advances have resulted in a step change in the potential to collect detailed ecological data at large scales. Unmanned Aerial Systems (UAS) are flying robots that can collect varied data, including colour and hyperspectral imagery allowing vegetation indices to be constructed, as well as environmental data (Nonami, 2007; Torres-Sánchez *et al*., 2014). Prior to the introduction of UAS, satellites and manned aircraft were the only way of capturing aerial imagery of landscapes, with numerous applications in the monitoring and management of ecological systems (Kerr & Ostrovsky, 2003; Pettorelli *et al*., 2005). There have been attempts to use imagery generated by such platforms to map weeds on the field scale but poor resolution of the imagery has previously limited its application (Lamb & Brown, 2001; Thorp & Tian, 2004).

A typical hobby grade UAS will have a pixel resolution of 2.8 cm pixel-1 when flown at 100 m altitude, flight time of 25 minutes and cost around €1000, therefore providing high resolution and low-cost imagery for small survey areas. Compared with field walking and conventional ecological monitoring techniques, this technology offers considerable potential for addressing the trade-off between resolution and extent. Consequently, there has been a substantial increase in interest in the use of UAS for mapping in the precision agriculture sector (Zhang & Kovacs, 2012).

Although UAS offer great potential to produce detailed data over large spatial extents, ultimately, they will be useful in research and management only if they can be shown to generate large amounts of reliable data. There have been attempts to use remote sensing to map populations in detail (Huang & Asner, 2009) but they have been limited in spatial and temporal scales (López-Granados, 2011; Rasmussen *et al*., 2013). Nevertheless, there is significant commercial interest in the applications of UAS in agriculture. However, this interest has largely not translated beyond a proof of principle with some of the commercial ventures over promising, under delivering and subsequently failing (Catapult Satellite Applications, 2016)

Ultimately, for UAS-derived imagery to be useful in weed monitoring, several conditions must be fulfilled. First, most importantly, it must be shown that imagery correlates closely with weed density. This is an obvious requisite for this technology to be practically useful. However, this is very difficult to test because to do so requires density data from many fields at fine spatial resolution to ground truth the imagery from UAS. As noted above, such data are difficult to acquire. Second, especially for management applications, the pipeline of data processing from image capture to weed density maps should include as few steps as possible. This is so that the technology is readily useable. Third, the platform and imagery should ideally be robust to variations between fields and observation conditions, so that minimal tuning or subjective interpretation by analysts are required. These conditions can be relaxed to varying degrees if additional local and context specific data are available. For example, variations in observation conditions (e.g. weather, light, soil, crop colour) can be accommodated by collecting ancillary data for calibration. However, this requires extra steps in data collection and analysis that may prove challenging or impractical in many applications. To date, although UAS are increasingly being used in field situations, the answers to these questions are largely unknown.

In this paper, we explore the potential for simple inexpensive UAS to acquire images that can be used in weed mapping. We focus on the use of readily available “off-the-shelf” systems that can be used by researchers, agronomists and farm managers for quantitative analysis of weed distributions. The first major question we address is whether imagery from such platforms is capable of measuring weed densities? To do this we combine imagery from UAS with an extensive dataset on weed populations across 26 fields. The second question is whether we can use models transferably across fields? We address this by developing statistical models that relate imagery and weed densities for one field and asking whether these accurately transfer to other sites. We show that in principle UAS-derived imagery closely relates to weed densities. However, we highlight various challenges in automating the collection and analysis of data.

**2.3 Materials and Methods**

*2.3.1* *Study system*

The weed *Alopecurus myosuroides* Huds. (black-grass) in winter wheat (*Triticum aestivum* L.) was chosen as a study system. This weed species has significant economic impacts on crop yields and is prevalent throughout northern Europe (Twining & Clarke, 2009). Black-grass has been shown to significantly reduce yields when present (Blair *et al*., 1999) and infests approximately 70% of wheat fields in the UK (Black-Grass Research Initiative, BGRI unpublished data).

We selected study sites that included both large and small farm sizes and differing crop rotations within each of the following five geographical regions in the UK: Oxfordshire, Bedfordshire, Norfolk, Lincoln and Yorkshire. Farm size varied from 80 to 3000 ha. Crop rotations varied from continuous cropping of winter wheat, to an 8-crop rotation. *Alopecurus myosuroides* populations were censused from the 1st of June 2015 to the 27th of July 2015, during which time the weeds were mature and visually distinct, corresponding to 87-89 on the BBCH scale respectively for the cereal crop (Lancahsire *et al*., 1991). In this period, 26 fields were surveyed across the five regions. This sample of 26 fields is by far the largest dataset on within-field weed distributions to have been used to assess the effectiveness of UAS technology. However not all 26 georectified maps were suitable for full analysis, due to poor data quality, resulting in 18 full fields suitable for full analysis.

*2.3.2 Weed population monitoring*

We used the density structured approach (Taylor & Hastings, 1998) implemented by Freckleton *et al* (2011) and Queenborough *et al*. (2011), in which five discrete density states (absent, low, medium, high, very high) were used to estimate *A. myosuroides* plant numbers. These discrete density-structured observations have been shown to be representative of counts of plants (Freckleton *et al*., 2011; Queenborough *et al*., 2011). The advantage of the density-structured approach over individual plant head counts is that it allows populations to be estimated very rapidly, permitting data to be collected at far greater spatial extent while also reducing fieldwork costs. Existing research suggests that misclassification between observers of density states is negligible (Collett, 2002).

Plots were 20 x 20 m, which is a convenient scale for monitoring (Queenborough *et al*., 2011). Surveys were performed by a team of 3 trained observers and the outcome of surveying on each field is a grid of density-state measurements of the whole field (see Fig. 1 for an example). The five density states were assigned using the quartiles of densities determined in the Farm Scale Evaluation of GM crop trials (Heard *et al*., 2003). The five density states counted *A. myosuroides* plants per 20 m2 in bands of: 0, 1-160, 161-450, 451-1450 and 1451+ respectively for Absent, Low, Medium, High and Very High-density state observations.

*2.3.3 Collection of UAS images*

To collect the UAS imagery data, we used a commercially available DJI Phantom 2 (Austin, 2010). Two cameras were used to collect images. Firstly, a modified GoPro Hero3 ("GoPro Official Website - Capture + share your world - HERO3+." https://gopro.com/update/hero3\_plus. Last accessed 24 February 2017) with a filter (https://event38.com/product/custom-ngb-filter-glass-for-diy-camera-conversion/) was used to capture modified colour aerial images and a 16.5 mm focal length, non-fisheye lens was installed to reduce the image distortion (RmodGB: blue, B: 390–520 nm; green, G: 470–570 nm; red-edge, Rmod 670–750 nm). Such images have been shown to be useful for mapping in an agricultural context (De Castro *et al*., 2015). Secondly, a Canon s100 ("Canon PowerShot S100 Black Refurbished | Canon Online Store." https://shop.usa.canon.com/shop/en/catalog/powershot-s100-black-refurbished. Last Accessed 24 February. 2017) was used to provide RGB images with focal length set to 24mm. Spectral data can be found via the respective online sources. The images were stored in RAW format and the cameras were triggered to capture images every 1.5 seconds via software control. White balance was set using a calibration card prior to each flight. The flights were flown autonomously in a grid pattern that used a 60% side and front overlap at a height of 100 m, this ensured optimal coverage of the target (Ballesteros *et al*., 2014). The average area covered over the 30 flights was 5.32 ha, an average Ground Sampling Distance (GSD) of 3.2 cm pixel-1 and an average flight time of 11 minutes.

*2.3.4 Data processing: image stitching*

Individually, each image represents a limited view of the field. For field scale analysis, it is necessary to combine these subsamples into one image of high quality. We used a commercial desktop solution for this Agisoft ("Agisoft PhotoScan." <http://www.agisoft.com/>. Last accessed 1 February 2017). We then cropped the UAS imagery to the extents of the accompanying density state grids using the georeferenced orthomosaics on a field-by-field basis. We manually applied a soil thresholding mask, to cut out pixels that were observed to be soil on a field-by-field basis to remove the pixels of soil that are present in the tramlines or patches of bare ground in the field that could introduce a bias. This was done in imageJ ("ImageJ - RSB Home Page." 2016. <https://imagej.nih.gov/ij/> last accessed 25 October 2016) by visual inspection of the amount of bare soil visible in each image. We then combined the data sets, so that every pixel had their respective 3 spectral band values, a location in geospace and an observed density state which was dependent on its location within the field.

*2.3.5 Data Analysis*

*2.3.6 Analysing correlations between weed surveys & imagery*

The objective of the first set of analyses was to assess the ability of the mean pixel values of the 20x20m plots for the respective spectral bands to capture explained variation in density states. A series of multiple linear regression models were fitted and then used to predict density states. A least squares model was fitted to the RGB data set using the spectral bands Red, Green and Blue as the predictors and observed density state as the response variable. A second regression model for Rmod used the spectral bands red-edge (Rmod), Green and Blue as predictors, and observed density state as the response variable.

*2.3.7 Testing predictive performance of images*

The second set of analyses was designed to test predictive performance of statistical models fitted to imagery. We used a random forest classifier to evaluate the spectral data’s ability to discern weed densities. A random forest model is an ensemble learning method that utilises Classification and Regression Tree (CART) analysis (Breiman, 2001). The model was applied in two ways: (i) to predict the presence/absence of black-grass and (ii) to discriminate between areas of High and Very High *A. myosuroides* observations. We used the same spectral bands as the linear models for the respective data sets and fitted the random forest model with 32000 trees. The spectral data sets were split into training and testing data at the 20 x 20 m scale, with the training data being used to build a random forest model and the testing data being used for predicting against. The data was split 80/20 respectively.

Area under the Curve (AUC) and Accuracy (ACC) were used as metrics to test the ability of the random forest model to predict the presence/absence of *A. myosuroides*. AUC is a measure of the area under a ROC (Receiver Operating Characteristic**)** curve and is an alternative measure of goodness of fit. ACC is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (Fawcett, 2006). This metric is important for assessing the predictive ability of the models.

*2.3.8 Field-to-field predictions*

The aim of this analysis was to measure the predictive performance of models by testing the extent to which a model developed in one field could be used to predict densities of weeds in other fields. We selected the Rmod data set for further analysis as this generated the best correlations between observed and fitted density from the linear regression models. To test the field-to-field predictive ability, we fitted a cubist model for each of the 8 individual fields for which Rmod data were available (Table 1). Cubist models are rule–based models that are created in a similar way to the random forest models but the terminal leaves contain linear regression models (Quinlan, 1992), thus allowing comparison to the initial analysis. Cubist models were chosen as they provide an ensemble classifier approach, resulting in an average prediction for the ensemble, as opposed to the single snapshot of the previous models. These models were constructed using all the data for each individual field and then used to predict the density states of the remaining fields. We assessed the performance of these models by recording the correlations between the predictions and the independent ground-truthed observations.

**Table 1** Explanatory power of imagery acquired by UAS to describe weed densities within fields. R2 values from the fitted regression models of density state as a prediction of the spectral bands are shown for the individual fields and for the entire data sets, RGB and Rmod respectively. The random forest results show the ability of the data to predict the Presence/Absence (P/A) of black-grass using the metrics Area Under a Curve (AUC) and Accuracy (ACC). A random forest model was also used to discriminate between High and Very High (H/VH) levels of black-grass using the same metrics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| RGB | Linear Model | | | Random Forest | | | |
| Field Number | R2 | *P*-value | df | P/A AUC | P/A Acc | H/VH AUC | H/VH Acc |
| 1 | 0.1568 | 0.0015 | 134 | 0.9140 | 0.8390 | N/A | N/A |
| 2 | 0.0344 | 0.4195 | 200 | 0.4354 | 0.4354 | N/A | N/A |
| 3 | 0.0308 | 0.8234 | 113 | 0.5167 | 0.5069 | N/A | N/A |
| 4 | 0.1670 | 0.0013 | 127 | 0.6923 | 0.5618 | N/A | N/A |
| 5 | 0.1549 | 0.0000 | 234 | 0.8357 | 0.6027 | 0.3333 | 0.8165 |
| 6 | 0.1305 | 0.0135 | 124 | 0.7452 | 0.5707 | N/A | N/A |
| 7 | 0.0836 | 0.0126 | 202 | 0.8743 | 0.6781 | 0.7598 | 0.5849 |
| 8 | 0.0270 | 0.6304 | 189 | 0.5654 | 0.5319 | N/A | N/A |
| 9 | 0.4555 | 0.0000 | 94 | 0.8397 | 0.8094 | N/A | N/A |
| Overall | 0.2937 | <2.2E-16 | 1481 | 0.8828 | 0.6827 | 0.9073 | 0.8658 |
| Rmod | Linear Model | | | Random Forest | | | |
| Field Number | R2 | *P*-value | df | P/A AUC | P/A Acc | H/VH AUC | H/VH Acc |
| 10 | 0.0596 | 0.1127 | 187 | 0.5807 | 0.5215 | N/A | N/A |
| 11 | 0.2533 | 0.0000 | 128 | 0.9281 | 0.6238 | 0.7346 | 0.6072 |
| 12 | 0.1528 | 0.0003 | 163 | 0.7547 | 0.5869 | N/A | N/A |
| 13 | 0.4577 | <2.2E-16 | 146 | 0.9152 | 0.7186 | 0.8692 | 0.6822 |
| 14 | 0.2372 | 0.0006 | 92 | 0.6908 | 0.6321 | 0.8153 | 0.5545 |
| 15 | 0.1289 | 0.0347 | 107 | 0.4635 | 0.4881 | N/A | N/A |
| 16 | 0.0729 | 0.1006 | 156 | 0.5759 | 0.5365 | N/A | N/A |
| 18 | 0.1365 | 0.0001 | 212 | 0.6899 | 0.5739 | N/A | N/A |
| Overall | 0.4132 | <2.2E-16 | 1247 | 0.8008 | 0.6373 | 0.9500 | 0.6069 |

**2.4 Results**

*2.4.1 Explanatory power of UAS imagery*

Examples of the different types of image that we used for building the models were produced by stitching the individual images together to form one analysable image (Fig. 1). Visual comparisons of Fig. 1(a, b, c) indicated that, visually at least, the variation was greater in the Rmod images. The grid overlay (Fig. 1c) represents the ground-truthed data that we use for training the models.

(a) (b) (c)

Figure 1. For illustrative purposes, this field was flown twice, and the camera was changed for each flight. With (a) greyscale colour enhanced Rmod and (b) RGB the resulting outputs. This allows for side by side visual comparison of the image data, with the same underlying level of black-grass, (c) overlay of the ground truthed observed density states. The legend corresponds to the accompanying density states that were recorded, ranging from Very High (v) to Absent (0).

We found that the variation within the images obtained from the UAS correlates with weed densities measured in the field. The accuracy of the method however varies with the dataset used (Fig. 2) and between fields (Table 1). The linear model can characterise the relationship broadly across all the spectral bands. The RGB data performs well in some fields, however overall the relationship between density states and the mean pixel values per 20 x 20 m plot is weaker, with a R2 value of 0.29 compared with the Rmod R2 value of 0.41 as evidenced in Fig. 2. Overall, we find that the Rmod data set has the largest R2 value (0.41) when fitted to the entire data set, as well as the best performing individual field (0.46).



Figure 2. Fits of density state, against ground truthed observed data for the Rmod (a) and RGB (b) data sets respectively. The models were trained on 80% of the data and then tested against the remaining data for the predictions. Fits were generated from the linear regression models (see text for details).

*2.4.2 Predictive ability*

We used the random forest models described in Table 1 to test whether we could distinguish between areas with (Presence) and without (Absence) *A. myosuroides*. The RGB data set performed best overall (AUC = 0.88, Acc = 0.68). We also tested the ability of the random forest models to differentiate between areas of High and Very High levels of *A. myosuroides*. Most fields being surveyed in the 2015 field season did not contain the full range (Absent to Very High) of *A. myosuroides* levels, thus we have relatively fewer data points to test this capability with. Nevertheless, the models still show a strong ability to distinguish between areas of High and Very High levels of *A. myosuroides*. The Rmod data set has the highest AUC (0.95) but the corresponding accuracy (0.61) is the lowest of the datasets, this is an important factor to consider due to the lack of data and potential for false-positives. The RGB data set shows a lower AUC (0.91) than Rmod but with a higher accuracy (0.87).

*2.4.3 Field-to-field predictions*

The heatmaps in Fig. 3 summarise the overall analysis of inter-field predictions. Each cell in the respective matrices represents a correlation coefficient of the observed density states and predicted states from the cubist models that have been trained on only one field’s worth of data. The results from this analysis are mixed. Although some correlations are relatively high, the average correlation for all the models was relatively weak (0.34). This suggests that the cubist models were locally over-fitting the relationship between density state and the spectral signal, resulting in poorer field-to-fieldtransferability of the models.



Figure 3. Heat map matrix, prediction correlation plots for a cubist model derived from Field 1 to Field 8 on the axis respectively from the Rmod dataset. High correlation values indicate higher prediction accuracies of density states. The darker the cell, the higher correlation between the models predictions and the observed density state. White cells indicate NA’s, these occur when the trained model did not predict a density state for every class that was present, this is due to the lack of that density state being present in its training data.

**2.5 Discussion**

Our main finding was that aerial images collected with a low cost UAS (<€1000) have the potential to be used to map populations of *A. myosuroides*. However, our results indicate that if this technology is to be applied at a large scale in an automated way, then there are several issues that need to be addressed. Secondly, our analyses of within-field variation using simple statistical models show that it is possible in principle to capture the variation in weed densities. However, models developed in one field rarely perform well when applied elsewhere, indicating that locally they were over-fitting the relationship between density state and the spectral signal. This means that currently the interpretation of such imagery is limited without supporting ground-truthed data; the ultimate objective of our research is to be able to generate estimates of densities from imagery without the need for detailed ecological surveys. Year on year transferability is currently being assessed. We have highlighted that there are challenges in generating robust predictive models that relate variation within images to weed densities within-fields, yet are applicable across multiple sites. Our work has revealed areas that need to be streamlined for the methodology to become more of a tool for management applications.

*2.5.1 Choice of spectral frequency*

We found the most informative spectral frequency to be Red-edge (Rmod). Of the sets of spectral bands we tested, Rmod captured the relationships between the pixel values and ground observations of *A. myosuroides* density state most accurately. There is an extensive literature on the uses of indices such as NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced vegetation index) with the use of satellite data (Xie *et al*., 2008; Pettorelli, 2013). These indices have been used in conjunction with some UAS studies, though they have mainly been proof of concept, due to technical limitations and their focus on small scale, high value crops, such as vines (Turner *et al*., 2011; Primicerio *et al*., 2012), whilst rarely addressing ecological monitoring problems.

*2.5.2 Choice of analysis*

Torres-Sánchez (Torres-Sánchez *et al*., 2014) used UAS to map weeds in an agricultural setting, although they primarily focused on capturing the aerial images in early season for the crops. This means that there are discernible rows of the crop from time of planting. Object Based Image Analysis (OBIA) has been the most commonly used methodology to detect weeds when studying this type of data (Pena *et al*., 2013). This approach is useful in the management of weeds in the early part of the season and has applications informing in-season decisions. Late in the season, rows are not discernible in crops like cereals, which have dense overlapping canopies when mature and hence these approaches are likely to be less useful.

Our approach focused on late season imaging of the crops. This reflects in part the purpose of our original modelling framework (Freckleton *et al*. 2011; Queenborough *et al*. 2011), which was to parameterise ecological models to project future weed densities. Monitoring late season weed numbers should give insight into where the weeds will emerge next year due to seed set. Rather than inform current management practices, such information can be used to make decisions in subsequent seasons, such as patch-spraying (Audsley, 1993) or variable sowing densities (Chauhan *et al*., 2011). The two approaches (late season versus early season monitoring) can work in conjunction with one another. One useful application of combining approaches would be to check the effectiveness of the management decisions previously made. However, the technical challenges of monitoring at different times are likely to be somewhat different.

*2.5.3 Automation*

To be of general use in both research and management, the process should be as automated as possible, requiring minimal interventions by the analyst. However, this requires that several key problems are solved. Most notably, as indicated by our results, images vary from field-to-field, so that the relationship between density and image intensity is not fully transferable from one field to another. Increasing the comparability of images is thus a key priority, for example through accounting for variable lighting and by standardising spectra.

A key assumption of image interpretation is that we are detecting *A. myosuroides*. In the current analysis, we have specifically focused on *A. myosuroides* and we have extensive ground-truth data to test the ability of imagery to detect this. In an automated system, we would ideally be able to proceed with minimal ground-truth data. The extent to which variation resulting from, for example, poor crop establishment, other weeds or disease, rather than the presence of *A. myosuroides* is unknown. In terms of in-field management this may not matter; output could still be informative to the farmer and agronomist. Variation in image intensity within field maps would act as ‘signposts’ to areas of the field that we have shown to be different from the normal crop. They would then be able to field-walk specific locations. This means that ground-truthing of the maps is still required to detect what the actual causes of the variation in the field are and automation would not be achieved. However, for the purposes of wide-scale mapping for larger areas or as a research tool, it will be important to examine how distinct factors can be distinguished. For example, yellowing of a wheat crop owing to disease such as *Puccinia striiformis* f.sp. *tritici* (yellow rust)(Moshou *et al*., 2004) should be distinguishable from *A. myosuroides* based on spectral characteristics.

When looking at ways to automate data collection, one important issue is setting a threshold for the detection of soil. In our current methodology, we manually set the threshold for each field by viewing the histogram of the pixel intensities in imageJ and then manually removing the pixels that fell below a cut off value. This analytical step could be improved using several approaches. For example, an OBIA system could be used to detect tramlines and then set an applicable boundary either side of each track to mask all the soil pixels from the analysis. Alternatively, a clustering-based image thresholding technique such as Otsu’s method could be applied (Shorter & Kasparis, 2009). The challenge is to determine how such an algorithm flexibly accounts for differences in soil colour between fields.

We find similarities between this work and that of Dvořák *et al* (2015), in that they used UAS to map alien invasive species using pixel based classification. They also highlighted the challenges of unstable scene illumination, an issue that individual field analysis presented. By compiling all the respective grid square data into one data set, rather than the current field by field data sets, we hope to mitigate some of the variation introduced by the unstable lighting conditions. New sensor technologies to combat issues such as this are constantly being released; one example is the recent announcement of an integrated imaging system and sun irradiance sense from MicaSense called Sequoia.

*2.5.4 Limitations and future work*

The limitations of this technology and methodology are that it is not completely independent of field walking to gather the ground-truthed data. The statistical methods used here are relatively unsophisticated and are potentially not utilising all the features of the current data. The current feature design of only using the mean pixel value for each 20 x 20 m grid is rudimentary, so in further studies we would include more features, like spatial correlation and field management histories. Nonetheless, this methodology has potential to amplify the work of field surveying, allowing data to be gathered on a scale that is currently unachievable for a small team. A team of field surveyors can produce a more accurate map of *A. myosuroides* than our current UAS method. Indeed, such data can be entered onto a computer at the time of mapping and a field-scale map generated that, if an internet connection is available, can be immediately uploaded and distributed. In contrast, the analysis of UAS-derived data requires several steps, including image stitching that can take several hours of computational time.

The advantages of using UAS are in terms of scale and a minimal analysis needed to assess *A. myosuroides* levels. There is generally expected to be a trade-off between extent of measurement and precision, and this is true for arable weeds (Marshall 1988). As we have shown recently, relatively coarse data can be extremely valuable for measuring weed populations, if they are available at sufficiently large scales (Queenborough *et al.* 2011; Freckleton *et al.* in revision). In the case of imagery from UAS, it is potentially possible to generate finer-scale maps than can be generated using techniques such as the field walking methods of Queenborough et al. (2011), and at greater speed. Hence there is the potential for UAS-derived imagery to allow a step change in the extent and accuracy of data collection.

There has been work to integrate the use of UAS into Site Specific Weed Management (SSWM) as the UAS allows for efficient and repeatable collection of spatial data (Torres-Sánchez *et al.*, 2013). Their study set out to describe the technical specifications and configuration of a UAS that can be used in SSWM. Farmers already use *A. myosuroides* maps, such as those generated by our ground-truthed data, to implement variable seed-rate planting (Helen Hicks pers. obs.). This allows farmers to plant crops at a higher seed density in areas known to have had high weed burdens in the previous year. The aim of this is to outcompete *A. myosuroides* in the early stages of germination, resulting in less *A. myosuroides* setting seed (Timmermann *et al*., 2003). The development of UAS-based weed mapping systems has the potential to provide weed maps more quickly and at a lower cost to the farmer. It is also important to understand that this work is tackling one of the most challenging issues in the field of weed mapping, identifying one mature grass within another mature grass and therefore there may be an upper limit in prediction accuracy.

In addition to developing technology that could be used for informing agronomic decision making, development of these data collection and processing techniques are important for research. A major factor in collection of population monitoring data is the difficulty in collecting enough data for model development within time and budget constraints *(*Bryson *et al*., 2014). The new methodology developed here, using UAS to collect highly detailed images of populations and building predictive statistical models, could potentially be applied to many population monitoring studies, such as rangeland and invasive weed mapping (Rango *et al*., 2009; Hung *et al*., 2014). However, our results indicate that there are obstacles to be overcome particularly if we are to avoid extensive ground-truthing and be able to readily apply such methodology to different fields and farms.

**2.6 Acknowledgements**

JL is funded by a studentship from the Grantham Centre for Sustainable Futures. Collection of field density data was funded by the BBSRC (BB/L001489/)

**2.7 References**

Audsley E (1993) Operational research analysis of patch spraying. *Crop Protection* **12**, 111-119.

Austin R (2010) Unmanned Aircraft Systems - UAVS Design, Development and Deployment. *Development* **54**, 332-348.

Ballesteros R, Ortega JF, Hernández D & Moreno MA (2014) Applications of georeferenced high-resolution images obtained with unmanned aerial vehicles. Part I: Description of image acquisition and processing. *Precision Agriculture,* **15**, 579-592.

Barnett DT, Stohlgren TJ, Jarnevich CS *et al*. (2007) The art and science of weed mapping. *Environmental Monitoring and Assessment* **132**, 235-252.

Blair AM, Cussans JW & Lutman PJW (1999) A biological framework for developing a weed management support system for weed control in winter wheat: Weed competition and time of weed control. In: *Proceedings of the 1999 Brighton Conference - Weeds.* 15-18 November 1999. Hilton Brighton Metropole Hotel, UK.

Breiman L (2001) Random Forests. *Machine Learning* **45**, 5-32.

Brenchley WE (1913) The Weeds of Arable Land. III. *Annals of Botany* **27**, 141-166.

Bryson M, Reid A, Hung C, Ramos F & Sukkarieh S (2014) Cost-Effective Mapping Using Unmanned Aerial Vehicles in Ecology Monitoring Applications. *Experimental Robotics SE* **35***,*  509-523.

Catapult Satellite Applications (2016) Unmanned Aerial Vehicles (UAVs) for UK Agriculture: Creating an Invisible Precision Farming Technology, 2016. https://sa.catapult.org.uk/wp-content/uploads/2016/07/White-paper-UAVs-and-agriculture\_Final2.pdf (last accessed 28 February 2017).

Chauhan BS, Singh VP, Kumar A & Johnson DE (2011) Relations of rice seeding rates to crop and weed growth in aerobic rice. *Field Crops Research* **121**, 105-115..

Collett D (2002) *Modelling binary data*. 2. CRC Press, New York, USA.

Craufurd PQ & Wheeler TR (2009) Climate change and the flowering time of annual crops. In: *Journal of Experimental Botany* **60**, 2529-2539.

Cuneo P, Jacobson CR & Leishman MR (2009) Landscape-scale detection and mapping of invasive African Olive (Olea europaea L. ssp. cuspidata Wall ex G. Don Ciferri) in SW Sydney, Australia using satellite remote sensing. *Applied Vegetation Science* **12**, 145-154.

De Castro AI, Ehsani R, Ploetz RC, Crane JH & Buchanon S (2015) Detection of laurel wilt disease in avocado using low altitude aerial imaging. *PLoS ONE* **10**, e0124642.

Dvořák P, Müllerová J, Bartaloš T & Brůna J (2015) Unmanned Aerial Vehicles for Alien Plant Species Detection and Monitoring. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **40**, 83.

Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognition Letters* **27**, 861-874..

Freckleton RP, Sutherland WJ, Watkinson AR & Queenborough SA (2011) Density-Structured Models for Plant Population Dynamics. *American Naturalist* **177**, 1-17.

Fried G, Norton LR & Reboud X (2008) Environmental and management factors determining weed species composition and diversity in France. *Agriculture Ecosystems and Environment* **128**, 68-76.

Gabriel D, Thies C & Tscharntke T (2005) Local diversity of arable weeds increases with landscape complexity. *Perspectives in Plant Ecology, Evolution and Systematics* **7**, 85-93.

Heard MS, Hawes C, Champion GT *et al*. (2003) Weeds in fields with contrasting conventional and genetically modified herbicide-tolerant crops. I. Effects on abundance and diversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **358**, 1819-1832.

Henckel, L. et al., 2015. Organic fields sustain weed metacommunity dynamics in farmland landscapes. *Proc. Biol. Sci.*, 282(1808), p.20150002. Available at: http://dx.doi.org/10.1098/rspb.2015.0002.

Huang CY & Asner GP (2009) Applications of Remote Sensing to Alien Invasive Plant Studies. *Sensors* **9**, 4869-4889.

Hung C, Xu Z & Sukkarieh S (2014) Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a UAV. *Remote Sensing* **6**,12037-12054.

Joseph LN, Field SA, Wilcox C & Possingham HP (2006) Presence-absence versus abundance data for monitoring threatened species. *Conservation Biology* **6**, 1679-1687.

Kerr JT & Ostrovsky M (2003) From space to species: ecological applications for remote sensing. *Trends in Ecology and Evolution* **8**, 299-305.

Lamb DW & Brown RB (2001) REVIEW PAPER: Remote-Sensing and Mapping of Weeds in Crops. *Journal of Agricultural Engineering Research* **78**, 117-125.

Lancahsire PD, BLEIHOLDER H, BOOM TVD *et al*. (1991) A uniform decimal code for growth stages of crops and weeds. *Annals of Applied Biology* **119**, 561-601.

Lawrence RL, Wood SD & Sheley RL (2006) Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sensing of Environment* **100**, 356-362.

Lintell Smith G, Freckleton RP, Firbank LG & Watkinson AR (1999) The population dynamics of Anisantha sterilis in winter wheat: comparative demography and the role of management. *Journal of Applied Ecology* **36**, 455-471.

López-Granados F (2011) Weed detection for site-specific weed management: Mapping and real-time approaches. *Weed Research* **51**, 1-11.

Marshall, E.J.P., 1988. Field-scale estimates of grass weed populations in arable land. *Weed Research*, 28(3), pp.191–198.

Marshall EJP (2009) The impact of landscape structure and sown grass margin strips on weed assemblages in arable crops and their boundaries. *Weed Research* **49**, 107-115.

Moshou D, Bravo C, West J, Wahlen T, McCartney A & Ramon H (2004) Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. *Computers and Electronics in Agriculture* **44**, 173-188.

Nonami K (2007) Prospect and Recent Research & Development for Civil Use Autonomous Unmanned Aircraft as UAV and MAV. *Journal of System Design and Dynamics*.

Nordmeyer H (2006) Patchy weed distribution and site-specific weed control in winter cereals. *Precision Agriculture* **1**, 120-128.

Pena JM, Torres-Sanchez J, de Castro AI, Kelly M & Lopez-Granados F (2013) Weed Mapping in Early-Season Maize Fields Using Object-Based Analysis of Unmanned Aerial Vehicle (UAV) Images. *PLoS ONE* **8**, e77151.

Pettorelli N (2013) The Normalized Difference Vegetation Index. *The Normalized Difference Vegetation Index, Oxford University Press, Oxford, UK*.

Pettorelli N, Vik JO, Mysterud A, Gaillard JM, Tucker CJ & Stenseth NC (2005) Using the satellite-derived NDVI to assess ecological responses to environmental change. In: *Trends in Ecology and Evolution* **20**, 503-510.

Preston, C.D. et al., 2002. *New atlas of the British & Irish flora*, Oxford University Press.

Primicerio J, Di Gennaro SF, Fiorillo E *et al*. (2012) A flexible unmanned aerial vehicle for precision agriculture. *Precision Agriculture* **13**, 517-523.

Queenborough SA, Burnet KM, Sutherland WJ, Watkinson AR & Freckleton RP (2011) From meso- to macroscale population dynamics: A new density-structured approach. *Methods in Ecology and Evolution* **2**, 289-302.

Quinlan JR (1992) Learning with continuous classes. *Machine Learning,* **92**, 343-348.

Rango a, Laliberte a, Herrick JE *et al*. (2009) Unmanned aerial vehicle-based remote sensing for rangeland assessment, monitoring, and management. *Journal of Applied Remote Sensing* **3***,* 608-611.

Rasmussen J, Nielsen J, Garcia-Ruiz F, Christensen S & Streibig JC (2013) Potential uses of small unmanned aircraft systems (UAS) in weed research. *Weed Research* **53**, 242-248.

Shorter N & Kasparis T (2009) Automatic vegetation identification and building detection from a single nadir aerial image. *Remote Sensing,* **1**, 731-757.

Storkey J, Stratonovitch P, Chapman DS, Vidotto F & Semenov MA (2014) A process-based approach to predicting the effect of climate change on the distribution of an invasive allergenic plant in europe. *PLoS ONE,* **9**, e88156.

Taylor CM & Hastings A (1998) Finding optimal control strategies for invasive species: a density-structured model for Spartina alterniflora. *Journal of Applied Ecology,* **41**, 1049-1057.

Thornton PK, Fawcett RH, Dent JB & Perkins TJ (1990) Spatial weed distribution and economic thresholds for weed control. *Crop Protection* **9**, 337-342.

Thorp KR & Tian LF (2004) A review on remote sensing of weeds in agriculture. *Precision Agriculture* **5**, 477-508.

Timmermann C, Gerhards R & Kühbauch W (2003) The economic impact of site-specific weed control. *Precision Agriculture* **4**, 249-260.

Torres-Sánchez J, López-Granados F, De Castro AI & Peña-Barragán JM (2013) Configuration and specifications of an Unmanned Aerial Vehicle (UAV) for early site specific weed management. *PLoS ONE* **8**, e58210.

Torres-Sánchez J, Peña JM, de Castro AI & López-Granados F (2014) Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. *Computers and Electronics in Agriculture* **103**, 104-113.

Tredennick, A.T., Hooten, M.B. & Adler, P.B., 2017. Do we need demographic data to forecast plant population dynamics? *Methods in Ecology and Evolution*, 8(5), pp.541–551.

Turner D, Lucieer A & Watson C (2011) Development of an Unmanned Aerial Vehicle (UAV) for hyper-resolution vineyard mapping based on visible, multispectral and thermal imagery. In: *34th International Symposium on Remote Sensing of Environment*, Sydney, Australia, p. 4.

Twining S & Clarke J (2009) Future of UK winter oilseed rape production. *Crop Protection Association Agricultural Industries Confederation: ADAS* 14. AIC, Peterborough , UK.

Westbury DB, Woodcock BA, Harris SJ, Brown VK & Potts SG (2008) The effects of seed mix and management on the abundance of desirable and pernicious unsown species in arable buffer strip communities. *Weed Research* **48**, 113-123.

Wilson BJ & Brain P (1991) Long-term stability of distribution of alopecurus-myosuroides huds within cereal fields. *Weed Research* **31**, 367-373.

Xie YC, Sha ZY & Yu M (2008) Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology-UK* **1**, 9-23.

Zhang C & Kovacs JM (2012) The application of small unmanned aerial systems for precision agriculture: a review. *Precision Agriculture* **13**, 694-712.

**Chapter 3**

**Testing the ability of Unmanned Aerial Systems and machine learning to map weeds at subfield scales: a test with the weed *Alopecurus myosuroides* (Huds).**

J P T Lambert, D Z Childs & R P Freckleton.

Correspondence:

JPT Lambert, Department of Animal & Plant Science, University of Sheffield, Western Bank, Sheffield, S10 2TN.

E-mail: Jptlambert1@sheffield.ac.uk

U.K. Tel: (+44) 114 222 0017

**3.1 Abstract**

BACKGROUND

It is important to map agricultural weed populations in order to improve management and maintain future food security. Advances in data collection and statistical methodology have created new opportunities to aid in the mapping of weed populations. We set out to apply these new methodologies (Unmanned Aerial Systems - UAS) and statistical techniques (Convolutional Neural Networks – CNN) for the mapping of black-grass, a highly impactful weed in wheat fields in the UK. We tested this by undertaking an extensive UAS and field-based mapping over the course of two years, in total collecting multispectral image data from 102 fields, with 76 providing informative data. We used these data to construct a Vegetation Index (VI), that we used to train a custom CNN model from scratch. We undertook a suite of data engineering techniques, such as balancing and cleaning to optimize performance of our metrics. We also investigate the transferability of the models from one field to another.

RESULTS

The results show that our data collection methodology and implementation of CNN outperform pervious approaches in the literature. We show that data engineering to account for “artefacts” in the image data increases our metrics significantly. We are not able to identify any traits that are shared between fields that result in high scores from our novel leave one field our cross validation (LOFO-CV) tests.

CONCLUSION

We conclude that this evaluation procedure is a better estimation of real-world predictive value when compared to past studies. We conclude that by engineering the image data set into discrete classes of data quality we increase the prediction accuracy from the baseline model by 5% to an AUC of 0.825. We find that the temporal effects studied here have no effect on our ability to model weed densities.

**Keywords:**

Unmanned Aerial Systems, weed mapping, Convolutional Neural Networks, black-grass, management

**3.2 Introduction**

The core objective of plant population ecology is to understand changes in numbers of individuals/organisms across time and space (1). Achieving this depends on methods that permit plants to be mapped and monitored at informative scales (2-4). Surveys of plant populations have been undertaken using a variety of different methods such as transect sampling, quadrat sampling and with Unmanned Ariel Systems (UAS) (5-7). Each of these methods has an inherent trade-off between the area that can be surveyed and the intensity at which the subjects can be studied in that area (8). Transect and quadrat sampling can be either used for small area, high intensity studies or large area, low intensity studies, but typically not both (9).

UAS present a unique opportunity for ecological monitoring because, potentially, they can yield data across both large spatial areas and at high survey intensity. This bridges the gap between local scales at which interactions matter, and larger landscape scales at which environmental variation is important (10). UAS have been applied in a range of ecological scenarios including mapping communities (11), population monitoring (12) and mapping individuals in small areas (13). However, few studies have focused on mapping populations at differing times and places, or the challenges of the homogeneous of the environment.

An economically important agricultural crop such as winter wheat (*Triticum aestivum L*.) may be significantly impacted by competition from weeds (14). Weed species add additional costs to the production of crops by increasing the need for agricultural inputs: e.g. in one national-scale audit, it was estimated that weeds cost the Australian economy A$3.5B a year (15). Monitoring data can reduce costs by facilitating precision application of inputs such as herbicides, or better-informed cultural management (16). Ecological monitoring depends on being able to locate and enumerate individuals or species within a given environment (17). Patches of weeds have shown to be persistent over 10 years, therefore mapping in one year represents a potential predictor of future occurrence (18). There are many challenges in the mapping of weeds such as their fast growth rates, and highly variable spatial and temporal distributions (19). Given the potential value of monitoring data, and the possibility of rapid large-scale acquisition of data using UAS, there is clear interest by researchers and farmers in applying this technology to measure weed populations (20).

Despite the potential for data derived from UAS to improve weed management, previous research has highlighted significant issues in the use of them to monitor weed populations (6). Specifically, images and models calibrated to measure weeds in one environment appear to perform poorly when transferred to another. There are several reasons for this limited transferability, for example, variation in weather conditions or different growth stages of the weed or crop. As crop plants grow over the field season their phenology changes, as does that of the weeds (21). This results in changes in the spectral properties of the crop and weed species, both in the visible spectrum and beyond (22, 23). Moreover, common crops are grown in many different varieties, each with their own unique phenology and physiology (24-26). The statistical methodology of random forests (RF) and a dataset of mean pixel values from UAS image plots, as used in our previous study of weed monitoring does not fully capture the extent of these variations, thus failing to generate highly transferable models (6).

Supervised machine learning is a statistical method that generates a classification output after being presented with an unclassified input, having previously been trained on data consisting of known inputs and outputs (27). All such models are trained using “features”. A feature is a numeric representation of the unclassified input. In the case of an image input, these can be engineered by researchers i.e. texture, colour, shape or they can be abstractly and randomly defined by the model and adapted over iterations. Here we highlight key network methods that are used in supervised machine learning.

Neural Networks (NN) conceptually mimic biological neurons in their node-like structure. Each node is interconnected to others and sends a “signal” if threshold values are passed. Threshold values are tuneable at each node and are adjusted automatically over the course of fitting the model. An important advantage of NN is that they can bypass the need for domain knowledge of the dataset (feature engineering), allowing more abstract and potentially useful features to be used. This does, however, make the model less interpretable, as the features that are used are selected without logical justification. As with most statistical methods, NN perform better when trained on more data.

CNN are a type of NN specifically applied to image data sets. Convolutional Neural Networks (CNN) have emerged as the most common, and frequently best performing, model for image classification tasks in the machine learning literature (28). CNN learn a sparser connection between regions of an image than traditional NN models by imposing spatial dependencies upon the pixels in the image (29). This may be of use when analysing weed distributions because these are spatially dependant (30-32). CNN do not use user defined features such as colour, shape or texture to learn from the data. Instead CNN create abstract feature maps and then through training/iterations, assigns importance to different feature maps (33) representing different states in the image. These components of a CNN make them a well-suited method for mapping weed populations, but the underpinning model correspondingly harder to interpret. Spatial information is retained, and automated abstract feature identification can identify common aspects among the classes of data that human feature selection would otherwise miss (34).

Here we investigate how images collected from UAS can be classified using CNN to predict weed densities in unseen images. We explore how data engineering can be undertaken to improve the results and account for the heterogenous nature of the environment. We also investigate the seasonal effects of mapping on our ability to correctly predict weed densities by comparing our models between years and the week of survey, thus addressing key limitations from past literature. Finally we assess true out of sample predictions of CNN models to assess their transferability across populations.

**3.3 Materials & Methodology**

*3.3.1 Description of dataset*

We studied *Alopecurus myosuroides* (black-grass) in populations of *Triticum aestivum L*. (winter wheat). 1.9 million hectares of wheat is cultivated per year in the UK, making it the most widely grown crop, with *A. myosuroides* becoming a significant problem throughout the UK (35).

Our field sites were part of an ongoing study by the Black Grass Resistance Initiative (BGRI) into herbicide resistance levels in the weed nationally. We surveyed 102 new fields across the arable regions of the UK. Late season monitoring (13rd June – 12th August in 2016 and 2017) was chosen as previous work shows that the weeds are distinguishable from the surrounding wheat crops at this time (6). This represents a BBCH weed growth stage of 87-89 (36).

Fields were subject to a range of differing management practices, across farms from 80 to 3000 ha. The populations of black-grass had previously been measured in fields using the methodology developed by (3, 35) to estimate plant density states in a plot. Plots of 20x20m were chosen as this allowed large amounts of contiguous ground-truthed data on the densities of black-grass in a field to be collected. The average field was 8ha with 110 plots per field, depending on the varying extents of the field. Five ordinal density states of black-grass were denoted: absent, low, medium, high, very high,(0, 1-160, 161-450, 451-1450 and 1451+, plants per 20m2 respectively). This method allows for multiple observers to be used, enabling large spatial scales to be covered with minimal misclassification error between observers.

*3.3.2 UAS platform*

A widely available commercial UAS platform was chosen to allow for low entry costs and high repeatability. We used the 3DR solo UAS[[1]](#footnote-1) as it permits third party imaging systems to be attached and operated. The Parrot Sequoia[[2]](#footnote-2) was chosen as the imaging sensor as this sensor has been specifically designed for use with UAS. This sensor records images in four discrete calibrated spectral channels: Green 550nm (*fg*), Red 660nm (*fr*), Red-Edge 735nm (*fre*) and Near Infrared 790nm (*fn*) at 1.2Mp. The sensor possesses a “sunshine sensor” that standardised against variable lighting conditions over the course of a flight by continuously recording the light conditions in each spectral channel and then automatically calibrating the outputs to the absolute values.

All flights were carried out following UK rules and regulations controlling the use of UAS for scientific research. Flights were conducted within 2 hours either side of solar noon to reduce the effect of sun angle. The optimum flight parameters to cover each field in the minimal amount of time were a flight height of 100m and an image overlap of 60% (37). Each flight generated thousands of subfield scale images that are stitched together to create a single orthomosaic image, encompassing an entire field using relatively few ground control points. For this Agisoft Photoscan was used. This software also creates Vegetation Indices (VIs) from the individual bands of the sequoia. The average ground sample distance (GSD) of all the flights was 8.27cm pixel-1.

Of the 102 fields that were flown, 76 generated data of high enough quality to analyse. The fields that were not suitable to be analysed were discarded for the following reasons: poor image quality, significant image stitching artefacts and sensor failure.

The calibrated spectral channels of the sequoia sensor allow for VIs to be calculated for each pixel. VIs are used as they reduce multiband observations to a single numerical index (38). We used Green Normalized Differential Vegetation Index GNDVI (equation 1) to classify images:

(1)

All subsequent references to the data, refers to the GNDVI dataset See appendix Table 5, for statistical measurements of the GNDVI dataset.

Our choice to base our analysis on GNDVI is because high biomass crops such as wheat cause saturation of chlorophyll levels in the red wavelength, resulting in poor performance when using Normalized Differential Vegetation Index NDVI (equation 2) (39).

(2)

Previous studies have focused on the NDVI owing to its correlation with plant vigour and growth (40). However, when needing to discriminate between invasive populations, vigour and growth rates with NDVI has shown to be uninformative in cases of high saturation of a spectral channel (41). Analysis based on UAS imagery has often overlooked this feature of NDVI , but is recognised in satellite remote sensing work (42-44).

The ground-truthed density data were overlaid on each georectified orthomosaic using GIS packages in R. Then the orthomosaic maps were split into 20x20m subplots, each geographically relating to the ground-truthed observations. This creates a dataset of images at the 20x20m scale, which our subsequent analysis area is based on. The resulting image dataset consists of 12,313 unique measurements of black-grass at 20x20m scale covering the full range of black-grass densities. The densities are however not evenly distributed. The breakdown as follows: Absent = 14.5% Low = 53.1% Med = 17.3% High = 8.2% Very High = 6.9%.

*3.3.3 Modelling approach and metrics*

We used a CNN to train a classifier on our black-grass image data. The model structure was taken from one of the top performing methods on the industry standard image database, ImageNet (45), called GoogLeNet (34). Whilst we use the structure of GoogLeNet, it is important to note that we do not use the pretrained model weights and biases that allowed the model to score so highly on ImageNet. Here we highlight four common components of our chosen model framework, that are then stacked together with other components such as batch normalisation and dropout to create a variety of different network structures:

1. Convolution: The convolutional step involves extracting features from an image whilst maintaining their spatial context, by using a filter to pass over an image and computing the dot product to create a generalised feature map.
2. Addition of Non-Linearity: Non-Linearity is introduced to the feature maps by applying a Rectified Linear Unit (ReLU), this speeds up the training process when compared to tanh/sigmoid activation functions. This means that model convergence will occur with a lower computational cost (46).
3. Pooling: Pooling of the feature map is used to reduce dimensionality. This reduces the parameter number in the network, a key stage in preventing overfitting. Pooling also makes the network more stable to distortions in the training images (47).
4. Fully connected final layer: This combines all the neurons of the previous layer and applies an activation function to determine the final classification of an image. The most common form of activation function is SoftMax and the predictions always sum to 1 (48).

CNN have been successfully applied to many datasets similar to ImageNet through a process known as transfer learning, whereby only the weights of the connected final layer of a pretrained model are altered (49). We do not use the process of transfer learning as our proposed dataset is significantly different from that of ImageNet. Instead, we use the GoogleLeNet structure and independently train all layers of our model.

To model a CNN three data sets are needed: training, validation and test sets. Each dataset comprises pairs of input images and target vectors. Target vectors act as a labelling method and are what the model tries to predict when given a new image. In our example the input image is a 20x20m image plot and the target vector represents the five different ordinal density states. CNN are trained using a variety of parameters. From our initial exploration of the modelling we settled on using the following as our standards: a decaying momentum beginning at 0.1 and halving every 32000 steps as our optimizer, categorical cross entropy as our loss function and a batch size of 128.

We report, where appropriate, three metrics for our models. These are: Multiclass AUC, Cohen’s kappa and weighted Cohen’s kappa. AUC refers to the Area Under the Receiver Operating Characteristic (ROC) curve, that is the true positive rate (Sensitivity) against the true negative rate (Specificity). AUC is used for its ability to differentiate between two groups, and is equal to the probability that the classifier will rank a randomly chosen positive example higher than a randomly chosen negative example (50). AUC values range from 0 – 1. We plot a diagonal line from (x=0, y=1) to (x=1, y=0) known as the line of equality or the random chance line (51). Points that fall below this line represent non-informative models where random classification would perform better. For the x-axis in our AUC plots we use 1 – Specificity.

The categorical predictions of a model and ground-truthed observations can be viewed as different raters. This allows us to assess the degree to which they agree or disagree and utilise Cohen’s kappa statistic (52) (equation 3):

(3)

Where ρo is the observed agreement and ρe is agreement due to chance. This results in a range from 1 indicating complete agreement between raters, through 0 indicating that agreement is only due to random allocation and -1 indicating complete disagreement.

AUC and kappa do not consider the ordinal structure of our data, with observations ranging from Absent to Very High in incrementing ordered categories. Therefore, an observation of Absent and a prediction of Low is closer to agreeing than if the prediction were Very High. We therefore used weighted Cohen’s kappa (equation 4):

(4)

Where κ is the number of categories, ωij, χij and *m*ijrepresentthe weight from the matrix. This allows us to count disagreements differently (53). The weighted kappa is on the same scale and distribution as the base Cohen’s kappa. We use a squared weighting matrix of 1, 4, 9, 16 and 25 ranging from agreement to significant disagreement, to penalise significantly wrong agreements.

*3.3.4.1 Model refinement: data balancing*

We checked the performance of the model in several respects. First, we analysed the effect of balancing the data in terms of the distribution of observations among density states. This is important because the dataset is heavily weighted towards the Low-density state, comprising over 50% of the dataset. Such imbalanced distributions can lead to lazy or biased classifiers, whereby the model can default to predicting the majority class but will nevertheless still score well in many metrics such as error or accuracy rate. To investigate this, we created balanced datasets and use metrics as outlined above. In our dataset the Very High class had the smallest representation with only 565 examples in the training set. We therefore randomly sampled 565 of each remaining density states, to create a balanced training set of 2825 images. The same balancing process was repeated for the validation and testing data sets resulting in 800 and 575 images respectively.

*3.3.4.2 Model refinement: data cleaning*

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

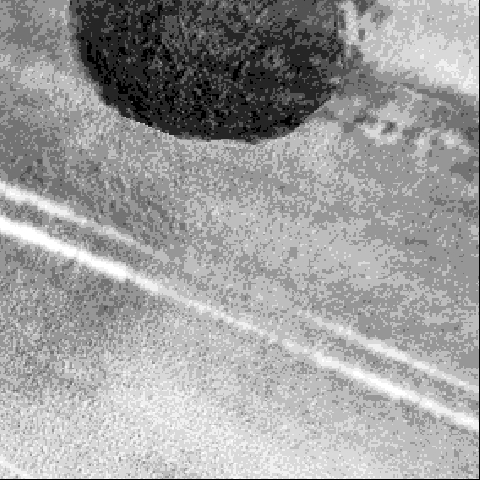
It is important to consider the quality of imaging data. Specifically, many of our 20x20m aerial plots contain “artefacts” that were not accounted for in our ground observations. Figure 1 shows examples of three such types of artefacts. In Figure 1 an overhanging tree, the tramline and the field hedgerow in the top right hand corner are introducing significant noise into the image that does not represent either wheat or black-grass. It is this excess noise/uncategorised data we aimed to remove.

Figure 1 Example of a Very High, 20x20m plot with significant non-black-grass “artefacts”, reducing the signal in the image coming from the Very High level of black-grass that was observed on the ground in this plot. The grid overlay represents the subsampling methodology used to break each image into 16 smaller representations of the entire plot. The subplots are referenced by their position relative to the bottom left hand corner (1,1) and top right corner (4,4).

To achieve this, we subsampled each individual 20x20m plot into 16 smaller images. Figure 1 demonstrates the outline of this subsampling grid. This yielded a dataset of 197 008 images. We then manually examined this dataset and set aside all subsamples that we determined to contain artefacts. In the case of Figure 1 only two subplots of “pure wheat” remained ((1, 2), (1, 3)), that were subsequently used in what we will refer to as the *Clean* dataset. This created a clean data set of 101,907 images and *Artefact* dataset of 95101 images. The training and test sets were the same as the previous experiments, but now “cleaned”. We use the Clean and Artefact datasets to build models and predict on the test data of the other dataset e.g. clean model on artefact test data, and vice versa. This allows us to test the influence of data cleaning.

To make a comparison with our ground observations, we must upscale the subplot predictions back up to the 20x20m scale at which ground observations were recorded. There is often variation in density within each plot, but this is not recorded. In a hypothetical situation this could mean that the model is perfectly fitting the subplot test data, but then being penalised as we are unable to ascertain the observed level of black-grass in that specific subplot, only the entire 20x20m plot. We therefore take the median prediction from each subplot of one 20x20m plot as the model observation. This gives us a prediction of only the areas of the image with wheat and/or black-grass in them, at a scale that allows for comparison to our ground truthed data.

*3.3.5.1 Model transferability: Field level Cross validation*

To test out-of-sample/new field performance we conducted leave-one-field-out cross validation (LOFO-CV) trails and created 76 models, i.e one per field. Each model was trained using the baseline model parameters and cleaned upscaled subplots from all the fields. One field was withheld from the training dataset to become the test set in each new model. We report back metrics at field level (i.e. not 20x20m plot level) as not all fields have the full five density states present.

*3.3.5.2 Modelling workflow – baseline model*

Having created the relevant datasets for each question we trained a model using our standard parameters. We began the analysis with a simple baseline test of how the models perform when 10% of the entire data is randomly selected as the test set. The model was then used to predict the ground-truthed observations of the relevant test set. We then calculated all relevant metrics and plot a ROC curve where appropriate. This assessed the performance of the CNN and established a baseline against which further analysis could be benchmarked. We investigated the effect of data balancing, data engineering and LOFO-CV against the baseline model.

To account for possible differences owing to variation in the date or survey or between years, we grouped the LOFO-CV models by years with 38 and 43 fields in 2016 and 2017 respectively and took the mean values of the AUC for each year. Each field season lasted 6 weeks and averaged the same number of fields each week. Consequently, we grouped the LOFO-CV models by week and took the mean values of AUC. Owing to the design of our field season we begin in the south and move north over the course of the season, so latitudinal effects will also be present but are not accounted for.

**3.4 Results**

*3.4.1 Baseline Model*

We find that the baseline model gives an AUC of 0.78, a weighted kappa of 0.59 and an average misclassification rate across all states of 17.8% as seen in Figure 2. We see that the Very High and Absent density states show the AUCs closest to x=1, y=1. This means that these density states are easier to distinguish for the model than the states in between.

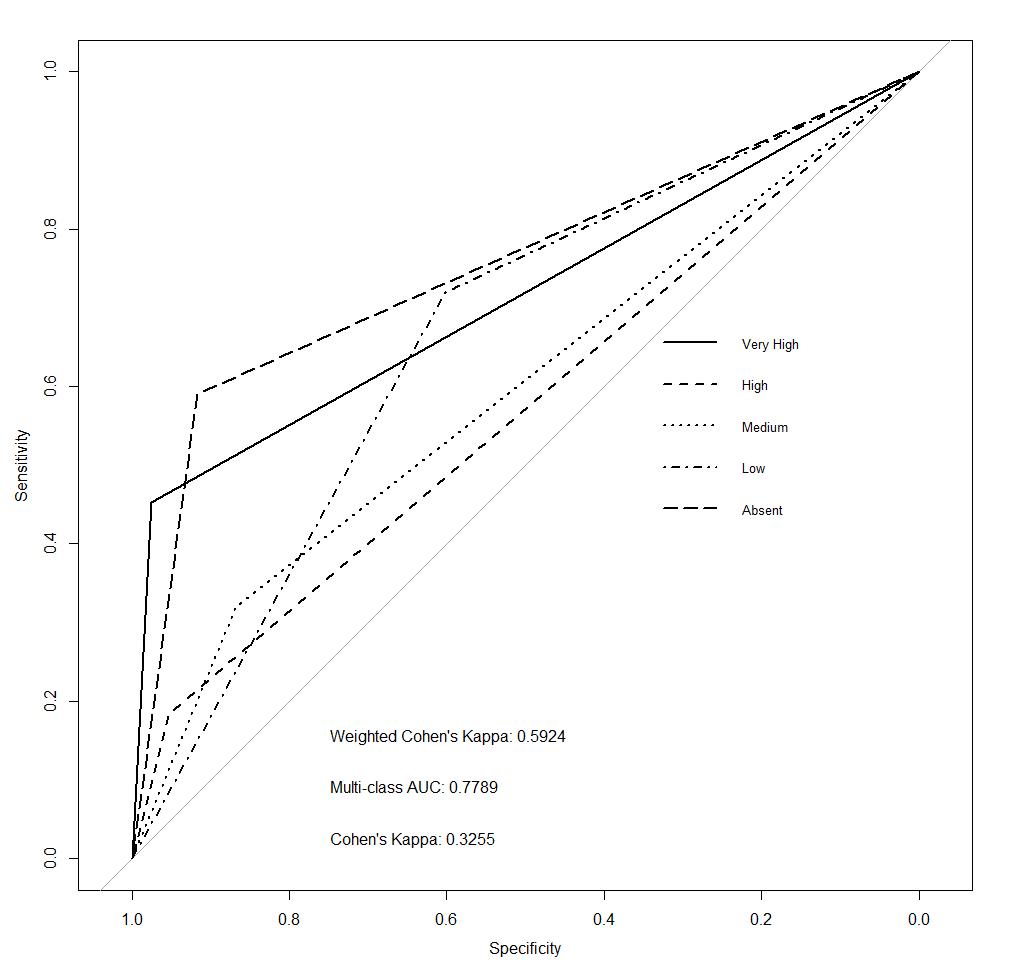


Figure 2 Baseline, ROC plot of a CNN trained using 90% of the dataset and used to predict the multiclass black-grass density state of the completely withheld random 10% of data.

*3.4.2 Data Balancing*

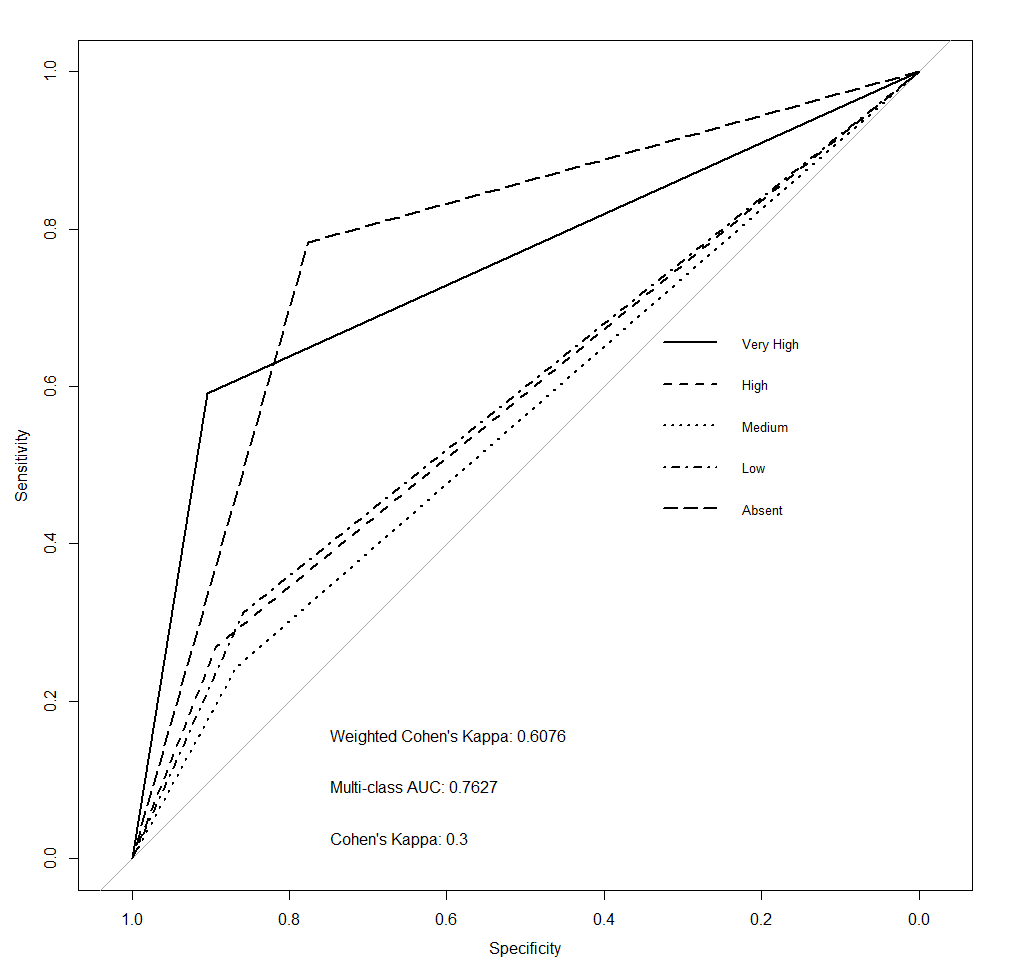
The same training and evaluation parameters were used to train a model for the data in which the proportions of the density states were balanced. We see that by balancing the data set we slightly reduced the AUC and Cohen’s kappa of the model (see Figure 3 for the ROC plot), whilst slightly increasing the weighted kappa and increasing the misclassification rate to 22.4%. This is most likely a consequence of the reduced number of training samples, leading to a poorer ability of the model to generalise features unique to each class. Table A1-A4 present statistical analysis on the differences between curves. (54). The results in Table A1 show that when the curves from Figure 2 (baseline model) are compared to those of Figure 3 (data balanced) that all but the Low density state curve are statistically non-significantly different. Balancing the dataset or not therefore does not affect the predictive performance of the models. We therefore continue to use the unbalanced dataset for the rest of our analysis.

Figure 3 ROC plot of a CNN trained using 90% of the balanced dataset used to predict the multiclass black-grass density state of the completely withheld random 10% of balanced data.

*3.4.3 Data Cleaning*

To examine how the data cleaning process (Figure 1) affects our models a new model was trained using the same parameters as the baseline model, but using the unbalanced, *Clean* dataset. Figure 4 shows us that the AUC increased by 4.6%, a significant improvement with a similar misclassification rate to the baseline of 17.5%. Table A2 presents the statistical breakdown of the individual comparisons of AUC to the baseline.

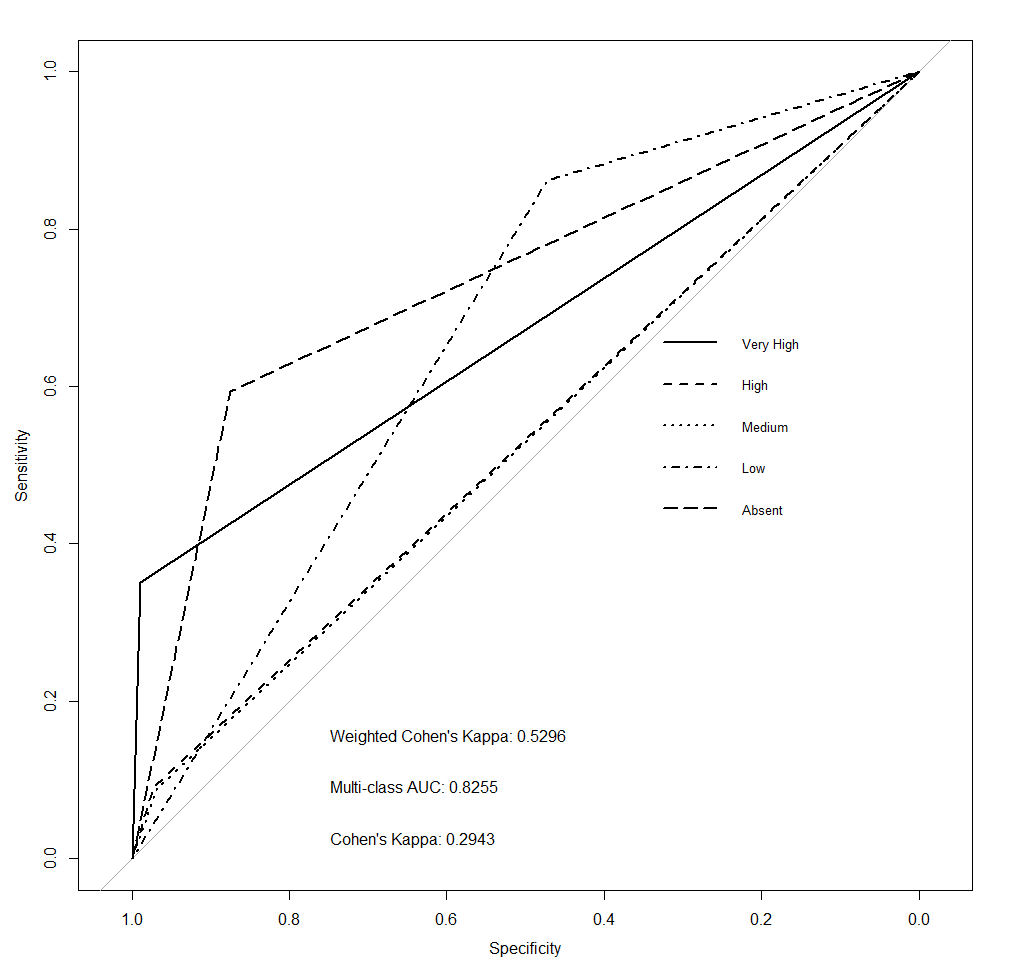


Figure 4 ROC plot of a CNN trained using 90% of the entire Clean subplot dataset used to predict the multiclass black-grass density state of the completely withheld random 10% of Clean data. The subplot predictions are then scaled back up to 20x20m plots for comparisons to our ground observations.

The images vary greatly in quality, with some having a large amount of high quality coverage, whilst in other cases only a small amount of the image is of good quality. We therefore divided the dataset according only to the percentage cover of good quality data of the original 20x20m plots remaining after the cleaning, regardless of black-grass level. Five equal categories of coverage of the 16 subplots, ranging from <20% (~3 subplots) to >80% (13-16 subplots) were established. Looking at the Multi-class AUC values for each plot in Figure 5, we see there is a ~6% difference in the lowest (0.67, <20%) and highest values (0.73, 60%-80%). We highlight the statistical differences between the categories with the highest and lowest AUCs in table A3. Showing that whilst the individual density states lines are not significantly different, the overall graphs are significant in conjunction with Figure 5.

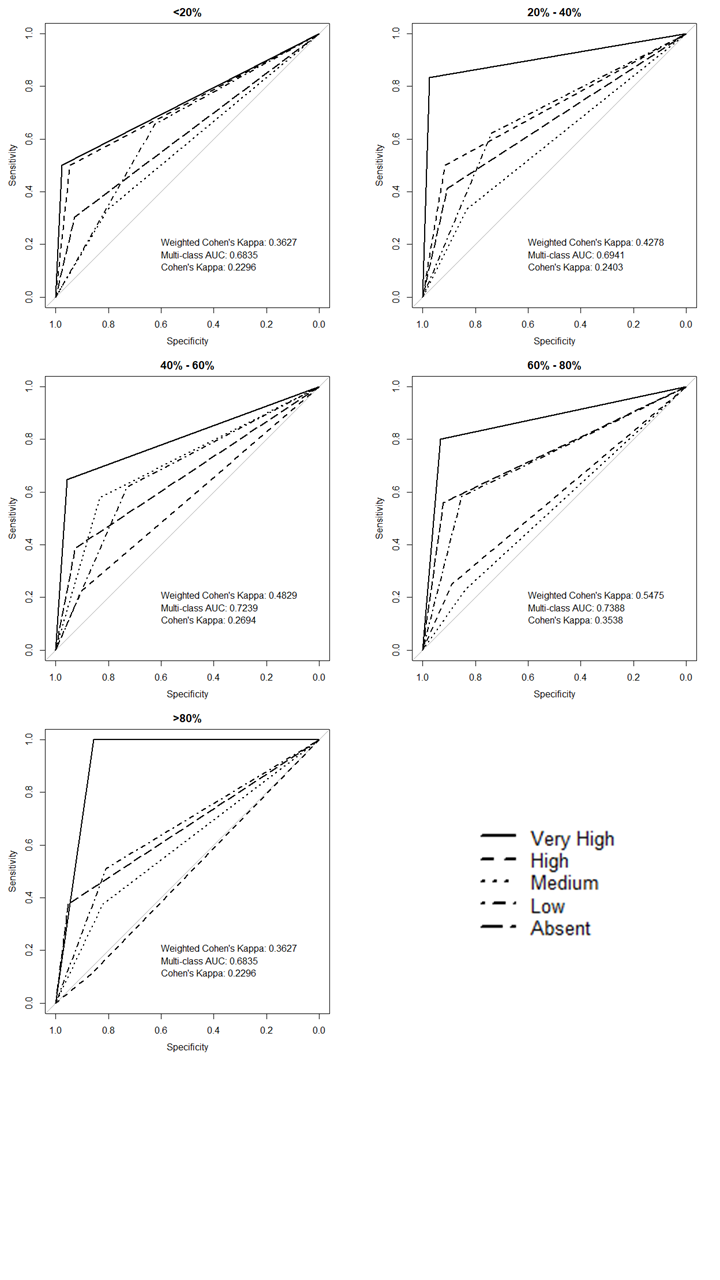


Figure 5 ROC plots showing how the percentage cover of the subplots in the *Clean* dataset affect performance (measured as AUC and kappa).

*3.4.4 Analysis artefact data*

Having shown in Figure 4 that cleaning and upscaling the data results in improved metrics from the baseline we next investigated the predictive performance of models fitted to the “artefact” images. To do this we used the 95101 artefact images set aside from the training set, predicted on the artefact images from cleaning the test data and then upscaled. Figure 6 suggests that the artefact plots still have features within them that allow us to classify black-grass as accurately as the Clean model (Figure 4). It also shows that with a higher weighted kappa and lower misclassification rate of 15.5%, it does better at not making large ordinal disagreements e.g. Very High observation Vs Absent prediction, when compared to the Clean model. The Clean model predicted Absent when a Very High was observed in 8.75% of cases, compared with the artefact model only predicting 6.3% of such cases.

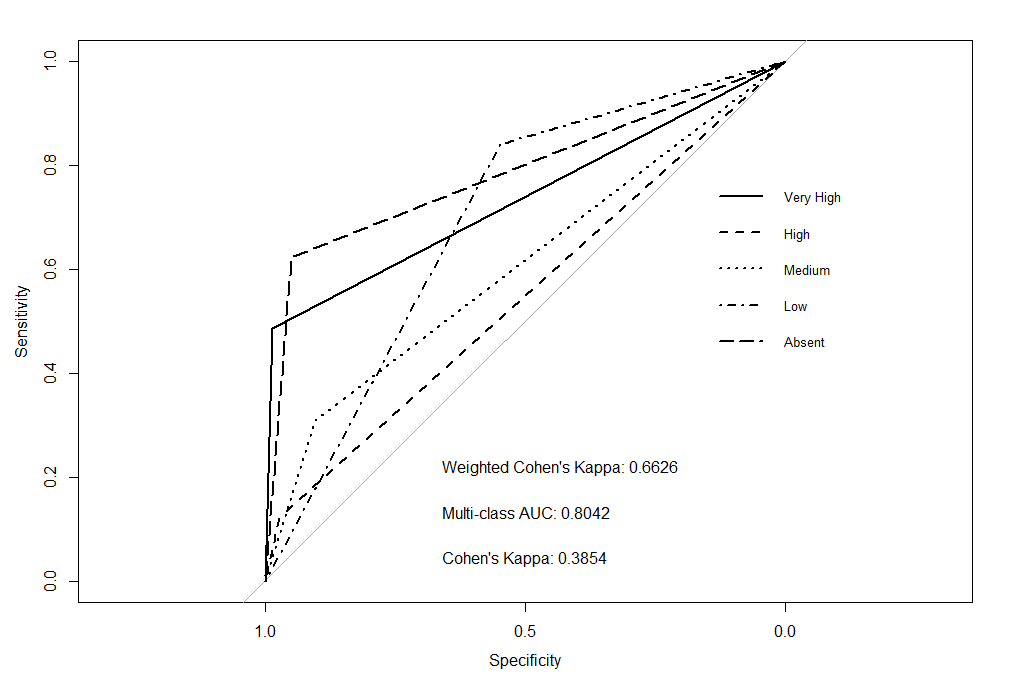


Figure 6 ROC plot of a CNN trained using 90% of the artefact subplot dataset used to predict the multiclass black-grass density state of the completely withheld random 10% of artefact data.

As shown in Figure 7, the clean model can predict the black-grass levels in the artefact dataset with some degree of accuracy, with an AUC of 0.61 and misclassification rate of 17.1%. However, the model for the artefact data is not able to predict the clean test dataset accurately, with an AUC of 0.463, a misclassification rate of 42.1% and the AUC for all density states were significantly different as shown in Table A4. This suggests that the features used by the artefact model are not conducive to black-grass identification. Therefore, the features in the model for Figure 6, must not be directly related to black-grass. This also suggests that our manual screening of the data may have been overly strict, and we are thereby missing data that could increase the ability of the model to generalise features for the identification of black-grass.

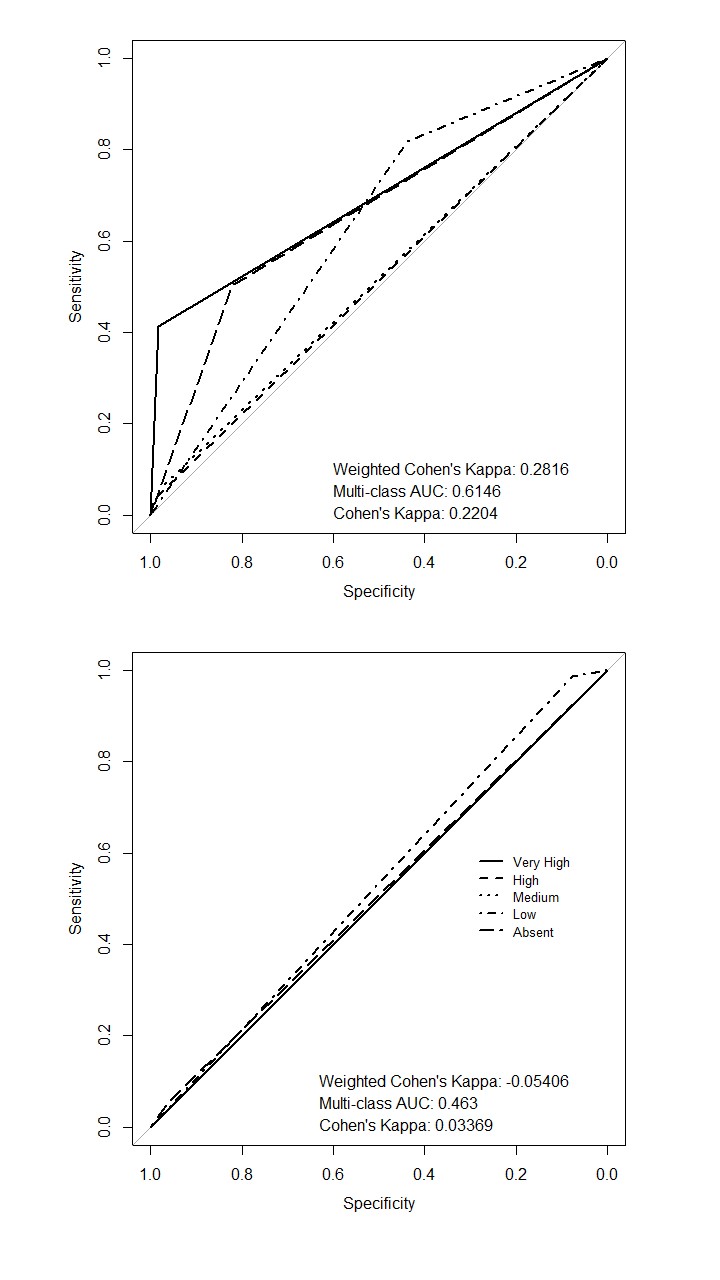
****

Figure 7 (top) ROC plot of a model trained using the Clean training set, then used to predict the five density level states in the artefact test set. (bottom) ROC plot of a model trained using the artefact training set, then used to predict the five density level states in the cleaned test set. The predictions are upscaled to plot level.

*3.4.5 Out of Sample predictions - LOFO-CV*

Here we examine the true out of sample prediction for the dataset. In all our previous models we have used an initial random 10% as our test dataset as described in our initial test set. Therefore, the model has been trained on a large sample of each individual field, allowing it to generalise features specific to that field, making it more sensitive to outliers. Thus, our reported results so far are not truly out of sample and may have limited repeatability in further studies, even when using the standardised data collection methodology described here.

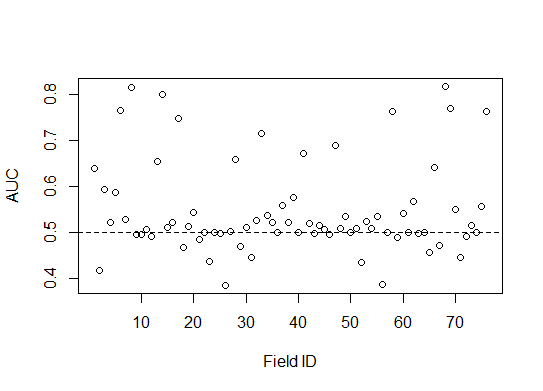


Figure 8 AUC of each field’s out of sample prediction. Each point represents a separate model that was trained on all but the Field ID in question which is used as the test set. Field ID is a randomised ordering of the field names across both survey years.

Figure 8 shows the mean AUC of the fields is 0.54 with a range of 0.38 0.81. This means that LOFO-CV predictions for these models are frequently no better than random. The kappa metrics were not used here as most of our out of sample fields did not contain the full range of black-grass densities and so are penalised for lack of agreement when there are no observations of a level.

*3.4.6 Temporal Effects*

To investigate temporal effects on the results of our out of sample predictions, we studied whether the year or the week we visited the field had any effect on the AUC. Figure 9 shows the mean and standard errors of the AUC for each year and week. Neither year nor week has a significant effect on the model performance measured by the AUC of the model, with adjusted R2 values of -0.011 and 0.008 respectively. This means that the temporal variation in the time surveying has not influenced our results.

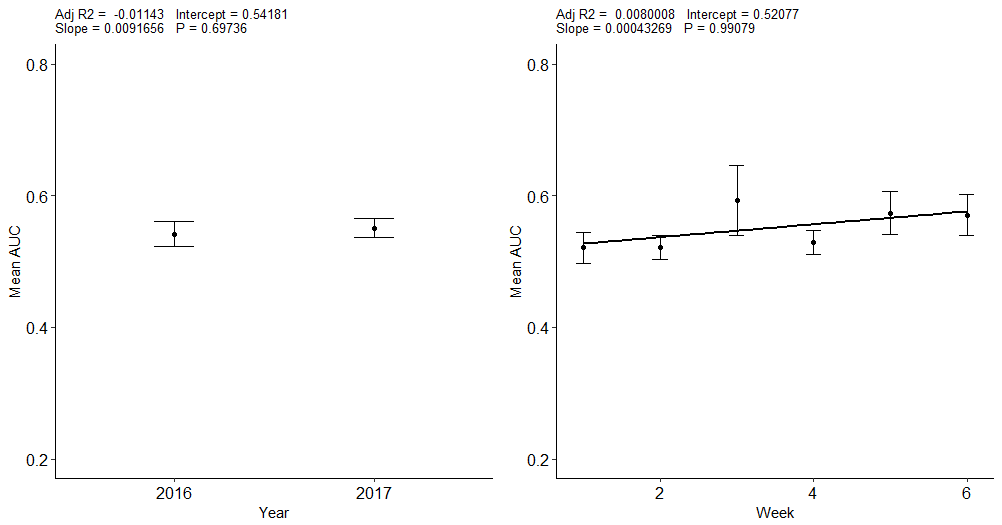


Figure 9 (left) Mean AUC for every model in each year. (right) Mean AUC for every model in each week.

**3.5 Discussion**

We set out to predict distributions of weed densities using UAS imagery and CNN. We have devised a standardised and repeatable UAS data collection methodology, applied it over multiple years across the major arable areas of the UK and utilised data engineering techniques to increase the quality of our datasets. Our main conclusion is that data engineering increases the performance of our metrics the most, relative to other methods attempted when given a sample of known states in a field. Increases in performance such as these are not common for CNN in the computer vision literature. There was no evidence that temporal factors such as year or time of sampling, affects the performance of the out of sample predictions.

However, when predicting on fields with no previous ground truthing (i.e. true out of sample data), the success as revealed by our metrics was highly variable. This may be due to the problem of dataset shift (55). Dataset or covariate shift occurs when there is a change in distribution of the classes between the training and test datasets. We know from our ground observations that on an individual field-by-field basis that it is rare to find fields with the full five density state distribution and there are no cases where all five are present in an equal distribution. One way of counteracting this issue in the literature is by constructing a density estimation of the labels in the test dataset and reweighting the training dataset accordingly (56). This approach is not applicable in a fully automated UAS system for the prediction of density states, as it is still dependant on ground-truthed observations from skilled observers.

Our study is the first to use repeated UAS surveys and deep learning statistical methodology to assess the impact of the significant heterogeneity in conditions across time and space on automated monitoring of weed densities. Anderson & Gaston (57) outline many areas in which UAS can be used in ecology and emphasise the need for temporally resolved studies, allowing for scale appropriate measurements using UAS that can be at user defined times and locations. This is a change in precedent from remote sensing work using satellite data, where data was only available at set times, resolutions and spectral frequencies. However, many previous studies using UAS have focused on repeated visits to one single site over time (58) or multiple sites at one time point (59). The use of trial plots in some studies does allow for a more detailed assessment of certain variables (60). However, in real world applications of methodologies and management decisions developed under these controlled settings, much more spatial and temporal variability when applied in agronomic use cases will be encountered, thus reducing the transferability and scope of the studies (61). Therefore, our focus of only using “live” uncontrolled agronomic scenarios, does result in reduced reported metrics but allows our work to be applied with a more realistic understanding of the results that would be seen in the field.

Neural networks have previously been used and compared to other statistical methods, to classify the state of weed populations at a range of spatial scales (62-64). Barrero (13), trained a NN with a user defined texture feature derived from NDVI to identify a weed species amongst a single rice paddy. They reported a 99% precision on test data, with no reported recall score. This is most likely an overstatement of the model performance and approach. However, this study only focused on the binary classification issue of presence/absence of a weed, a much simpler and less informative on-farm metric, and only considered predictions from a single field at a single time point, suggesting that the performance is being overstated with no LOFO-CV being attempted. It is to be expected that our metrics (AUC, Cohen’s kappa and weighted Cohen’s kappa) are lower than the equivalent ones reported in the NN study, due to our focus on multiple fields spanning a wide variety crop conditions and for the more advanced use of density state predictions. Therefore, our results are more representative and transferable than these studies due to our LOFO-CV analysis, for methodologies involving UAS and machine learning to map weed populations going forward. However, our results indicate a more extensive and controlled analysis of the transferability of models is still needed.

The process of manually screening the datasets for artefacts is a slow and non-reproducible or scalable task. In the future we propose to train a classifier to automatically partition an entire dataset into clean and artefact sections. This approach is comparable to work that quantifies the data quality of video using a CNN (65). This would allow us to expand our analysis into other Vegetation Indices by improving and standardising the data processing pipeline.

With the artefact dataset predicting to the same if not higher standards in our metrics than the clean dataset, it stands to reason then that a composite modelling approach could be undertaken to channel the clean and artefact subplots to their respective models and then recombined at the upscaling stage. This is a concept similar to ensemble based classifiers, where multiple differing model types are trained on the same data set and aggregate their predictions for the test set (66). Our approach described here would use this concept but instead of differing model types on the same dataset, we propose the same model on differing datasets and aggregating their predictions. This would reduce the amount of data loss and combine the differing feature sets of the models to aid in the detection of arable weeds.

*3.5.1 Concluding remarks*

We have demonstrated here how data engineering of UAS imagery and use of CNN can be used to classify weed densities. We highlight the methodological improvements resulting in increased prediction accuracy compared to past research using a variety of metrics, statistics and data collection procedures that provide a more detailed assessment of true model performance. All our models apart from the LOFO-CV are composed of a random 10% of individual subplots for the test set. This means that the models will have most likely been exposed to some in-field examples of the test set, and therefore can generate features that are specific and not generalised to the detection of the weed. We can conclude that when only considering the images of a new field and no other data, we cannot be highly confident in the ability of most of our models to map the black-grass in the field. Whilst we don't show a significant improvement in LOFO-CV testing with no apparent factors that make an individual field be predicted well or poorly. We believe that the robustness of this evaluation procedure is a greater estimation of real-world predictive value when compared to past studies, which consequently overestimate their applicability. Therefore, the methodology set out in this paper represents a new standard in the area of weed mapping with UAS due to the expanded capabilities of data collection, statistical methods and evaluation procedures.

**3.6 Acknowledgments**

JL was funded by a studentship from the Grantham Centre for Sustainable Futures. Collection of field density data was funded by the BBSRC (BB/L001489/).

**3.7 Appendix**

Table 1 Non-Equal dataset AUC’s compared to the Equal datasets. Used (54) to test the statistical difference of the AUC of each Density state.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Density State | AUC 1 | AUC 2 | D | p-value |
| Abs | 0.75 | 0.77 | -0.94 | 0.34354 |
| Low | 0.66 | 0.58 | 2.87 | 0.004032 |
| Med | 0.59 | 0.55 | 1.48 | 0.138393 |
| High | 0.56 | 0.58 | -0.46 | 0.643204 |
| V High | 0.71 | 0.74 | -1.00 | 0.313908 |

Table 2 Non-Equal dataset AUC compared to the Cleaned dataset.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Density State | AUC 1 | AUC 2 | D | p-value |
| Abs | 0.75 | 0.73 | 0.76 | 0.445291 |
| Low | 0.66 | 0.66 | -0.34 | 0.727911 |
| Med | 0.59 | 0.52 | 3.72 | 0.000193 |
| High | 0.56 | 0.53 | 1.81 | 0.069353 |
| V High | 0.71 | 0.67 | 1.34 | 0.177991 |

Table 3 Worst performing bracket AUC from data quality testing (20% <)(AUC 1) compared to the best performing AUC bracket (60% - 80%) (AUC 2).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Density State | AUC 1 | AUC 2 | D | p-value |
| Abs | 0.61 | 0.74 | -2.25 | 0.023968 |
| Low | 0.64 | 0.71 | -1.47 | 0.140146 |
| Med | 0.56 | 0.53 | 0.23 | 0.813781 |
| High | 0.72 | 0.56 | 1.12 | 0.261717 |
| V High | 0.73 | 0.86 | -0.70 | 0.479862 |

Table 4 Clean model AUC’s from predicting on the artefact dataset compared to the Artefact model AUC’s from predicting on the clean dataset.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Density State | AUC 1 | AUC 2 | D | p-value |
| Abs | 0.66 | 0.51 | 12.22 | 2.43E-34 |
| Low | 0.62 | 0.53 | 11.62 | 2.94E-31 |
| Med | 0.52 | 0.5 | 4.70 | 2.56E-06 |
| High | 0.51 | 0.5 | 2.88 | 0.003877 |
| V High | 0.69 | 0.5 | 9.26 | 1.92E-20 |

Table 5 Statistical measurements of the GNDVI pixel values for each vegetation group.

|  |  |  |
| --- | --- | --- |
| GNDVI | Mean | Standard Deviation |
| Black-grass | 0.336 | 0.007 |
| Winter wheat | 0.304 | 0.011 |

**3.8 References**

1. Gibson DJ. Methods in comparative plant population ecology: Oxford University Press; 2014.

2. Harper JL. Population biology of plants. Population biology of plants. 1977.

3. Queenborough SA, Burnet KM, Sutherland WJ, Watkinson AR, Freckleton RP. From meso‐to macroscale population dynamics: a new density‐structured approach. Methods in Ecology and Evolution. 2011;2(3):289-302.

4. Symonides E. On the ecology and evolution of annual plants in disturbed environments. Vegetatio. 1988;77(1-3):21-31.

5. Burnham KP, Anderson DR, Laake JL. Estimation of density from line transect sampling of biological populations. Wildlife monographs. 1980(72):3-202.

6. Lambert J, Hicks H, Childs D, Freckleton R. Evaluating the potential of Unmanned Aerial Systems for mapping weeds at field scales: a case study with Alopecurus myosuroides. Weed research. 2018;58(1):35-45.

7. McIntyre G. A method for unbiased selective sampling, using ranked sets. Australian Journal of Agricultural Research. 1952;3(4):385-90.

8. Rondinini C, Wilson KA, Boitani L, Grantham H, Possingham HP. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecology letters. 2006;9(10):1136-45.

9. Braunisch V, Suchant R. Predicting species distributions based on incomplete survey data: the trade‐off between precision and scale. Ecography. 2010;33(5):826-40.

10. Brown JH. On the relationship between abundance and distribution of species. The american naturalist. 1984;124(2):255-79.

11. Laliberte AS, Herrick JE, Rango A, Winters C. Acquisition, orthorectification, and object-based classification of unmanned aerial vehicle (UAV) imagery for rangeland monitoring. Photogrammetric Engineering & Remote Sensing. 2010;76(6):661-72.

12. Hardin PJ, Jackson MW. An unmanned aerial vehicle for rangeland photography. Rangeland Ecology & Management. 2005;58(4):439-42.

13. Barrero O, Rojas D, Gonzalez C, Perdomo S, editors. Weed detection in rice fields using aerial images and neural networks. Signal Processing, Images and Artificial Vision (STSIVA), 2016 XXI Symposium on; 2016: IEEE.

14. Lemerle D, Verbeek B, Cousens R, Coombes N. The potential for selecting wheat varieties strongly competitive against weeds. Weed Research. 1996;36(6):505-13.

15. Sinden J, Jones R, Hester S, Odom D, Kalisch C, James R, et al. The economic impact of weeds in Australia. Technical Series. 2004;8.

16. Pedersen SM, Fountas S, Have H, Blackmore B. Agricultural robots—system analysis and economic feasibility. Precision agriculture. 2006;7(4):295-308.

17. Kremen C, Merenlender AM, Murphy DD. Ecological monitoring: a vital need for integrated conservation and development programs in the tropics. Conservation biology. 1994;8(2):388-97.

18. Gerhards R, Christensen S. Real‐time weed detection, decision making and patch spraying in maize, sugarbeet, winter wheat and winter barley. Weed research. 2003;43(6):385-92.

19. Freckleton RP, Sutherland WJ, Watkinson AR, Stephens PA. Modelling the effects of management on population dynamics: some lessons from annual weeds. Journal of Applied Ecology. 2008;45(4):1050-8.

20. Pena JM, Torres-Sánchez J, de Castro AI, Kelly M, López-Granados F. Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. PloS one. 2013;8(10):e77151.

21. Xiao D, Tao F, Liu Y, Shi W, Wang M, Liu F, et al. Observed changes in winter wheat phenology in the North China Plain for 1981–2009. International Journal of Biometeorology. 2013;57(2):275-85.

22. Steven M, Malthus T, Demetriades-Shah T, Danson F, Clark J. High-spectral resolution indices for crop stress. High-spectral resolution indices for crop stress. 1990:209-27.

23. Wang N, Zhang N, Dowell FE, Sun Y, Peterson DE. Design of an optical weed sensor usingplant spectral characteristics. Transactions of the ASAE. 2001;44(2):409.

24. Lawless C, Semenov M, Jamieson P. A wheat canopy model linking leaf area and phenology. European Journal of Agronomy. 2005;22(1):19-32.

25. Sakamoto T, Van Nguyen N, Ohno H, Ishitsuka N, Yokozawa M. Spatio–temporal distribution of rice phenology and cropping systems in the Mekong Delta with special reference to the seasonal water flow of the Mekong and Bassac rivers. Remote Sensing of Environment. 2006;100(1):1-16.

26. Vina A, Gitelson AA, Rundquist DC, Keydan G, Leavitt B, Schepers J. Monitoring maize (Zea mays L.) phenology with remote sensing. Agronomy Journal. 2004;96(4):1139-47.

27. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning: Springer series in statistics Springer. Berlin; 2001.

28. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. Neural computation. 1989;1(4):541-51.

29. Sünderhauf N, McCool C, Upcroft B, Perez T, editors. Fine-Grained Plant Classification Using Convolutional Neural Networks for Feature Extraction. CLEF (Working Notes); 2014.

30. Freckleton RP, Hicks HL, Comont D, Crook L, Hull R, Neve P, et al. Measuring the effectiveness of management interventions at regional scales by integrating ecological monitoring and modelling. Pest management science. 2017.

31. Mortensen D, Dieleman JA, Johnson G. Weed spatial variation and weed management. 1998.

32. Perry N, Hull R, Lutman P, editors. Stability of weed patches. 12th European Weed Research Symposium, Wageningen, The Netherlands; 2002.

33. Zeiler MD, Fergus R, editors. Visualizing and understanding convolutional networks. European conference on computer vision; 2014: Springer.

34. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al., editors. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.

35. Hicks HL, Comont D, Coutts SR, Crook L, Hull R, Norris K, et al. The factors driving evolved herbicide resistance at a national scale. Nature ecology & evolution. 2018;2(3):529.

36. Lancashire PD, Bleiholder H, Boom T, Langelüddeke P, Stauss R, WEBER E, et al. A uniform decimal code for growth stages of crops and weeds. Annals of applied Biology. 1991;119(3):561-601.

37. Ballesteros R, Ortega J, Hernández D, Moreno M. Applications of georeferenced high-resolution images obtained with unmanned aerial vehicles. Part I: Description of image acquisition and processing. Precision Agriculture. 2014;15(6):579-92.

38. Wiegand C, Richardson A, Escobar D, Gerbermann A. Vegetation indices in crop assessments. Remote Sensing of Environment. 1991;35(2-3):105-19.

39. Gitelson AA, Kaufman YJ, Merzlyak MN. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote sensing of Environment. 1996;58(3):289-98.

40. Sripada RP, Heiniger RW, White JG, Weisz R. Aerial color infrared photography for determining late-season nitrogen requirements in corn. Agronomy Journal. 2005;97(5):1443-51.

41. Underwood E, Ustin S, DiPietro D. Mapping nonnative plants using hyperspectral imagery. Remote Sensing of Environment. 2003;86(2):150-61.

42. da Silva Junior CA, Nanni MR, Teodoro PE, Silva GFC, de Lima MG, Eri M. Comparison of mapping soybean areas in Brazil through perceptron neural networks and vegetation indices. African Journal of Agricultural Research. 2016;11(43):4413-24.

43. Peña JM, Torres-Sánchez J, Serrano-Pérez A, de Castro AI, López-Granados F. Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. Sensors. 2015;15(3):5609-26.

44. Torres-Sánchez J, López-Granados F, Peña JM. An automatic object-based method for optimal thresholding in UAV images: Application for vegetation detection in herbaceous crops. Computers and Electronics in Agriculture. 2015;114:43-52.

45. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision. 2015;115(3):211-52.

46. Krizhevsky A, Sutskever I, Hinton GE, editors. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems; 2012.

47. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:12070580. 2012.

48. Simard PY, Steinkraus D, Platt JC, editors. Best practices for convolutional neural networks applied to visual document analysis. null; 2003: IEEE.

49. Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging. 2016;35(5):1285-98.

50. Fawcett T. An introduction to ROC analysis. Pattern recognition letters. 2006;27(8):861-74.

51. Carter JV, Pan J, Rai SN, Galandiuk S. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. Surgery. 2016;159(6):1638-45.

52. Fleiss JL, Cohen J, Everitt B. Large sample standard errors of kappa and weighted kappa. Psychological bulletin. 1969;72(5):323.

53. Cohen J. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin. 1968;70(4):213.

54. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988:837-45.

55. Moreno-Torres JG, Raeder T, Alaiz-RodríGuez R, Chawla NV, Herrera F. A unifying view on dataset shift in classification. Pattern Recognition. 2012;45(1):521-30.

56. Gretton A, Smola AJ, Huang J, Schmittfull M, Borgwardt KM, Schölkopf B. Covariate shift by kernel mean matching. 2009.

57. Anderson K, Gaston KJ. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. Frontiers in Ecology and the Environment. 2013;11(3):138-46.

58. JONES IV GP, Pearlstine LG, Percival HF. An assessment of small unmanned aerial vehicles for wildlife research. Wildlife Society Bulletin. 2006;34(3):750-8.

59. Getzin S, Wiegand K, Schöning I. Assessing biodiversity in forests using very high‐resolution images and unmanned aerial vehicles. Methods in Ecology and Evolution. 2012;3(2):397-404.

60. Holman FH, Riche AB, Michalski A, Castle M, Wooster MJ, Hawkesford MJ. High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. Remote Sensing. 2016;8(12):1031.

61. Concepción ED, Díaz M, Baquero RA. Effects of landscape complexity on the ecological effectiveness of agri-environment schemes. Landscape Ecology. 2008;23(2):135-48.

62. Irmak A, Jones J, Batchelor W, Irmak S, Boote K, Paz J. Artificial neural network model as a data analysis tool in precision farming. Transactions of the ASABE. 2006;49(6):2027-37.

63. LÓPEZ‐GRANADOS F, PEÑA‐BARRAGÁN JM, JURADO‐EXPÓSITO M, Francisco‐FERNÁNDEZ M, Cao R, ALONSO‐BETANZOS A, et al. Multispectral classification of grass weeds and wheat (Triticum durum) using linear and nonparametric functional discriminant analysis and neural networks. Weed Research. 2008;48(1):28-37.

64. Mansourian S, Darbandi EI, Mohassel MHR, Rastgoo M, Kanouni H. Comparison of artificial neural networks and logistic regression as potential methods for predicting weed populations on dryland chickpea and winter wheat fields of Kurdistan province, Iran. Crop Protection. 2017;93:43-51.

65. Le Callet P, Viard-Gaudin C, Barba D. A convolutional neural network approach for objective video quality assessment. IEEE Transactions on Neural Networks. 2006;17(5):1316-27.

66. Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2012;42(4):463-84.

**Chapter 4**

**Are Humans better than Convolutional Neural Networks across multiple Vegetation indexes for the identification of Black-grass in UAS imagery?**

**4.1 Introduction**

Weed species pose significant threats to agricultural production worldwide (Radosevich *et al*., 2007, Shabbir and Bajwa, 2006, DiTomaso *et al*., 2017). The threat of reduced agricultural production from weed populations presents farmers and countries with challenges to food security (FAO, 2012). The spread and increase of populations has been aided by human activity (Hulme, 2009). Monitoring the ecological dynamics of these populations would assist farmers to adapt their management practices in the field (Freckleton *et al*., 2017, Fernandez‐Quintanilla, 1988, Heijting *et al*., 2007). The effectiveness of management practices is often dependant on the accuracy and timeliness of data (Hodgson *et al*., 2018), thus advancements in data collection methods could enhance the insights generated on-farm and management outcomes thus improving food security.

Widely-used conventional approaches to monitoring ecological dynamics have relied upon quadrat or transect methodologies (Ludwig *et al*., 1988). This limits the extent of monitoring owing to time and financial pressures as ground based observers are relatively expensive and do not scale efficiently. Satellite remote sensing counters the issues of extent from conventional methods (Pettorelli *et al*., 2005), but often lack the required spatial resolution to monitor farm scale ecological dynamics, or do not provide a consistent reliable data source due to clouds and other climatic events (Thorp and Tian, 2004). Unmanned Arial Systems (UAS) have become a popular option for ecologists to collect data (Koh and Wich, 2012). UAS benefit from their adaptability to carry a range of sensor payloads, and ability to rapidly monitor at ecologically relevant scales (Anderson and Gaston, 2013). Therefore, UAS have been used in a wide variety of ecological settings including surveying Geese (Chabot and Bird, 2012), seals (Krause *et al*., 2017), weeds (Lambert *et al*., 2018) and forests (Zhang *et al.*, 2016).

Agricultural crops and weeds are potentially well suited for UAS mapping with spectral reflectance sensors (Moran *et al.*, 1997, Lamb and Brown, 2001). This is because of factors such as their immobile nature, differentiating spectral characteristics that permit species identification (Zwiggelaar, 1998) and the relatively homogeneous controlled environments resulting from farming practise. A variety of statistical methods have been applied to assess this potential such as object-based image analysis (OBIA) (Getzin *et al*., 2012) Random Forests (RF) (Lambert *et al*., 2018) and Convolutional Neural Networks (CNN) (Cruzan *et al*., 2016).

Vegetation indices (VI) are used for the quantification and evaluation of plant based characteristics such as vigour, vegetation cover and leaf area index *(Xue and Su, 2017, Li et al*., 2014). VI are made from a combination of spectral reflectance bands. Previous research has shown that different bands may reveal unique aspects of plant physiology: near Infrared band for water and carbohydrate content (Batten, 1998) and red band for thermal regulation and abiotic/biotic stress (Oerke *et al.*, 2014). These bands and others can then be combined and mixed to form a large variation of VI’s each capturing different properties of the plant (Bannari *et al*., 1995). Whilst RGB is considered a VI, it contains limited spectral information for observers and analysis, but observers may not take the lack of spectral information into consideration (Jackson and Huete, 1991). There has been little research into the effects of different VI for the classification of weeds, with most studies only focusing on one VI (Chapter 2**).**

As the previous chapters of this thesis highlight, there are challenges in the use of UAS imagery to map black-grass. Notably, the need for ground based observers to provide some reference data from each field for the models to then incorporate, and transferability of the models from field to field was also highlighted. A key drawback from the use of CNN models is their lack of transparent features for the classification of weed density states. This means that troubleshooting the transferability of the models is a greater challenge. Human observers however, would be able to be interviewed to identify the reasoning for their classifications, thus generating more informative descriptions of the “features” they identify. Computer-based problem solving rapidly advanced owing to an increase in data, computational resources and modelling approaches (Cireşan *et al*., 2010). Applications in image classification have used these features to outperform humans across all metrics on the computer vision industry standard benchmark data set, ImageNet (Wu *et al*., 2015).

Citizen scientists and statistical models have been used in studies to generate classifications of canopy phenology and land cover, from remotely sensed images (Richardson *et al*., 2009, Schowengerdt, 2006). There are however few examples of comparing results from models and human observations from remotely sensed data in the field of agriculture(Newman *et al*., 2012). Past approaches have been underpinned by precise ground truthed data from teams of skilled researchers using field walking methodologies to produce maps of weed densities (Queenborough *et al*., 2011). Whilst precise, the field-based researcher approach does not efficiently scale to the spatial extent of modern agricultural needs (Qi *et al*., 2008). Therefore, further examination of ways to increase the spatial extent of these weed mapping methodologies should be assessed. We propose to combine the skills/precision of the researchers and the ability of UAS to generate images on a more relevant spatial extent to provide optimum coverage and observer skill.

Here we assessed the ability of experienced human surveyors to quantify the presence or absence of the arable weed black-grass (*Alopecurus myosuroides*) in UAS imagery. We then compare these results with our models, derived from the same data set. We hypothesise that: i) the RGB VI would result in the highest accuracy of the VI’s for observers; ii) Our statistical models would be more accurate and faster than the trained observers. We show that human observers get the highest AUC score on RGB images. However, our models get higher AUC’s across all the tested VI’s, with the model for GNDVI scoring particularly well. We conclude that the GNDVI model is the best performing setup of those tested to quantify the presence or absence of black-grass in UAS imagery.

**4.2 Materials & Methodology**

*4.2.1 Description of dataset*

We used the UAS imagery dataset and ground-truthed data defined and collected from the field seasons of 2016 & 2017 as outlined in chapter 3 of the thesis (Testing the ability of Unmanned Aerial Systems and machine learning to map weeds at subfield scales: a test with the weed Alopecurus myosuroides (Huds)) to test our hypotheses. Weed densities in this dataset are categorised into five ordinal density states ranging from Absent to Very High. We refactored the density states to create a binary dataset in which weeds were present or absent. We then balanced the two classes, so class frequency would not influence our conclusions. To do this, we randomly sampled the presence data set to select an equal number (3375) of observations of absences from our data set. This resulted in a dataset of both absent and present density states, totalling a dataset of 6750 images.

*4.2.2 Vegetation Index creation*

We chose four VI to assess a range of index types. RGB data allows us to use a full colour index, (i) corrected and modified conventional index, (ii) conventional index and (iii) soil reflectance adjusted index. We used the calibrated spectral bands from the previous study to construct the following three VI:

1. GNDVI – Green Normalised Difference VI (Gitelson *et al*., 1996).

(1)

1. NDVI – Normalised Difference VI (Rouse Jr *et al*., 1974).

(2)

iii) SAVI – Soil Adjusted VI (Huete, 1988).

(3)

Where *fn* is the near Infrared frequency, *fg* is the Green frequency, *fr* is the Red frequency and L is the soil conditioning index, ranging from 0 – 1 where 1 represents high vegetation coverage. We set L to 0.8, due to the timing of our surveying and corresponding vigour resulting in low levels of soil visibility on average.

*4.2.3 Skilled human observer approach*

Twelve individuals were selected, owing to their experience of monitoring black-grass using quadrat-based counts and field-walking methods. The number of observers (12) was chosen to ensure sufficient statistical power for our analysis (Freckleton and Watkinson, 2001). Full ethics approval and consent was gained from the skilled observers prior to their involvement in assessing the images. Each observer was sent our custom computer application with full graphical user interface (Figure 1). At each iteration, the user is shown three random images in which black-grass is present or absent to provide example data. A different randomly chosen image from either class is then positioned in the centre as the test image. The user is then asked to assess whether the test image contains black-grass or not, via a checkbox. The user is not informed of the correct or incorrect answer for each test image. This is repeated for 50 random different images, with each new image provided alongside new labelled examples of present and absent images at the sides. Once completed they are instructed to change the VI they were assessing via a drop down menu, at which point they assess another 50 new images. This process of assessing 50 images per VI is repeated across all four VI’s for each observer. The program records an anonymised user ID, input, time taken per image and the ID of the image they were assessing. This generated 2400 test results across all observers and VI’s.

*4.2.4 Statistical methods*

Our analyses were designed to investigate how observer Area Under the Curve (AUC) scores for each VI changed. The AUC refers to the curve in the Receiver Operating Characteristic (ROC) plot. This metric is a measure of how well two groups can be distinguished from one another based on independent predictions. AUC values range from 0 – 1, with 1 being a perfect classifier. The binary nature of our data and predictions allows us to compute a confusion matrix with four possible assignments: True Negative (TN), False Negative (FN), False Positive (FP) and True Positive (TP). These can be combined into single metrics: Sensitivity or True Positive Rate (TPR) (equation 4) and Specificity or True Negative Rate (TNR) (equation 5). For the X-axis in our AUC plots we use 1 – Specificity (TNR). A line representing Specificity = Sensitivity = 0.5 measures how random chance would perform. Any data below this line is considered worse than random chance. As we know the true observation of each image, we are able to compare it to the user’s or model’s response, allowing us to create confusion matrixes for each combination of VI, user and model.

(4)

(5)

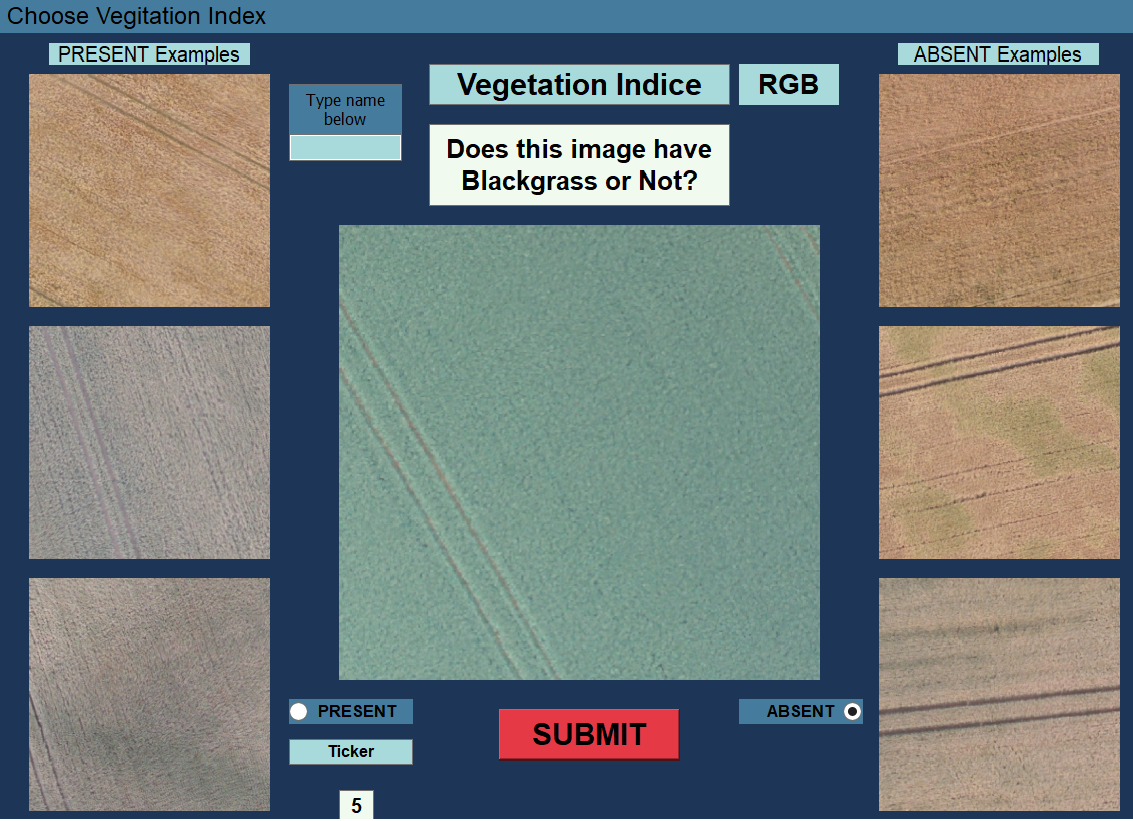
To assess if the results from the AUCs are statistically different from one VI to another or between observers and models, we used the methodology outlined by DeLong *et al* (1988). This is a nonparametric approach and utilises the covariance of the two curves, therefore correlated curves with similar looking AUCs can still be significantly different (Robin *et al*., 2011, Hajian-Tilaki, 2013). We report the outcome of this test when comparing one VI vs all.

Figure 1. Graphical User Interface for the assessment of the presence or absence of black-grass in UAS imagery using the RGB index, by skilled observers.

*4.2.5 CNN modelling approach*

Previous research has shown Convolutional Neural Networks (CNN) perform better than random forests and linear models at modelling weed densities in images. This is due to their usage of the inherent spatial dependencies of pixels in an image (Sünderhauf *et al*., 2014, Szegedy *et al*., 2015, Perry *et al*., 2002).We therefore used the same model structure and parameters as our previous research. In that work, we trained a model on GNDVI images to classify five density states, but here we use our simpler binary presence or absence dataset instead. Consequently, we now only predict two classes instead of the previous five.

*4.2.6 Data Engineering*

In chapter 2 it was shown that it is important to account for the “artefacts” in the images. We applied the same data cleaning methodology as developed in chapter 3. This approach was shown to be the most effective way of maximising the predictive capability of the model, as measured by the AUC score. However, we improved the methodology by implementing a CNN model that would screen for “artefacts” in the images and assign the images into the respective datasets. This increases the reproducibility of the methodology whilst also speeding up the process. The approach is similar to the inductive learning approach taken by Huang and Jensen (1997) used to build an automated knowledge base for the classification of “clean” and “artefact” data. Initially, we manually “cleaned” one dataset (GNDVI). We used imageJ to split each 20x20m plot into 16 smaller images, these subplots were then analysed and grouped into “clean” or “artefact” subsets. We used these subsets to generate labels for each category so that we could train a new CNN to automatically screen the remaining VI datasets. We then undertook quality control checks on the “clean” and “artefact” datasets to make sure they were constructed as intended and followed the upscaling procedure when assessing the test datasets.

*4.2.7 Reasons for misclassification*

To assess if there are any common features leading to the correct or erroneous prediction of the presence or absence of black-grass in an image by human observers, we identified all images that were seen by two or more observers (1054 of 2400 images), this resulted in 482 unique images. We then calculated the mean accuracy across observers for each image. We identified all images where the mean accuracy was either 0 or 1, indicating all observers were wrong (79 images) or correct (133 images) respectively. We applied a chi square test to these data sets and assess the residuals to see if any VI was over or under represented. We tested whether there was an effect of average data quality or time taken to answer on these two groups by performing a two-sample t-test. We also test whether there are statistical differences between the times taken across observers to classify each VI. We do this by performing multiple pairwise-comparisons between the means of each VI using a Tukey multiple comparison of means test.

**4.3 Results**

*4.3.1 Trained Observers*

We found that collectively, the observers performed best in distinguishing the presence of black-grass in the RGB images (AUC = 0.62), as shown in Figure 2. However, observers are able to distinguish black-grass in GNDVI images at a rate only 1% better than random. We found that difference in AUC values between RGB and all other VI’s to be statistically significant as seen in Table 1.

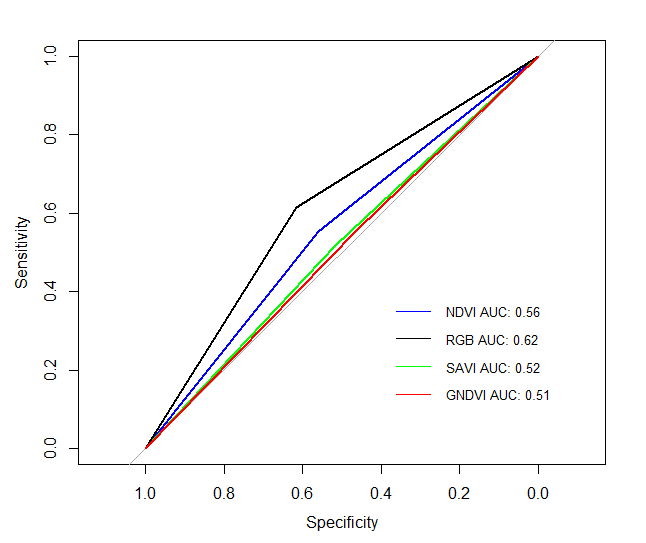


Figure 2 AUC plot to show the ability of skilled observers to assess the presence or absence of black-grass in UAS imagery across multiple VI’s.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| VI | AUC 1 | AUC 2 | D | p-value |
| GNDVI | 0.615 | 0.509 | 3.6906 | 2.338e-4 |
| NDVI | 0.615 | 0.556 | 2.0498 | 0.0406 |
| SAVI | 0.615 | 0.516 | 3.4463 | 5.881e-4 |

Table 1 displays the results from the DeLong statistical significance test of AUC plots, here the score of all the observers combined for the RGB index (AUC 1) is being compared to their score for the remaining VI’s (AUC 2). D is the difference calculated, D=(AUC1-AUC2)/s, where s is the standard deviation of the bootstrapped differences.

*4.3.2 CNN model*

We found that our CNN models outperform skilled observers across all VI’s. That GNDVI records the highest AUC of 0.81 and that NDVI scores the lowest with an AUC of 0.58 in Figure 3. Table 2 shows that the GNDVI AUC is significantly different to all the other VI’s.

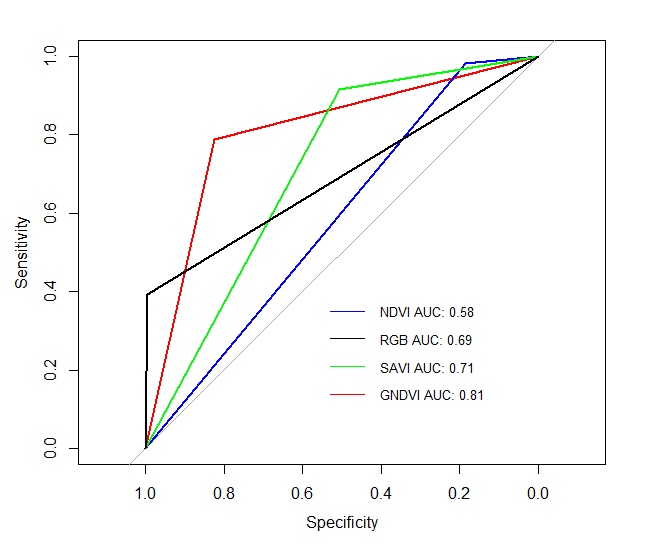


Figure 3 AUC plot of four CNN models, trained on different VI’s to assess the presence or absence of black-grass in UAS imagery.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| VI | AUC 1 |  |  | AUC 2 | D | p-value |
| RGB | 0.806 |  |  | 0.693 | 6.8784 | 8.12e-12 |
| NDVI | 0.806 |  |  | 0.584 | 14.245 | 2.2e-16 |
| SAVI | 0.806 |  |  | 0.710 | 5.3041 | 1.259e-07 |

Table 2 displays the results from the DeLong statistical significance test of AUC plots, here the score of the CNN model for the GNDVI index (AUC 1) is being compared to the scores for the remaining VI’s (AUC 2) from the models. D is the difference calculated, D=(AUC1-AUC2)/s, where s is the standard deviation of the bootstrapped differences.

4.3.3 *Error analysis*

Table 3 shows our chi-squared results when comparing all examples where observers were either all correct or all wrong about the same image. This test compares the evenly distributed expected counts with the observed counts for each VI in each observation scenario. When observers were all wrong about an image, they performed worse than expected on the SAVI index (2.593) and better than expected on the RGB index (-3.064). When observers were all correct about an image, they performed better than expected on the RGB index (2.743) and worse on the GNDVI index (-1.232).

Table 3. Chi-squared residuals for observers when grouped by their prediction of an image that was seen by multiple observers and their resulting predictions being either all wrong or all correct for the image. Pearson residuals are reported here, covering the differences between the observed and expected values. Positive values for all wrong observations indicated worse than expected performance. Conversely positive values for all correct predictions indicate a greater number of observations than expected.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Observations | GNDVI | NDVI | RGB | SAVI |
| All wrong (79) | 0.393 | 0.079 | -3.064 | 2.593 |
| All correct (133) | -1.232 | -0.511 | 2.734 | -0.991 |

We found that when all observers were correct about an image, they had an average answer time of 4.15 seconds and when wrong, a time of 3.98 seconds as seen in Figure 3.Our two-sample t-test showed this difference to not be statistically significant with a p-value of 0.341. We found that the data quality (see chapter 3 for definition) of 20x20m plots where all observers were wrong, was on average 68.7% and where all observers were correct 71.7%.



Figure 3, plot of mean time taken against per image when all observers are either correct or incorrect about an image.

When assessing differences in time taken to provide an answer for observers across the four VI, we found that the NDVI index took 4.82 seconds on average (Figure 4). We use a one-way ANOVA test to show that some of the group means are significantly different (p-value = 2.99e-05), then we show with the Tukey HSD test, that the NDVI result is statistically significant when compared to the remaining VI’s (Table 4) with all combinations of NDVI resulting in adjusted p-values below 0.005 .

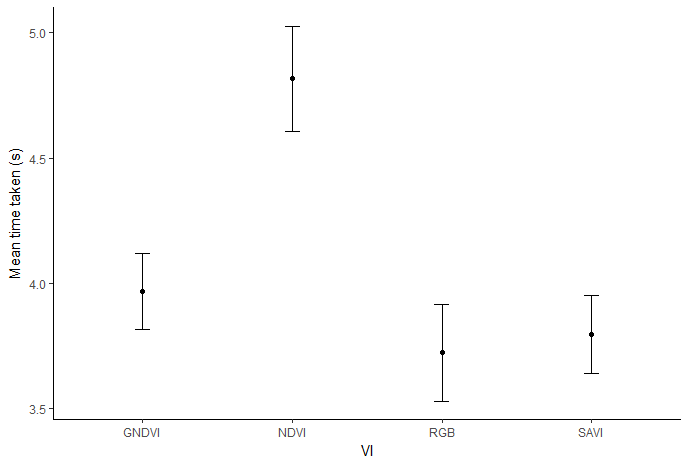


Figure 4, plot of mean time taken for all observers against each VI.

Table 4, pairwise-comparison between the mean time to answer for each VI and the corresponding adjusted p-value.

|  |  |
| --- | --- |
| VI combo | Adjusted p-value |
| NDVI-GNDVI | 0.0046 |
| RGB-GNDVI | 0.7687 |
| SAVI-GNDVI | 0.9060 |
| RGB-NDVI | 0.0001 |
| SAVI-NDVI | 0.0003 |
| SAVI-RGB | 0.9915 |

**4.4 Discussion**

We set out to investigate the use of alternative VI’s for the classification of *A. myosurodies* (black-grass) in UAS imagery. We also assessed the potential for the use of skilled observers to inform this classification process, when compared to our best performing models. We addressed the optimum methodology for our classification by setting up 12 skilled surveyors to be shown 50 random examples representing the presence or absence of the weed in an image for each VI and recorded their response. Our main finding was that skilled observers do not add additional precision or insight to the process of the identification of black-grass from UAS imagery. Additionally, GNDVI is the best performing index while using the CNN method and that RGB is the best performing metric for skilled observers. This study shows that GNDVI with a CNN is the optimum VI and methodology for the classification of black-grass.

4.4.1 *Implications of findings*

The implications of these results have important consequences for farmers who use off-the-shelf drones or commercial services that map farms using RGB sensors (McKinnon., *et al* 2017). This limits the informative spectral information that can be generated from UAS mapping and has the potential to prevent meaningful use of UAS in the field of precision agriculture. This is due to many commercial service providers in this sector, overpromising and underdelivering on promised breakthroughs to farmers and thus tainting farmers’ experience of UAS on farms (personal communication). This work shows how the use of the Sequoia sensor, a commercial grade calibrated multispectral camera with additional RGB sensor, improves upon the use of “modified” spectral sensors that have been used in previous studies (Lambert *et al*., 2018). The use of an off-the-shelf sensor makes the methodology more reproducible and reduces the potential of upstream data quality issues, before the image is even captured (Jensen., *et al* 2013).

Our use of 12 skilled observers is an attempt at a small scale citizen science study. UAS images and citizen science projects have been used before for the counting of seabirds and Lesser Snow Geese (Bowley *et al*., 2017, Hodgson *et al*., 2018). These studies have shown that CNN and generalised linear models respectively, are better than human observers over many metrics but have only focused on RGB data. Therefore, with our use of CNN and multiple VIs we conclude that statistical approaches are better suited than human observers, we also show that this is true across multiple VI. Whilst on a smaller scale than these past studies, our use of “skilled” observers, was designed to sufficiently identify effects in the study group, see Appendix Figure 1 for an observer by observer breakdown of their AUC scores. This shows that there is no smaller group or individual from our test set who compares to our CNN models. A scenario in which only a few observers were as good or better than the models would be informative so that interviews could be undertaken to assess the heuristics of their decision making for the classification.

There are studies that look at multiple VIs (Khan *et al*., 2018), but they are then limited in the spatial extent of their investigation, in this case only looking at one field over the course of a season whereas ours covered 102 fields across 2 years. However, the study looked to convert an RGB image into a VI via the use of a CNN. They show that via their method, the RGB data performed better. This contradicts with our results where we used the calibrated spectral bands to construct the VI and resulted in all but the NDVI model performing better than the RGB data.

4.4.2 Conclusions

This work provides evidence corroborating the assumptions that have been made throughout our past work, notably on the choice of VI and use of statistical models over the use of skilled observers for the optimum classification of black-grass states from UAS imagery.

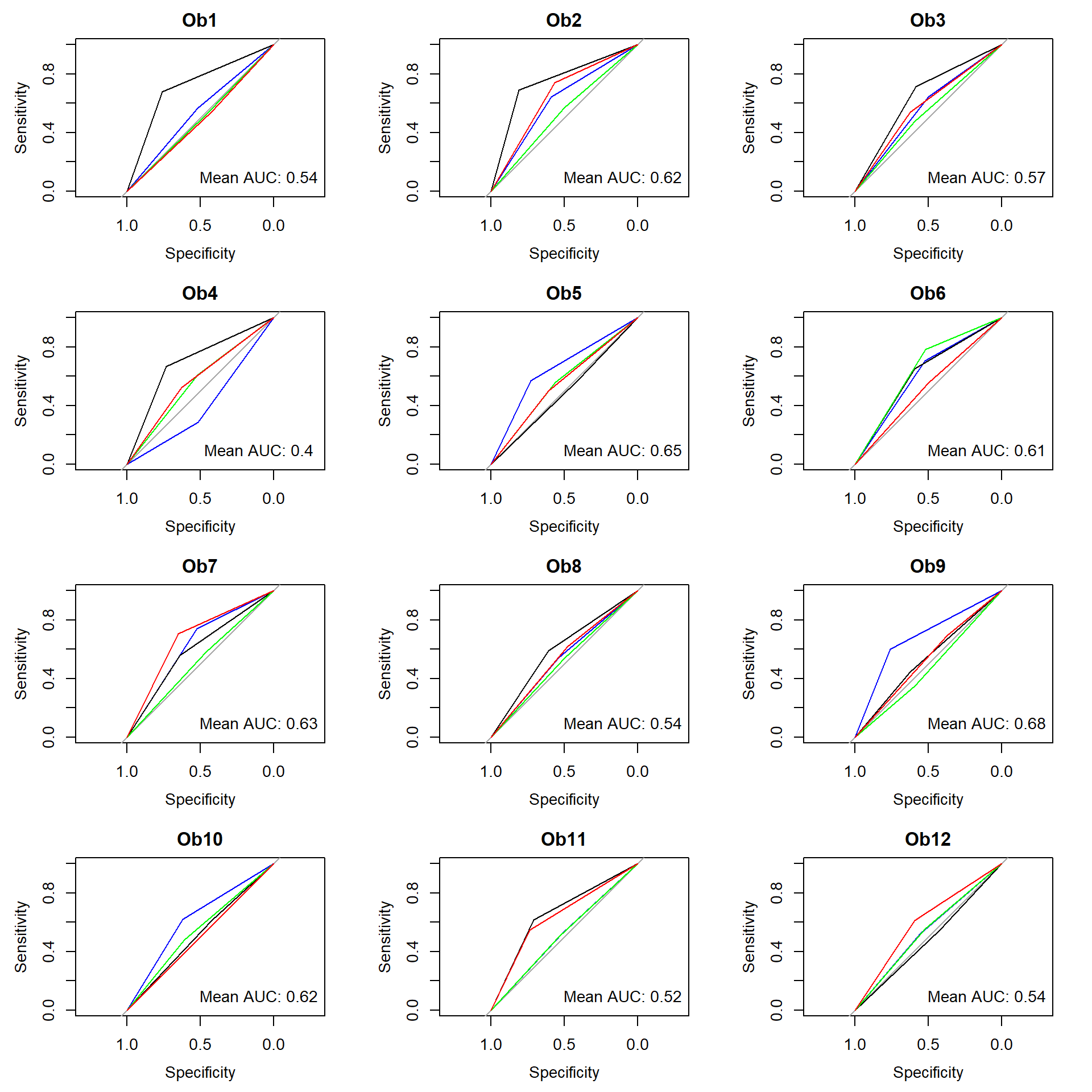
**4.5 Appendix**

Figure 1, AUC plots for each VI for each observer and mean AUC.

**4.6 References**

ANDERSON, K. & GASTON, K. J. 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment,* 11**,** 138-146.

BANNARI, A., MORIN, D., BONN, F. & HUETE, A. 1995. A review of vegetation indices. *Remote sensing reviews,* 13**,** 95-120.

BATTEN, G. 1998. Plant analysis using near infrared reflectance spectroscopy: the potential and the limitations. *Australian Journal of Experimental Agriculture,* 38**,** 697-706.

BOWLEY, C., MATTINGLY, M., BARNAS, A., ELLIS-FELEGE, S. & DESELL, T. Toward using citizen scientists to drive automated ecological object detection in aerial imagery. e-Science (e-Science), 2017 IEEE 13th International Conference on, 2017. IEEE, 99-108.

CHABOT, D. & BIRD, D. M. 2012. Evaluation of an off-the-shelf unmanned aircraft system for surveying flocks of geese. *Waterbirds,* 35**,** 170-174.

CIREŞAN, D. C., MEIER, U., GAMBARDELLA, L. M. & SCHMIDHUBER, J. 2010. Deep, big, simple neural nets for handwritten digit recognition. *Neural computation,* 22**,** 3207-3220.

CRUZAN, M. B., WEINSTEIN, B. G., GRASTY, M. R., KOHRN, B. F., HENDRICKSON, E. C., ARREDONDO, T. M. & THOMPSON, P. G. 2016. Small unmanned aerial vehicles (micro‐UAVs, drones) in plant ecology. *Applications in plant sciences,* 4**,** 1600041.

DELONG, E. R., DELONG, D. M. & CLARKE-PEARSON, D. L. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics***,** 837-845.

DITOMASO, J. M., VAN STEENWYK, R. A., NOWIERSKI, R. M., VOLLMER, J. L., LANE, E., CHILTON, E., BURCH, P. L., COWAN, P. E., ZIMMERMAN, K. & DIONIGI, C. P. 2017. Enhancing the effectiveness of biological control programs of invasive species through a more comprehensive pest management approach. *Pest management science,* 73**,** 9-13.

FAO, F. 2012. Cereal Supply and Demand Brief. October.

FERNANDEZ‐QUINTANILLA, C. 1988. Studying the population dynamics of weeds. *Weed Research,* 28**,** 443-447.

FRECKLETON, R. P., HICKS, H. L., COMONT, D., CROOK, L., HULL, R., NEVE, P. & CHILDS, D. Z. 2017. Measuring the effectiveness of management interventions at regional scales by integrating ecological monitoring and modelling. *Pest management science*.

FRECKLETON, R. P. & WATKINSON, A. R. 2001. Nonmanipulative determination of plant community dynamics. *Trends in Ecology & Evolution,* 16**,** 301-307.

GETZIN, S., WIEGAND, K. & SCHÖNING, I. 2012. Assessing biodiversity in forests using very high‐resolution images and unmanned aerial vehicles. *Methods in Ecology and Evolution,* 3**,** 397-404.

GITELSON, A. A., KAUFMAN, Y. J. & MERZLYAK, M. N. 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote sensing of Environment,* 58**,** 289-298.

HAJIAN-TILAKI, K. 2013. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian journal of internal medicine,* 4**,** 627.

HEIJTING, S., VAN DER WERF, W., STEIN, A. & KROPFF, M. 2007. Are weed patches stable in location? Application of an explicitly two‐dimensional methodology. *Weed Research,* 47**,** 381-395.

HODGSON, J. C., MOTT, R., BAYLIS, S. M., PHAM, T. T., WOTHERSPOON, S., KILPATRICK, A. D., RAJA SEGARAN, R., REID, I., TERAUDS, A. & KOH, L. P. 2018. Drones count wildlife more accurately and precisely than humans. *Methods in Ecology and Evolution,* 9**,** 1160-1167.

HUANG, X. & JENSEN, J. R. 1997. A machine-learning approach to automated knowledge-base building for remote sensing image analysis with GIS data. *Photogrammetric engineering and remote sensing,* 63**,** 1185-1193.

HUETE, A. R. 1988. A soil-adjusted vegetation index (SAVI). *Remote sensing of environment,* 25**,** 295-309.

HULME, P. E. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of applied ecology,* 46**,** 10-18.

JACKSON, R. D. & HUETE, A. R. 1991. Interpreting vegetation indices. *Preventive veterinary medicine,* 11**,** 185-200.

JENSEN, S., Plale, B., Aktas, M.S., Luo, Y., Chen, P. and Conover, H., 2013. Provenance capture and use in a satellite data processing pipeline. IEEE Transactions on Geoscience and Remote Sensing, 51(11), pp.5090-5097.

KHAN, Z., RAHIMI-EICHI, V., HAEFELE, S., GARNETT, T. & MIKLAVCIC, S. J. 2018. Estimation of vegetation indices for high-throughput phenotyping of wheat using aerial imaging. *Plant methods,* 14**,** 20.

KOH, L. P. & WICH, S. A. 2012. Dawn of drone ecology: low-cost autonomous aerial vehicles for conservation. *Tropical Conservation Science,* 5**,** 121-132.

KRAUSE, D. J., HINKE, J. T., PERRYMAN, W. L., GOEBEL, M. E. & LEROI, D. J. 2017. An accurate and adaptable photogrammetric approach for estimating the mass and body condition of pinnipeds using an unmanned aerial system. *PloS one,* 12**,** e0187465.

LAMB, D. & BROWN, R. B. 2001. Pa—precision agriculture: Remote-sensing and mapping of weeds in crops. *Journal of Agricultural Engineering Research,* 78**,** 117-125.

LAMBERT, J., HICKS, H., CHILDS, D. & FRECKLETON, R. 2018. Evaluating the potential of Unmanned Aerial Systems for mapping weeds at field scales: a case study with Alopecurus myosuroides. *Weed research,* 58**,** 35-45.

LI, F., MISTELE, B., HU, Y., CHEN, X. & SCHMIDHALTER, U. 2014. Reflectance estimation of canopy nitrogen content in winter wheat using optimised hyperspectral spectral indices and partial least squares regression. *European Journal of Agronomy,* 52**,** 198-209.

LUDWIG, J. A., QUARTET, L. & REYNOLDS, J. F. 1988. *Statistical ecology: a primer in methods and computing*, John Wiley & Sons.

MCKINNON, T. and Hoff, P., 2017. Comparing RGB-based vegetation indices with NDVI for agricultural drone imagery.

MORAN, M. S., INOUE, Y. & BARNES, E. 1997. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote sensing of Environment,* 61**,** 319-346.

NEWMAN, G., WIGGINS, A., CRALL, A., GRAHAM, E., NEWMAN, S. & CROWSTON, K. 2012. The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment,* 10**,** 298-304.

OERKE, E.-C., MAHLEIN, A.-K. & STEINER, U. 2014. Proximal sensing of plant diseases. *Detection and Diagnostics of Plant Pathogens.* Springer.

PERRY, N., HULL, R. & LUTMAN, P. Stability of weed patches. 12th European Weed Research Symposium, Wageningen, The Netherlands, 2002. 398-399.

PETTORELLI, N., VIK, J. O., MYSTERUD, A., GAILLARD, J.-M., TUCKER, C. J. & STENSETH, N. C. 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in ecology & evolution,* 20**,** 503-510.

QI, A., PERRY, J., PIDGEON, J., HAYLOCK, L. & BROOKS, D. 2008. Cost‐efficacy in measuring farmland biodiversity–lessons from the Farm Scale Evaluations of genetically modified herbicide‐tolerant crops. *Annals of Applied Biology,* 152**,** 93-101.

QUEENBOROUGH, S. A., BURNET, K. M., SUTHERLAND, W. J., WATKINSON, A. R. & FRECKLETON, R. P. 2011. From meso‐to macroscale population dynamics: a new density‐structured approach. *Methods in Ecology and Evolution,* 2**,** 289-302.

RADOSEVICH, S. R., HOLT, J. S. & GHERSA, C. M. 2007. *Ecology of weeds and invasive plants: relationship to agriculture and natural resource management*, John Wiley & Sons.

RICHARDSON, A. D., BRASWELL, B. H., HOLLINGER, D. Y., JENKINS, J. P. & OLLINGER, S. V. 2009. Near‐surface remote sensing of spatial and temporal variation in canopy phenology. *Ecological Applications,* 19**,** 1417-1428.

ROBIN, X., TURCK, N., HAINARD, A., TIBERTI, N., LISACEK, F., SANCHEZ, J.-C. & MÜLLER, M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics,* 12**,** 77.

ROUSE JR, J. W., HAAS, R., SCHELL, J. & DEERING, D. 1974. Monitoring vegetation systems in the Great Plains with ERTS.

SCHOWENGERDT, R. A. 2006. *Remote sensing: models and methods for image processing*, Elsevier.

SHABBIR, A. & BAJWA, R. 2006. Distribution of parthenium weed (Parthenium hysterophorus L.), an alien invasive weed species threatening the biodiversity of Islamabad. *Weed Biology and Management,* 6**,** 89-95.

SÜNDERHAUF, N., MCCOOL, C., UPCROFT, B. & PEREZ, T. Fine-Grained Plant Classification Using Convolutional Neural Networks for Feature Extraction. CLEF (Working Notes), 2014. 756-762.

SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D., ERHAN, D., VANHOUCKE, V. & RABINOVICH, A. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 1-9.

THORP, K. & TIAN, L. 2004. A review on remote sensing of weeds in agriculture. *Precision Agriculture,* 5**,** 477-508.

WU, R., YAN, S., SHAN, Y., DANG, Q. & SUN, G. 2015. Deep image: Scaling up image recognition. *arXiv preprint arXiv:1501.02876*.

XUE, J. & SU, B. 2017. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors,* 2017.

ZHANG, J., HU, J., LIAN, J., FAN, Z., OUYANG, X. & YE, W. 2016. Seeing the forest from drones: Testing the potential of lightweight drones as a tool for long-term forest monitoring. *Biological Conservation,* 198**,** 60-69.

ZWIGGELAAR, R. 1998. A review of spectral properties of plants and their potential use for crop/weed discrimination in row-crops. *Crop Protection,* 17**,** 189-206.

**Chapter 5**

**5. Discussion**

**5.1 Summary**

The aim of this thesis was to evaluate the use of imagery generated from UAS to monitor variation in densities of *A. myosuroides* (black-grass) across multiple field sites in the UK over several years. I have demonstrated that UAS imagery can to some extent be used to map density states, and that the application of machine learning statistical methods advances the potential of UAS as a data collection platform. Each chapter examined elements of this broad objective as outlined in the introduction. Here I discuss my findings and their implications for researchers and farmers, as well as some possible directions for future work.

*Chapter 2 – Evaluating the potential of Unmanned Aerial Systems for mapping weeds at field scales: a case study with Alopecurus myosuroides.*

I established the underlying methodologies and tested the ability of UAS generated imagery against observations generated by field walking methods from a team of observers. I show that linear models and random forests trained on UAS imagery with basic data engineering, are correlated with ground-truthed observations for the detection of either the presence or absence of black-grass. However, the models show little practical ability to detect black-grass if applied to new fields and the results revealed challenges in automating several stages of the methodology to increase reproducibility.

*Chapter 3 – Testing the ability of Unmanned Aerial Systems and machine learning to map weeds at subfield scales: a test with the weed Alopecurus myosuroides (Huds).*

I refined the data collection methodology, and utilised calibrated multispectral image data. This data collection methodology was applied across two field seasons to generate the largest known dataset of labelled, remotely sensed weed imagery. I show how state of the art machine learning methods, when applied to this larger dataset result in the best performance, whilst modelling the full five-density states. I also examined if there are factors that limit the transferability of the models and assess if the date of mapping the field impacts model performance. I showed that time of visit between years and week of mapping has no effect on the performance of the model, suggesting that extensive late in-season monitoring can potentially be undertaken.

*Chapter 4 – An assessment of the optimum VI for the classification of black-grass in UAS imagery and the ability of skilled observer’s compared to statistical models.*

I aimed to compare how the VI I had been using for Chapter 3 compared to alternatives that had been selected in other literature. I reveal that GNDVI is the best performing, reinforcing the conclusions of the previous chapter. When comparing skilled observers to my models, I show that the models perform better than all observers across all Vegetation Indexes (VI) tested.

**5.2 Single field vs multi field**

A major novel element of the research presented here is the scope at which it was undertaken. The majority of past studies have focused on single “controlled” field studies and studies are an exception if they look at two to five fields as highlighted throughout the chapters. My work covered 142 fields over three field seasons in total. When implementing new technologies, I believe they should be tested at appropriate scales and I calculate that we imaged over 1136 hectares of UK wheat fields during the field seasons, which its self is ~0.06% of the 1.8 million hectares of wheat cultivated annually in the UK. Therefore, I believe that even more data needs to be collected to further test the effectiveness of the use of UAS and machine learning to automate agriculture, beyond the proof of concept and development of methodologies in the work I have presented here. Further research is being undertaken using the methodologies developed here and applied to a greater range of spectral values.

By considering 142 separate fields I was able to investigate the relationships between black-grass density states and UAS imagery, showing an inconsistent ability to predict across fields and counties. This indicates that the features identified by the model in one field are not generalised enough to currently be applied in agronomic scenarios, even though the results reported here are significant improvements over past literature and methodologies. Recent work has shown (Metcalfe, 2017) that incorporating meta data such as pH, soil moisture content and topography into models can lead to improvements in predictions of black-grass locations in subsequent years. However, these meta data require high sampling intensity within a field, an aspect that my work is attempting reduce. There is literature on the identification of such edaphic properties from remotely sensed satellite data (Ben-Dor, 2002, Nanni and Demattê, 2006, Peng *et al*., 2017), so it is conceivable that this data could be integrated with our UAS mapping approach for better results.

It is also worth commenting on the setup of the field season to map in fields with *known* black-grass populations. This inherently biases the data collection towards the presence of the weed in the data set. I believe the incorporation of a combination of fields that prioritise the variation or diversity of the level of weed density states and “clean” (fields without black-grass in) fields into future field seasons would help remove this data imbalance and would be more representative of fields on a national scale.

The results/metrics that we report back are not necessarily the best performing in the agronomic literature, but we believe this is due to the narrow focus of all previous research in the area and the misuse of statistics such as Barrero *et al* (2016). We also highlight how many other papers in the literature may have overstated the results for applications of these technologies on farm, performance of their stated methodologies and reported metrics from in-sample rather than out-of-sample testing (Getzin *et al*., 2012, Holman *et al*., 2016, JONES IV *et al*., 2006). Our LOFO-CV approach is a good first step towards trying to understand the variability between fields. This is a key component for the roll out of this methodology into commercial settings as the approach currently requires some ground truthed data from the field to “prime” the model. Whilst this is not impractical for an individual to collect during a UAS flight, it does limit the potential scope of the use of UAS as a fully autonomous data collection platform.

**5.3 Usage of Vegetation Indexes**

One of the key decisions to make when undertaking remote sensing of vegetation is the choice of VI to use. Over the course of this thesis there have been advancements in sensor technology, notably the miniaturisation and cost reduction of multi spectral cameras. As UAS are limited by the weight they can fly with the miniaturisation has been the most significant factor for their increased usage. Initially in the first field season (2015) a modified GoPro was used, primarily because of its size rather than its well-documented usage in agricultural mapping. (Goli *et al., 2015*) This limited the spectral information which we were able to obtain, preventing a full analysis of VI usage. To our knowledge we were the first academics to implement the multispectral Sequoia sensor with calibrated spectral channels in the subsequent field seasons. This relatively low cost sensor with identical dimensions to the GoPro, allowed us to capture greater amounts of high quality data, thus allowing a full examination of VI for our objectives. Future developments of sensor technology, I believe will see an increase in spectral resolution and the inclusion of more spectral frequency. Future directions of research in this area could be to focus on hyperspectral imaging systems in a lab environment. This would allow wheat and black-grass to be cultivated and then imaged in highly controlled settings. Potentially a spectral band could then be identified that is able to distinguish between the weed and crop with high levels of accuracy. These bands could then be incorporated into UAS appropriate sensor to create a black-grass mapping sensor.

The work presented here disagrees with the majority held belief in NDVI as the optimum index for cereal crop mapping. Literature has already stated that the index saturates at high vegetation levels (Mutanga and Skidmore, 2004a, Wu *et al*., 2008, Mutanga and Skidmore, 2004b), exactly the scenario in which attempting to identify one grass species within another will reduce capabilities. We show that the GNDVI, a corrected and modified conventional index outperforms, full colour, conventional and soil reflectance adjusted indexes.

**5.4 Late season vs early season**

Our approach is based on late season monitoring, given that we assess the weed at a BBCH growth stage of 87-89. Agronomically the weed maps generated from surveying at this time are only of use in the subsequent seasons of weed management as it is too late in the season to affect either weed density or crop yield (Rew *et al*., 2001). However, research shows that maps of weed from this period often don’t exactly reflect the presence of weed seedlings in the next cultivation cycle (Walter *et al*., 2002) and that buffer zones should be incorporated into patch spraying methods. Concurrently early season, post-emergence mapping with UAS involves crude estimators of weed presence by defining a crop row and assigning everything not in a row as a weed (Pena *et al*., 2013). This method has also not been extended beyond a trial 140x100m plot. But the weed maps generated here do allow in-season adaptation of management practices. Therefore, future applications of these methodologies suggest that the optimal weed mapping solution would not be an either or scenario between early or late season mapping but the incorporation of both to form continuous weed mapping.

**5.5 Statistical approaches**

As chapter 3 in particular has shown, the application of more advanced statistical models and what is classified as deep learning techniques such as CNN have advanced our ability to classify weed populations from UAS imagery. With this increased performance there has been a loss in model interpretability as the features that CNN models use are not necessarily human eligible. This is an issue when we then look into troubleshooting the transferability of models between fields.

An interesting paper on the transferability of CNN features, (Yosinski *et al*., 2014) states that transferability is negatively affected by two issues: the specialisation of higher layer neurons and the difficulties related to splitting between co-adapted neurons. They concluded that if the test dataset is small and the number of parameters in the model is large then the finetuning may result in overfitting. Applying this to our work, our test dataset in the LOFO-CV tests are small (~110 plots) and the number of parameters in the Inception V3 model architecture that we use is relatively large (23 million) as defined by Szegedy *et al* (2015). This suggests that our models may be overfitting at the field level.

Over the course of the past four years of this research the field of machine learning and in particular CNN has progressed significantly. With a combination of factors such as larger datasets, greater computational resources, more efficient frameworks and more researchers focusing on the field, there is sure to be even more fundamental advancements to come in the following years.

**5.6 Conclusion**

Large scale population monitoring is crucial for preserving, maintaining and extending the capabilities of natural systems with the rapid pace of global change. Continued development of methods for data collection and statistical analysis are essential for the large-scale management of agricultural systems. However, it is important that methods are rigorously tested, and the limitations fully discussed in the literature before full usage in applied and commercial settings.

**5.7 References**

BARRERO, O., ROJAS, D., GONZALEZ, C. & PERDOMO, S. Weed detection in rice fields using aerial images and neural networks. Signal Processing, Images and Artificial Vision (STSIVA), 2016 XXI Symposium on, 2016. IEEE, 1-4.

BEN-DOR, E. 2002. Quantitative remote sensing of soil properties.

GETZIN, S., WIEGAND, K. & SCHÖNING, I. 2012. Assessing biodiversity in forests using very high‐resolution images and unmanned aerial vehicles. *Methods in Ecology and Evolution,* 3**,** 397-404.

GOLI, N.R. & LANDRUM, B., 2015. Development of an Integrated UAS for Agricultural Imaging Applications. In *15th AIAA Aviation Technology, Integration, and Operations Conference* (p. 3286).

HOLMAN, F. H., RICHE, A. B., MICHALSKI, A., CASTLE, M., WOOSTER, M. J. & HAWKESFORD, M. J. 2016. High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sensing,* 8**,** 1031.

JONES IV, G. P., PEARLSTINE, L. G. & PERCIVAL, H. F. 2006. An assessment of small unmanned aerial vehicles for wildlife research. *Wildlife Society Bulletin,* 34**,** 750-758.

METCALFE, H. 2017. *Modelling the spatial variation in Alopecurus myosuroides for precision weed management.* University of Reading.

MUTANGA, O. & SKIDMORE, A. K. 2004a. Hyperspectral band depth analysis for a better estimation of grass biomass (Cenchrus ciliaris) measured under controlled laboratory conditions. *International journal of applied earth observation and geoinformation,* 5**,** 87-96.

MUTANGA, O. & SKIDMORE, A. K. 2004b. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing,* 25**,** 3999-4014.

NANNI, M. R. & DEMATTÊ, J. A. M. 2006. Spectral reflectance methodology in comparison to traditional soil analysis. *Soil Science Society of America Journal,* 70**,** 393-407.

PENA, J. M., TORRES-SÁNCHEZ, J., DE CASTRO, A. I., KELLY, M. & LÓPEZ-GRANADOS, F. 2013. Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. *PloS one,* 8**,** e77151.

PENG, J., LOEW, A., MERLIN, O. & VERHOEST, N. E. 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics,* 55**,** 341-366.

REW, L.J. and Cousens, R.D., 2001. Spatial distribution of weeds in arable crops: are current sampling and analytical methods appropriate?. *Weed Research*, *41*(1), pp.1-18

SZEGEDY, C., LIU, W., JIA, Y., SERMANET, P., REED, S., ANGUELOV, D., ERHAN, D., VANHOUCKE, V. & RABINOVICH, A. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 1-9.

WALTER, A., CHRISTENSEN, S. & SIMMELSGAARD, S. E. 2002. Spatial correlation between weed species densities and soil properties. *Weed Research,* 42**,** 26-38.

WU, C., NIU, Z., TANG, Q. & HUANG, W. 2008. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and forest meteorology,* 148**,** 1230-1241.

YOSINSKI, J., CLUNE, J., BENGIO, Y. & LIPSON, H. How transferable are features in deep neural networks? Advances in neural information processing systems, 2014. 3320-3328.

1. "Solo - The Smart Drone | Commercial Drone Platform." <https://3dr.com/solo-drone/>. Accessed 11 Jan. 2018. [↑](#footnote-ref-1)
2. "Sequoia - MicaSense." https://www.micasense.com/sequoia/. Accessed 11 Jan. 2018. [↑](#footnote-ref-2)