

# **Sustainable Intensification of Arable Agriculture: The Role of Earth Observation in Quantifying the Agricultural Landscape**

**Merryn Hunt**

This thesis is submitted in partial fulfilment of the requirement for the award of the degree of  
Doctor of Philosophy



June 2020

## **Abstract**

By 2050, global food production must increase by 70% to meet the demands of a growing population with shifting food consumption patterns. Sustainable intensification has been suggested as a possible mechanism to meet this demand without significant detrimental impact to the environment. Appropriate monitoring techniques are required to ensure that attempts to sustainably intensify arable agriculture are successful. Current assessments rely on datasets with limited spatial and temporal resolution and coverage such as field data and farm surveys. Earth Observation (EO) data overcome limitations of resolution and coverage, and have the potential to make a significant contribution to sustainable intensification assessments. Despite the variety of established EO-based methods to assess multiple indicators of agricultural intensity (e.g. yield) and environmental quality (e.g. vegetation and ecosystem health), to date no one has attempted to combine these methods to provide an assessment of sustainable intensification.

The aim of this thesis, therefore, is to demonstrate the feasibility of using EO to assess the sustainability of agricultural intensification. This is achieved by constructing two novel EO-based indicators of agricultural intensity and environmental quality, namely wheat yield and farmland bird richness. By combining these indicators, a novel performance feature space is created that can be used to assess the relative performance of arable areas. This thesis demonstrates that integrating EO data with *in situ* data allows assessments of agricultural performance to be made across broad spatial scales unobtainable with field data alone. This feature space can provide an assessment of the relative performance of individual arable areas, providing valuable information to identify best management practices in different areas and inform future management and policy decisions. The demonstration of this agricultural performance assessment method represents an important first step in the creation of an operational EO-based monitoring system to assess sustainable intensification, ensuring we are able to meet future food demands in an environmentally sustainable way.

## **Table of Contents**

<b>List of Figures .....</b>	<b>v</b>
<b>List of Tables .....</b>	<b>vii</b>
<b>Acknowledgements .....</b>	<b>viii</b>
<b>Declaration .....</b>	<b>ix</b>
<b>Statement of authorship for multi-authored chapters .....</b>	<b>x</b>
<b>1. Introduction.....</b>	<b>1</b>
<b>2. Monitoring the Sustainable Intensification of Arable Agriculture: the Potential Role of Earth Observation – Literature Review .....</b>	<b>6</b>
Abstract.....	6
2.1 Introduction.....	7
2.2 Key Concepts/Definitions.....	8
2.2.1 Agricultural Intensification (AI).....	8
2.2.2 Sustainable Intensification (SI) .....	9
2.2.3 Assessing Sustainable Intensification.....	10
2.3. Assessing the Intensity of Arable Production.....	12
2.3.1 Current Approaches .....	12
2.3.2 Potential for Expanding the Use of EO to Assess Agricultural Intensity .....	14
2.4. Assessing the Environmental Sustainability of Arable Systems .....	15
2.4.1 Current Approaches .....	15
2.4.2 Potential Applications of EO-based Methods for Assessing Environmental Sustainability.	18
2.5. Discussion.....	25
2.5.1 Opportunities for an EO-based SI Assessment System .....	27
2.6. Conclusion .....	33
<b>3. High Resolution Wheat Yield Mapping Using Sentinel-2.....</b>	<b>35</b>
Abstract.....	35
3.1 Introduction.....	36
3.2 Field Sites .....	39
3.3 Data and Methods .....	40
3.3.1 Wheat Yield Data .....	41
3.3.2 Sentinel-2 Data .....	44
3.3.3 Environmental Data .....	46
3.3.4 Random Forest Regression.....	47
3.3.5 Accuracy Assessment .....	50
3.3.6 Establishing a Baseline .....	50

3.4 Results .....	51
3.4.1 Baseline Data .....	51
3.4.2 Random Forest Model Comparison .....	52
3.4.3 Mapping Within-field Wheat Yield Variability .....	57
3.4.4 Mapping Within-field Wheat Yield Variation at Landscape-scale .....	60
3.5 Discussion .....	62
3.5.1 Benefits of Random Forest .....	62
3.5.2 Optimum Processing Resolution .....	64
3.5.3 Variability in Accuracy Through the Season .....	67
3.5.4 Future Developments .....	68
3.6 Conclusion .....	69
<b>4. Satellite-derived Environmental Heterogeneity and Productivity as Indicators of Bird Diversity 70</b>	
Abstract .....	70
4.1 Introduction .....	72
4.1.1 Bird Diversity Monitoring in GB .....	74
4.2 Data and Method .....	75
4.2.1 Bird Data .....	76
4.2.2 Habitat Heterogeneity Variables .....	77
4.2.3 Habitat Productivity Variables .....	79
4.2.4 Random Forest Models and Variable Selection .....	80
4.2.5 Feature Contribution .....	82
4.3 Results .....	82
4.3.1 Model Accuracies and Estimative Maps .....	82
4.3.2 Feature Contribution (for the Refined Models) .....	86
4.4 Discussion .....	91
4.4.1 Comparison to Other Studies .....	91
4.4.2 Influence of the Landscape on Bird Diversity .....	92
4.4.3 Contribution of Satellite Data to Biodiversity Monitoring .....	93
4.4.4 Future Developments .....	94
4.5 Conclusion .....	96
<b>5. Towards Monitoring the Sustainable Intensification of Arable Agriculture Using Satellite-derived Indicators of Farm Performance .....</b>	<b>97</b>
Abstract .....	97
5.1 Introduction .....	99
5.2 Method .....	101
5.2.1 Training Data for Wheat Yield Modelling .....	104



5.2.2 Estimator Variables for Wheat Yield Modelling.....	104
5.2.3 Training Data for Farmland Bird Richness Modelling.....	106
5.2.4 Estimator Variables for Farmland Bird Richness Modelling .....	107
5.2.5 Random Forest Regression .....	110
5.2.6 Gap Calculation .....	111
5.2.7 Performance Feature Space .....	113
5.3 Results .....	113
5.3.1 Wheat Yield and Farmland Bird Richness Gap Data .....	113
5.3.2 Performance Feature Space and Map.....	116
5.4 Discussion .....	120
5.5 Conclusion .....	124
<b>6. Research Outcomes .....</b>	<b>126</b>
6.1 Key Contributions of Thesis.....	126
6.2 Estimation Uncertainty .....	129
6.3 Performance Assessment Scale .....	133
6.4 Realising the Potential of EO to Assess Sustainable Intensification .....	135
6.5 Summary .....	137
<b>References .....</b>	<b>138</b>
<b>Appendices .....</b>	<b>159</b>
Appendix 1: Supplementary Material for Chapter 2 .....	159
Appendix 2: A basic introduction to Random Forest and its implementation .....	193
Appendix 3: Supplementary Material for Chapter 3 .....	203
Appendix 4: Supplementary Material for Chapter 4 .....	204
Appendix 5: Supplementary Material for Chapter 5 .....	214

## **List of Figures**

Figure 3.1 Location of study sites for yield estimation training data.....	40
Figure 3.2 Overview of the method used to estimate yield at high resolution on a landscape-scale.....	41
Figure 3.3 Summary of the criteria for combine harvester data cleaning.....	42
Figure 3.4 Example yield data points for one field showing a) gaps in the data arising from the data collection and cleaning process and b-c) the stages in the buffering process used to remove these gaps.....	43
Figure 3.5 Example of the distribution of yield data points relative to a) 10m and b) 20m resolution interpolated yield data.....	44
Figure 3.6 Ten-fold RMSE values from Random Forest yield estimation analysis calculated using the training data set and RMSE values from the validation dataset.....	54
Figure 3.7 Observed yield interpolated from the combine harvester data and estimated yield from the S2_Env RF model for a selection of fields within the training area and the difference between the observed and estimated yield.....	58
Figure 3.8 Frequency distributions for observed and estimated yield using the validation data set for the S2_Env model for all fields and a sample of individual fields.....	59
Figure 3.9 Linear regression between observed and estimated yield for the validation data set from the S2_Env model.....	60
Figure 3.10 Landscape-scale wheat yield estimation based on S2_Env RF model.....	62
Figure 3.11 Evidence of crops ripening between successive Sentinel-2 images for June and July.....	67
Figure 4.1 Overview of the method used to estimate bird diversity distribution across GB.....	76
Figure 4.2 Variance explained (%) for the bird diversity distribution models including different numbers of estimators (by ranking).....	81
Figure 4.3 Estimated farmland bird richness and diversity (Shannon Index) maps at 1km resolution.....	85
Figure 4.4 Estimated woodland bird richness and diversity (Shannon index) maps at 1km resolution.....	86
Figure 4.5 Feature contribution plots for the farmland bird richness estimation model.....	87
Figure 4.6 Feature contribution plots for the farmland bird diversity estimation model.....	89
Figure 4.7 Feature contribution plots for the woodland bird richness estimation model.....	90
Figure 4.8 Feature contribution plots for the woodland bird diversity estimation model.....	91

Figure 5.1 Study area location for agricultural performance assessment.....	102
Figure 5.2 Overview of the method used to create the feature space to assess relative farm performance.....	103
Figure 5.3 Map of a) average wheat yield per 1km, b) wheat yield gap, c) farmland bird richness, and d) farmland bird richness gap for the study area.....	115
Figure 5.4 Performance feature space created by plotting values for the wheat yield gap and farmland bird richness gap for each 1km square within the landscape.....	118
Figure 5.5 Map indicating the relative performance of 1km squares within the landscape.....	120

## **List of Tables**

Table 2.1 Key EO-derived indicators for assessing agricultural intensity.....	13
Table 2.2 Potential EO-based indicators which could be used to assess environmental sustainability.....	18
Table 2.3 Examples of possible Essential Sustainable Intensification Variables (ESIVs).....	29
Table 3.1 Central wavelength and spatial resolution for the Sentinel-2 bands used for yield estimation.....	45
Table 3.2 Explanatory variables used in Random Forest regression analysis for yield estimation.....	45
Table 3.3 Vegetation indices calculated using Sentinel-2 imagery.....	46
Table 3.4 Data combinations tested in Random Forest analysis for yield estimation.....	49
Table 3.5 Yield estimation RMSE and R-squared values calculated from the validation dataset for linear and Random Forest regression using vegetation indices calculated for each month.....	51
Table 3.6 RMSE and R-squared values calculated from the validation dataset for Random Forest regression yield estimation using NDVI data accumulated over the growing season.....	52
Table 3.7 Results of Random Forest analysis for yield estimation.....	53
Table 3.8 Yield estimation RMSE values for individual fields using the validation data set for the S2_Env model.....	60
Table 4.1 R-squared values for bird diversity distribution models for (i) RF models containing all variables, (ii) RF models containing variables categorised as important by the minimal depth selection, and (iii) the final refined models.....	84
Table 5.1 Details of the Sentinel-2 bands used for yield estimation.....	105
Table 5.2 Farmland bird species groupings based on the BTO/JNCC/RSPB wild bird indicators for the UK and England.....	107
Table 5.3 Broad land cover classes used in this study and the corresponding original LCM2015 subclasses.....	108
Table 5.4 Details of the habitat structure metrics derived from LCM2015 using FRAGSTATS.....	109
Table 5.5 RMSE and R-squared values for wheat yield and farmland bird richness Random Forest models.....	114

## **Acknowledgements**

Thank you to Clare and Alan, my wonderful supervisors, who guided me through the, sometimes daunting, process of completing this thesis. I could not have asked for two better supervisors, your guidance, advice and encouragement has been invaluable over the last three and a half years, and I have thoroughly enjoyed working with you both.

I am also grateful to the Graduate School for the Environment for funding this PhD, and to UKCEH and all the people there for providing me with such a welcoming work environment.

Thank you to Luis Carrasco for all his guidance and advice as I made my first foray into the world of *R* and particularly for his insights into working with the bird richness data. Thank you also to John Redhead for his assistance with accessing the crop yield data, and Gavin Siriwardena for his input into the bird richness mapping.

Thank you also to those people and organisations who have assisted me with provision of data, without which this thesis would not have been possible: the farmers and farming companies who provided the crop yield data and the ASSIST project that made access to this data possible; the British Trust for Ornithology who supplied the Breeding Bird Survey data; Copernicus and USGS for provision the Sentinel-2 and Landsat data respectively; and UKCEH for providing access to the Land Cover Maps.

And last, but by no means least, thank you to my incredible family and friends for all their support and encouragement, not just in the last three and a half years, which helped me to persevere and get to where I am today.

## **Declaration**

I declare that, other than where the contribution of others is specified, this thesis is entirely my own work and has not been submitted for the award of any other degree at this or any other university.

A handwritten signature in black ink, appearing to read 'Merryn Hunt', written in a cursive style.

Merryn Hunt

## **Statement of authorship for multi-authored chapters**

### **Chapter 2: Monitoring the Sustainable Intensification of Arable Agriculture: the Potential Role of Earth Observation – Literature Review**

***Hunt, M.L., Blackburn, G.A., & Rowland, C.S. (2019). Monitoring the Sustainable Intensification of Arable Agriculture: the Potential Role of Earth Observation. International Journal of Applied Earth Observation and Geoinformation, 81, 125-136.***

The scope and structure of this review was decided upon through a series of discussions between the authors. As the first author, I was solely responsible for conducting the literature review at the heart of this chapter, and producing an initial draft, including all figures, tables and supplementary material. I then revised this draft following feedback from co-authors and referees, and was responsible for the submission of the finished manuscript for publication.

### **Chapter 3: High Resolution Wheat Yield Mapping Using Sentinel-2**

***Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., & Rowland, C.S. (2019). High resolution wheat yield mapping using Sentinel-2. Remote Sensing of Environment, 233, 111410.***

The concept and methodological approach for this chapter was decided upon through a series of discussions between the authors. As the first author of the above publication, I extracted the relevant explanatory variables from the satellite data and environmental data and carried out the yield estimation and subsequent analysis. From the results of the analysis, I produced the draft text of the manuscript, including all tables and figures. Following comments of co-authors and referees, I revised this draft, and was responsible for the submission of the finished manuscript for publication.

### **Chapter 4: Satellite-derived Environmental Heterogeneity and Productivity as Indicators of Bird Diversity**

***Hunt, M.L., Blackburn, G.A., Siriwardena, G., Carrasco, L., & Rowland, C.S. (2020). Satellite-derived environmental heterogeneity and productivity as indicators of bird diversity. Yet to be submitted for publication.***

The concept and methodological approach for this chapter was decided upon through a series of discussions between the authors. As first author, I was responsible for extracting the relevant explanatory variables from the satellite and Land Cover Map datasets, and calculating the necessary

richness and diversity response variables. I then carried out the Random Forest modelling and feature contribution analysis. Once the results were obtained, I produced an initial draft of the paper, including all figures and tables. This was revised following feedback from the other authors.

## **Chapter 5: Towards Monitoring the Sustainable Intensification of Arable Agriculture Using Satellite-derived Indicators of Farm Performance**

***Hunt, M.L., Blackburn, G.A., Carrasco, L., Siriwardena, G., Redhead, J.W., & Rowland, C.S. (2020). Towards Monitoring the Sustainable Intensification of Arable Agriculture Using Satellite-derived Indicators of Farm Performance. Yet to be submitted for publication.***

The concept and methodological approach for this chapter was decided upon through a series of discussions between the authors. As first author, I was solely responsible for all analysis involved in this paper including extraction of explanatory variables, creation of the farmland bird richness and wheat yield datasets, and the creation of the agricultural performance space and map. Once the analysis was complete, I produced a draft manuscript, including all figures and tables. This draft was then revised in response to feedback from the other authors.

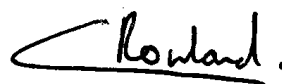
I confirm that the above information on the authorship of these chapters and the contribution of Merryn Hunt is accurate.



Merryn Hunt



Co-authors: G. Alan Blackburn



Clare S. Rowland



## **1. Introduction**

By 2050, global food production must increase by an estimated 70% to meet growing demand resulting from projected population increases, shifting food consumption patterns and increases in income (Dillon *et al.*, 2016; FAO, 2009; Garnett and Godfray, 2012; Tilman *et al.*, 2011, 2002). Globally we are seeing an abrupt decline, and sometimes complete stagnation, in the rate of yield increase on existing farmland (Grassini *et al.*, 2013). If yield growth continues at its current trajectory, projected food demand will not be met without large expansion of crop area, which would have wide-ranging environmental consequences (Cassman and Grassini, 2020).

Sustainable intensification of agriculture has been proposed as a way to address the need for increased food production on existing agricultural land, while maintaining or enhancing environmental quality (Dillon *et al.*, 2016). Various methods of sustainable intensification are being developed such as Integrated Pest Management and Conservation Agriculture (Pretty and Bharucha, 2014). However, the success of these methods is context-dependent (Pretty and Bharucha, 2014), and the potential to achieve sustainable intensification varies from local field or farm scales, to regional, national and global scales (Cassman and Grassini, 2020). As such, any attempts at sustainable intensification must be monitored carefully to ensure they are achieving their intended outcomes.

To date, attempts to monitor the sustainable intensification of agriculture have been largely non-existent. A number of studies have explored related themes, however, these have tended to focus on assessing either the intensity (e.g. Herzog *et al.*, 2006) or the sustainability (e.g. Dillon *et al.*, 2010) of agriculture, rarely have they combined these elements to assess sustainable intensification. In addition, these studies tend to focus on specific farms, due to their reliance on traditional *in situ* data such as field and farms surveys. These data are typically costly and time consuming to collect, resulting in data of limited spatial and temporal resolution, which restricts the scale of these studies.

Meeting future food demands, without destroying the environment, is a global issue and therefore requires a global solution. As such, we need to understand how agriculture is performing in all locations, not simply in a few specific places. The current lack of a system for quantifying sustainable intensification potential in terms of both production and environmental performance across various scales has been identified as one of the critical missing links in the current global research portfolio (Cassman and Grassini, 2020). Gaining an understanding of agricultural performance at relevant scales requires a long-term data set of high resolution (field-level) data collected on a global scale. Creating such a data set with traditional data collection techniques on their own is unfeasible. We must therefore look beyond the traditional to more modern solutions if we are to close this knowledge gap. This thesis aims to demonstrate, for the first time, one possible solution to this data problem by exploring the potential of Earth Observation to assess sustainable intensification.

Earth Observation (EO) provides repeat measurements and a long-term dataset with global coverage over a range of spatial and temporal resolutions. As such, EO offers the potential to scale up *in situ* data to allow various indicators of agricultural intensity (e.g. yield) and aspects of environmental sustainability (e.g. biodiversity) to be monitored across the globe at a range of spatial and temporal scales. While EO has been used to monitor a variety of these indicators, so far no one has attempted to combine these indicators to provide a large-scale assessment of agricultural performance. The aim of this thesis, therefore, is to explore the potential contribution that EO could make to assessing the sustainable intensification of arable agriculture. This is achieved through four subsidiary aims:

1. Review evidence from the literature in order to scope out the potential role of EO in monitoring the sustainable intensification of arable agriculture;
2. Quantify agricultural intensity on a landscape-scale using wheat yield variability derived from EO data as an indicator;

3. Assess environmental quality on a landscape-scale using farmland bird richness derived from EO data as a proxy for environmental quality;
4. Combine the EO-derived measures of agricultural intensity and environmental quality to create a feature space to assess agricultural performance on a landscape-scale and for specific arable areas.

This thesis comprises of six chapters (plus appendices). In chapter 1, the aims and structure of the thesis are outlined. Chapter 2 provides an introduction to the concept of and need for sustainable intensification and the current issues regarding how it is monitored. A detailed review is then presented of the contribution that Earth Observation could make in addressing these monitoring issues, exploring the potential of EO to monitor indicators of agricultural intensity and environmental quality; the latter is split into indicators of vegetation health, soil quality, water quality and availability, biodiversity and ecosystem health. This chapter concludes with an overview of the opportunities for an EO-based sustainable intensification assessment system and provides a set of recommendations for its implementation including the development of a set of Essential Sustainable Intensification Variables (ESIVs). This review was published in the *International Journal of Applied Earth Observation and Geoinformation*; a full reference can be found in the statement of authorship.

Following this review, three analytical chapters are presented. The first of these has been published (see statement of authorship for full reference and authorship details) and the other two have been prepared in the style appropriate for submission to a scientific journal. An overall reference list, collated from the reference lists for each individual chapter, is provided at the end of this thesis.

Chapter 3 demonstrates the potential to map within-field wheat yield variability on a landscape scale by combining satellite data with environmental data (e.g. meteorological data). This

work explored the impact of data spatial resolution and availability on the accuracy of yield estimation through the use of different combinations of input data for different periods throughout the growing season. This work demonstrated how satellite data can be used to scale up precision yield measurements from a limited number of fields to provide a high-resolution map of yield on a landscape scale. This analysis helped to identify the best combination of satellite and environmental data to maximise accuracy and coverage of yield estimation. This knowledge was used to create the wheat yield dataset used as an indicator of agricultural intensity in chapter 5.

Chapter 4 explores the use of satellite data to assess environmental quality, using bird species diversity as a proxy for environmental quality. This work demonstrates that satellite-derived measures of habitat productivity and heterogeneity can be used to model distributions of farmland and woodland bird diversity and richness for the whole of Great Britain. Furthermore, this analysis provides an insight into the contributions of the most important explanatory variables, demonstrating the non-linear nature of these relationships. Understanding the nature of these relationships is important to identify the characteristics required to promote bird richness, which can inform management practices. The results of this work were used to inform the creation of the farmland bird richness dataset used as an indicator of environmental quality in chapter 5.

Chapter 5 demonstrates the potential for using EO to assess agricultural performance on a landscape-scale. This is achieved by combining the indicators of agricultural intensity (yield) and environmental quality (bird species richness) derived in chapters 3 and 4 to create a novel performance feature space to assess relative agricultural performance. The analysis demonstrates how this feature space can be used to assess performance of arable areas within the landscape allowing identification of the best and worst performing locations to inform changes in management practice.

Chapter 6 then concludes this thesis, summarising the main research contributions from each chapter and discussing some of the wider considerations, such as issues of uncertainty and

scale, when assessing agricultural performance and how these may be addressed. Finally, this chapter outlines the future work that is required to ensure the potential contribution of EO to sustainable intensification assessments is fully realised.

## **2. Monitoring the Sustainable Intensification of Arable Agriculture: the Potential Role of Earth Observation – Literature Review**

Merryn L. Hunt, George Alan Blackburn & Clare S. Rowland

This chapter is a replication of a constituent paper of this research that was published in International Journal of Applied Earth Observation and Geoinformation.

Hunt, M.L., Blackburn, G.A., & Rowland, C.S. (2019). Monitoring the Sustainable Intensification of Arable Agriculture: the Potential Role of Earth Observation. *International Journal of Applied Earth Observation and Geoinformation*, 81, 125-136.

### **Abstract**

Sustainable intensification (SI) has been proposed as a possible solution to the conflicting problems of meeting projected increase in food demand and preserving environmental quality. SI would provide necessary production increases while simultaneously reducing or eliminating environmental degradation, without taking land from competing demands. An important component of achieving these aims is the development of suitable methods for assessing the temporal variability of both the intensification and sustainability of agriculture. Current assessments rely on traditional data collection methods that produce data of limited spatial and temporal resolution. Earth Observation (EO) provides a readily accessible, long-term dataset with global coverage at various spatial and temporal resolutions. In this paper we demonstrate how EO could significantly contribute to SI assessments, providing opportunities to quantify agricultural intensity and environmental sustainability. We review an extensive body of research on EO-based methods to assess multiple indicators of both agricultural intensity and environmental sustainability. To date these techniques have not been combined to assess SI; here we identify the opportunities and initial steps required to achieve this. In this context, we propose the development of a set of Essential Sustainable Intensification Variables (ESIVs) that could be derived from EO data.

## **2.1 Introduction**

With a projected population increase of 2.3 billion by 2050, increases in income and shifting food consumption patterns, global food production must increase by an estimated 70% to meet growing demand (Campbell *et al.*, 2014; Caviglia and Andrade, 2010; Dillon *et al.*, 2016; FAO, 2009; Garnett *et al.*, 2013; Lampkin *et al.*, 2015; Schut *et al.*, 2016; Tilman *et al.*, 2011, 2002). Both agricultural expansion (clearing additional land for crop production) and intensification (increasing productivity of existing agricultural land) could provide necessary crop production increases (Godfray and Garnett, 2014; Tilman *et al.*, 2011). Current competition for land restricts the potential for conversion of new land to agriculture, thus limiting the viability of expansion in many cases (Godfray *et al.*, 2010; Pretty *et al.*, 2011). In addition, expansion is thought to have a greater detrimental impact on the environment than intensification, with potential for significant greenhouse gas release through land conversion and major biodiversity losses affecting essential ecosystem service provision (Garnett *et al.*, 2013; Godfray and Garnett, 2014). Consequently, future demands must be met through production increases on current agricultural land alongside shifts in diet and the reduction in food waste (including transport and consumption). Previous agricultural intensification has been achieved through changes in management practices including increased agrochemical inputs, cropping intensity and irrigation, and adoption of monoculture practices (Benton *et al.*, 2003; Crowder and Jabbour, 2014; Meeus, 1993; C. Stoate *et al.*, 2001). However, it is now widely recognised that such intensification measures detrimentally impact the environment, through over exploitation of natural resources for inputs and emission of pollution and waste (Pretty *et al.*, 2011). This raises concerns over the long-term ability to maintain intensive agricultural practices, with intensification-induced environmental degradation having negative feedbacks on sustained crop productivity (Bommarco *et al.*, 2013a; Foley *et al.*, 2005; Matson *et al.*, 1997). It is clear therefore that a sustainable method of agricultural intensification is required.

One possible solution is sustainable intensification (SI), which involves increasing production efficiency to achieve higher agricultural outputs with the same or fewer inputs, while simultaneously

significantly reducing or eliminating environmental degradation (Dillon *et al.*, 2016). However, no definitive mechanisms of SI exist, with the success of different methods dependent on situation-specific conditions. As such, to ensure any attempts at SI are successful suitable methods are required to assess the sustainability of intensification efficiently over diverse landscapes and spatial scales on a long-term basis.

The purpose of this review is to outline the current state of SI assessments and explore the potential contributions EO could make. While it is true that various studies have used EO to assess either agricultural intensity or various indicators of environmental sustainability, to date no one has attempted to combine established EO-based methods to provide an actual assessment of sustainable intensification. Hence, this review explores the basis for the development of an operational SI monitoring system that uses EO data. This review is structured as follows. Section 2.2 provides an overview of the key concepts of agricultural intensification (AI) and SI, as well as briefly introducing SI assessment. Section 2.3 and Section 2.4 present a more detailed outline of the current approaches used to assess agricultural intensity and agricultural environmental sustainability, respectively, highlighting ways in which EO data is presently used and further contributions it could make. The review concludes with a discussion of the opportunities of EO to contribute towards an operational SI monitoring system applicable for a range of spatial and temporal scales. This review focuses on the intensification arable agriculture; as such methods for monitoring pastoral agriculture are not explicitly discussed.

## **2.2 Key Concepts/Definitions**

### **2.2.1 Agricultural Intensification (AI)**

Agricultural intensification (AI) is the “increase in agricultural production per unit of inputs”, where inputs may include labour, land, time, seed, fertiliser, feed or cash (FAO, 2004). Intensification can refer to maintenance of current production with decreased inputs, and/or increased production through higher input productivity (FAO, 2004). Methods of intensification include: increased



agrochemical inputs; increased cropping intensity (e.g. double or triple cropping); increased crop density; removal of linear and point features such as hedgerows and ponds (landscape simplification and field enlargement); decreased crop diversity (monoculture adoption); and increased irrigation (e.g. Crowder and Jabbour, 2014; Donald *et al.*, 2001; Newton, 2004; Stoate *et al.*, 2001).

Detrimental environmental impacts of AI are wide-ranging, covering a range of spatial (local to global) and temporal (short-term to long-term or permanent) scales. These factors are driving the growing interest in alternative, more sustainable methods for meeting growing food demand. Figure A1 in Appendix 1 highlights some of the key environmental impacts of various mechanisms of agricultural intensification. The potential for environmental degradation resulting from AI activities is intensified by the complexity of the agricultural environment, with numerous interactions, connections and feedbacks within the system, and multiple causal relationships. Such degradation could have significant impacts both within the immediate vicinity of intensification and over wider spatial scales. The wide range of potential impacts and system complexity poses a challenge when trying to devise a monitoring system that can accurately measure all required elements.

### **2.2.2 Sustainable Intensification (SI)**

Sustainable intensification (SI) has been proposed as an alternative to conventional intensification providing necessary yield increases, whilst ensuring environmental degradation is kept at a sustainable level (Tilman *et al.*, 2011). The concept originated in the 1990s (Buckwell *et al.*, 2014), with much debate since over the exact definition of the term “sustainable intensification”. A common definition describes SI as a form of production wherein greater yields are produced with the same or fewer inputs, while adverse environmental impacts are simultaneously reduced or eliminated and contribution to natural capital and ecosystem service flow is increased (Barnes and Thomson, 2014; Dillon *et al.*, 2016; Garnett and Godfray, 2012; Godfray *et al.*, 2010; Pretty, 2008; Pretty *et al.*, 2011). The theoretical model of sustainable intensification is illustrated in figure A2 in Appendix 1. SI as a concept prescribes no particular development paths or methods; the aim is simply to create resource efficient agriculture with significantly better environmental performance

than conventional intensification (Buckwell *et al.*, 2014). Instead, a framework is provided facilitating exploration of the optimum mix of approaches based on existing situation-specific biophysical, social, cultural and economic contexts (Buckwell *et al.*, 2014; Garnett *et al.*, 2013; Garnett and Godfray, 2012; Pretty and Bharucha, 2014). The suitability of different methods varies depending on conditions, as well as current agricultural productivity and environmental performance of the system (Buckwell *et al.*, 2014; Garnett and Godfray, 2012).

Possible interventions to achieve SI include: Integrated Pest Management (IPM), use of on- and off-farm biodiversity to manipulate pest ecologies; Agroforestry Systems, for example intercropping; Precision Agriculture; and Conservation Agriculture (Pretty and Bharucha, 2014).

### **2.2.3 Assessing Sustainable Intensification**

As there is no definitive mechanism for SI, realising the goal of resource efficient agriculture requires suitable methods to assess the sustainability of intensification efficiently over diverse landscapes and spatial scales on a long-term basis. At present, there is no routine monitoring of sustainable intensification, with availability of suitable data representing a major issue (Cassman and Grassini, 2020; Siebrecht, 2020). Current assessment attempts rely largely on farm surveys (questionnaires and interviews), field data, national government statistics and other traditional data sources. Data collection is often costly and time consuming, limiting the spatial and temporal scale and extent, with consequent impacts on the representativeness of both the data and the assessments. As such, the current reliance on interpolation of point data and average statistics severely restricts timely provision of accurate sustainability assessments for all agricultural areas. Deficiencies such as these highlight the need for a new, more efficient assessment technique.

Generally, studies focus on either agricultural sustainability, with no explicit attempt to quantify intensification (e.g. Dillon *et al.*, 2010; Rasul and Thapa, 2004), or on agricultural intensity at a specific point in time, with no assessment of sustainability (e.g. Herzog *et al.*, 2006; Niedertscheider *et al.*, 2016). Studies that assess the sustainability of intensification are largely

conducted on farms where management practices are known to have shifted towards more intensive measures; hence, no attempt is made to quantify the degree or rate of intensification. The few studies that do include a measure of intensity (e.g. crop yield) commonly ignore change over time and use a single point in time. Intensification is a process rather than a fixed end state; looking at intensity at one specific time is therefore not sufficient (Elliott *et al.*, 2013; Firbank *et al.*, 2013). To get a full picture of the sustainability of any intensification in agricultural production, the change that has occurred over time must be investigated and trends identified to determine the actual environmental impacts. The potential of a quantified SI assessment approach at the individual farm level, using data held by farmers from two different years, was explored by Firbank *et al.* (2013). Results demonstrated the ability to assess sustainability of farms adopting different management strategies, but the data source restricted the completeness of the assessment, limiting measurable indicators and identifiable spatial and temporal variation. The limited temporal resolution of the data restricted full exploration of the intensification and subsequent effects on yield and environmental sustainability. The study constitutes an important step in the development of an operational SI assessment method, but more comprehensive data is required.

Ideally, field measurements would be taken at all locations to provide data on which SI could be assessed. However, as this is not feasible, data sources with greater spatial and temporal coverage and lower acquisition costs must be sought that can be used in conjunction with field measurements. Incorporation of satellite data into sustainability assessments could allow greater flexibility, over various spatial and temporal scales, providing more accurate and representative results at lower costs. Recent decades have seen considerable increases in Earth Observation (EO) data use, with applications in diverse research areas. A number of international monitoring systems have been developed incorporating satellite data for crop condition monitoring and yield forecasting over regional, national and global scales. Such systems include the Group on Earth Observations Global Agricultural Monitoring system (GEOGLAM) (GEO, 2019), the USDA Foreign Agricultural Service (FAS) Global Agricultural Monitoring (GLAM) System (USDA FAS, 2019), the Chinese Academy

of Sciences Crop Watch Program (Bingfang, 2006), and the Monitoring Agriculture by Remote Sensing (MARS) project developed by the Joint Research Centre of the European Commission (European Commission, 2016). The operational status of these systems demonstrates the value of EO for agricultural monitoring. Outputs from these systems could be used to monitor intensification, but currently this is not done explicitly. The potential contribution of these EO-based systems to SI assessment has so far not been fully explored or realised, with little or no evidence of the use of satellite-derived data within agricultural sustainability assessments.

To date EO researchers have not explicitly attempted to quantify SI and so intensity and environmental sustainability have not been assessed for the same sites from EO. This review treats aspects of sustainability and intensification separately to provide a comprehensive overview of current research and the potential contribution of Earth Observation. The next two sections provide a more detailed overview of the methods used to assess both agricultural intensity and sustainability, highlighting the ways in which satellite data is used at present and the opportunities moving forwards.

## **2.3. Assessing the Intensity of Arable Production**

### **2.3.1 Current Approaches**

The types of indicators used to assess agricultural intensity differ between studies. Some focus on indicators which reflect the increase in land productivity caused by human intervention (Dietrich *et al.*, 2012) such as yield per ha (e.g. Singh *et al.*, 2002). Others focus on indicators which measure the change in inputs or other factors of management (Shriar, 2000) including total nitrogen (fertiliser) input (e.g. Temme and Verburg, 2011), number of pesticide applications (e.g. Herzog *et al.*, 2006), and inputs costs per ha (e.g. Teillard *et al.*, 2012). As such, the range of indicators used to assess agricultural intensity can be split into two general groups: agricultural input indicators (e.g. input cost per ha, crop acreage) and agricultural output indicators (e.g. production per area and time). Examples of indicators used in EO-based studies can be found in table 2.1. Some direct

indicators of agricultural intensity, such as fertiliser and pesticide input cannot be measured using EO, but may be detectable indirectly from, for example, changes in yield. Typically either a single indicator is adopted to assess intensity (e.g. Biradar and Xiao, 2011; Mingwei *et al.*, 2008), or multiple indicators are aggregated to produce an intensity index (e.g. Kerr and Cihlar, 2003; Shriar, 2000). Aggregated indicators simplify complex situations into a single element, but this is done at the expense of interpretability and transparency. Whether a single indicator or an index is appropriate will vary depending on the purpose of the study (Herzog *et al.*, 2006).

*Table 2.1: Key EO-derived indicators for assessing agricultural intensity. Examples of methods to derive these EO-based indicators can be found in table A2 in Appendix 1.*

<b>Agricultural Intensity</b>	
• Crop yield (e.g. tonnes/ha)	• Cropping area (e.g. acres, km <sup>2</sup> )
• Multi-cropping: <i>Number of harvests within a single year (i.e. growing season)</i>	
• Cropping intensity: <i>Number of cropping cycles per year or number of years a field is sown with crops and actually reaches harvest</i>	
• Cropping frequency: <i>Number of years a pixel was cropped over an observation</i>	
• Crop duration ratio: <i>Ratio of time period (during growing season) for which a pixel was cropped and the total length of the growing season</i>	
• Fallow cycles: <i>Recurring periods of fallow cropland</i>	
<i>Further information in table A2 (Appendix 1)</i>	

Data sources for agricultural intensity assessments include interviews, government statistics, field and farm surveys, aerial photographs and satellite data (see table A1 in Appendix 1 for further examples of data sources and indicators used by various authors to investigate agricultural intensity). EO techniques are fairly common within this area, with a range of satellites appearing within the literature including MODIS, AVHRR, Landsat and Sentinel. EO-based methods used to investigate agricultural intensity vary depending on the indicator of interest; some examples of specific methods can be found in table A2 in Appendix 1.

EO has a range of advantages over other data sources for intensity assessments. The use of EO allows better cross-country comparisons to be made; country boundary restrictions do not apply

in the same way as to government statistics data, thus improving data consistency (Herzog *et al.*, 2006). Additionally, low-cost methods using relatively simple technology can be developed, allowing application of EO-based methods in areas where costs of traditional data collection methods prohibit reliable intensity assessments (Ferencz *et al.*, 2004). Furthermore, EO-based cropping indicators perform well for broad-scale agricultural monitoring, suggesting they could complement (potentially) more accurate sample-based ground-data, by providing wall-to-wall observations of agricultural management (Estel *et al.*, 2016). Lack of spatially distributed information on key environmental and agronomic variables tends to limit application of crop simulation models for regional scale yield estimation (Moriondo *et al.*, 2007). EO data can alleviate this problem, by providing estimates of relevant variables over a range of spatial and temporal scales.

### **2.3.2 Potential for Expanding the Use of EO to Assess Agricultural Intensity**

As previously noted, the use of EO within agricultural intensity studies is already fairly common, but it is still not typically routine and operational. In addition, with continued advancements in sensor technology, the launch of new satellites and the development of new methods, the full potential of EO has yet to be realised. Moving forward research should continue to focus on the creation of high resolution, global products which can be provided regularly (annually), consistently and in a timely manner (Claverie *et al.*, 2018; Egorov *et al.*, 2019; Roy *et al.*, 2010). In the past, the production of high resolution operational EO-derived products was hindered by a lack of suitable cloud-free imagery and the time and computing power required to process the vast number of images needed to provide global coverage. However now, with an increase in the number of moderate and high resolution satellites, improvements in gap-filling and sensor integration techniques, and the advent of cloud computing systems that can facilitate more rapid processing, the potential for producing high resolution EO products for assessing agricultural intensity on a global scale has never been greater. Work is already underway to produce a variety of large-scale high-resolution (30m) products, for example, cropland extent maps on a country- (e.g. Teluguntla *et al.*, 2018) and continental-scale (e.g. Xiong *et al.*, 2017). International programmes such as

GEOGLAM and commercial organisations such as OneSoil (OneSoil, 2018) are also working to map various agricultural parameters, including crop type and crop condition/development, on a global scale. The timely production of high resolution, global products, which provide an accurate representation of the diverse agrosystems around the world, is likely to require either the development of generic, transferable models or an increase in the collection and provision of *in situ* data. In reality, the solution will probably involve some combination of the two.

The accessibility of appropriate field data for calibration and validation is a major constraint on the development of an operational EO-based system for assessing intensification. A step-change in monitoring capability could be provided by having EO as part of an integrated system that makes *in situ* data routinely available. Similarly, the creation and adoption of better data fusion methods may help with interpretation of the EO data, limiting the impact of confounding factors, and improving assessment potential. However, despite the various challenges which exist at present, EO data provides an important, practical and viable approach for regional and global monitoring of land surface dynamics, including variations in agricultural intensity (Yan *et al.*, 2014). The value of EO for agricultural intensity assessments lies, in part, in the spatially explicit nature of the data, consistency across political borders and the systematic acquisition setup (Kuemmerle *et al.*, 2013).

## **2.4. Assessing the Environmental Sustainability of Arable Systems**

### **2.4.1 Current Approaches**

Sustainability has three dimensions: economic, social and environmental sustainability (Allahyari *et al.*, 2016). EO has the potential to contribute valuable information on the environmental dimension however, to provide a comprehensive assessment of sustainability, EO must be used alongside datasets covering economic and social variables. Accepted socio-economic indicators are regularly (e.g. yearly) monitored by governments and international organisations (e.g. United Nations, World Health Organisation) on a regional, national and global scale. This provides a readily available, long-term dataset for economic and social sustainability, something that is not currently

available for environmental sustainability. Consequently, in this review the focus is on the environmental dimension of sustainability.

Assessing environmental sustainability is a complex process fraught with challenges and pitfalls. Selection of appropriate indicators, weighting and aggregation methods for specific situations and requirements are essential to the successful assessment of sustainability. A number of frameworks have been developed to aid in the selection and aggregation of appropriate indicators, to provide a single score by which sustainability of agricultural systems can be assessed. These frameworks differ in their definition of sustainability, indicator selection approach, and aggregation and validation methods. Frameworks include

- SAFA: Sustainability Assessment of Food and Agriculture Systems (FAO, 2014, 2013)
- IDEA: Indicateur de Durabilité des Exploitations Agricoles (Zahm *et al.*, 2008)
- ISAP: Indicator of Sustainable Agricultural Practice (ISAP) (Rigby *et al.*, 2001)
- RISE: Response-Inducing Sustainability Evaluation (RISE) (Häni *et al.*, 2003, 2006)
- SAFE: Sustainability Assessment of Farming and the Environment (Van Cauwenbergh *et al.*, 2007)
- SSP: Sustainability Solution Space for Decision Making (Wiek and Binder, 2005)
- Sustainable Intensification Assessment Framework (Musumba *et al.*, 2017; Snapp *et al.*, 2018)

Issues surrounding the use of indicators and frameworks are explored in other publications (e.g. Binder *et al.*, 2010; Binder and Wiek, 2006; Gómez-Limón and Sanchez-Fernandez, 2010; Roy and Chan, 2012; Singh *et al.*, 2009; Smith *et al.*, 2017; Stein *et al.*, 2001) so no further discussion will be presented here.

Current agricultural sustainability studies generally rely on a mixture of primary and secondary data including: questionnaires, field data collection (e.g. soil sampling, spatial



information) and government statistics (e.g. Gómez-Limón and Sanchez-Fernandez, 2010; Rasul and Thapa, 2004; Rodrigues *et al.*, 2010; Zhen *et al.*, 2005). There are however a number of challenges specifically arising as a consequence of current data sources and collection methods which limit the ability to accurately and efficiently assess environmental sustainability:

- (1) **No baseline data** – Sustainability studies frequently lack baseline data, using data from a single point in time. This prevents analysis of temporal variability. Multi-temporal datasets would enable more comprehensive and therefore more reliable assessments.
- (2) **Uncertainty from sample data interpolation** – Reliance on sample data and interpolation or averaging techniques limit the potential to accurately assess spatial variability. Ideally all points would be sampled, providing (near-) continuous coverage, however, the cost and time required to complete such a task, clearly makes this impossible using traditional data collection techniques. Current assessments assume the sample data are representative of the wider study area, which may introduce error.
- (3) **Subjective data** – Frequent use of questionnaires, farm surveys and interviews arguably affects the objectivity of many studies. Apart from obvious issues over the truth of answers, subjectivity of qualitative collection methods limits the extent to which data can be integrated and compared. Use of objective, quantified data would likely improve assessment capabilities.
- (4) **Limited data** – Data availability, resolution and coverage (both spatial and temporal) are often limited by costly and time-consuming collection methods. Bias may also exist, with data less readily available in more inaccessible and poorer areas. Sustainability assessments are important in all areas if food demands are to be met and environmental quality maintained or improved; lack of necessary data due to traditional data collection methods hinders this.

EO has the potential to reduce some of these issues when assessing environmental sustainability.

## 2.4.2 Potential Applications of EO-based Methods for Assessing Environmental Sustainability

The potential applications for EO-based methods are described by splitting environmental sustainability into five key areas: (1) vegetation health; (2) soil quality; (3) water quality and availability; (4) biodiversity; and (5) ecosystem health (table 2.2). Environmental sustainability does encompass other aspects however the ability to assess elements such as air quality using EO, at scales relevant to agricultural systems, is restricted. This review therefore focuses on the areas considered to have most potential for assessment via EO-derived indicators for agriculture-based studies. Research efforts have not been evenly split between the areas covered in this section. However, to maintain a consistent structure and attempt to provide a balanced overview, each has approximately equal coverage in this section.

*Table 2.2: Potential EO-based indicators which could be used to assess environmental sustainability. Examples of EO-based methods to derive these indicators can be found in tables A3 to A7 in Appendix 1.*

<b>Environmental Sustainability</b>		
<b>Vegetation health</b>		
<ul style="list-style-type: none"> <li>• Crop condition</li> <li>• Biophysical traits <i>inc. biomass, fraction of absorbed photosynthetically active radiation (fAPAR), photosynthetic activity</i></li> <li>• Structural traits <i>inc. crop/canopy height, leaf area index (LAI), biomass, canopy morphology</i></li> <li>• Biochemical traits <i>inc. chlorophyll (Ch), water content, nitrogen (N) and phosphorous (P)</i></li> </ul>		
<i>Further information in table A3 (Appendix 1)</i>		
<b>Soil Quality</b>		
<ul style="list-style-type: none"> <li>• Soil organic carbon (SOC)</li> <li>• Soil salinity</li> </ul>	<ul style="list-style-type: none"> <li>• Soil organic matter (SOM)</li> <li>• Crop residue/conservation tillage density</li> </ul>	<ul style="list-style-type: none"> <li>• Soil moisture content</li> <li>• Nitrogen status/availability</li> </ul>
<i>Further information in table A4 (Appendix 1)</i>		
<b>Soil erosion/protection</b>		
<ul style="list-style-type: none"> <li>• Vegetation cover</li> </ul>	<ul style="list-style-type: none"> <li>• Erosion feature detection</li> </ul>	<ul style="list-style-type: none"> <li>• Erosion modelling <i>e.g. USLE</i></li> </ul>
<i>Further information in table A4 (Appendix 1)</i>		

Table 2.2 continued

<b>Water Quality</b>	
<ul style="list-style-type: none"> <li>• Water Quality Indices <i>derived from different spectral band combinations</i></li> <li>• Physical water quality parameters <i>inc. total suspended solids (TSS), turbidity, suspended sediment concentration (SSC), chlorophyll concentration, temperature and water clarity</i></li> <li>• Chemical water quality parameters <i>inc. concentration of total nitrogen, NO<sub>3</sub>-N (nitrate as nitrogen) and total phosphorous</i></li> <li>• Water quality proxy <i>e.g. health of vegetation alongside water bodies</i></li> </ul>	
<i>Further information in table A5 (Appendix 1)</i>	
<b>Water Availability</b>	
<ul style="list-style-type: none"> <li>• Water body area and configuration</li> <li>• Water level and volume</li> </ul>	<ul style="list-style-type: none"> <li>• Water use efficiency and crop water stress</li> </ul>
<i>Further information in table A5 (Appendix 1)</i>	
<b>Biodiversity</b>	
<ul style="list-style-type: none"> <li>• Direct mapping of individuals and associations</li> <li>• Habitat suitability <i>based on known habitat requirements of specific species</i></li> <li>• Species Richness</li> <li>• Invasive Species</li> </ul>	<ul style="list-style-type: none"> <li>• Plant (and animal) species diversity</li> <li>• Landscape structure <i>inc. composition, isolation and complexity</i></li> </ul>
<i>Further information in table A6 (Appendix 1)</i>	
<b>Ecosystem Health</b>	
<ul style="list-style-type: none"> <li>• Vigour               <ul style="list-style-type: none"> <li>- Net Primary Productivity (NPP) &amp; Gross Primary Productivity (GPP)</li> <li>- Fractional cover of green vegetation, non-photosynthetic vegetation (NPV) and bare soil</li> <li>- Biochemical properties <i>inc. nitrogen, phosphorous and chlorophyll</i></li> </ul> </li> <li>• Organisation               <ul style="list-style-type: none"> <li>- Species richness and biodiversity</li> <li>- Vegetation structural traits</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Resilience</li> <li>• Ecosystem Services as a Proxy for Ecosystem Health</li> </ul>
<i>Further information in table A7 (Appendix 1)</i>	

#### 2.4.2.1 Vegetation Health

Environmental quality depends in part on the presence of healthy, diverse and abundant vegetation to provide ecosystem services including soil protection, carbon sequestration and flood prevention (Crossman *et al.*, 2013; Hein, 2014). Vegetation health, in turn, relies on a healthy environment to provide essential resources, including stable soil substrate and nutrients. This interdependence suggests that agricultural and non-agricultural vegetation health can be used as an indicator of environmental quality within the agricultural system. Various aspects of vegetation health have been assessed using EO over a range of spatial (e.g. field-/plot-scale to tens/hundreds km<sup>2</sup> and global scale) and temporal scales (e.g. single date assessments to decadal variation), in a diverse range of environments (e.g. grasslands, shrublands, forests, rainforests, mountainous

regions), including agricultural systems (e.g. corn farms, irrigated maize). Many of these studies were conducted using freely available satellite data including Landsat, MODIS and AVHRR, suggesting established methods exist, which can be readily applied to a range of environments. Examples of EO-based methods used to assess vegetation health can be found in table A3 in Appendix 1.

Empirical models are commonly used to assess a variety of vegetation health-related properties. The frequency and timing of image acquisition affects the strength of relationships between specific vegetation indices (VIs) (e.g. NDVI) and related variables (e.g. net primary productivity) (Tebbs *et al.*, 2017). In some situations it may therefore be preferable to use coarser spatial resolution satellites which provide daily coverage (e.g. MODIS), instead of finer spatial resolution satellites with less frequent data acquisition (e.g. Landsat) (Jackson *et al.*, 2004). The successful application of Radiative Transfer Models (RTMs) demonstrates the potential to develop (simple) algorithms to predict various plant traits from satellite data, spanning a range of vegetation types (Myneni *et al.*, 1997; Trombetti *et al.*, 2008).

#### **2.4.2.2 Soil Quality**

Deterioration of soil quality through intensive use of agricultural land has far-reaching impacts affecting plant productivity, water and air quality (Doran and Zeiss, 2000). Soil quality is therefore an important indicator of sustainability and has been studied over a range of scales, from 10s to 1000s km<sup>2</sup>, and across diverse landscapes including cultivated, semi-natural and natural vegetation areas using EO data. Previous studies clearly demonstrate the great potential for soil quality assessment in agricultural areas. However, disparity between achievable spatial resolutions with current freely available satellite data (e.g. Landsat) and typical agricultural field sizes limits the ability to conduct field-scale assessments in some parts of the world. Examples of methods used to assess soil quality and erosion/protection using EO data can be found in table A4 in Appendix 1.

Soil reflectance is a function of the soil's physical properties such as soil moisture and soil organic carbon, but also tillage practices, crop residue and row orientation. Multispectral imagery is

therefore best suited for application to farms with uniformly tilled fields and constant soil moisture conditions at the time of image acquisition; such conditions increase the dominance of the property of interest in the spectral response (Barnes and Baker, 2000). Microwave sensors allow direct soil moisture estimates by exploiting the relationship between moisture content and the dielectric constant of the soil (Wagner *et al.*, 2007). Polarisation and study site conditions influence detection of soil parameter variation. For example, at low moisture conditions, vertical polarisation offers higher sensitivity to salinity; at high moisture levels horizontal polarisation exhibits slightly higher sensitivity (Lasne *et al.*, 2008; Shoshany *et al.*, 2013).

#### **2.4.2.3 Water Quality and Availability**

Agricultural intensification can negatively affect both quality and availability of water resources, so sustainability assessments should consider both. EO-based water quality and availability assessments have focused predominantly on large water bodies such as lakes, reservoirs and coastal environments, for non-agriculture-related investigations. Although the methodology may theoretically be transferable to agricultural environments, the applicability of EO methods to sustainability assessments depends partly on the scale of the study. Water bodies of interest for field- or farm-based assessments are likely to be much smaller than can easily be detected by current satellites due to limiting spatial resolutions; as sensors continue to develop the potential to adopt such methods will increase. At present, EO-based water-related assessments are likely to be most suitable for catchment- or regional-scale investigations, where the impact of multiple farms on larger water bodies is considered using coarser resolution data. Some of the common EO-based methods used to assess water quality and availability are outlined in table A5 in Appendix 1.

Water quality assessments commonly employ empirical models to estimate physical and chemical water quality parameters. Transference of empirical methods to areas and circumstances for which they were not formulated adds uncertainty and may not be appropriate. For example, current generation satellite sensors have limited ability to map chlorophyll content in mesotrophic

and oligotrophic water, despite successful application for eutrophic and hyper-eutrophic water (Gons *et al.*, 2008). However, understanding the limitations of the empirical relationships means they can be applied appropriately. For example a set of robust relationships established between suspended sediment concentration and MODIS spectral reflectance data can be applied, but only for rivers larger than 500m (Martinez *et al.*, 2008).

The operational feasibility of applying a single water body detection algorithm over diverse environmental and climatic conditions has been demonstrated, producing results with a high-degree of accuracy (Mueller *et al.*, 2016). The spatial resolution of satellite altimetry is about 1.7 to 3km (for calm waters), so water level assessments have typically targeted large lakes or rivers, however, satellite altimetry-based techniques have been successfully applied to medium-sized water bodies (200-800m wide), with some potential for application to small-sized water bodies (40-200m wide) (Sulistioadi *et al.*, 2015).

#### **2.4.2.4 Biodiversity**

Agricultural intensification has been linked with biodiversity losses on local, regional and global scales. Biodiversity helps maintain ecosystem health and provision of ecosystem services, so its loss can have serious implications for human well-being (Pettorelli *et al.*, 2014; Turner *et al.*, 2015). Preservation or ideally enhancement of biodiversity is essential for sustainable agriculture, hence the spatial and temporal variability of biodiversity is an important indicator of sustainability. The use of EO to monitor biodiversity has been reviewed by a number of authors, who have outlined key methods, challenges and opportunities (e.g. Gillespie *et al.*, 2008; Kerr and Ostrovsky, 2003; Kuenzer *et al.*, 2014; Mairota *et al.*, 2015; Nagendra, 2001; Rocchini *et al.*, 2010; Turner *et al.*, 2003; Wang *et al.*, 2010). Consequently, in this review we provide only a brief overview covering examples of indicators and related methods (table 2.2). Examples of some EO-based methods used to assess biodiversity can be found in table A6 in Appendix 1.

Biodiversity can be assessed using both direct and indirect techniques. Individual plants or associations of single species may be directly mapped using (very) high spatial resolution satellite images. Where direct mapping of biodiversity indicators is not feasible, indirect approaches that rely on environmental parameters as proxies may be adopted (Turner *et al.*, 2003). Both species abundance and richness are considered to be fundamentally affected by landscape heterogeneity; more heterogeneous landscapes can host more diverse species than homogeneous landscapes due to greater niche availability (Feng *et al.*, 2010; Honnay *et al.*, 2003). EO-derived measures of landscape structure such as composition, complexity and isolation can therefore be used to predict species distribution and diversity. Habitat fragmentation and removal of connectivity elements (e.g. hedgerows) negatively affect species richness and distribution; this relationship allows species richness predictions based on relatively simple landscape metrics (Griffiths and Lee, 2000; Honnay *et al.*, 2003). Lack of consideration of environmental factors such as temperature and disturbance, which affect biodiversity, may reduce accuracy of land cover- and landscape metric-based predictions (Griffiths and Lee, 2000; Z. Li *et al.*, 2014).

Satellite-derived habitat maps can be used in conjunction with known habitat requirements to model potential distribution and abundance of individual species (e.g. Nagendra, 2001; Weiers *et al.*, 2004); habitat suitability parameters such as the existence of suitable water bodies can be derived to assist this modelling (e.g. Weiers *et al.*, 2004). Habitat-based species distribution modelling requires land cover data of sufficient spatial and thematic resolution to ensure all habitats that the target species could potentially occupy are identified; *in situ* or ancillary data is almost certainly necessary to meet requirements for predicting actual distributions of many species (Kerr and Ostrovsky, 2003). Where species do not occupy all suitable habitats, only potential rather than actual distribution can be predicted (Davis *et al.*, 2007; Kerr and Ostrovsky, 2003). A limitation of using habitat maps to predict species distribution and diversity is the fact that this provides no insight into within-habitat variation (Nagendra, 2001).

#### **2.4.2.5 Ecosystem Health**

Ecosystem health affects the ability of ecosystems to provide essential services. A sustainable ecosystem is one which has the ability to maintain its function (or vigour) and structure (or organisation) over time and is resilient even with the application of external stress (Costanza and Mageau, 1999). The potential of EO has been acknowledged and reviewed previously by other authors (e.g. Andrew *et al.*, 2014; Feng *et al.*, 2010; Z. Li *et al.*, 2014). Assessments generally focus on three aspects of ecosystem health: *vigour*, a measure of a system's activity, metabolism and primary productivity; *organisation*, the number and diversity of interactions between a system's components; and *resilience*, a system's ability to maintain its pattern and structure in the presence of stress (Costanza and Mageau, 1999; Z. Li *et al.*, 2014; Li and Guo, 2012). Common indicators of ecosystem health are GPP or NPP (for vigour), species richness and diversity (for organisation) and the ratio of an ecosystem health indicator pre- and post-disturbance (for resilience) (Costanza and Mageau, 1999; Z. Li *et al.*, 2014; Li and Guo, 2012) (table 2.2). Examples of some possible EO-based methods for assessment of ecosystem health can be found in table A7 in Appendix 1.

Ecosystem health can also be assessed using the supply of ecosystem services (ES) as a proxy. ES represent the goods and services supplied by organisms and their activities, controlled by the abiotic characteristics of the system and the anthropogenic impacts it experiences (Andrew *et al.*, 2014). Agricultural systems are both a source and beneficiary of ES (Balbi *et al.*, 2015; Power, 2010). EO-based ES mapping has been reviewed by various authors (e.g. Andrew *et al.*, 2014; Crossman *et al.*, 2013). ES supply has been mapped at various scales (e.g. sub-national to global) using both direct and indirect techniques. Whether direct or indirect techniques should be used depends on the information needs and characteristics of the ES of interest (Andrew *et al.*, 2014). EO-based methods can improve ES supply estimates through their ability to depict subtle spatial variations in the plant functional traits and soil properties known to influence the supply of many services (Andrew *et al.*, 2014).



It is worth noting that apart from microwave imagery, EO has limited capability for penetrating the vegetation canopy to extract useful information about any lower strata such as herbs or shrubs when the overstory is dense (Nagendra, 2001). Similarly, EO is incapable of providing direct information on below ground components of the ecosystems; these must be inferred from above ground data (Feng *et al.*, 2010). EO-based assessments may therefore only provide a partial picture of diversity and soil-based ES, for example, restricting the potential to comprehensively assess ES provision. This restriction of EO data highlights the importance of creating an integrated framework of EO and *in situ* data collection, not only to provide data for training and validation of EO products, but also to provide information on the areas that cannot be adequately assessed using EO.

## **2.5. Discussion**

When using satellite data there are a number of considerations to take into account, including data gaps and image frequency. The availability, or lack thereof, of cloud-free images at suitable timescales is a critical challenge for many EO-based applications. Research suggests revisit frequencies ranging from <1 day to exactly 8 days (depending on location and time of year) would be necessary to achieve a view at least 70% clear within 8 days (Whitcraft *et al.*, 2015). Previously, such revisit times were only achievable using coarse resolution data, which restricted the ability to perform fine-scale assessments of agricultural intensity. However, the number of medium-high resolution satellites has increased in recent years, for example with the launch of those within the Copernicus Programme. Combining freely available data from these various systems increases the number of images available for each point on earth. For example, by combining Landsat 8 with Sentinel-2A/B data, one image should be available every 3 days on average (Li and Roy, 2017), increasing the likelihood of obtaining sufficient cloud-free images. The creation of a Harmonised Landsat and Sentinel-2 reflectance data set (Claverie *et al.*, 2018) alongside the free availability of Sentinel-1 data (Torres *et al.*, 2012), which is not affected by cloud cover, the launch of satellite

fleets (e.g. Planet Labs) and the increased use of UAVs are all helping to mitigate the impact of cloud-cover on EO-based applications.

Various techniques also exist to deal with gaps and noise in the data. These techniques include pixel unmixing (Zhang *et al.*, 2017), data fusion (Gevaert and García-Haro, 2015; Senf *et al.*, 2015; Wang and Atkinson, 2018), best-pixel selection (Griffiths *et al.*, 2013; Hermosilla *et al.*, 2018), data interpolation (Inglada *et al.*, 2017; Vuolo *et al.*, 2017), climatology fitting (Verger *et al.*, 2013), temporal smoothing (Kandasamy *et al.*, 2013; Shao *et al.*, 2016; Tan *et al.*, 2011) and temporal aggregation (Loveland *et al.*, 2000). The introduction of computing platforms such as Google Earth Engine (Gorelick *et al.*, 2017) allows these complex algorithms to be applied to large volumes of data, by providing access to greater computing power and satellite datasets on a global-scale (Carrasco *et al.*, 2019). These platforms greatly reduce the time and cost associated with image processing, which increases the viability of such gap-filling techniques and facilitates development of new approaches for producing necessary cloud-free datasets.

Another key issue is the development of universally applicable EO-based monitoring techniques. Methods, such as crop yield models, that rely on largely location- and sensor-specific empirical relationships to retrieve indicators are common within many EO-based applications. However, lack of historic and large-scale *in situ* data for all areas limits the potential for calibration and validation of such methods (Doraiswamy *et al.*, 2005; Estel *et al.*, 2016). Suitable validation and calibration data may exist, but availability to researchers may be restricted due to commercial confidentiality, among other factors. The frequent reliance on empirical relationships, and lack of *in situ* data for all locations, limits the transportability of these methods to different sensors and study areas (Andrew *et al.*, 2014; Z. Li *et al.*, 2014); development of more generalised models is required. Less empirical models still often have parameters that need calibrating, to enable them to be applied appropriately to new geographical areas, new crops, or new animal species.

The creation of generic, transferable and widely applicable models must consider a variety of factors including the scale and resolution over which specific relationships apply (Foody, 2004; Gillespie *et al.*, 2008; Gustafson, 1998) and the variation in the relative dominance of different variables within space and time (A. K. Prasad *et al.*, 2006). Additionally, the impact that landscape characteristics can have on the accuracy with which indicators can be retrieved needs to be considered. For example, the size, shape and orientation of objects, such as narrow rivers (Sulistioadi *et al.*, 2015) and small habitat patches (Luoto *et al.*, 2002), can limit their detectability. Underlying environmental conditions can also affect the accuracy with which variables such as crop residue (Pacheco and McNairn, 2010) and soil moisture (Lakhankar *et al.*, 2009) can be retrieved. As such, comparison over diverse agricultural environments must be done with care, and with an understanding of the underlying differences in landscape characteristics.

To create models that can be widely applied, either empirical models are required that are integrated with ground collection efforts (e.g. Boryan *et al.*, 2011), or model-based methods that need no calibration to be applied to new areas must be developed. In practice, the solution is likely to involve a mixture of these two options.

### **2.5.1 Opportunities for an EO-based SI Assessment System**

Despite the potential importance of SI for securing future food supplies, methods for assessing the success of SI attempts are currently lacking. Gaps in these assessments arise, in part, due to the reliance on data that typically lacks the temporal and spatial coverage and resolution necessary to make a reliable assessment of SI. EO provides a range of opportunities for the development of an operational SI monitoring system that can be applied from field-scale to global-scale, at various temporal resolutions. The global coverage and consistency (e.g. image resolution, data quality and processing standards) afforded by EO data facilitates multi-scale analysis and comparisons between countries with diverse farming practices and field data availability. Additionally, the ability to derive different indicators from the same EO data sources ensures

coherence between measures of sustainability and intensification; this facilitates easier, more reliable integration of indicators into an assessment framework. The availability of long-term data (30-40 years from Landsat data archive) and repeat measurements allows the establishment of a baseline against which long-term and short-term changes can be assessed. Combined with the increasing availability of free satellite data, these factors allow a more flexible, adaptable and cost-effective approach to SI assessment to be developed. Furthermore, the ease, speed and efficiency with which assessments can be conducted is increased by the digital nature of EO data which allows, for example, simpler data input and processing. Recent advances in computer processing power/capability, such as cloud computing, further improve the situation, enhancing our ability to process and analyse large datasets.

As this review demonstrates, the basic EO-based assessment techniques for indicators of environmental sustainability and agricultural intensity are already established. This means that future work can focus on amalgamating existing work and creating a framework to integrate relevant indicators. An important step in the development of a comprehensive EO-based assessment framework is the creation of a set “Essential Sustainable Intensification Variables” (ESIVs) to form the basis for a global monitoring program. Having a set of “essential variables” helps to prioritise efforts by outlining a minimum set of essential measurements (Pereira *et al.*, 2013) required to capture major dimensions of agricultural and environmental change, allowing the sustainability of intensification to be assessed. Identifying the relevant variables to be included in this list will ultimately require the establishment of a clear selection framework to ensure all relevant variables are identified and their selection is justified (although the development of such a framework falls beyond the scope of this study). The development of this list should build on the selection frameworks created for the Essential Climate Variables (ECVs) and Essential Biodiversity Variables (EBVs) of the Global Framework for Climate Services (GCFs) and Group on Earth Observations Biodiversity Observation Network (GEO BON) respectively. In brief, variable selection will require an open, inter-disciplinary process, involving the engagement of scientific, policy and other

communities (Pereira *et al.*, 2013). This will ensure it builds on existing activities such as the ECVs, EBVs and GEOGLAM. Variables should be identified based on key criteria such as relevance, feasibility, scalability, temporal sensitivity and cost effectiveness (Bojinski *et al.*, 2014; Pereira *et al.*, 2013). As Pettoirelli *et al.* (2016) highlight for the ECVs and EBVs, the identification of suitable ESIVs will be an evolving process. The list of indicators will need to be periodically updated as technology advances, and as sensor availability and observation priorities change. Table 2.3 provides an example list of possible “essential” EO-based indicators. This list is based on indicators already incorporated in the EBVs and ECVs, and previous SI frameworks including the Sustainable Intensification Assessment Framework (Musumba *et al.*, 2017). As this review is focused on the role of EO within SI assessment, this list only includes possible indicators that could be derived using EO data. The final list of essential variables would need to include indicators derived from a variety of data sources to ensure key aspects of sustainable intensification can be monitored.

*Table 2.3: Examples of possible Essential Sustainable Intensification Variables (ESIV). \*indicates variables which are already included within the list of EBVs. \* indicates variables which are already included within the list of ECVs. NB: the OneSoil data is not currently available for download, but it shows the potential for the creation of an operational global crop-type map.*

ESIV examples	Relevance	Existing Operational Products	Ideal Product Coverage
<b>Agricultural intensity</b>			
Crop type	Essential product required to be able to accurately monitor/derive crop yield and area, but not an indicator in its own right.	<ul style="list-style-type: none"> <li>• OneSoil crop-type map 2016, 2017, 2018</li> <li>• Country-level products e.g. CEH Land Cover plus: Crops 2015 (partial coverage), 2016, 2017, 2018 (GB)</li> </ul>	• Annually (possibly more often for multi-cropping systems)
Crop yield	Needed to quantify agricultural intensity	Currently no operational products exist	• Annually
Crop yield gap	Needed to help identify areas that could be farmed more intensively	Currently no operational products	• Annually
Cropping area	Needed to quantify agricultural extent	<ul style="list-style-type: none"> <li>• OneSoil crop-type map 2016, 2017, 2018</li> <li>Crops are a subset of land cover maps:</li> <li>• Country-specific Land Cover maps e.g. CEH GB Land Cover Map 2015 30m</li> <li>• Pan-European: CORINE Land Cover minimum mapping unit 25ha</li> <li>• Global land cover map 30m (Chen <i>et al.</i>, 2015)</li> </ul>	• Annually

Table 2.3 continued

ESIV examples	Relevance	Existing Operational Products	Ideal Product Coverage
<b>Environmental sustainability</b>			
NPP/GPP*	Provides a measure of the health/degradation of the ecosystem; underpins all production-based ecosystem services.	<ul style="list-style-type: none"> <li>• NASA MODIS yearly 500m/1km 2000 to present</li> <li>• Copernicus GDMP 10-day 1km 1999 to present &amp; 10-day 300m 2014 to present</li> </ul>	<ul style="list-style-type: none"> <li>• Patches of natural/semi-natural habitat within the farmed area</li> <li>• Annually</li> </ul>
Soil moisture *	Indicator of soil quality. Increased fertiliser use can increase water consumption and deplete soil moisture. A decline in soil moisture may be an indication of unsustainable intensification.	<ul style="list-style-type: none"> <li>• SMOS daily/3-day/10-day/monthly 15/25km 2009 to present</li> <li>• Copernicus METOP/ASCAT daily 0.01° 2007 to present</li> <li>• Copernicus Sentinel-1 daily 1km 2015-2017 (Europe only)</li> </ul>	<ul style="list-style-type: none"> <li>• Farmed area</li> <li>• Annually</li> </ul>
Soil erosion	Removal of interstitial features (e.g. hedgerows) and increased runoff (due to soil compaction) increases soil erosion. An increase in soil erosion may be an indication of unsustainable intensification.	Currently no operational products exist	<ul style="list-style-type: none"> <li>• Farmed area plus surrounding area</li> <li>• Annually</li> </ul>
Soil organic carbon *	Indicator of soil quality. Increased irrigation and soil erosion lead to a decline in organic matter content. A decrease in soil organic carbon may be an indication of unsustainable intensification.	Currently no operational products exist	<ul style="list-style-type: none"> <li>• Farmed area</li> <li>• Annually</li> </ul>
Water clarity/turbidity	Increased agrochemical inputs and increased soil erosion reduce water quality. Lower clarity/higher turbidity may be an indication of unsustainable intensification.	<ul style="list-style-type: none"> <li>• Copernicus ENVISAT/MERIS 10-day 300m/1km 2002 to 2012</li> <li>• Copernicus Sentinel-3/OLCI 10-day 300m/1km 2017 to present</li> </ul>	<ul style="list-style-type: none"> <li>• Nearby water bodies</li> <li>• Annually</li> </ul>
Landscape structure*	Removal of interstitial features e.g. hedges and increased field sizes cause simplification of habitat structure and loss of ecosystem connectivity. Knock-on effect on species populations and diversity (e.g. birds).	<ul style="list-style-type: none"> <li>• Country-specific Land Cover maps e.g. CEH GB Land Cover Map 2015 30m</li> <li>• Pan-European: CORINE Land Cover Minimum mapping unit 25ha</li> <li>• Global land cover map 30m (<i>Chen et al., 2015</i>)</li> </ul>	<ul style="list-style-type: none"> <li>• Surrounding area</li> <li>• Annually</li> </ul>
Species richness/diversity*	Increase in agrochemical inputs, irrigation and landscape structure simplification alter species composition. A decrease in species richness and diversity may be an indication of unsustainable intensification.	<ul style="list-style-type: none"> <li>• European Atlas of Forest Tree Species 1km</li> </ul>	<ul style="list-style-type: none"> <li>• Surrounding area</li> <li>• Annually</li> <li>• Flora and fauna species</li> </ul>

Once a comprehensive list of ESIVs has been generated, careful consideration must be given to the selection of appropriate methods to assess each variable. This could be achieved by adopting an open process of algorithm inter-comparison and selection similar to that used by ESA in their Climate Change Initiative (Hollmann *et al.*, 2013), 2013) and the Sen2-Agri project (Bontemps *et al.*, 2015). To allow reliable comparison of different algorithms and to ensure their relevance at local scales and widely varying agricultural systems at global scales, an open test dataset similar to that used in the Sen2-Agri project should be developed. This dataset was created through acquisition of satellite and *in situ* data from the same season over sites representative of global agricultural system diversity (Bontemps *et al.*, 2015). Application of the potential algorithms to this data set ensured an objective and transparent algorithm selection method, which should be mirrored in the creation of the ESIV data products.

An important part of building the ESIV products will be a comprehensive assessment of the associated uncertainties and a clear communication of these uncertainties to the end-user. Kissling *et al.* (2018) set out a workflow of 11 steps used to operationalise the process of building EBV data products, including the quantification and communication of uncertainties in terms of data, model algorithms and parameters. Consideration must be given to uncertainties associated with the underlying raw data, from both satellite and *in situ* sources, and from processing methods (e.g. gap-filling techniques) and models applied to this data. Kissling *et al.* (2018) highlight the need to develop high-throughput processing tools for quantifying uncertainties; the same will be true for the ESIVs.

Another key element will be the development of a framework through which the ESIVs can be utilised to provide an assessment of the sustainability of agricultural intensification. Previous assessment frameworks have utilised indicators in a number of ways: (i) individually, expressed in units, (ii) as part of a set, or (iii) in a composite index, whereby scores of individual indicators are combined into a single, dimensionless number, or sustainability score (Dantsis *et al.*, 2010; Farrell

and Hart, 1998; Mitchell *et al.*, 1995; Van Passel and Meul, 2012). An example of a widely used composite indicator for sustainability assessment is the Ecological Footprint which combines various indicator footprints including carbon, forest, crop land, & built up land footprint to provide a measurement of human demand for land and water areas (Toderoiu, 2010 in Čuček *et al.*, 2012; Galli *et al.*, 2012). Some studies choose not to aggregate their indicators, adopting instead the use of sustainability polygons, webs and radars, which removes the need for aggregation across different scales by displaying scores for different index components simultaneously (Rigby *et al.*, 2001). The development of a suitable assessment framework has been identified as one of the key strands of SI research (Siebrecht, 2020) and is currently being actively explored (see recent reviews by Cassman and Grassini, 2020; Siebrecht, 2020). However, at present there is little investigation into the incorporation of EO data into such frameworks. The creation of an EO-based framework, and a decision about the best way to utilise the ESIVs within this, will require expert knowledge, interdisciplinary/international collaboration and consultation with researchers and intended users.

The success of the proposed EO-based assessment framework will rely heavily on the development of an integrated system of routine collection and provision of *in situ* data. *In situ* data is required to perform a number of roles including calibration and validation of EO-derived products and assessment of elements of the environment that cannot efficiently/effectively be monitored using EO data, for example below-ground properties and processes. A comprehensive assessment of SI will also require economic and social data that cannot be provided by EO.

An EO-based assessment framework could be implemented at different scales and at different levels of detail. For example at country-level, assessments are more likely to have access to environmental and farming data sets that would enable more detailed assessments of environmental sustainability. National-scale assessments are also more likely to have access to additional economic and social data that would enable more comprehensive SI assessments to be conducted. However, globally less detailed *in situ* data is likely to be available, although programs



such as GEOSS and GEOGLAM have shown that good quality reference data sets can be collected for some areas to help the development of more global solutions. The development of detailed nationally-based methods, and less detailed globally-based methods, is likely to occur in different ways. A globally-based system would develop most sensibly through integration with existing global initiatives such as GEOSS, which is already generating products capable of agricultural and environmental monitoring. Whereas national-scale solutions are likely to develop from existing country-level environmental and agricultural monitoring schemes. However, for both the global and national-scale, the more integrated the EO and other strands of environmental monitoring are the better system will be.

## **2.6. Conclusion**

One element of meeting the future food demands of a growing population, with shifting food consumption patterns, will be the intensification of agricultural production. To ensure long term environmental degradation is avoided, any increases in food production must be undertaken in a sustainable manner. The lack of any prescribed methods of sustainable intensification mean that to successfully achieve this goal a comprehensive method of assessing the sustainability of intensification endeavours must be developed. Various frameworks exist at present; however, these commonly rely on traditional data sources that do not provide adequate coverage, resolution, or frequency of data to generate reliable results for all agricultural systems. The potential for an EO-based assessment system is clear, with an extensive body of research into EO methods for monitoring earth surface properties and their spatial and temporal variation.

The element that is currently missing is a system for combining these indicators to provide a comprehensive assessment of the sustainability of agricultural intensification. Such a system could build on the approaches used to develop the EBVs and ECVs and global agricultural monitoring schemes such as GEOGLAM. Determining the optimum format for this system will require a multi-

disciplinary, multi-organisation working group involving farmers, researchers, government bodies and other stakeholders.

Irrespective of the exact nature of the final system, EO offers the opportunity to obtain more spatially and temporally representative data, over scales and resolutions unobtainable with conventional data collection methods. An EO-based system, however, does not exclude the need for *in situ* data, rather it will supplement current systems facilitating more efficient and consistent multi-scale assessments over a range of temporal resolutions at a lower cost. Integration of EO and *in situ* data on national and global scales, will be provide a step change in our ability to provide regular, consistent and timely assessments. This is essential if we are to meet future production demands without causing significant, irreparable damage to the environment.

### **3. High Resolution Wheat Yield Mapping Using Sentinel-2**

Merryn L. Hunt, George Alan Blackburn, Luis Carrasco, John W. Redhead & Clare S. Rowland

This chapter is a replication of a constituent paper of this research that was published in Remote Sensing of Environment.

Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., & Rowland, C.S. (2019). High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment*, 233, 111410.

#### **Abstract**

Accurate crop yield estimates are important for governments, farmers, scientists and agribusiness. This paper provides a novel demonstration of the use of freely available Sentinel-2 data to estimate within-field wheat yield variability in a single year. The impact of data resolution and availability on yield estimation is explored using different combinations of input data. This was achieved by combining Sentinel-2 with environmental data (e.g. meteorological, topographical, soil moisture) for different periods throughout the growing season. Yield was estimated using Random Forest (RF) regression models. They were trained and validated using a dataset containing over 8000 points collected by combine harvester yield monitors from 39 wheat fields in the UK. The results demonstrate that it is possible to produce accurate maps of within-field yield variation at 10m resolution using Sentinel-2 data (RMSE 0.66 tonnes/ha). When combined with environmental data further improvements in accuracy can be obtained (RMSE 0.61 tonnes/ha). We demonstrate that with knowledge of crop-type distribution it is possible to use these models, trained with data from a few fields, to estimate within-field yield variability on a landscape scale. Applying this method gives us a range of crop yield across the landscape of 4.09 to 12.22 tonnes/ha, with a total crop production of approx. 289000 tonnes.

### **3.1 Introduction**

Crop yield is a key agricultural variable. Accurate crop yield estimates serve a range of important purposes helping to make agriculture more productive and more resilient. Reliable yield estimates can be used to identify yield-limiting factors to guide development of site-specific management strategies (Diker *et al.*, 2004; Jin *et al.*, 2017b). Building a time-series of yield estimates allows producers and consultants to understand how management strategies affect crop productivity, guiding future practices (Birrell *et al.*, 1996; Grisso *et al.*, 2002; Lobell, 2013). Accurate estimates also provide valuable information about mean yields and variability of yields at the field-scale, which are required for insurance and land market decisions (Lobell *et al.*, 2015). Despite its importance, crop yield information is currently patchy within and between countries, in part due to commercial sensitivities. Various organisations are rapidly addressing this issue for present day yield estimates. Activities such as GEOGLAM (GEO, 2018; Whitcraft *et al.*, 2015) are assessing crop condition on a country/global-scale, while commercial companies are offering predictive services at a field/farm-scale. However, as these organisations typically focus on assessing current conditions rather than retrospective estimation, there is currently no facility to build up a long-term time series of field-scale crop yields. There are also a lack of estimates of within-field yield variability at the landscape-scale, which is of most concern to scientists assessing the sustainability of agriculture and its impact on the environment.

Agricultural monitoring has been a key focus of Earth Observation (EO) activity since the first terrestrial satellites were launched (Anuta and MacDonald, 1971; Draeger and Benson, 1972; Horton and Heilman, 1973). However, the potential of EO has been limited by image costs and limited repeat frequency, which combined with cloud means that key phases in crop growth are missed. This is all changing with the opening of the Landsat archive (Wulder *et al.*, 2012), the launch of the Sentinel satellites (Drusch *et al.*, 2012; Torres *et al.*, 2012) and readily accessible cloud-computing platforms like Google Earth Engine (Gorelick *et al.*, 2017). EO systems are increasingly able to support the operational production of data products, however, it is still important to choose the

most appropriate data set and method for mapping a particular variable. This is particularly true for agricultural monitoring, where key validation data, specifically crop yield, is held by individual farmers. Such data is often deemed commercially sensitive, making it difficult to collate large data sets to enable development and validation of EO-based methods. The EO work to date has therefore been constrained by the type and scale of validation data available.

Various studies have explored the possibility of using EO data to map yield at the field-level, with particular focus on yield variability within smallholdings (Burke and Lobell, 2017; Jain *et al.*, 2016; Jin *et al.*, 2017a). While results of these studies have been promising, many of them rely on commercial EO data (Burke and Lobell, 2017) or a combination of commercial and freely available EO data (Jin *et al.*, 2017a). Costs of very high resolution (<5m) commercial satellite data are decreasing, particularly with the increase in the number of “cubesat” companies (Burke and Lobell, 2017). However, the fact that there is still a cost associated with obtaining the data means that it will not be universally accessible, particularly in developing countries. If similar accuracies can be achieved using slightly lower resolution freely available data, as provided by Sentinel-2, then this provides a more practical option for yield mapping. Previous studies have highlighted the potential of Sentinel-2 to play a key role in estimating crop yield (Battude *et al.*, 2016; Lambert *et al.*, 2017; Skakun *et al.*, 2017), but so far the potential for mapping within-field variability in yield has yet to be fully explored.

Lack of high resolution yield data for training and validation is a common problem for EO-based studies seeking to map yield at high resolution. Yield data are often collected in the field through crop cuts on sample plots and farm surveys. Lack of accurate location data and concerns over yield data accuracy mean this data is typically aggregated to the field level (Burke and Lobell, 2017; Lambert *et al.*, 2017) or to the district level (Jin *et al.*, 2017a). Various studies have demonstrated the relatively high yield estimation accuracy obtainable using high resolution satellite images for aggregated spatial units, and high resolution maps have been produced (e.g. 1m: Burke

and Lobell, 2017). However, due to the common practice of aggregating crop yield data past studies have typically been unable to verify the accuracy of within-field variability shown.

In recent years, there have been a number of innovations in farming technology to allow farmers to observe, measure and respond to spatial and temporal variation in crops. Such “precision farming” approaches aim to ensure accurate targeting of agricultural interventions and reduce waste and detrimental impacts. A key component of precision farming has been the incorporation of high-accuracy GPS technology into farm machinery, including combine harvesters. Coupled with on-board yield monitors, this offers the potential for accurate, fine-resolution mapping of within-field variation in crop yields. High resolution data collected by yield monitors on-board combine harvesters has been used to assess the capability of EO to estimate crop yield, with positive results (Kayad *et al.*, 2016; Yang *et al.*, 2009). So far, however, high resolution yield data has not been combined with Sentinel-2 data to estimate yield, beyond the initial exploration of the correlation between Sentinel-2 NDVI and spring barley yield data by Jurecka *et al.* (2016). As such, the present study seeks to explore the ability of Sentinel-2 data to estimate within-field yield variability using combine harvester data for training and validation.

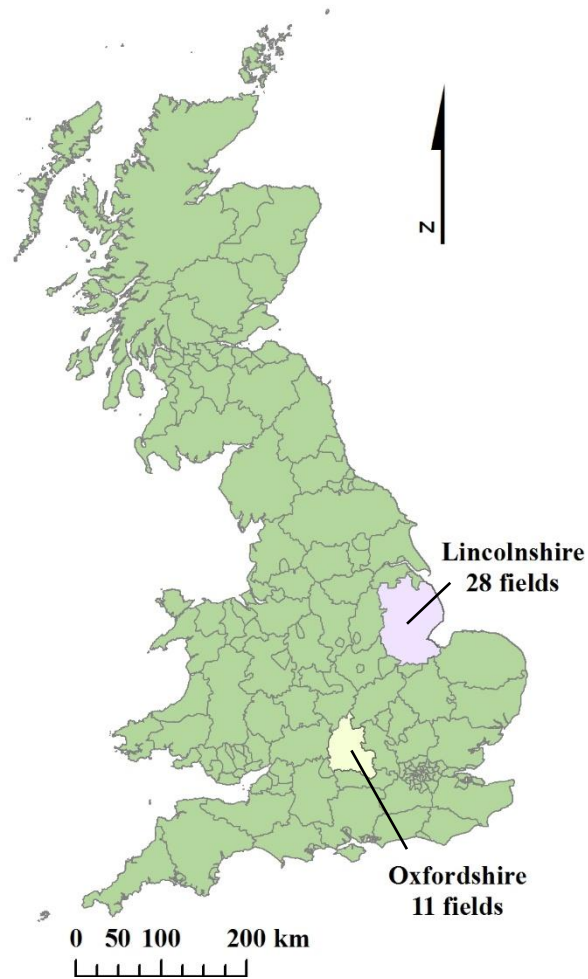
In this study, the capability of Sentinel-2 to estimate within-field wheat yield variability was assessed. The aim was to produce an empirical model calibrated with combine harvester data to estimate yield. A method was developed that can be applied for a given year at high spatial resolution at the landscape scale, when suitable training data are available. Random Forest (RF) models were trained and validated using data from yield monitors on-board combine harvesters. The combine harvester dataset contained over 8000 points collected in 39 wheat fields within the UK. The analysis was structured around 5 key questions designed to explore how different combinations of data, in terms of both type and temporal coverage, impact the accuracy of wheat yield estimation.

- Question 1: How does Sentinel-2 spatial resolution affect the accuracy of yield estimation?
- Question 2: Does calculating separate vegetation indices (VIs) contribute any extra information to the estimation model?
- Question 3: How do different combinations of Sentinel-2 data and environmental data affect estimation accuracy?
- Question 4: Which single-date Sentinel-2 image provides the most accurate estimation?
- Question 5: How does estimation accuracy vary with accumulation of data throughout the growing season for Sentinel-2 data only (Qu 5a), Sentinel-2 and environmental data combined (Qu 5b), and environmental data only (Qu 5c)?

The paper concludes by applying the optimal RF model to estimate within-field yield variability on a landscape scale.

### **3.2 Field Sites**

This study was conducted using data from 39 conventionally farmed wheat fields in the UK. The data were spread over two different regions, with 28 fields in Lincolnshire and 11 fields in Oxfordshire covering a total of 438.2ha and 224.2ha respectively (figure 3.1). Lincolnshire is relatively flat and, at 75% arable, is the most intensively farmed county in the UK, whereas Oxfordshire is less flat, with more of a mix of arable (52%) and grassland (32%) (Rowland et al., 2017a). The average annual rainfall in Lincolnshire, from 1981-2010, was 614mm and for Oxfordshire 659mm. Annual average temperatures ranged from 6.3 to 13.5°C and 6.9 to 14.6°C for Lincolnshire and Oxfordshire respectively (Met Office, 2018). In 2016 the average wheat yield at the Lincolnshire sites was 10.27 tonnes/ha, and at the Oxfordshire sites 9.79 tonnes/ha (based on cleaned and interpolated combine harvester yield data at 10m resolution).



*Figure 3.1: Location of study sites.*

### **3.3 Data and Methods**

Figure 3.2 provides an overview of the method used in this study, outlining how the combine harvester data, satellite data and environmental data were processed and combined to estimate yield. The specific details of the data and data processing techniques are outlined in sections 3.3.1-3.3.3, and the analysis techniques are outlined in sections 3.3.4-3.3.6.



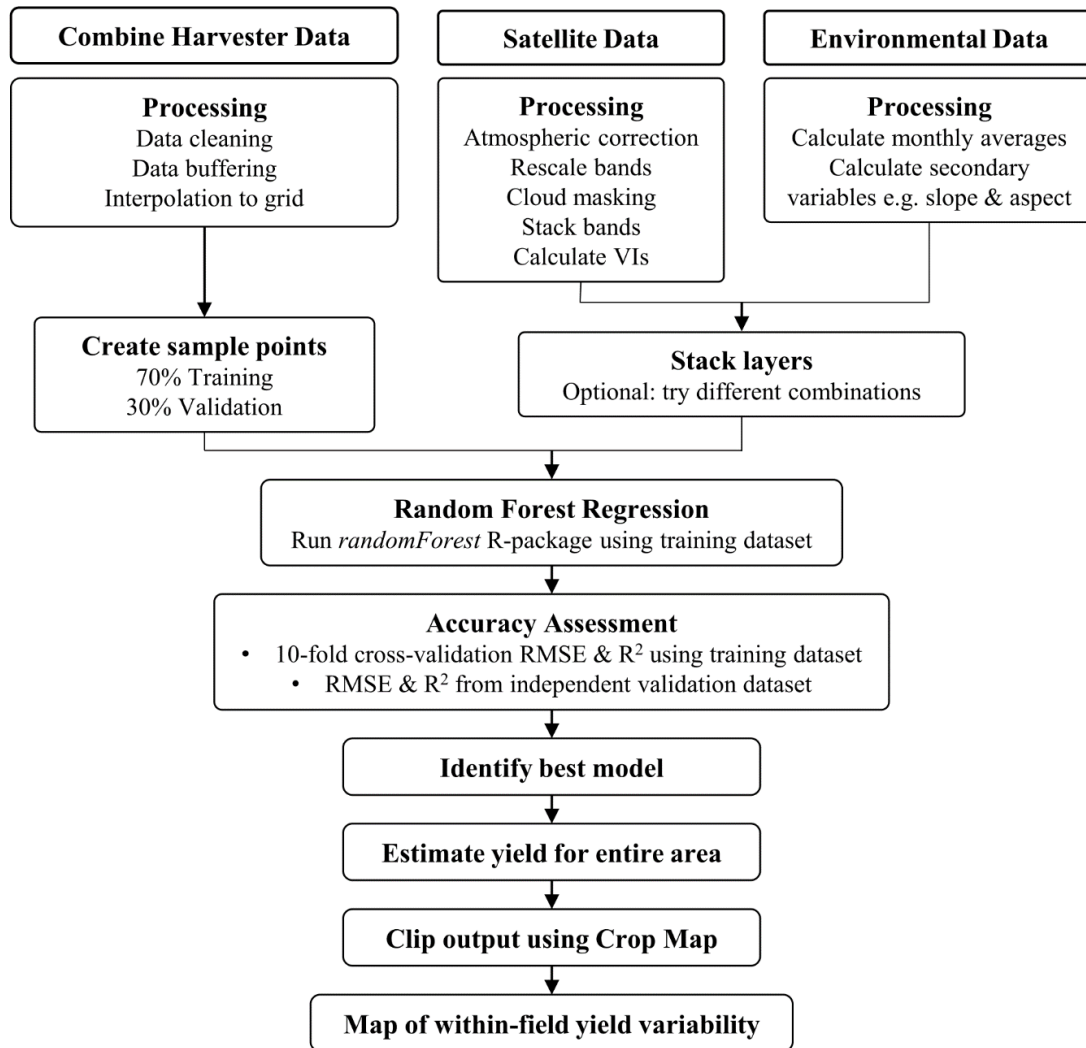
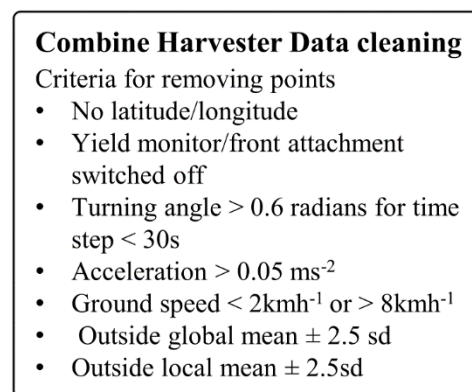


Figure 3.2: Overview of the method used to estimate yield at high resolution on a landscape scale.

### 3.3.1 Wheat Yield Data

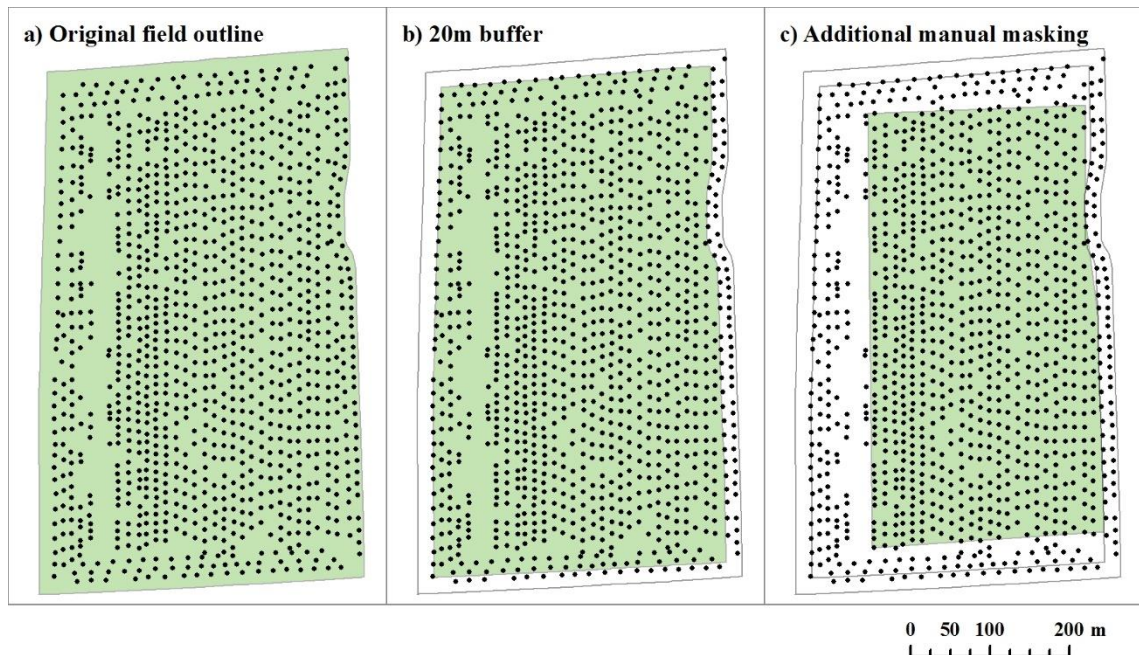
High resolution wheat yield data was downloaded from CLAAS telematics, a web-based vehicle fleet management data analysis system (CLAAS, 2018). The wheat yield data were acquired during the 2016 harvest period between 6<sup>th</sup> August and 9<sup>th</sup> September using combine harvesters equipped with a GPS and optical yield monitor. Wheat was chosen as the crop of interest for this study due to its high prevalence within the available dataset. In the UK, winter wheat crops are typically planted in October and harvested in August (AHDB, 2018). The raw data were cleaned to remove inaccurate grain yield measurements arising, for example, from the harvesting dynamics of the combine harvester and the accuracy of positioning information (AHDB, 2016; Lyle *et al.*, 2014).

Simple cleaning steps included removing data points for which no latitude/longitude were recorded and points where the yield monitor or front attachment were not switched on. Additionally, a check was applied to ensure each field was harvested by a single combine harvester, as different combines will have differently calibrated yield monitors. A series of threshold-based cleaning steps were then applied to remove values recorded while the combine harvester was turning (turning angle  $> 0.6$  radians for time step  $< 30$ s), accelerating or decelerating (accel.  $> 0.05 \text{ ms}^{-2}$ ), or when the speed fell outside the optimum limits to accurately measure the yield (ground speed  $< 2 \text{ kmh}^{-1}$  or  $> 8 \text{ kmh}^{-1}$ ). Finally, data were cleaned on a per field basis removing yield values which fell outside the global mean  $\pm 2.5 \text{ sd}$  or the local mean  $\pm 2.5 \text{ sd}$  (based on the closest 3 points). A summary of the criteria for data cleaning can be found in figure 3.3.



*Figure 3.3: Summary of the criteria for data cleaning.*

To avoid any mixed pixels in the satellite data, a 20m buffer around the inward edge of the field was applied to the cleaned data. Further to this, additional areas were manually masked out to remove large gaps arising in the dataset as a result of the data collection and cleaning process. These gaps typically occurred at the edge of the fields and in areas where the combine harvester turned. Figure 3.4 shows an example of the data gaps in one field and the stages in the buffering process used to remove them. Post-buffering the data covered an area of 252.2ha (c.f. 438.2ha) in Lincolnshire and 100.4ha (cf. 224.2ha) in Oxfordshire.



*Figure 3.4: Example yield data points for one field showing a) gaps in the data arising from the data collection and cleaning process and b-c) the stages in the buffering process used to remove these gaps.*

The cleaned and buffered yield data were resampled to resolutions of 10m and 20m using an Inverse Distance Weighting function. Yield was mapped at these resolutions to align with the Sentinel-2 data used within this study, and to allow an assessment of the optimum resolution for yield estimation to be made. The appropriateness of mapping at these resolutions was supported by the relative uniformity of points (figure 3.5) and the mean nearest neighbour distance of 11m for the yield points. Additionally, when considering yield data, a major factor limiting the spatial resolution is the width of the cutting head on the combine harvester, which will determine the minimum acceptable resolution. The cutting widths for the combine harvesters used in this study ranged from 4.95m to 12.27m, thus providing further justification for mapping yield at 10m and 20m resolution. Sample points were generated in the centre of each interpolated raster cell. To reduce the impact of auto-correlation between pixels only alternate pixels were used, producing a sample dataset containing 8794 values. The sample data was then randomly split into training (70%) and validation (30%) datasets.

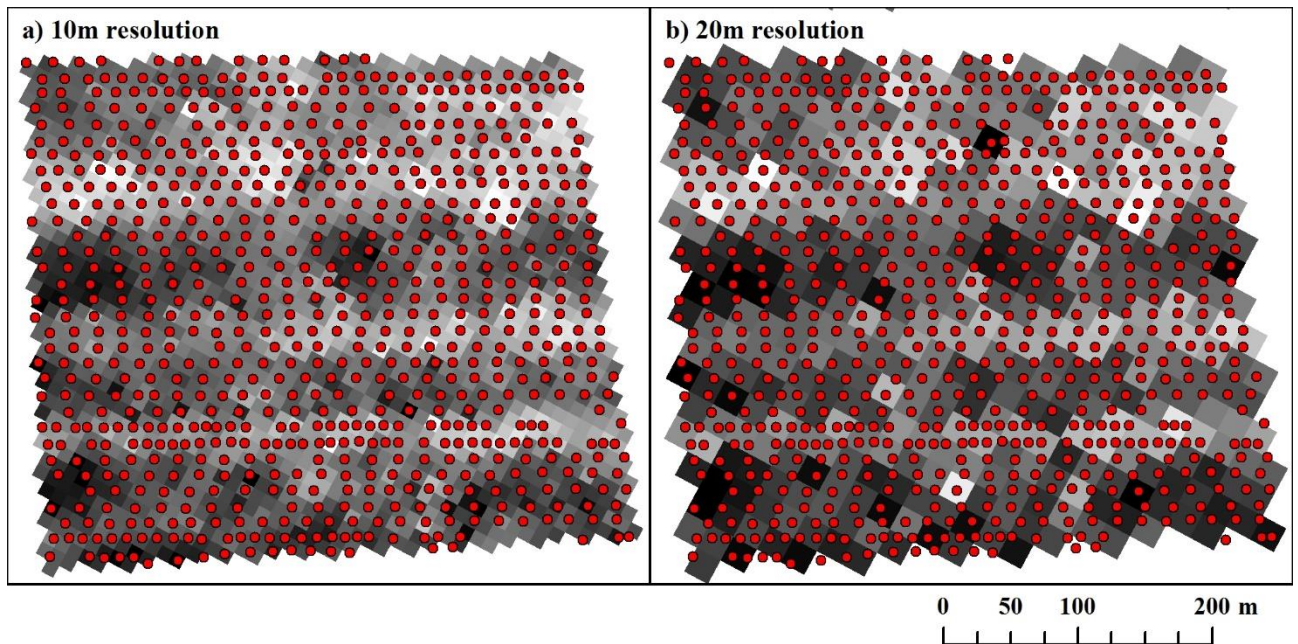


Figure 3.5: Example of the distribution of yield data points relative to a) 10m and b) 20m resolution interpolated yield data.

### 3.3.2 Sentinel-2 Data

#### 3.3.2.1 Sentinel-2 Image Processing

Predominantly cloud-free Sentinel-2 images (Claverie *et al.*, 2018; Drusch *et al.*, 2012) for tiles 30UXC and 30UXD were downloaded from the Copernicus Open Access Hub (ESA, 2018); only bands at 10 or 20m resolution were used in this study. Details of the bands used within this study can be found in table 3.1. Relatively cloud-free images were available over the growing season for the 29<sup>th</sup> December 2015, 20<sup>th</sup> April 2016, 6<sup>th</sup> June 2016 and 19<sup>th</sup> July 2016 (table 3.2); figure A3 in Appendix 3 shows where these images fit in relation to the growth stages of wheat. The four suitable images available from Sentinel-2 compare favourably to Landsat-8, which would have provided only one suitable cloud-free image for the 2016 growing season. All bands were atmospherically corrected using the Sen2Cor processor and bands at 20m resolution were rescaled to 10m before the bands were stacked. Cloud was then manually masked out of the April and December images, because the current Sentinel-2 cloud masking is not completely accurate (Coluzzi *et al.*, 2018).

Table 3.1: Central wavelength and spatial resolution for the Sentinel-2 bands used in this study (Drusch et al., 2012).

Spectral Band	Central Wavelength (nm)	Spatial Resolution (m)
Band 2 Blue	490	10
Band 3 Green	560	10
Band 4 Red	665	10
Band 5 Vegetation red edge	705	20
Band 6 Vegetation red edge	740	20
Band 7 Vegetation red edge	783	20
Band 8 NIR	842	10
Band 8a Narrow NIR	865	20
Band 11 SWIR	1610	20
Band 12 SWIR	2190	20

Table 3.2: Explanatory variables used in Random Forest regression analysis.

Variable type	Dataset	Pixel size	Temporal coverage
Sentinel-2	Sentinel-2 Level 1C bands: 2, 3, 4, 5*, 6*, 7*, 8, 8a*, 11*, 12*	10m (*20m rescaled to 10m)	29 <sup>th</sup> Dec 2015 20 <sup>th</sup> April 2016 6 <sup>th</sup> June 2016
Vegetation indices	GCVI, GNDVI, NDVI, SR and WDRVI calculated from Sentinel-2 data	10m	19 <sup>th</sup> July 2016
Environmental	Precipitation	UKCP09 gridded observation dataset – Total precipitation amount over the calendar month (mm)	5km Dec 2015 – July 2016
	Temperature	UKCP09 gridded observation dataset – Average of daily mean air temperature over the calendar month (°C)	5km
	SWI	Monthly average Soil Water Index calculated using SCAT-SAR SWI T01 data	500m
	DTM	NEXMap Digital Terrain Model	10m
	Aspect	Calculated using the NEXMap	10m
	Slope	DTM	10m
			Created using data collected in 2002 & 2003

### 3.3.2.2 Vegetation Indices Calculation

Five vegetation indices (VIs) that have been used in previous yield estimation studies (e.g. Jin et al., 2017a; Shanahan et al., 2001; Yang et al., 2009, 2000; Yang and Everitt, 2002) were calculated

from the Sentinel-2 imagery, specifically GCVI, GNDVI, NDVI, SR and WDRVI (see table 3.3 for equations).

*Table 3.3: Vegetation indices calculated using Sentinel-2 imagery, where R is Red (B4), G is green (B3) and NIR is near-infrared (B8a)*

VI	Abbreviation	Equation	Reference
Green Chlorophyll Vegetation Index	GCVI	$GCVI = \left( \frac{NIR}{G} \right) - 1$	Gitelson <i>et al.</i> , 2003
Green Normalised Difference Vegetation Index	GNDVI	$GNDVI = \frac{NIR - G}{NIR + G}$	Gitelson <i>et al.</i> , 1996
Normalised Difference Vegetation Index	NDVI	$NDVI = \frac{NIR - R}{NIR + R}$	Rouse <i>et al.</i> , 1973
Simple Ratio	SR	$SR = \frac{NIR}{R}$	Jordan, 1969
Wide Dynamic Range Vegetation Index	WDRVI	$WDRVI = \frac{0.2 * NIR - R}{0.2 * NIR + R}$	Gitelson, 2004

### 3.3.3 Environmental Data

#### 3.3.3.1 Precipitation and Temperature

Monthly 5km gridded UKCP09 data sets of total rainfall (mm) and mean air temperature (°C) were downloaded from the UK Met Office (Met Office, 2017). Monthly data was downloaded for December 2015 to July 2016 to match the period covered by the Sentinel-2 images (table 3.2). 5km is coarse and ideally higher resolution data would have been utilised. Unfortunately such data were not available for the study sites at the required dates. However, given the spatial distribution of the fields across the study areas of Lincolnshire and Oxfordshire data from 54 of the 5km squares was used. This distribution allowed spatial variation in precipitation and temperature across the study area to be detected despite the coarse resolution of the data.

#### 3.3.3.2 Soil Water Index

The Soil Water Index (SWI), first proposed by Wagner *et al.* (1999), is an indicator of the soil moisture profile. SWI values for December 2015 to July 2016 were obtained from the SCAT-SAR SWI



T01 dataset (Scatterometer – Synthetic-Aperture-Radar Soil Water Index) created by the TU Wien Department of Geodesy and Geoinformation (table 3.2). This data is derived from radar data observed by the MetOp-A/B ASCAT and Sentinel-1 SAR satellite sensors. SWI images have a pixel spacing of 500m which correspond to a resolution of 1km. Monthly mean values were calculated from the SWI giving a percentage ranging from completely dry soil (0%) to completely saturated soil (100%).

#### **3.3.3.3 Topographic Variables**

A 10m resolution digital terrain model (DTM) was obtained from NEXTMap Britain (table 3.2). The DTM was created by Intermap Technologies Inc. based on airborne radar data collected during 2002 and 2003 (Intermap Technologies, 2009). This data was used to calculate aspect and slope variables at 10m resolution.

#### **3.3.4 Random Forest Regression**

Random Forest was trained and applied to estimate wheat yields over the satellite image extent. Random Forest (RF; Breiman, 2001) is a machine learning algorithm that can be used to estimate a continuous response variable using regression analysis. The RF algorithm first creates a pre-defined number of new training sets with random sampling and then builds a different tree for each of these bootstrapped datasets. In each tree, a random subset of explanatory variables is used to recursively split the data at each node into more homogenous units (Breiman, 2001; Everingham *et al.*, 2016; A. M. Prasad *et al.*, 2006). The trees are fully grown and the mean fitted response from all the individual trees provides the estimated value of a continuous response (Everingham *et al.*, 2016). Further details on Random Forest and its implementation can be found in Appendix 2. Previous studies have used RF to estimate yields for a variety of crops including sugarcane (Everingham *et al.*, 2016), corn (Kim and Lee, 2016), wheat, maize and potato tuber (Jeong *et al.*, 2016).

In this study RF analysis was carried out using a modified version of the “randomForestPercentCover” script produced by Horning (2014), which uses the R “randomForest” package developed by Liaw and Wiener (2002). The original script was designed to explore continuous vegetation cover, so modification was required to provide mean yield per pixel as opposed to percentage vegetation cover. The default settings of the randomForest package were used: one third of all available explanatory variables were used to split the data at each node and the number of trees was 500 (Liaw and Wiener, 2002).

The RF model was trained to estimate crop yield using the variables outlined in table 3.2 as explanatory variables. The impact of different data combinations and different temporal coverages on estimation accuracy were explored using the layer combinations shown in table 3.4.



Table 3.4: Data combinations tested in Random Forest analysis. All Sentinel-2 data is at 10m resolution (except for the S2\_20 combination). All environmental data were resampled to 10m. For individual layer details see table 3.2.

Combination	Data layers
<b>Question 1</b>	
S2	Sentinel-2 data
S2_20	Sentinel-2 data resampled to 20m
<b>Question 2</b>	
S2	Sentinel-2
S2_VI	Sentinel-2, VIs
VI	VIs
<b>Question 3</b>	
S2	Sentinel-2
S2_Met	Sentinel-2, Precipitation, Temperature
S2_SWI	Sentinel-2, SWI
S2_Topo	Sentinel-2, DTM, Aspect, Slope
S2_Env	Sentinel-2, Precipitation, Temperature, SWI, DTM, Aspect, Slope
<b>Question 4</b>	
D	Sentinel-2 data December only
A	Sentinel-2 data April only
Jn	Sentinel-2 data June only
J	Sentinel-2 data July only
<b>Question 5a</b>	
D	Sentinel-2 data December only
DA	Sentinel-2 data December and April
DAJ	Sentinel-2 data December, April and June
DAJJ (S2)	Sentinel-2 data December, April, June and July
<b>Question 5b</b>	
D-S2_Env	Sentinel-2 and Environmental data December only
DA-S2_Env	Sentinel-2 data December and April Environmental data up to end of April
DAJ-S2_Env	Sentinel-2 data December, April and June Environmental data up to end of June
DAJJ-S2_Env (S2_Env)	Sentinel-2 data December, April, June and July Environmental data up to end of July
<b>Question 5c</b>	
D-Env	Environmental data December only
DA-Env	Environmental data up to end of April
DAJ-Env	Environmental data up to end of June
DAJJ-Env	Environmental data up to end of July

### 3.3.5 Accuracy Assessment

The performance of the models built from each layer combination were compared using the coefficient of determination ( $R^2$ ) and the root mean squared error (RMSE, eq. 6).

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (E_i - O_i)^2}{n}} \quad [6]$$

Where  $O$  represents the observations in the test data sets,  $E$  the estimated yield, and  $n$  is the number of samples. These accuracy measures (RMSE &  $R^2$ ) were calculated using two different datasets: (i) ten-fold cross-validation (using the training data) and (ii) an independent validation dataset not used to train the RF models. In 10-fold cross-validation, the data is divided into 10 nearly equally sized subsets. Ten iterations of training and validation are performed such that within each iteration a different subset of the data is withheld for validation, while the remaining 9 subsets are used to train the model. The RMSE and  $R^2$  values for each iteration are then averaged to provide an overall estimate of model accuracy (Refaeilzadeh *et al.*, 2009). The standard deviation in accuracy measures over the ten iterations were used to produce error bars to aid comparison of models. The accuracy measures were calculated for the cross-validation and independent validation datasets to ensure that the models were not overfitting the training data. Model accuracy was considered to be dependably different if accuracy error bars did not overlap.

### 3.3.6 Establishing a Baseline

To set this study within the wider context of yield estimation methodologies, a baseline was established against which to compare the models created. As yield has often been estimated using simple (linear) regression applied to a variety of VIs, this method was used to provide the baseline. Linear and Random Forest (RF) regression were applied to a variety of single-date VIs derived from the available Sentinel-2 imagery. As well as using single-date VIs, previous studies have also used multi-date VI data accumulated throughout the growing season. The variation in accuracy with accumulation of VI data was therefore assessed, using RF regression and the NDVI as an example.

### 3.4 Results

#### 3.4.1 Baseline Data

From the baseline data analysis, linear regression produced RMSE values between 1.68 to 2.00 tonnes/ha ( $R^2$  0.01 to 0.29), while RMSE values from RF ranged from 1.54 and 2.01 tonnes/ha ( $R^2$  0.12 to 0.44) (table 3.5). Of the combinations of month and VI assessed the NDVI and WDRVI for July offered the highest accuracy (RMSE 1.54 tonnes/ha). Compared to this baseline, all further models created in this study displayed improved yield estimation accuracy (table 3.7; figure 3.6). The baseline results also suggest that the accuracy of yield estimation improves throughout the growing season, with reductions in RMSE as NDVI data accumulates from December to July (table 3.6).

*Table 3.5: RMSE and R-squared values calculated from the validation dataset for linear and Random Forest regressions using vegetation indices calculated for each month.*

Month	VI	Linear Regression		Random Forest Regression	
		RMSE	RSQ	RMSE	RSQ
December	GCVI	1.86	0.12	1.87	0.20
	GNDVI	1.87	0.12	1.90	0.19
	NDVI	1.87	0.12	1.87	0.20
	SR	1.82	0.16	1.85	0.21
	WDRVI	1.84	0.14	1.86	0.21
April	GCVI	2.00	0.01	1.93	0.18
	GNDVI	1.99	0.02	1.90	0.19
	NDVI	1.97	0.04	2.01	0.12
	SR	1.99	0.03	2.01	0.12
	WDRVI	1.98	0.03	2.01	0.13
June	GCVI	1.68	0.28	1.82	0.24
	GNDVI	1.70	0.27	1.79	0.25
	NDVI	1.79	0.19	1.91	0.15
	SR	1.78	0.20	1.98	0.13
	WDRVI	1.79	0.19	1.96	0.13
July	GCVI	1.74	0.25	1.59	0.41
	GNDVI	1.70	0.28	1.59	0.41
	NDVI	1.69	0.29	1.54	0.44
	SR	1.78	0.22	1.55	0.44
	WDRVI	1.71	0.28	1.54	0.44

*Table 3.6: RMSE and R-squared values calculated from the validation dataset for Random Forest regressions using NDVI data accumulated over the growing season.*

<b>NDVI</b>	<b>RMSE</b>	<b>RSQ</b>
December	1.86	0.23
December + April	1.37	0.54
December + April + June	1.24	0.62
December + April + June + July	0.96	0.77

### **3.4.2 Random Forest Model Comparison**

Validation of the RF models was conducted in two ways, using the 10-fold cross-validation from RF (using the training data) and also in a separate validation using a small data set that was not used for training. In general, the validation RMSEs fall within the error bars for the training RMSEs (table 3.7; figure 3.6). This suggests the accuracy reported using the training data is relatively reliable and RF is not overfitting the data. Where this is not the case (S2\_20, S2\_SWI, DA-Env, DAJ-Env), the validation RMSE is only 0.01 tonnes/ha outside the error bar, suggesting only minimal discrepancy. This difference may be due to the relatively small size of the validation dataset.

Table 3.7: Results of Random Forest analysis.

Combination	RMSE (training data – 10-fold cross validation)	RMSE (validation data)	R <sup>2</sup> (training data – 10-fold cross validation )	R <sup>2</sup> (validation data)
<i>S2 (DAJJ)</i>	0.64	0.66	0.90	0.89
<i>S2_20</i>	0.78	0.70	0.85	0.88
<i>S2_VI</i>	0.64	0.66	0.90	0.89
<i>VI</i>	0.88	0.87	0.81	0.81
<i>S2_Met</i>	0.63	0.65	0.90	0.89
<i>S2_SWI</i>	0.58	0.62	0.91	0.91
<i>S2_Topo</i>	0.60	0.63	0.91	0.90
<i>(DAJJ-) S2_Env</i>	0.59	0.61	0.92	0.91
<i>D</i>	1.01	1.01	0.74	0.74
<i>A</i>	0.94	0.96	0.78	0.77
<i>Jn</i>	0.88	0.88	0.80	0.81
<i>J</i>	0.89	0.90	0.80	0.80
<i>DA</i>	0.78	0.78	0.85	0.85
<i>DAJ</i>	0.70	0.69	0.88	0.88
<i>D-S2_Env</i>	0.64	0.67	0.89	0.89
<i>DA-S2_Env</i>	0.60	0.63	0.91	0.90
<i>DAJ-S2_Env</i>	0.60	0.62	0.91	0.91
<i>D-Env</i>	0.69	0.71	0.88	0.87
<i>DA-Env</i>	0.66	0.69	0.89	0.88
<i>DAJ-Env</i>	0.65	0.69	0.89	0.88
<i>DAJJ-Env</i>	0.67	0.69	0.89	0.88

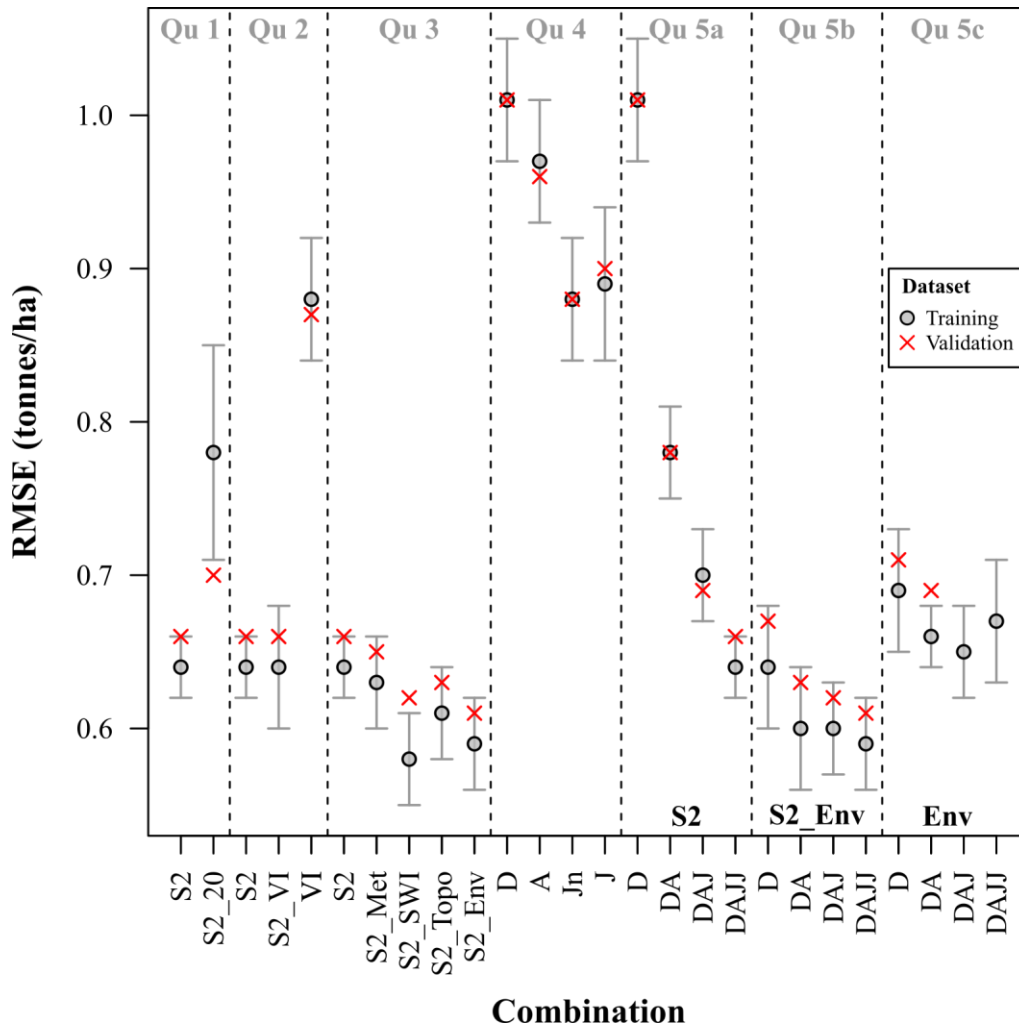


Figure 3.6: Ten-fold RMSE values from Random Forest analysis calculated using the training dataset and RMSE values from the validation dataset. Error bars produced using the standard deviation in ten-fold RMSE iterations. Specific data values can be found in table 3.7. For question 5, S2 is the Sentinel-2 only data, S2\_Env is the Sentinel-2 and environmental datasets, whilst Env is just the environmental data sets (see table 3.4 for more details).

This study centred on 5 key questions designed to investigate how inclusion of different datasets affects the accuracy of yield estimation. The results of the RF analysis are outlined in the following sections.

**Question 1:** How does resampling the spatial resolution of Sentinel-2 data affect the accuracy of yield estimation?

As Sentinel-2 has bands with differing resolutions (10m, 20m), the data will typically be resampled to either 10m or 20m for analysis. Comparison of RF using 10m (*S2*) and 20m resolution (*S2\_20*) Sentinel-2 data demonstrates that yield estimation is more accurate for the 10m model (figure 3.6).

**Question 2:** Does calculating separate VIs contribute any extra information to the estimation model?

The RMSE is very similar between the *S2* and *S2\_VI* models, although the uncertainty increases for *S2\_VI*, while using the VI data on its own produces lower accuracy (figure 3.6). This shows that adding VIs to the basic Sentinel-2 data does not improve the accuracy of yield estimation.

**Question 3:** How do different combinations of Sentinel-2 data and environmental data affect estimation accuracy?

The model results demonstrate that yield estimation can be improved by the introduction of environmental data to the Sentinel-2-based RF model. However, the results differ depending on the type of data added, i.e. meteorological, topographical, soil moisture or a combination of all three. Compared to the *S2* combination, *S2\_SWI* and *S2\_Env* produce higher accuracy estimations, while *S2\_Met* and *S2\_Topo* do not offer any definite improvement (figure 3.6). This suggests that adding either soil moisture data or a combination of all available environmental data to Sentinel-2 data can improve yield estimations.

**Question 4:** Which single-date Sentinel-2 image provides the most accurate estimation?

The availability of spectral data varies between years and locations. In places particularly prone to cloud cover, such as the UK, only 1 or 2 cloud-free images may be available over the growing season. How the accuracy of yield estimation from single-date images varies throughout the year is therefore an important question. Comparison of the available Sentinel-2 images demonstrates that estimation accuracy increases substantially from December to June (figure 3.6). From June onwards however there is no clear difference in accuracy.

**Question 5:** How does estimation accuracy vary with accumulation of data throughout the growing season for Sentinel-2 data only (Qu 5a), Sentinel-2 and environmental data combined (Qu 5b), and environmental data only (Qu 5c)?

***5a: Sentinel-2 data***

The accumulation of Sentinel-2 data over the year improves estimation accuracy throughout the growing season. Clear decreases in RMSE are observed as successive Sentinel-2 images are added to the estimation model (figure 3.6). The biggest improvement occurs from December to April.

***5b: Sentinel-2 plus environmental data***

The addition of environmental data to Sentinel-2 data improves estimation accuracy across all date combinations compared to the Sentinel-2 only combinations (Qu 5a) (figure 3.6). Combining Sentinel-2 data and environmental data from December alone (*D-S2\_Env*) provides similar accuracy to the full Sentinel-2 data set combined (*DAJJ S2*). RMSE does not vary substantially as successive data are added to the *S2\_Env* combinations. This suggests little improvement with accumulation of data over the growing season.



### **5c: Environmental data**

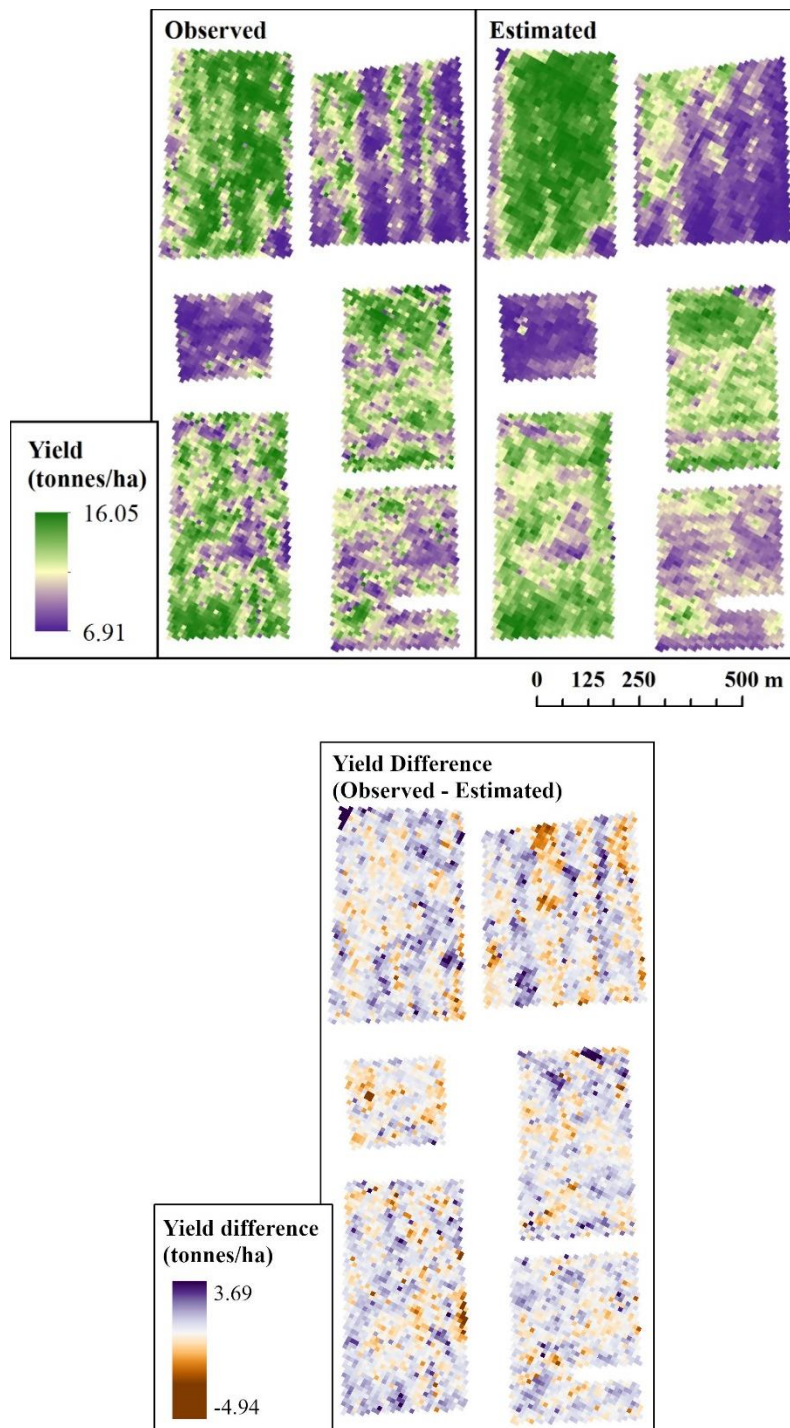
Environmental data for December alone provides a yield estimation accuracy comparable to the *DAJ* Sentinel-2 data combination (Qu 5a) (figure 3.6). Accumulation of environmental data over the growing season has little impact on estimation accuracy.

The environmental data contains two types of data: those which are static over the growing season (topography), and those which are dynamic (precipitation, temperature, SWI). Considering these separately, the topographic data appear contribute more to the estimation accuracy (RMSE  $1.18 \pm 0.05$  tonnes/ha) than the other environmental variables (RMSE  $1.32-1.34 \pm 0.02-0.05$  tonnes/ha depending on temporal coverage). However, the topographic data alone does not match the high accuracy achieved when the two types of environmental data are combined (regardless of temporal coverage).

In general, most of the combinations containing only environmental data provide less accurate estimates than having a combination of Sentinel-2 data and environmental data.

### **3.4.3 Mapping Within-field Wheat Yield Variability**

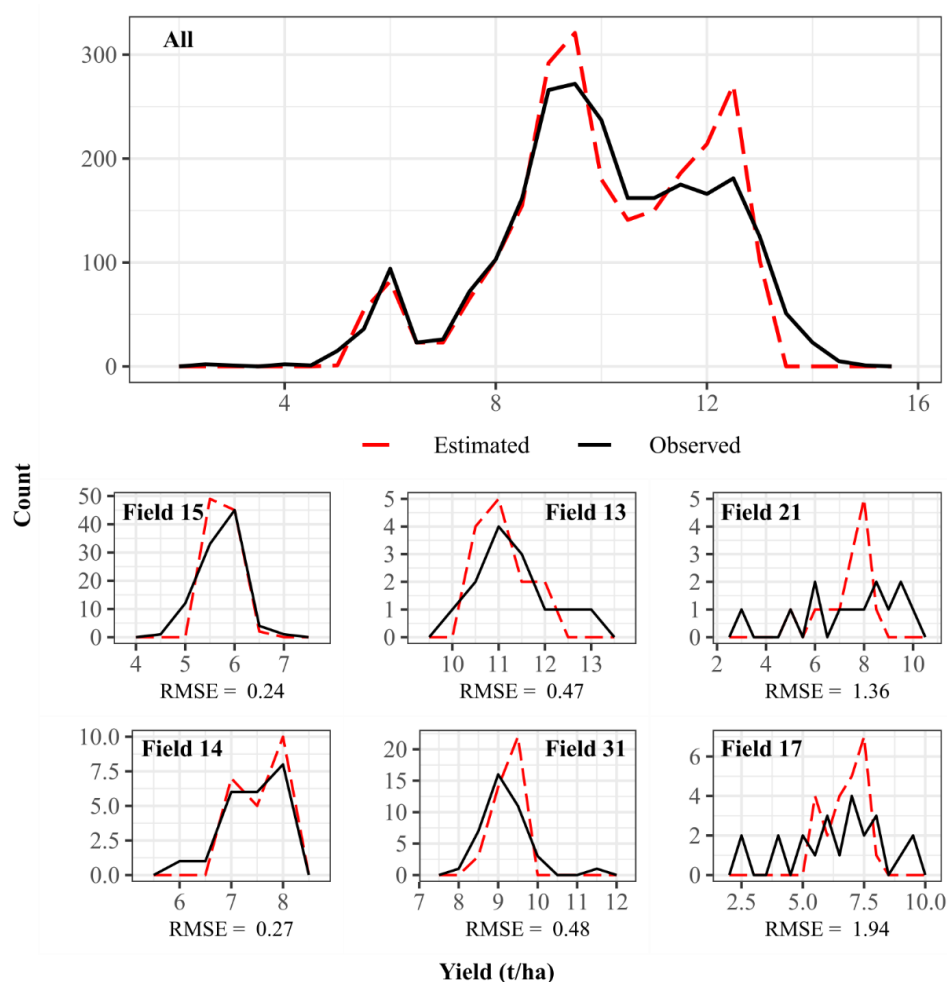
The results from the 5 questions demonstrate that within-field yield variability can be estimated relatively accurately, with an RMSE between 0.61 and 1.01 tonnes/ha, depending on the data combination. This accuracy is reflected when comparing the observed and estimated yields, which show that the estimated yields reflect the general patterns of yield variability within individual fields (figure 3.7). The difference map (figure 3.7) shows no clear pattern of over- or under-estimation of yield values, suggesting there are no systematic spatial errors in the estimated yield values.



*Figure 3.7: Observed yield interpolated from the combine harvester data (top left), estimated yield from the S2\_Env RF model (top right) for a selection of fields within the training area, and the difference between the observed and estimated yield (bottom).*

Comparing frequency distributions of observed and estimated yield for each field suggests that the ability of the RF models to detect within-field variability varies between fields (figure 3.8

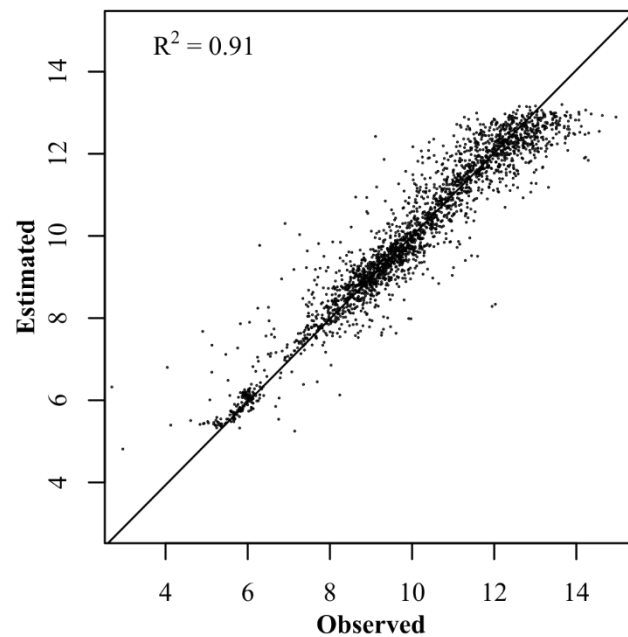
shows the frequency distributions for the best RF model: *S2\_Env*). The shape of the yield distribution varies between fields, with some exhibiting simple unimodal distributions (e.g. field 15 (figure 3.8)) and others more complex bimodal distributions (e.g. field 21 (figure 3.8)). Comparing the two distributions for both individual fields and all fields combined there appears to be a tendency for overestimation of the frequency of modal values, and underestimation of the highest and lowest values. Despite these tendencies, the model appears to provide relatively accurate estimates of within-field yield variability for individual fields with RMSE values between 0.24 and 1.94 tonnes/ha (table 3.8). Additionally, the regression graph confirms the trends shown in the frequency distributions (figure 3.9).



*Figure 3.8: Frequency distributions for observed and estimated yields using the validation data set for the *S2\_Env* model for all fields and a sample of individual fields. Individual fields chosen were those with the two highest (fields 15 and 14), two middle (fields 13 and 31), and two lowest (fields 21 and 17) RMSE values to provide a representative selection.*

*Table 3.8: RMSE values for individual fields using the validation data set for the S2\_Env model. NB: this model was run using data from 34 fields, rather than the full 39, due to missing data values for some satellite images.*

Field number	RMSE	Field number	RMSE	Field number	RMSE
1	0.45	13	0.47	25	0.56
2	0.37	14	0.27	26	0.72
3	0.61	15	0.24	27	0.59
4	0.7	16	0.43	28	0.65
5	0.4	17	1.94	29	0.29
6	0.29	18	0.45	30	0.63
7	0.32	19	0.58	31	0.48
8	0.61	20	0.49	32	0.3
9	0.41	21	1.36	33	0.46
10	0.47	22	0.87	34	0.37
11	0.28	23	0.77		
12	0.89	24	0.79		



*Figure 3.9: Linear regression between observed and estimated yield for the validation data set from the S2\_Env model.*

#### **3.4.4 Mapping Within-field Wheat Yield Variation at Landscape-scale**

Satellite data enables scaling-up of yield estimation across the wider landscape area using data from a few fields. To demonstrate this potential, the S2\_Env RF model was used to estimate yield for the area covered by the Sentinel-2 image (figure 3.10 shows a portion of this map). Fields

containing wheat were identified using the 2016 *Land Cover Plus®: Crops* map. To remove mixed boundary pixels from the dataset, field boundaries in the crop map were buffered in by 20m. In this study, the yields estimated in all fields across the entire area fell within the range of values in the training data, increasing the likelihood of the yield estimations being accurate. Additionally, the fields from which the training and validation data were obtained mark the southwest and northeast extent of the landscape-scale map. Extrapolation outside the input data range, both spatially and in terms of yield values, would be less reliable.

High resolution yield maps make it is possible to look at within-field and between-field yield differences, and identify wider landscape patterns. For example, in the area covered in this study yield ranges from 4.09 to 12.22 tonnes/ha, with a mean value of 9.02 tonnes/ha (mean per field 5.83 to 11.21 tonnes/ha) and a total yield production of approx. 289000 tonnes. Using such maps it is possible, among other things, to identify clusters of higher or lower yielding fields within the same climate region. For example, in figure 3.10 there is a cluster of higher yielding fields in the northwest corner of the map and a cluster of lower yielding fields in the east of the image. Knowledge of such clusters facilitates further investigation into the causes of yield variation within the landscape, such as differences in crop management practices and environmental conditions. Furthermore, using information on yield variability it is possible to identify different management zones and yield-limiting factors to improve the efficiency of farming practices in different areas (Diker *et al.*, 2004).

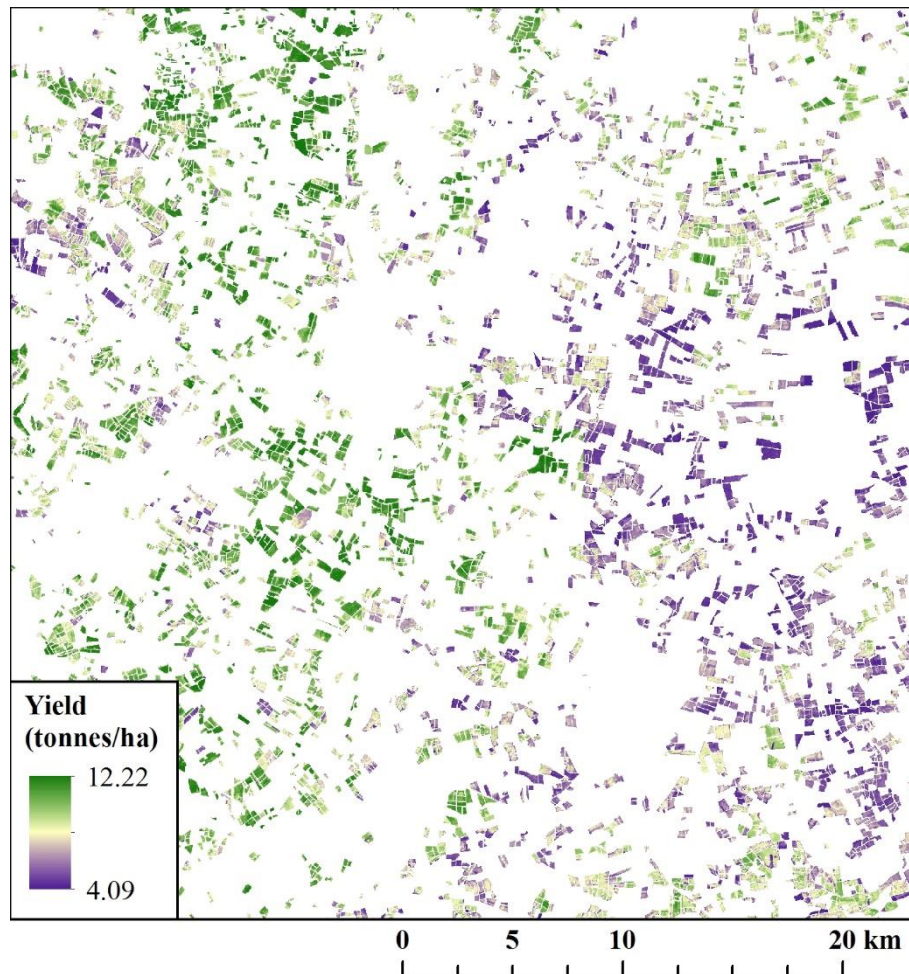


Figure 3.10: Landscape-scale wheat yield estimation based on S2\_Env RF model.

### **3.5 Discussion**

#### **3.5.1 Benefits of Random Forest**

All the multi-variable RF regression models developed in this study outperformed the single-date VI-based linear regression and RF models used as a baseline. This demonstrates the superior ability of RF and multi-variable models in general. While RF is now widely used for image classification, its use for yield estimation is not so common with studies generally relying on traditional regression models. However, RF has a number of key advantages over traditional regression models for yield estimation, some of which are demonstrated by the results of this study. Firstly, using RF may increase the amount of data available for training. RF randomly selects a subset of the calibration dataset that it reserves for assessing model accuracy rather than model training

(Jeong *et al.*, 2016). In this study, the additional step was taken of also splitting the data into training and validation datasets outside of RF to provide a means of checking whether the model was overfitting the data. The results suggest overfitting was not an issue in this study. If holding back some data for validation is less important for RF than for traditional regression models, this would increase the volume of data available to train the model, which will likely improve its estimative capability.

Secondly, it appears RF is able to utilise relationships between explanatory variables to control for confounding factors. Of the data combinations explored in this study, the integration of environmental data with Sentinel-2 data provided the most accurate yield estimation. Environmental data has been used alongside satellite data to support crop yield estimation in numerous studies, commonly through the use of crop simulation models (Azzari *et al.*, 2017; Doraiswamy *et al.*, 2005; Jin *et al.*, 2017b; Lobell *et al.*, 2015; Moriondo *et al.*, 2007). Despite the clear advantages of including environmental data such as the SWI in the RF model, linear regression reveals no obvious relationship between SWI and crop yield ( $R^2$  of 0.004-0.11 depending on the month). It therefore appears that the improvement in accuracy arises not from a direct relationship between soil moisture and yield, but from an underlying relationship between SWI and spectral reflectance. It may be that the inclusion of SWI data enables RF to control for the impact on spectral reflectance of soil moisture variability between Sentinel-2 images. RF appears to be able to identify and unpick relationships between explanatory variables and to use these to account for confounding factors, which could reduce accuracy. The ability of RF to cope with multi-variate relationships between data of different types and resolutions is a key advantage over methods such as linear regression, which can only address uni-variate relationships.

Further to this, the apparent ability of RF to detect underlying relationships can also reduce the number of explanatory variables required to provide an accurate estimation. Previous studies have commonly utilised a variety of VIs to estimate yield by inferring relationships between VIs and

yield (Liaqat *et al.*, 2017; Lopresti *et al.*, 2015; Ren *et al.*, 2008), or to derive relationships with surface parameters such as LAI and fAPAR, which can be used to estimate yield (Boschetti *et al.*, 2014; Nigam *et al.*, 2017). In this study, using VIs and the original Sentinel-2 data together provided no improvement in accuracy. This may indicate that RF is able to infer the relevant information for yield estimation normally provided by VIs from the individual Sentinel-2 bands themselves. Whether this is the case or not, the fact that RF does not require separate VIs could have significant benefits. By removing the need to calculate separate indices, RF may simplify processing and reduce processing time.

### **3.5.2 Optimum Processing Resolution**

This study demonstrates that Sentinel-2 data has the potential to provide relatively accurate estimates of within-field yield variability in the UK. In this study, yield estimation is more accurate at 10m resolution than 20m resolution. Conversely, Yang *et al.* (2009) found accuracy increased as resolution decreased; SPOT 5 pixels rescaled to 30m resolution explained 15% more of the yield variability than the original 10m pixels. The reason for this disagreement may be found in the nature of the different datasets used in each study. Pre-rectification, SPOT 5 images have a locational accuracy of 30m (Yang *et al.*, 2009), while Sentinel-2 images have a locational accuracy of 20m (Drusch *et al.*, 2012). Such differences in spatial precision could partly account for the discrepancy in the image resolution-yield accuracy relationship seen in these two studies.

In addition, the accuracy of the yield data used within different studies will vary as data will be collected at different times, for different crops and using different yield monitors and combine harvesters. Yield monitors are susceptible to a number of potential errors including time delays, calibration errors and combine operational errors (Grisso *et al.*, 2002). The exact yield monitor used and the way in which these errors are assessed and adjusted for will affect the final accuracy of the yield data. While various studies have been conducted into the different options for data correction (Lyle *et al.*, 2014), there is currently no universally accepted procedure. It is therefore likely that the



corrections applied and the thresholds used will differ between studies, affecting the relative accuracy of the training data.

Our findings showed higher yield estimation accuracy at 10m than 20m. This may be because advances in satellite sensor design and data processing, alongside improved processing methods for combine harvester data, provide higher quality image data and reference data that enable accurate yield estimation at high resolution. This suggests that it is important to optimise the resolution and the match between the satellite data and the reference data. Testing a number of different resolutions may be the best way of identifying the optimum resolution, as it may not be obvious from the point density and resolution of the satellite data.

The high frequency of cloud cover within the UK restricts the number of optical satellite images available for crop yield estimation (Armitage *et al.*, 2013). Satellites with a lower spatial resolution and higher temporal resolution, such as MODIS, have the potential to provide a greater number of cloud-free images throughout the growing season. The availability of more cloud-free images would allow crop growth dynamics to be tracked more accurately over the growing season. This might allow more generic solutions for using satellite data to estimate within-field yield variability. However, the typically small field-sizes (approx. 2ha to 175ha for wheat) and high within-field variability within the UK mean that using lower resolution images would not be suitable, with large numbers of mixed pixels being produced. Assessment of within-field variability within the UK therefore requires satellite data with a higher spatial resolution, even if it means allowances have to be made for image frequency and availability of cloud-free images.

While this study uses Sentinel-2 data, it is important to remember that higher resolution data is available from various commercial sources (e.g. RapidEye, Planet Labs). Such higher resolution data could potentially allow more detailed assessment of within-field variability. However, previous work highlights the limits to the spatial precision of the combine harvester data, because of the way the sensors and combine harvesters work (Lyle *et al.*, 2014). The yield spatial

resolution and precision is system dependent, as it is a function of the monitoring equipment, the cutting head and the software. For example Lyle *et al.* (2014) found a spatial resolution of about 20-25m appropriate for the system they investigated. This suggests that the key constraints on the highest spatial resolution that yield can be mapped and validated at may be determined by the combine systems rather than the satellite data. As such, whether there is any benefit to using higher resolution commercial satellite data for the spatial resolution it offers will depend on the exact nature of the sensor used. There may, however, be a benefit from the high repeat frequency that could capture key periods of the growing season, even if the data cannot be used to estimate yield at higher resolutions than Sentinel-2. Since the precision and spatial 'footprint' of yield monitor data is determined largely by header width, future advances may be driven by research purposes that require more spatially precise data, through for example, use of plot combine harvesters with smaller header widths than commercial combine harvesters (Marchant *et al.*, 2019). However, similar advances are unlikely for commercial yield monitors due to the impact smaller header widths would have on harvesting times and efficiency.

Despite the difference in spatial resolution between the Sentinel-2 data (10m) and the temperature and precipitation data (5km), the results suggest that inclusion of these environmental variables did in fact increase the accuracy of the results. This is likely due to the fact that the 39 fields used for training the RF models were widely dispersed over the Oxfordshire and Lincolnshire study areas. This meant that data from 54 of the 5km squares was used to build the RF model, despite the relatively small area covered by the fields themselves (476 ha), allowing some variation across the study area to be detected. It is likely that the inclusion of higher resolution data would increase the accuracy further by allowing better detection on finer scale variations in temperature and precipitation across the study area. Future work could look at methods for downscaling the data to make it more suitable for field-scale yield assessment.

### 3.5.3 Variability in Accuracy Through the Season

The accuracy of yield estimation based on single-date Sentinel-2 images generally improves throughout the growing season. The biggest improvement occurs between the December and April images, with a further, smaller increase by June. There are a few possible explanations for this. Firstly, the signal-to-noise ratio will vary throughout the growing season, with differences in sun angle and incoming radiation intensity, which will affect the estimation accuracy. Secondly, towards the beginning of the growing season (e.g. December) the canopy may not have developed enough to give a good characterisation of the spatial variability in growth. Later in the growing season (e.g. April), the canopy will be more fully developed allowing more accurate detection of spatial variability. Visual interpretation of the Sentinel-2 images (figure 3.11) suggests the lack of improvement from June to July may be due to the crops ripening, or beginning to ripen, over this period. This will likely affect the accuracy of yield estimation from Sentinel-2 data alone.



Figure 3.11: Evidence of crops ripening between successive Sentinel-2 images for June (left) and July (right).

### 3.5.4 Future Developments

Future work should explore the contribution Sentinel-2 can make to crop models used to estimate yield. Crop models are widely used to estimate and predict crop yields and are known to provide relatively accurate results for specific crops. Previous crop model studies have commonly relied on freely available data from satellites such as Landsat (Lobell *et al.*, 2015; Xie *et al.*, 2017), MODIS (Doraiswamy *et al.*, 2005; Ines *et al.*, 2013) and AVHRR (Moriondo *et al.*, 2007). The low to moderate resolution of such data has limited the ability to assess within-field yield variability, with yield estimation studies mostly focusing on farm- (Sehgal *et al.*, 2005), regional- (Huang *et al.*, 2015; Padilla *et al.*, 2012) and county-scales (Ju *et al.*, 2010). The ability to detect within-field yield variability using Sentinel-2 demonstrated by this study suggests future work should explore the benefit of incorporating Sentinel-2 data into current crop models. Battude *et al.* (2016) demonstrated the theoretical potential using SPOT4-Take5 data, which was designed to simulate the spatial and temporal sampling of Sentinel-2, within the Simple Algorithm For Yield (SAFY) crop model to estimate maize yields. Further work is needed to ascertain whether this potential can be realised with actual Sentinel-2 data, and whether this translates to other crop models.

Additionally, an exploration of the key Sentinel-2 bands for yield estimation could prove useful. Knowledge of which bands are most valuable for estimating yield could allow models to be streamlined, removing the bands which contribute the least to yield estimation. Such work would require consideration of study sites in a variety of countries with a range of environmental conditions to ensure that any patterns of band importance apply generally and are not limited to specific sites. Building on this, future work could also compare the ability of Landsat and Sentinel wavebands to estimate yield. Such a comparison could provide valuable information on the requirements of satellite sensors for yield estimation, and, for example, whether the inclusion of the Sentinel-2 vegetation red edge bands contributes any useful information. Understanding band importance for different applications is valuable for the remote sensing community as it can inform the development of future satellites.

In this study, no attempt was made to extrapolate beyond the available data, so yield estimation was constrained by three factors: firstly, by the upper and lower limits of the yield data, with all estimated yield values falling within the range of the training dataset; secondly, by the geographical location of the study sites, which marked the north-eastern and south-western-most extent of the landscape-scale yield estimations; finally, by focussing on wheat fields only. Future work should test the transferability of the method used in this study (figure 3.2) to other areas, environmental conditions and crop types.

### **3.6 Conclusion**

This study demonstrates that Sentinel-2 data is capable of providing relatively accurate estimates of within-field yield variability (RMSE 0.66 tonnes/ha) when combine harvester data are available to calibrate against. Combining Sentinel-2 with environmental data provides more accurate estimates than using Sentinel-2 data or environmental data individually (RMSE 0.61 tonnes/ha). Furthermore, RF appears to provide higher yield estimation accuracy than commonly used simple VI-based linear regression. This study has also proposed a method that can be adapted to other crops and locations, when suitable training data are available. The method is applied to estimate yield at the landscape scale and produce a landscape-level estimate of crop yield.

#### **4. Satellite-derived Environmental Heterogeneity and Productivity as Indicators of Bird Diversity**

Merryn L. Hunt, George Alan Blackburn, Gavin Siriwardena, Luis Carrasco & Clare S. Rowland

##### **Abstract**

Birds are useful indicators of general biodiversity, which, despite targets to reduce its loss, has continued to decline globally. Understanding the spatial drivers which affect species diversity patterns at different scales is essential to predict how species may respond to future environmental changes and inform the creation of effective conservation strategies. To achieve this, comprehensive information is required regarding species distribution and changes over time. However, current bird population indicators are typically aggregated at country or regional levels making it difficult to analyse the data spatially. An opportunity exists to better explore the spatial drivers of bird diversity by developing methods using satellite data to model the diversity data to produce higher resolution maps of bird diversity. This would provide greater spatial detail to aid efforts to reduce or halt biodiversity decline.

This paper explores methods to model bird diversity distribution across Great Britain using satellite data. Random Forest regression, trained using Countryside Survey 2000 data, was used to explore the extent to which a combination of satellite-derived measures of habitat heterogeneity and habitat productivity could explain the variation of bird diversity across Great Britain. This study focused on farmland and woodland birds, grouped according to the BTO/JNCC/RSPB wild bird indicator designations for the UK and England. Refined RF models were produced using variable selection techniques to reduce the number of variables. Feature contribution analysis was also performed to explore the nature of the relationships between the response and estimator variables in the refined RF models. The results of this study demonstrate that it is possible to estimate farmland and woodland bird diversity relatively accurately ( $R^2$  0.64 to 0.72) using just a few satellite-

derived measures of habitat productivity and heterogeneity. The variable selection and feature contribution analysis highlight a number of important spatial drivers of species richness/diversity including arable land area for farmland birds, woodland patch edge length for woodland birds, and high productivity grassland during spring for both groups.

## **4.1 Introduction**

Despite targets to reduce its loss, global biodiversity has continued to decline, with no significant reductions in rate, as pressures on biodiversity have increased (Butchart *et al.*, 2010). Loss of species diversity has been linked to large effects on primary productivity and decomposition, with a knock-on effect on the provision of essential ecosystem services (Hooper *et al.*, 2012). An understanding of the spatial drivers determining species richness and diversity patterns at different spatial scales is essential to predict how species may respond to future environmental changes and inform the creation of effective conservation strategies. To achieve this, comprehensive information is required regarding species distribution and changes over time. However, field assessment may not be practical or feasible for large areas (Heywood, 1995; Link and Sauer, 1997), as techniques are often difficult, time-consuming and expensive. One means by which this issue can be overcome is through the integration of satellite data into biodiversity monitoring schemes.

Satellite images provide a readily accessible, global dataset at various spatial and temporal resolutions from which indicators of species diversity may be derived (e.g. Kerr and Ostrovsky, 2003; Nagendra, 2001; Turner *et al.*, 2003). Previous studies have mapped bird diversity – an important indicator of global biodiversity patterns (Furness and Greenwood, 2013) – using satellite-derived measures of two key factors affecting species diversity: (1) spatial heterogeneity, including measures of habitat structure, composition and connectivity (Carrasco *et al.*, 2018; Coops *et al.*, 2009b; Griffiths and Lee, 2000; Luoto *et al.*, 2004); (2) environmental productivity, measured using fraction of absorbed photosynthetically active radiation (fAPAR) (Coops *et al.*, 2009a), gross and net primary productivity (GPP/NPP) (Phillips *et al.*, 2010, 2008) and normalised difference vegetation index (NDVI) (Duro *et al.*, 2014; Foody, 2005; Seto *et al.*, 2004). Studies tend to focus on either spatial heterogeneity or environmental productivity measures, although there are examples of these measures being integrated to map species diversity in places including North America (e.g. Hurlbert and Haskell, 2003) and Canada (e.g. Coops *et al.*, 2009b).



While Landsat data has been around for many years, coarser resolution MODIS data has typically been the favoured source for satellite-derived measures of environmental productivity to estimate bird diversity on a national scale (e.g. Bonthoux *et al.*, 2018; Coops *et al.*, 2009b; Hobi *et al.*, 2017). Use of higher resolution data would allow these measures to be derived at scales more suitable for detecting smaller-scale, within-habitat variations, of environmental productivity thus improving the accuracy with which bird diversity can be mapped. Historically, however, use of such data has been hindered by the processing power and time required for analysis on a national scale. Now, with the availability of machine learning and cloud computing platforms such as Google Earth Engine (Gorelick *et al.*, 2017), this analysis is possible. Recent years have also seen a focus on improving the quality and consistency of satellite data sets for automated analysis, with improved georeferencing and cloud-masking (Roy *et al.*, 2014). These developments mean that it is now timely to assess the spatial drivers of bird richness/diversity at a higher spatial resolution of 30m. While this does not alter the actual resolution at which bird diversity itself can be mapped, something that is determined by the typically coarse (e.g. >1km) resolution of the field survey-based training data, it will allow the determining factors of bird richness to be better represented, thus improving map accuracies.

Birds are useful indicators of general biodiversity and ecosystem health (Furness and Greenwood, 2013), and preserving and enhancing bird diversity is important (Whelan *et al.*, 2008). However, increases in agricultural intensity and land use conversion over the decades has led to severe declines in many farmland bird populations across much of Europe (Chamberlain *et al.*, 2000; P. F. Donald *et al.*, 2001). This study focusses on mapping bird species diversity within Great Britain. Reports based on the Wild Birds Population Indicators (DEFRA, 2018), derived from British Breeding Bird Survey (BBS) data, suggest that farmland and woodland bird populations decreased by 56% and 24% respectively between 1970 and 2016 (DEFRA, 2018). The Wild Birds Population Indicators, and indices like them, have been critical in quantifying declines and identifying causal factors. However, they are typically aggregated at country or regional levels making it difficult to analyse the data

spatially. An opportunity exists to better explore the spatial drivers of bird diversity declines, and potentially increases in bird diversity, by developing methods to model the bird diversity data to produce higher resolution maps of bird diversity. This would provide greater spatial detail to aid efforts to reduce or halt this decline.

Therefore, in this paper we explore methods to model bird diversity distributions across GB using satellite data. To do this, Random Forest regression, trained using Countryside Survey 2000 data, was used to explore the extent to which a combination of satellite-derived measures of habitat heterogeneity and habitat productivity could explain the variation of bird diversity across GB. Variable selection techniques were used to reduce the number of variables and produce a set of refined RF models. These models were then used to produce national scale estimative maps of farmland and woodland bird diversity. To guide conservation schemes, it is also important to have an understanding of the key spatial drivers of species diversity. Finally, therefore, feature contribution analysis was used to explore the nature of the relationships between the response and estimator variables in the refined RF models. While this study focuses on mapping bird diversity in GB, the approach developed could theoretically be applied in any country where similar bird diversity data exists.

#### **4.1.1 Bird Diversity Monitoring in GB**

Bird species richness in GB has previously been estimated using measures of habitat heterogeneity derived from the Land Cover Map 2000 (LCM2000; Fuller *et al.*, 2002) and training data from bird counts in the Countryside Survey 2000 (CS2000; Wilson and Fuller, 2002). Rhodes *et al.* (2015) demonstrated that, while they cannot match the accuracy provided by field surveys (i.e. CS2000), satellite-derived measures of broad habitat can provide a reasonable estimation of bird richness for a variety of species. The results suggest that the spatial and temporal coverage offered by increasingly freely available satellite data could allow a more cost effective and practical approach for mapping bird richness on a national and global scale. Carrasco *et al.* (2018) explored

the relationships between bird richness and spatial environmental heterogeneity variables, using LCM2000 to scale-up field survey data to provide a map of bird species richness for Great Britain. The current study builds on this work to incorporate not only satellite-derived measures of habitat heterogeneity, but also habitat productivity. It is hoped that inclusion of productivity measures will provide a more nuanced picture of bird diversity patterns, allowing variations in factors such as management practices and climate to be detected, rather than simply landscape structure.

## **4.2 Data and Method**

Figure 4.1 provides an overview of the method used in this study, outlining how the bird count data and satellite data were processed and combined to estimate bird diversity distributions across GB.

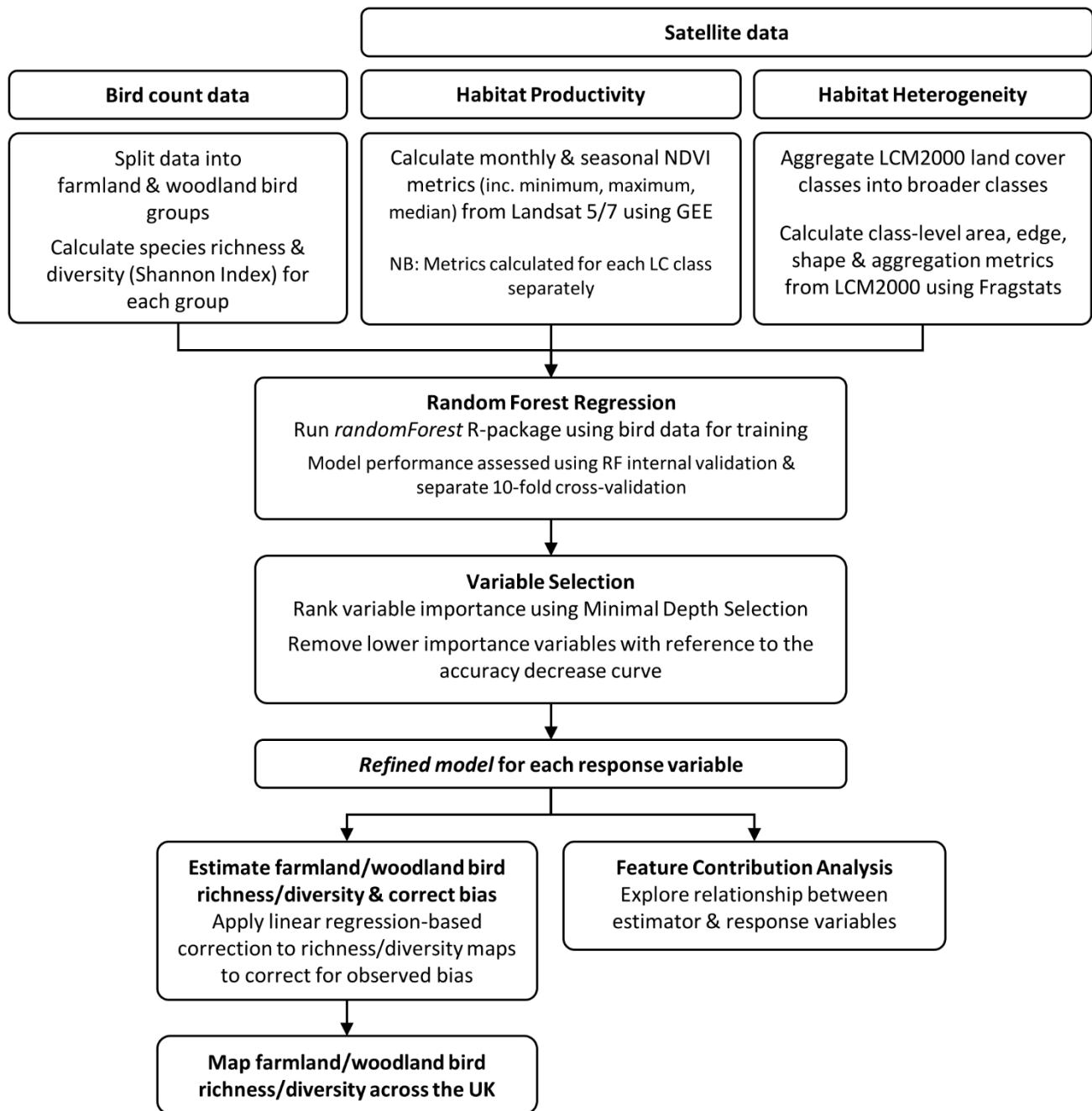


Figure 4.1: Overview of the method used to estimate bird diversity distributions across GB.

#### 4.2.1 Bird Data

Bird count data were collected between April and June of 2000 in 335 Countryside Survey 2000 (CS2000) squares (Wilson and Fuller, 2002). These squares are part of a randomly stratified sample of 569 nationally representative 1km squares of rural GB used to compile detailed

information on the landscape. Bird counts were recorded on two separate visits during the early and late breeding season, using up to 4km of line transect counts within each square. The surveying methodology followed that of the BTO/JNCC/RSPB Breeding Bird Survey (Harris *et al.*, 2019). In this study, species richness and the Shannon diversity index were used as measures of farmland and woodland bird species diversity. The designation of farmland and woodland species was based on the groupings used by the BTO/JNCC/RSPB wild bird population indicators for the UK and England (Eaton and Noble, 2018); these groupings can be seen in table A8 in Appendix 4. Farmland and woodland bird richness were calculated for all squares by counting all the different species within these groups recorded along any of the four transects.

#### 4.2.2 Habitat Heterogeneity Variables

While there are many measures of environmental heterogeneity (Stein *et al.*, 2014), in this study the focus is on land cover heterogeneity, referring to between-habitat heterogeneity (Stein *et al.*, 2014), which can be readily measured using satellite-derived land cover maps. Measures of habitat heterogeneity including patch area and edge length within each 1km square were derived from the UK Land Cover Map 2000 (LCM2000; Fuller *et al.*, 2002) using FRAGSTATS v4 (McGarigal *et al.*, 2012). LCM2000 is made up of 26 LCM Subclasses. To calculate the landscape variables the land cover classes were aggregated into a smaller set of broader land cover classes (details of class groupings can be found in table A9 in Appendix 4). The various FRAGSTATS metrics were only calculated for the arable, broadleaved, coniferous, grassland and semi-natural broader land cover classes. Details of the 17 class-level area, edge, shape and aggregation metrics calculated for each land cover class can be found in table A10 in Appendix 4. Below are the details for the metrics which appear in the final models built for each response variable; details of variable selection techniques are given in section 4.2.4. Metric descriptions adapted from McGarigal (2015).

- **Percentage of landscape (PLAND):** the percentage of each square comprised of a particular class.

- **Effective mesh size (MESH):** Quantifies habitat fragmentation based on the probability that two randomly chosen points in the region under interest are located in the same non-fragmented patch (Jaeger, 2000). The probability is multiplied by the total area of the landscape unit. The more barriers (e.g. roads, railroads) in the landscape, the lower the probability that the two locations will be located in the same patch, and the lower the effective mesh size.
- **Contiguity index (CONTIG):** Measure of spatial connectedness/contiguity of cells within a grid-cell given as the mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class. An index value of 0 represents a one-pixel patch, increasing to a limit of 1 as connectedness increases.
- **Related circumscribing circle (CIRCLE):** Measure of overall patch elongation using the ratio of patch area to the ratio of the smallest circumscribing circle given as mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class. Highly convoluted but narrow patches give a low index value, while narrow and elongated patches have a high index value.
- **Patch cohesion (COHESION):** Provides a measure of the physical connectedness of the corresponding class. COHESION approaches 0 as the proportion of the focal class decreases and becomes increasingly subdivided, and therefore less physically connected.
- **Largest Patch Index (LPI):** Quantifies the percentage of the total landscape area comprised by the largest patch of a class. The LPI approaches 0 when the largest patch of the corresponding type is increasingly small. An LPI value of 100 indicates the entire landscape is made up of a single patch of the corresponding class. Measured in percent (%).
- **Total edge (TE):** Absolute measure of total edge length of a particular class. Measured in metres (m).
- **Edge density (ED):** Edge length of a particular class standardised to a per unit area basis. Measured in metres per hectare (m/ha).

### 4.2.3 Habitat Productivity Variables

In this study, the Normalised Difference Vegetation Index (NDVI) was used as a proxy for habitat productivity. There are some constraints of using the NDVI, including saturation of the index in highly productive areas (Box *et al.*, 1989; Huete *et al.*, 2002) and its susceptibility to soil reflectance influence in regions with low vegetation (Huete, 1989). However, studies have demonstrated a strong positive correlation between NDVI and net primary productivity (NPP) at latitudes and habitat types similar to those that occur in Britain (e.g. Boelman *et al.*, 2003; Evans *et al.*, 2005; Kerr and Ostrovsky, 2003; Tebbs *et al.*, 2017), hence the NDVI was used.

The NDVI-based habitat productivity metrics were calculated in Google Earth Engine (Gorelick *et al.*, 2017) using data from Landsat 5 and Landsat 7. A greenest pixel composite was produced for each month from March to September for both Landsat 5 and Landsat 7 using images from 1999 to 2002; these were used to produce monthly NDVI images. The Landsat 5 and Landsat 7 data sets were then merged by taking the maximum NDVI value for each pixel from the Landsat 5 and Landsat 7 monthly NDVIs; this created monthly Maximum Value Composites (MVCs). Atmospherically corrected Landsat 5 and Landsat 7 images have been shown to produce similar NDVI measurements (Thieme *et al.*, 2020; Vogelmann *et al.*, 2001), hence no cross-calibration was required before the data sets were merged. Combining Landsat 5 and Landsat 7 reduced the impact of cloud cover, resulting in an average of approximately 6.1 cloud-free images per pixel for each month (figure A4 Appendix 4). This allowed monthly MVCs with near-complete cloud-free coverage for the whole of GB to be produced (figure A5 in Appendix 4).

From the monthly MVCs, individual NDVI metrics were calculated for each land cover class within each square, for each month and for the growing season (March-September) as a whole; areas of the different land cover classes were identified using LCM2000. The metrics chosen were mean, standard deviation, coefficient of variation, median, minimum, maximum, range, 20<sup>th</sup> percentile, 80<sup>th</sup> percentile, interquartile range and sum (growing season only) of the NDVI values.

These metrics were only calculated for the arable, broadleaved, coniferous, grassland and semi-natural broad land cover classes derived from LCM2000 (table A9 in Appendix 4). An example of the Google Earth Engine script used to extract the NDVI metrics can be found in Appendix 2.

#### **4.2.4 Random Forest Models and Variable Selection**

Random Forest (RF) regression (Breiman, 2001) was used to determine which measures of habitat productivity and habitat heterogeneity, in combination, provide the best estimation of farmland and woodland bird species richness and diversity. The *randomForest* package (Liaw & Wiener, 2002) in *R* was used to build and analyse the RF models with 1000 trees. Further details on Random Forest and its implementation can be found in Appendix 2. RF was chosen because of its ability to handle non-linear responses and complex interactions between variables (Breiman, 2001; A. M. Prasad *et al.*, 2006). Previous studies have used RF to assess the response of bird species richness to environmental heterogeneity variables (Carrasco *et al.*, 2018) and estimate rare species distribution in undersampled areas (Mi *et al.*, 2017).

Minimal depth selection (Ishwaran *et al.*, 2010) was used to rank the explanatory power (importance) of each estimator variable within the RF models, using the *randomForestSRC* package in *R* (Ishwaran and Kogalur, 2019). Minimal depth assumes that variables that tend to split nearest to the root node have a higher impact on the estimation than variables that split nodes further down the tree. While it is possible for non-estimative variables to split close to the root node and not impact estimation, such occurrences are rare in a large forest of trees and averaging minimises their effects (Ishwaran *et al.*, 2010). Variables which were not deemed important by the minimal depth selection were excluded from the RF models.

Feature selection was used to further reduce model complexity and correlation between estimators, and simplify interpretation of the RF models. To identify the number of variables which should be included in each model, lower importance variables, as determined by the minimal depth



selection, were excluded progressively and the change in the variance explained by the model was assessed. This information was used to produce an accuracy curve (figure 4.2) after removing all of the variables one by one (Ishwaran *et al.*, 2010). The number of variables included in each of the final refined RF models was determined based on the point at which a decrease in accuracy (variance explained) was first observed in the accuracy curves following the addition of another estimator variable to the model. Graphs for this analysis can be seen in figure 4.2. This analysis revealed that the top 4 most important variables, according to the minimal depth selection, were required for the farmland bird richness, the top 10 for farmland bird diversity, the top 3 for woodland bird richness, and the top 4 for woodland bird diversity. Subsequent analyses are based on these refined RF models.

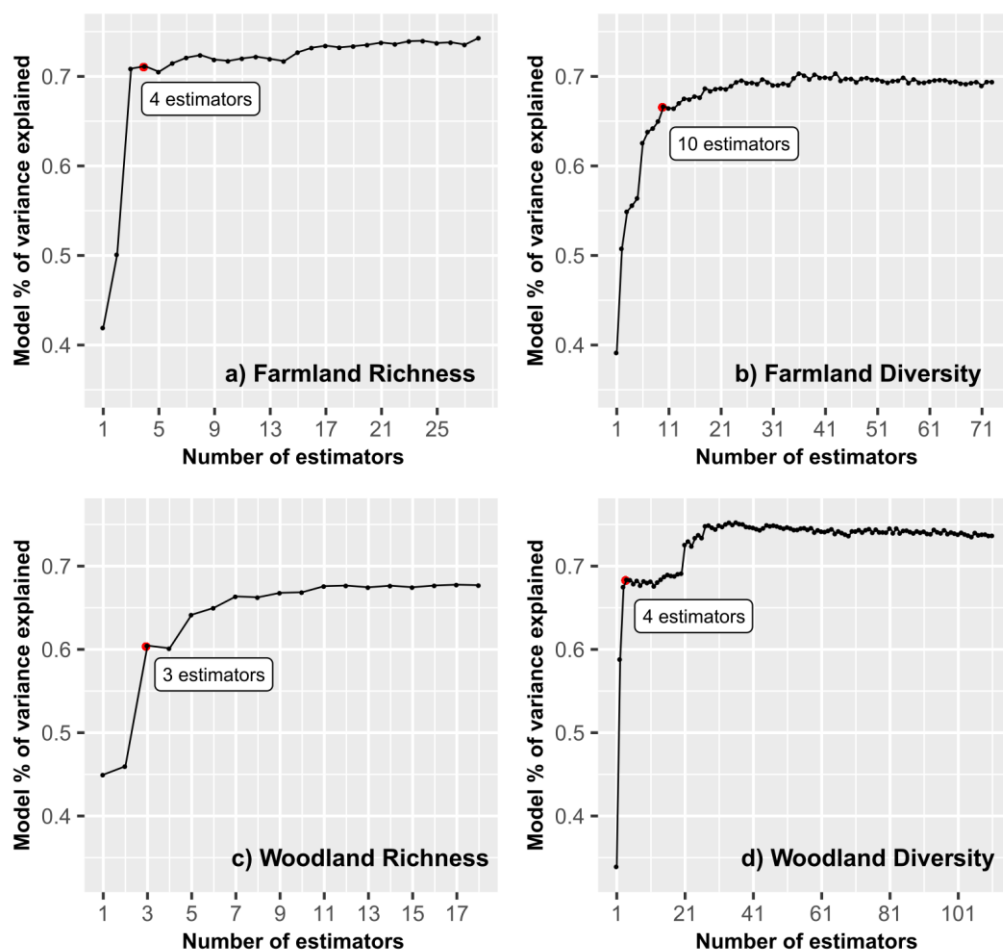


Figure 4.2: Variance explained (%) for the models including different numbers of estimators (by ranking). The red dot indicates the point immediately before a decrease in accuracy is first observed as another estimator is added, indicating the number of estimators to be included in each refined model.

The refined models were used to estimate farmland and woodland bird species richness and diversity across GB at 1km resolution. All models were trained using data from the 335 CS2000 squares with bird count data. As the data were not collected in squares with greater than 54.6% urban cover, squares with more urban cover than this were excluded. A tendency was observed for RF to underestimate the maximum value and overestimate the minimum value for each of the response variables. This is a recognised problem and was solved by applying a linear regression-based correction to the maps produced by RF to adjust for this bias (Zhang and Lu, 2012).

The performance of the models built at each different stage of the variable selection process were compared using  $R^2$  values calculated in two different ways: (i) internal validation carried out by the randomForest package, and (ii) a separate 10-fold cross-validation using the full training data set.

#### **4.2.5 Feature Contribution**

To explore the relationship between the estimator and response variables in each of the final models, feature contribution analysis was performed (Palczewska *et al.*, 2014) using the *forestFloor* package in *R* (Welling *et al.*, 2016). To calculate feature contribution, the increments of the estimated response are recorded after each node split by a given variable. FC is determined by summing the increments for each observation for each variable, then dividing by the number of trees. The effect of a studied variable in isolation on variations in the response variable estimations can be effectively separated and visualised by plotting the feature contribution against the value of each variable.

### **4.3 Results**

#### **4.3.1 Model Accuracies and Estimative Maps**

$R^2$  values for RF models constructed using individual habitat productivity and heterogeneity measures can be found in table A11 in Appendix 4. The RF models containing all variables had 10-

fold cross validation  $R^2$  values between 0.66 and 0.76 for farmland and woodland bird richness and diversity; while models containing only the important variables (identified using minimal depth selection) produced  $R^2$  values between 0.68 and 0.74 (table 4.1).

The refined models (with the number of variables determined using the accuracy curve) contained the following variables in order of importance:

- **Farmland bird richness:** percentage of arable land, effective arable mesh size, March 80<sup>th</sup> percentile NDVI for grassland, April maximum NDVI for grassland
- **Farmland bird diversity:** arable contiguity index, September NDVI range for arable land, June 80<sup>th</sup> percentile NDVI for grassland, April maximum NDVI for grassland, April 80<sup>th</sup> percentile NDVI for grassland, March 80<sup>th</sup> percentile NDVI for grassland, April median NDVI for semi-natural land, March maximum NDVI for grassland, September maximum NDVI for grassland, arable related circumscribing circle (area-weighted)
- **Woodland bird richness:** April 80<sup>th</sup> percentile NDVI for grassland, total broadleaved woodland edge length, broadleaved woodland edge density
- **Woodland bird diversity:** September maximum NDVI for broadleaved woodland, April 80<sup>th</sup> percentile NDVI for grassland, semi-natural patch cohesion, semi-natural largest patch index

These refined models had 10-fold cross validation  $R^2$  values between 0.64 and 0.72 (table 4.1). While this suggests they offer a slightly lower accuracy, the considerable reduction in the number of variables required has significant benefits, reducing complexity of the models and simplifying interpretation.

*Table 4.1: R-squared values for (i) RF models containing all variables, (ii) RF models containing variables categorised as important by the minimal depth selection, (iii) the final refined RF models. R-squared values are given for the internal RF validation and a separate 10-fold cross validation.*

Response Variable	R <sup>2</sup> values							
	<i>Full RF model (all variables)</i>		<i>Model containing only important variables (# of variables)</i>			<i>Refined RF model (# of variables)</i>		
	Internal RF validation	10-fold cross validation	Internal RF validation	10-fold cross validation		Internal RF validation	10-fold cross validation	
Farmland bird richness	0.72	0.73	0.74	(28)	0.74	0.72	(4)	0.72
Farmland bird diversity	0.66	0.66	0.69	(73)	0.70	0.67	(10)	0.67
Woodland bird richness	0.70	0.72	0.68	(18)	0.68	0.60	(3)	0.64
Woodland bird diversity	0.75	0.76	0.74	(111)	0.74	0.68	(4)	0.72

The refined RF models were used to produce maps of farmland and woodland bird richness and diversity in 2000 for GB at a 1km resolution (figure 4.3 and 4.4). Figure 4.3a and 4.3b show that farmland bird richness and diversity are highest in the east of England and lower in the west of England and in Scotland, a pattern which broadly matches the distribution of arable land across GB. Figure 4.4a and 4.4b show that levels of woodland bird richness and diversity appear to be more distributed around GB, although there appears to be a prevalence for higher species richness/diversity in lowland areas and lower richness/diversity in highland areas; this broadly reflects the distribution of broadleaved and coniferous woodland in GB. The fact that the distributions of land cover and bird richness/diversity do not match completely highlights the importance of elements other than simple percentage land cover in determining bird richness/diversity.

Although a bias correction was applied, this did not completely account for the under-estimation of maximum values and over-estimation of minimum values. For example, the minimum value of woodland bird richness in the observed data was 0, while the lowest estimated value was 3. The fact that the estimated values do not currently reflect what we see in the CS data may suggest

that the current estimator variables do not capture all the important environmental factors affecting bird species richness/diversity; further discussion of this is presented in section 4.4.4. Despite this slight uncertainty, there was no evidence of systematic error spatially, with no pattern in terms of the location (e.g. NS/EW, upland/lowland) of over- or under-estimated pixels, or in the magnitude of the error (figures A6 and A7 in Appendix 4).

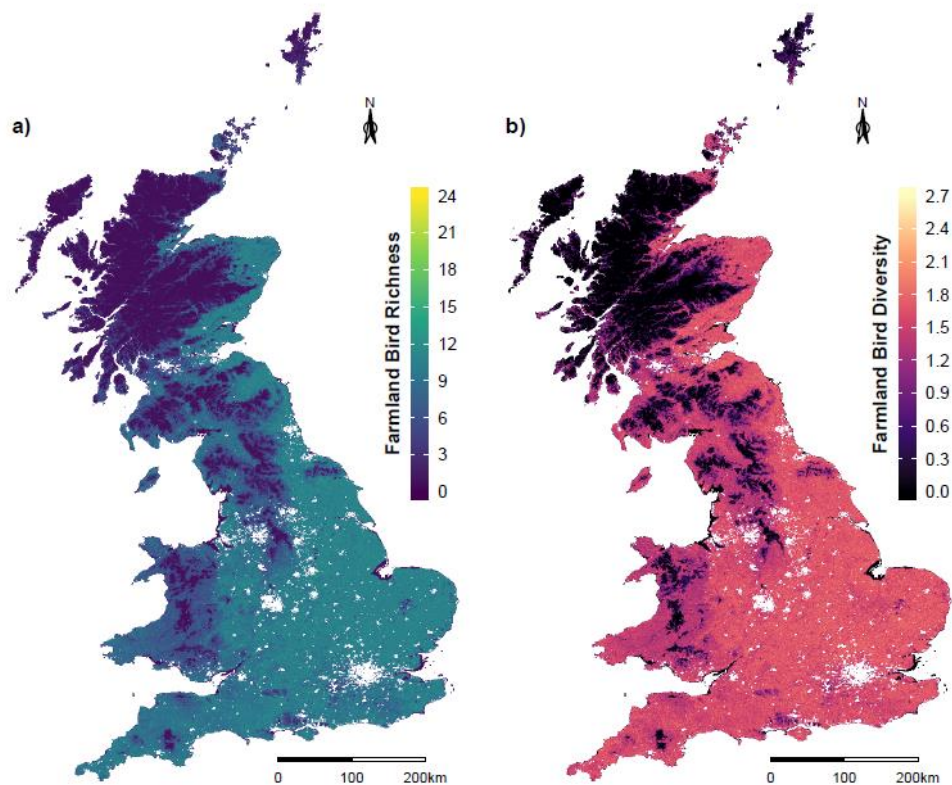


Figure 4.3: Estimated farmland bird (a) richness and (b) diversity (Shannon Index) maps at 1km resolution.

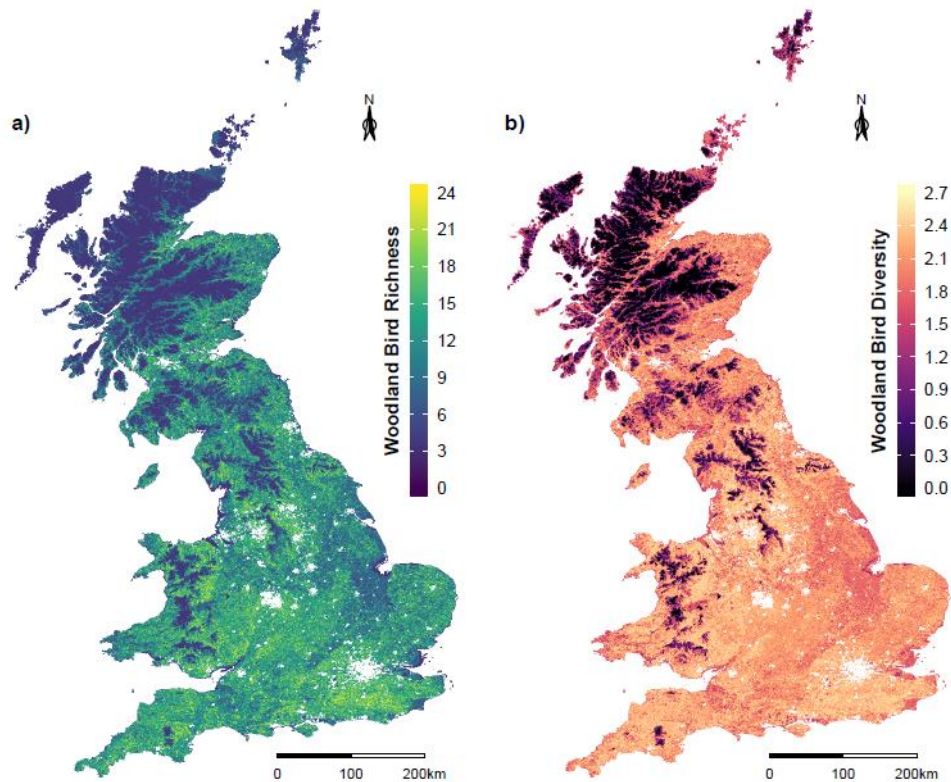


Figure 4.4: Estimated woodland bird (a) richness and (b) diversity (Shannon Index) maps at 1km resolution.

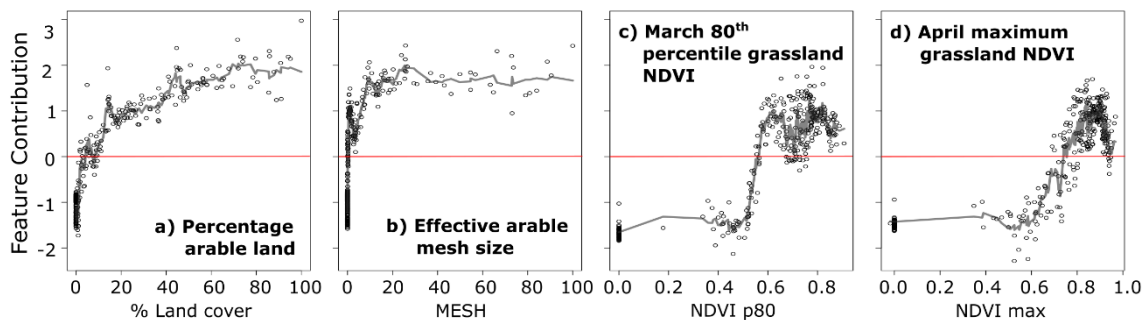
#### 4.3.2 Feature Contribution (for the Refined Models)

The feature contribution analysis showed different response shapes for the different habitat productivity and heterogeneity measures included in the refined models for each response variable (figure 4.5-4.8). The following section breaks down the feature contributions for each of the explanatory variables in each refined model for farmland bird richness and diversity, and woodland bird richness and diversity in turn. Feature contribution values above 0 indicate that the specific estimator value has a positive effect on the response variable; feature contribution values below 0 indicate a negative effect.

#### 4.3.2.1 Farmland Bird Richness

From the feature contribution analysis (figure 4.5) it appears that for farmland bird richness:

- Arable land covering more than 5% of the areas has positive effect on richness (figure 4.5a). The intensity of this effect increases with the percentage of arable land, until around 70% when it becomes asymptotic.
- An effective mesh size of more than 0 has a positive effect on richness (figure 4.5b). The intensity of this effect increases as effective mesh size increases from 10 to 30, before reaching a plateau and decreasing slightly.
- The presence of high productivity grassland in March and April ( $\text{NDVI p80}_{\text{Mar}} > 0.5$  and  $\text{NDVI max}_{\text{Apr}} > 0.7$ ) contributes positively towards richness (figure 4.5c and 4.5d). The intensity of the positive effect for each of these variables increases rapidly, with a clear step change for the March variable, to a maximum before flattening out and fluctuating slightly thereafter.



*Figure 4.5: Feature contribution plots for the farmland bird richness estimation model. The y-axis represents the change of estimated bird richness for a given variable value, measured with the cross-validated feature contribution. The x-axis represents the studied variable. The fitted line is based on the k-nearest neighbour (knn) estimations. The red line indicates the point of zero feature contribution.*

#### **4.3.2.2 Farmland Bird Diversity**

From the feature contribution analysis (figure 4.6) it appears that for farmland bird diversity:

- The presence of large contiguous patches of arable land is important, with high contiguity index values ( $>0.8$ ) having a positive effect and low values having little to no effect (figure 4.6a).
- Broad variation in the productivity of arable land in September ( $\text{NDVI range}_{\text{Sep}} > 0.25$ ) has a positive effect on richness, with a small gradual increase to a peak around an  $\text{NDVI range}_{\text{Sep}}$  value of 0.7 (figure 4.6b). Very low NDVI range values have a negative effect.
- The presence of high productivity grassland throughout the growing season ( $\text{NDVI max}_{\text{Mar}}/\text{max}_{\text{Apr}}/\text{max}_{\text{Sep}} > 0.6-0.8$ ;  $\text{NDVI p80}_{\text{Mar}}/\text{p80}_{\text{Apr}}/\text{p80}_{\text{June}} > 0.5-0.6$ ) contributes positively towards diversity (figure 4.6c-f, 4.6h-i). The intensity of this effect tends to increase rapidly towards a maximum before reaching a plateau.
- The effect of the productivity of semi-natural habitats in April appears to be generally weak, although there is a peak around an  $\text{NDVI median}_{\text{Apr}}$  value of 0.6 where it appears to have a relatively significant positive effect (figure 4.6g).
- The presence of very elongated, narrow arable patches do not appear to promote high diversity, with high CIRCLE ( $>0.7$ ) values having a negative effect (figure 4.6j); CIRCLE values between 0.4 and 0.6 have a weak positive effect.



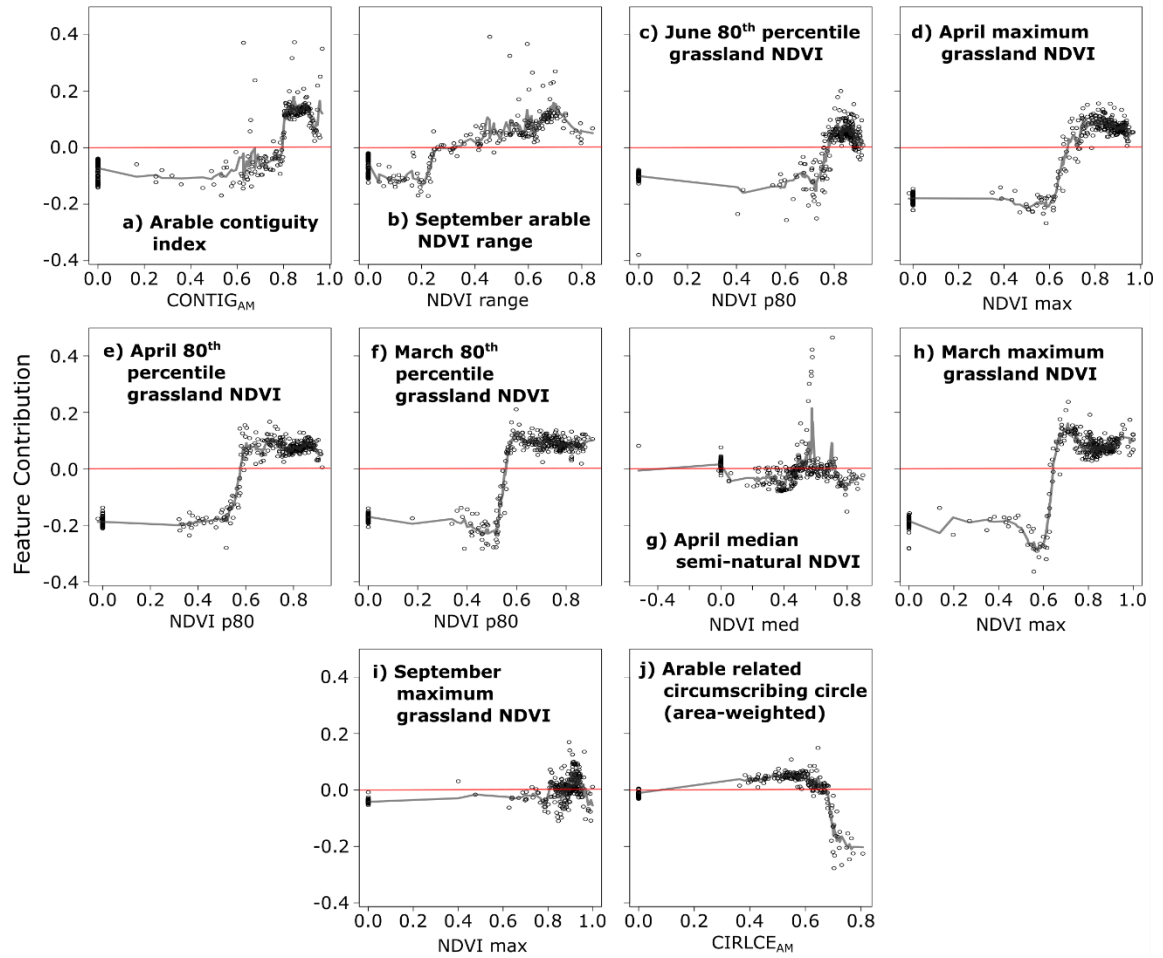


Figure 4.6: Feature contribution plots for the farmland bird diversity estimation model. The y-axis represents the change of estimated bird diversity for a given variable value, measured with the cross-validated feature contribution. The x-axis represents the studied variable. The fitted line is based on the k-nearest neighbour (knn) estimations. The red line indicates the point of zero feature contribution.

#### 4.3.2.3 Woodland Bird Richness

From the feature contribution analysis (figure 4.7) it appears that for woodland bird richness:

- The presence of high productivity grassland in April ( $\text{NDVI p80}_{\text{Apr}} > 0.6$ ) contributes positively towards richness, reaching a peak in intensity around an  $\text{NDVI p80}_{\text{Apr}}$  value of 0.75 (figure 4.7a).

- Total broad-leaved woodland edge length greater than 1000m and edge density greater than 10m/ha contribute positively towards richness, reaching peak intensity around 6000m and 60m/ha respectively, and flattening out thereafter (figure 4.7b and 4.7c). This may indicate that the presence of more, smaller patches of broad-leaved woodland or patches with more complex shapes helps to promote high richness.

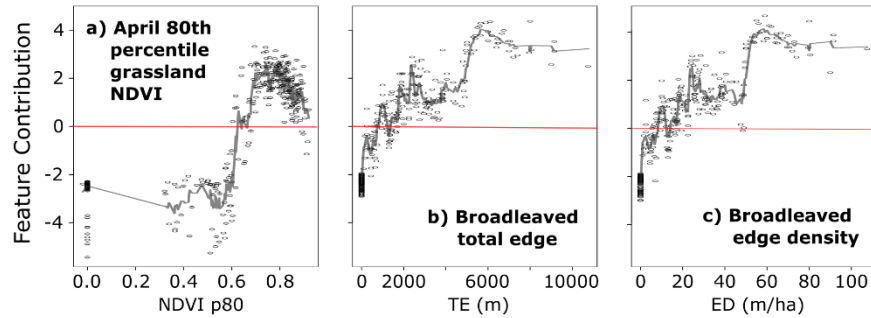


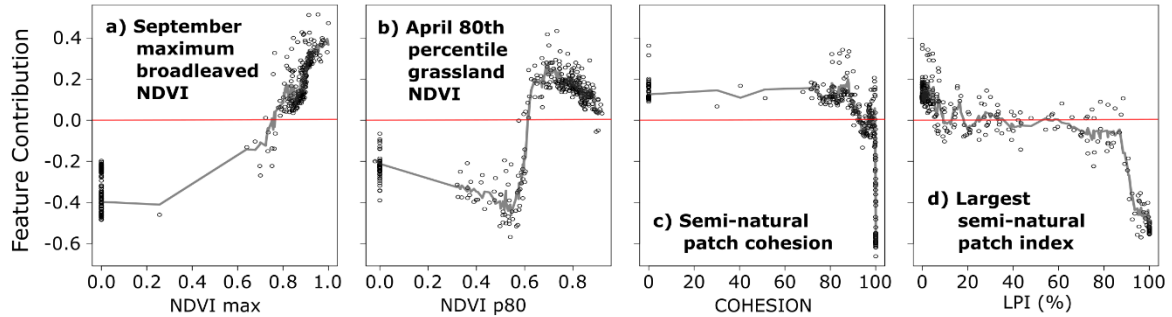
Figure 4.7: Feature contribution plots for the woodland bird richness estimation model. The y-axis represents the change of estimated bird richness for a given variable value, measured with the cross-validated feature contribution. The x-axis represents the studied variable. The fitted line is based on the k-nearest neighbour (knn) estimations. The red line indicates the point of zero feature contribution.

#### 4.3.2.4 Woodland Bird Diversity

From the feature contribution analysis (figure 4.8) it appears that for woodland bird diversity:

- The presence of high productivity broad-leaved woodland in September ( $\text{NDVI max}_{\text{Sep}} > 0.8$ ) and high productivity grassland in April ( $\text{NDVI p80}_{\text{Apr}} > 0.6$ ) both have a positive effect on diversity (figure 4.8a and 4.8b). The intensity of the effect of broad-leaved woodland productivity increases continuously, while for the grassland the intensity peaks around an  $\text{NDVI p80}_{\text{Apr}}$  value of 0.7 and then decreases slightly.
- High connectivity between semi-natural habitats does not appear to promote high diversity, with COHESION values greater than 90 having a strong negative effect; values below 90 have a comparatively weak positive effect (figure 4.8c).

- As the percentage of the landscape composed of the largest semi-natural patch (LPI) increases the effect becomes increasingly negative (figure 4.8d). Although there is some fluctuation up until LPI of 60%, the highest positive effects are seen for LPI of 20%.



*Figure 4.8: Feature contribution plots for the woodland bird diversity estimation model. The y-axis represents the change of estimated bird diversity for a given variable value, measured with the cross-validated feature contribution. The x-axis represents the studied variable. The fitted line is based on the k-nearest neighbour (knn) estimations. The red line indicates the point of zero feature contribution.*

## 4.4 Discussion

### 4.4.1 Comparison to Other Studies

The results presented here compare favourably to previous studies which have explored the estimation of bird species diversity using satellite data. Bonthoux *et al.* (2018), for example, managed to explain 35% and 45% of the variation in farmland and woodland bird species richness in France respectively by combining MODIS-derived single-date NDVIs with climate data in Generalised Additive Models (GAM). This study was conducted on a similar scale to the present study (4km<sup>2</sup> cf. 1km<sup>2</sup> sample squares) suggesting the factors important for species richness may be similar. The fact that the present study explained a higher percentage of variation (72% for farmland birds; 67% for woodland birds) may be due to the higher resolution of the Landsat data used (30m cf. 250m). The higher resolution data may allow for better discrimination of habitat composition and heterogeneity variation across the landscape, meaning relationships between species richness and landscape

factors such as resource availability can be better detected. The difference in model accuracy may also result from differences in the training data, both in terms of how it was collected (sampling method, distribution, etc.) and which species were classified as farmland and woodland birds in each case.

Carrasco *et al.* (2018) explored the relationships between bird species richness and a variety of environmental heterogeneity variables based on data from the UK Countryside Survey 2000. The RF model created explained 71% of the variance in bird species richness. To produce a map of bird richness for GB at 1km resolution, the top-ranked environmental heterogeneity variables were up-scaled using data from the Landsat-derived Land Cover Map 2000 (Fuller *et al.*, 2002) and the ITE Land Classification (Bunce *et al.*, 1991). These modelled variables were used to feed the national scale estimative RF models. The present study builds on this, incorporating not only measures of environmental heterogeneity but also habitat productivity. The results of this study demonstrate that it is possible to obtain similar levels of accuracy using satellite-derived measures of habitat heterogeneity and productivity without the need for upscaling of environmental heterogeneity variables derived from field data. Removing the need for the up-scaling technique used in Carrasco *et al.* (2018) simplifies processing and reduces field data requirements for model training; reduced requirements for training data could reduce costs of future data collection.

#### **4.4.2 Influence of the Landscape on Bird Diversity**

Woodland bird richness was higher in areas with higher woodland edge. These results support previous studies that showed that bird richness and diversity was greatly enhance by the presence of open habitats within woodlands, and by increased woodland hedges (Fuller *et al.*, 2007; Terraube *et al.*, 2016). Higher plant diversity and composition and habitat structural diversity at woodland edges have been pointed out as possible drivers of diversity for birds living within these habitats (Terraube *et al.*, 2016).

This study indicates that the presence of high amounts of arable land increased farmland bird richness, although the positive effect of arable land percentage stabilized after 70%. Because the presence of grassland was another important estimator of farmland species richness, it is reasonable to infer that a combination of high coverage of arable land with other habitat patches that increase the landscape heterogeneity might be optimal for maintaining a high number of farmland species. However, these associations might be strongly dependent on the surrounding areas. For instance, high percentages of arable land could be enhancing farmland bird richness where arable lands are less abundant in nearby landscapes, while areas with low percentages might still be supporting high richness in areas surrounded by intensively cultivated regions (Robinson *et al.*, 2001).

The results also showed that the presence of highly productive grasslands was an important estimator for farmland and woodland richness and diversity. Although an intense management of grasslands can reduce the suitability of these habitats to provide breeding and foraging resources for birds (Vickery *et al.*, 2001), the presence of these grasslands could be a key component to increase the heterogeneity of British landscapes (Fuller *et al.*, 2004). For areas with high coverage of arable land or woodland, the presence of improved grasslands could be providing extra ecological niches for farmland and woodland birds, respectively.

#### **4.4.3 Contribution of Satellite Data to Biodiversity Monitoring**

Knowledge of species richness and distribution, and relationships between species and their habitats, is essential to guide conservation and land management actions aimed at preventing biodiversity loss. Field-based methods, while providing indispensable information, are typically limited in coverage and spatial/temporal resolution due to the time and cost associated with data collection. Integration of field data with satellite data has the potential to provide wall-to-wall continuous mapping and therefore more suitable information to guide conservation actions on both local and national scales. As this study demonstrates, satellite data can be used not only to estimate

species richness, but also to provide insight into the factors affecting species richness. This information can inform the creation of more comprehensive conservation schemes which can be tailored to specific areas.

Used alongside field-based data, satellite data can also offer a cost effective method for large-scale, long term monitoring to ensure that conservation plans are meeting their intended goals. The availability of long-term data (30-40 years from Landsat data archive) allows the establishment of a baseline against which long-term and short-term changes can be monitored, facilitated by repeat satellite measurements. The introduction of cloud computing platforms such as Google Earth Engine (Gorelick *et al.*, 2017) provides access to greater computing power and global-scale satellite datasets, allowing species richness to be monitored at higher temporal resolutions at a lower cost. Greater availability and frequency of species richness data on a range of scales would allow more accurate monitoring of short-term changes in biodiversity allowing more rapid responses to any observed changes.

Field data of course retains an essential role in the monitoring of biodiversity, providing essential training and validation data for estimation models. Moving forward, the integration of field data and satellite data into a biodiversity monitoring framework could provide essential information to help reduce predicted biodiversity losses as a result of anthropogenic drivers such as land use change and changing climate.

#### **4.4.4 Future Developments**

In this study, Landsat data was used to derive a range of metrics designed to detect spatial variation in habitat productivity and habitat heterogeneity. However, there are other metrics, which are important for determining bird species richness, that were not captured due to the nature and resolution of the data used. For example, Goetz *et al.* (2007) demonstrated the value of LiDAR-derived measures of canopy structural diversity for estimating species richness. Additionally, the

resolution of Landsat data (30m) meant it was not possible to detect linear landscape features such as hedgerow habitats. Previous studies (Aue *et al.*, 2014; Carrasco *et al.*, 2018; Hinsley and Bellamy, 2000; Morelli *et al.*, 2014; Rhodes *et al.*, 2015) have demonstrated the importance of such features for some species (e.g. woodland generalists), particularly in open landscapes, by increasing the number of ecological niches. Sullivan *et al.* (2017) found that the inclusion of a national-scale model of linear features created using airborne SAR improved estimates of bird species abundance. Future work should explore how the incorporation of additional measures of habitat heterogeneity and structure affects the accuracy of bird diversity estimation.

Exploration of how the relationships between response and estimator variables vary with spatial and temporal scale and landscape context would also be useful. Bonthoux *et al.* (2018), for example, noted that the results of their study, which found single-date NDVIs to be more effective than the Dynamic Habitat Index (DHI), did not agree with previous studies conducted using larger sample units of investigation (e.g. Coops *et al.*, 2009a; Hobi *et al.*, 2017) that found the cumulative DHI had the highest estimative capability. Relationships have also been found to vary between different landscape types (e.g. upland and lowland) and in different ecoregions. For example, Petit *et al.* (2004) found that in lowland areas linear features are crucial for promoting species diversity, while in upland ecosystems quality may be more important. Exploration of the variability in relationships at different spatial and temporal scales, and in different landscape contexts, would allow a more comprehensive understanding of the factors affecting bird species richness. This would allow more accurate estimations to be made at scales relevant to specific monitoring schemes and conservation management needs. An understanding of how these relationships vary across time and space would also allow an assessment to be made as to the applicability/transferability of empirical approaches, such as the one used in this study, to other areas and time periods.

## **4.5 Conclusion**

In conclusion, this study demonstrates that it is possible to estimate farmland and woodland bird richness and diversity relatively accurately ( $R^2$  0.64 to 0.72) using just a few satellite-derived measures of habitat productivity and heterogeneity. The variable selection and feature contribution analysis highlight a number of important spatial drivers of species richness/diversity including arable land area for farmland birds, woodland patch edge length for woodland birds, and high productivity grassland during spring for both groups. Work such as this, which integrates satellite-derived metrics with species diversity field data, has the potential to provide important insights for monitoring and conservation. Further work could explore the spatial and temporal variation in the effect of these individual spatial drivers and whether inclusion of additional datasets would allow variation in important environmental characteristics to be better detected.



## **5. Towards Monitoring the Sustainable Intensification of Arable Agriculture Using Satellite-derived Indicators of Farm Performance**

Merryn L. Hunt, George Alan Blackburn, Luis Carrasco, Gavin Siriwardena, John W. Redhead & Clare S. Rowland

### **Abstract**

Global population increases are driving the demand for food and leading farmers to intensify their agricultural practices. Appropriate monitoring methods are important to ensure that this intensification is sustainably achieved. Current reliance on traditional data collection methods, which produce agricultural intensity and environmental quality data of limited spatial and temporal resolution, hampers our ability to meet this need. Earth Observation has the potential to improve our monitoring capabilities, providing a readily accessible, long-term dataset with global coverage at various spatial and temporal resolutions. However, so far no one has attempted to explore this potential.

This study uses satellite-derived indicators of agricultural intensity and environmental quality to assess the relative performance of arable farming. Wheat yield is used as a measure of agricultural intensity, with farmland bird richness used as a proxy for environmental quality. The satellite-derived indicators are produced from *in situ* data on wheat yield and farmland bird richness, Landsat-8 and Sentinel-2 satellite data and empirical modelling techniques (Random Forest regression). The agricultural intensity and environmental quality data are then combined to produce a novel feature space which provides a relative assessment of agricultural performance on a landscape scale. This feature space allows areas of differing agricultural intensity and environmental quality to be identified, making it possible to see how individual arable areas are performing relative to the surrounding landscape. Such knowledge of farm performance could be used to monitor the impacts of different management practices and could help identify the optimum practices in specific areas. This would help a range of stakeholders, including farmers moving towards sustainable agricultural intensification. As this assessment of farm performance is derived from EO data, there is

considerable potential for using the approach to guide sustainable intensification in arable landscapes across the globe. This unique assessment of relative agricultural performance demonstrates that Earth Observation has a significant role to play in monitoring sustainable intensification and ensuring it meets our future needs.

## **5.1 Introduction**

Global food production must increase by 70% by 2050 to meet the demands of a growing population with shifting food consumption patterns and increases in income (Dillon *et al.*, 2016; FAO, 2009). Competition for land means intensification of production on current agricultural land will play a key role in this, alongside reductions in food waste and shifts in diet. However, previous intensification endeavours, for example through increased cropping intensity and agrochemical inputs, are now recognised as having a detrimental impact on the environment (Pretty *et al.*, 2018; C Stoate *et al.*, 2001; Tilman *et al.*, 2001; Tschamntke *et al.*, 2005). The increased emission of pollution and waste, over exploitation of natural resources for inputs and the associated negative environmental feedbacks, call into question the ability to maintain such intensive agricultural practices for sustained crop productivity (Bommarco *et al.*, 2013b; Foley *et al.*, 2005).

Sustainable intensification has been proposed as a way to address the conflicting requirements of preserving environmental quality and meeting projected increases in food demand. Sustainable intensification involves achieving higher agricultural outputs with the same or fewer inputs through increased production efficiency on the same amount of agricultural land, while simultaneously reducing or eliminating environmental degradation (Dillon *et al.*, 2016; Garnett *et al.*, 2013). Unfortunately, due to the situation-specific success of different approaches, resulting from the spatial and temporal variability in environmental variables, there are no definitive mechanisms to achieve sustainable intensification. To ensure any attempts at sustainable intensification are successful, it is therefore essential that suitable methods are in place to efficiently assess the sustainability of intensification on a long-term basis over diverse landscapes and spatial scales.

Currently, assessments of sustainable intensification rely on traditional data sources such as farm surveys, field data and national government statistics (e.g. Dillon *et al.*, 2010; Firbank *et al.*, 2013; Rasul and Thapa, 2004). However, such data are typically costly and time consuming to collect, resulting in data of limited spatial and temporal scale and extent. The inherent constraints of such

data sources affect our ability to provide accurate assessments of productivity and environmental performance for all agricultural areas in a timely manner. Earth Observation (EO) may offer a solution to this data problem. EO provides a readily accessible, long-term dataset with global coverage at various spatial and temporal resolutions. Incorporation of these data into assessments of agricultural productivity and environmental performance could help to provide more accurate, spatially explicit results at lower costs.

EO has been used for agricultural monitoring since the launch of the first terrestrial satellites (Anuta and MacDonald, 1971; Draeger and Benson, 1972). With the advent of readily accessible cloud computing platforms like Google Earth Engine (Gorelick *et al.*, 2017), the launch of the Sentinel satellites (Drusch *et al.*, 2012; Torres *et al.*, 2012) and the opening of the Landsat archives (Wulder *et al.*, 2012), EO systems are increasingly able to generate operational data products to support agricultural management. The use of EO to monitor various measures of agricultural intensity such as crop yield (e.g. Becker-Reshef *et al.*, 2010; Doraiswamy *et al.*, 2005) and cropping intensity (e.g. Jain *et al.*, 2013; L. Li *et al.*, 2014) in a range of countries over various scales is fairly common practice; although additional work is required to make this routine and operational. The use of EO to assess environmental quality is currently less established, but there is a clear opportunity to monitor indicators including vegetation health (e.g. crop condition) and ecosystem health (e.g. net primary productivity), soil quality (e.g. soil organic carbon), water quality (e.g. water clarity), and biodiversity (e.g. species diversity). A full review of this potential can be found in Hunt *et al.* (2019b, Chapter 2).

EO data are already being used in several international agricultural monitoring systems for crop condition monitoring and yield forecasting over regional, national and global scales; such systems include the Group on Earth Observations Global Agricultural Monitoring System (GEOGLAM) (Parihar *et al.*, 2012). While these systems demonstrate the value of EO for agricultural monitoring, currently they are not explicitly being used to monitor agricultural intensity, and no attempts have been made to routinely monitor environmental quality. Hence, the aim of this paper is to combine

satellite-derived measures of agricultural intensity and environmental quality to provide a relative assessment of the performance of arable agriculture on a landscape scale.

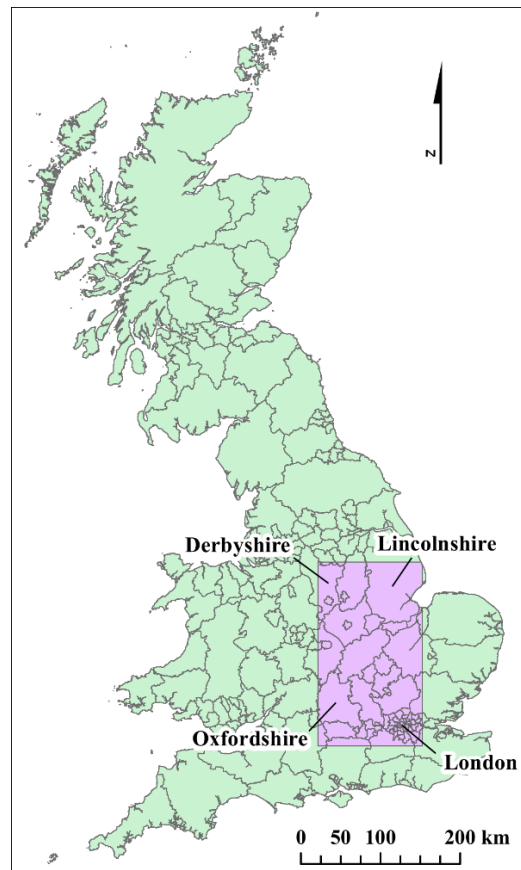
In this study, wheat yield is used as an indicator of agricultural intensity and farmland bird species richness is used as a proxy for environmental quality. While wheat is not the only crop grown in the study area, it has the largest planted area of all cereal crops, covering 1.8Mha in 2016 (DEFRA, 2019). As such, wheat yield is used as an indicator of general intensity in a given area as it maximises the sample size for assessing intensity. The choice of farmland bird species richness as a proxy for environmental quality recognises the utility of birds as an indicator of general biodiversity (Furness and Greenwood, 2013) and environmental quality (Butler *et al.*, 2012). Therefore, to fulfil the aim of the study, to provide a relative assessment of agricultural performance, the following objectives were addressed:

1. Estimate wheat yield and farmland bird richness across the 20,000 km<sup>2</sup> study area for 2016 from satellite data using Random Forest regression models trained on *in situ* observations.
2. Identify the potential (maximum) yield and richness within individual environmental zones and calculate the gap between potential and actual values for all 1km squares across the study area.
3. Create a performance feature space by plotting values for the yield and richness gaps against one another.

## **5.2 Method**

The focus of this study is on an area covering 20,000km<sup>2</sup> (100 x 200km) in the east of England, bounded by Lincolnshire to the northeast, London to the southeast, Oxfordshire to the southwest and Derbyshire to the northwest (figure 5.1). This site was selected based on the availability of high resolution crop yield data within this area (see section 5.2.1) and the coverage of the appropriate Sentinel-2 tiles (see section 5.2.2.1). The average total monthly rainfall for the area from December 2015 to July 2016, the period covered in this study, ranged from 25.4 to 95.4mm,

and average monthly mean temperature ranged from 5.0 to 17.4°C (Met-Office, 2017). This area is dominated by agriculture with climatic conditions suitable for growing wheat, among other crops. Agriculture within this area falls into two broad categories: (1) large scale industrialised farms in flatter areas, and (2) smaller-scale farms in upland areas.



*Figure 5.1: Study area location. The rectangular area shows the extent of the study area.*

Figure 5.2 provides an overview of the method used in this study, outlining how the wheat yield and farmland bird richness maps were produced and then used to assess relative agricultural performance in the study area. Details of the data used are outlined in sections 5.2.1-5.2.2 (wheat yield) and 5.2.3-5.2.4 (farmland bird richness); the analysis techniques along with production and use of the performance feature space are outlined in sections 5.2.5-5.2.8.

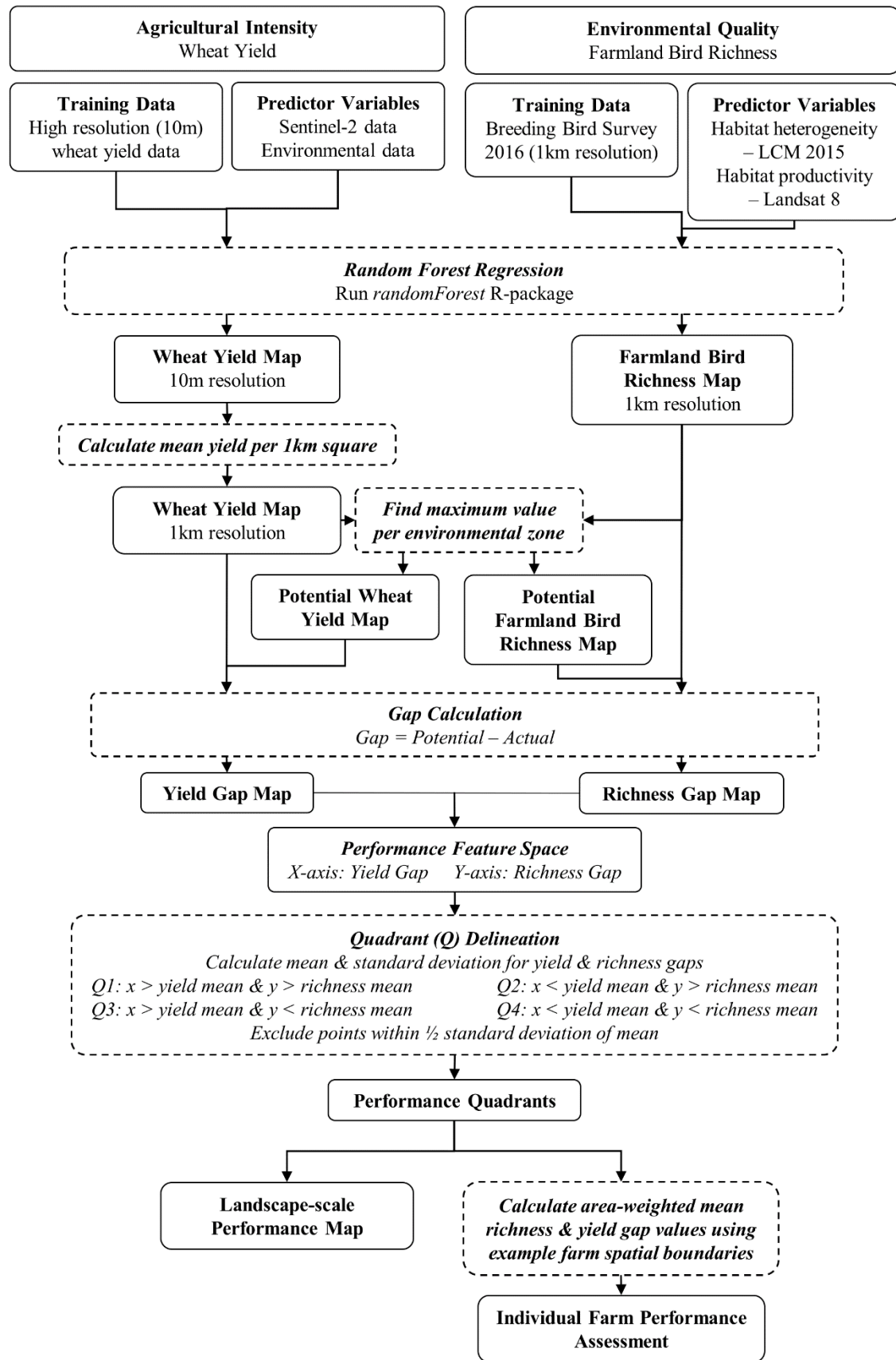


Figure 5.2: Overview of the method used to create the feature space to assess relative farm performance.

### 5.2.1 Training Data for Wheat Yield Modelling

High resolution wheat yield data, collected during the 2016 harvest period between 6<sup>th</sup> August and 9<sup>th</sup> September by combine harvesters equipped with GPS and optical yield monitors, were downloaded from CLAAS telematics (CLAAS, 2018). The methods for processing these data are described here, but additional details are given in Hunt *et al.* (2019a, Chapter 3). The data were spread over two different regions, with 28 fields in Lincolnshire and 11 fields in Oxfordshire, marking the north-east and south-west extent of the study area. A series of simple and threshold-based cleaning steps were applied to the raw data to remove inaccurate grain yield measurements resulting, for example, from when the combine was turning (AHDB, 2016; Hunt *et al.*, 2019a, Chapter 3; Lyle *et al.*, 2014).

A 20m buffer was applied to the cleaned data around the inward edge of each field to avoid areas of the field where the satellite pixels would span field boundaries (mixed pixels). Gaps occurring in the dataset due to the data collection and cleaning process were then removed using additional manual masking. These gaps typically appeared at the edge of the fields and in areas where the combine harvesters turned.

An Inverse Distance Weighting function was applied to the cleaned and buffered point data to convert it to 10m resolution raster data to align with the Sentinel-2 data (see section 5.2.2.1). The impact of auto-correlation between pixels was reduced by using only alternate pixels, producing 8794 data points which were randomly split into training (70%) and validation (30%) datasets.

### 5.2.2 Estimator Variables for Wheat Yield Modelling

Previous work (Hunt *et al.*, 2019a, Chapter 3), found that using a combination of satellite data and environmental data, such as precipitation and temperature, provided a more accurate estimate of yield than using satellite data on its own. Hence, in this study a combination of such data were used to estimate wheat yield across the landscape. Details of these data are outlined in the following sections (5.2.2.1-5.2.2.4).



### 5.2.2.1 Sentinel-2 Data

The Sentinel-2 images (Level 1C Top-of-Atmosphere reflectance product; see Claverie *et al.*, 2018; Drusch *et al.*, 2012) for tiles 30UXC and 30UXD for the 6<sup>th</sup> June 2016 were downloaded from the Copernicus Open Access Hub (ESA, 2018). The decision to use the image for June was based on its high estimative capability compared to available images from other months (Hunt *et al.*, 2019a, Chapter 3) and the fact that it offered maximum coverage of the study site, not hampered by cloud cover or missing data. In this study only Sentinel-2 bands at 10 or 20m resolution were used; details of these bands can be found in table 5.1. These bands were atmospherically corrected using the Sen2Cor processor and bands at 20m resolution were resampled to 10m prior to stacking.

*Table 5.1: Details of the Sentinel-2 bands used in this study (Drusch et al., 2012).*

Spectral band	Central wavelength (nm)	Spatial resolution (m)
Band 2 blue	490	10
Band 3 green	560	10
Band 4 red	665	10
Band 5 vegetation red edge	705	20
Band 6 vegetation red edge	740	20
Band 7 vegetation red edge	783	20
Band 8 NIR	842	10
Band 8a narrow NIR	865	20
Band 11 SWIR	1610	20
Band 12 SWIR	2190	20

### 5.2.2.2 Precipitation and Temperature

Monthly total rainfall (mm) and mean air temperature (°C) 5km gridded UKCP09 datasets for December 2015 to July 2016 were downloaded from the UK Met Office (Met-Office, 2017). This time period was chosen to cover the main 2016 growing season.

### 5.2.2.3 Soil Water Index

Soil Water Index (SWI) values, indicating the soil moisture profile, were obtained from the SCAT-SAR SWI T01 dataset created by TU Wien Department of Geodesy and

Geoinformation (Bauer-Marschallinger *et al.*, 2018). SWI images have a pixel spacing of 500m, which correspond to a resolution of 1km. The data are derived from the Sentinel-1 SAT and MetOp-A/B ASCAT satellite sensors. Data for December 2015 to July 2016 were used to calculate monthly mean values, giving a percentage ranging from 0% (completely dry soil) to 100% (completely saturated soil).

#### **5.2.2.4 Topographic Variables**

The 10m resolution NEXTMap Britain digital terrain model (DTM), created by Intermap Technologies Inc., was used to calculate aspect and slope variables at 10m resolution. These data are based on airborne radar data collected during 2002 and 2003 (Intermap-Technologies, 2009).

### **5.2.3 Training Data for Farmland Bird Richness Modelling**

Bird count data from the 2016 Breeding Bird Survey (BBS) were obtained from the British Trust for Ornithology (BTO) (Harris *et al.*, 2017). These surveys are conducted by volunteers who record all bird species encountered while walking two 1km transects across their square during two early-morning visits in the April-June Survey period. The locations of the 1km BBS survey squares are selected through stratified random sampling, with 1km squares from the National Grid assigned randomly within BTO regions (Harris *et al.*, 2017). The focus of this study was on farmland bird species due to their reliance on arable land, making them good indicators of the environmental quality of these areas. Farmland bird species were identified using the groupings from the BTO/JNCC/RSPB wild bird population indicators for the UK and England (Eaton and Noble, 2019); a list of these farmland species can be seen in table 5.2. In this study, species richness was used as a measure of farmland bird species diversity; this was calculated by counting all the different farmland species recorded within each square.

Table 5.2: Farmland bird species groupings based on the BTO/JNCC/RSPB wild bird indicators for the UK and England (Eaton and Noble, 2019).

Farmland Birds Species			
Corn Bunting	<i>Emberiza calandra</i>	Skylark	<i>Alauda arvensis</i>
Goldfinch	<i>Carduelis carduelis</i>	Starling	<i>Sturnus vulgaris</i>
Greenfinch	<i>Carduelis chloris</i>	Stock Dove	<i>Columba oenas</i>
Grey partridge	<i>Perdix perdix</i>	Tree Sparrow	<i>Passer montanus</i>
Jackdaw	<i>Corvus monedula</i>	Turtle Dove	<i>Streptopelia turtur</i>
Kestrel	<i>Falco tinnunculus</i>	Yellow Wagtail	<i>Motacilla flava</i>
Lapwing	<i>Vanellus vanellus</i>	Yellowhammer	<i>Emberiza citrinella</i>
Linnet	<i>Linaria cannabina</i>	Whitethroat	<i>Sylvia communis</i>
Reed Bunting	<i>Emberiza schoeniclus</i>	Woodpigeon	<i>Columba palumbus</i>
Rook	<i>Corvus frugilegus</i>		

This study used data for 3695 squares spread across England, Wales and Scotland; of these squares, 616 fall within the study area. Unlike the yield data, the bird data were not split into separate training and validation datasets due to the lower number of available data points. To maximise the amount of data available for training, the full dataset of 3695 squares was used to train Random Forest to create a map of farmland bird richness for the whole of Great Britain (details about the Random Forest modelling are given in section 5.2.5).

#### 5.2.4 Estimator Variables for Farmland Bird Richness Modelling

Previous studies have used satellite-derived measures of two key factors affecting bird species diversity to map its spatial variation: (1) habitat heterogeneity, including measures of spatial distribution and habitat connectivity (e.g. Carrasco *et al.*, 2018; Coops *et al.*, 2009b); (2) habitat productivity measured, for example, using gross and net primary productivity (Phillips *et al.*, 2010) and normalised difference vegetation index (NDVI) (e.g. Duro *et al.*, 2014; Foody, 2005; Seto *et al.*, 2004). In this study, land cover heterogeneity was used as a measure of habitat heterogeneity (Stein *et al.*, 2014), as it can be readily measured using satellite-derived land cover maps; section 5.2.4.1 outlines the methods used to extract these data. The NDVI was used as a proxy for habitat productivity; section 5.2.4.2 details the methods used to derive these data. Studies have demonstrated a strong positive correlation between net primary productivity (NPP) and the NDVI in

habitat types and latitudes similar to those found in Britain (e.g. Boelman *et al.*, 2003; Evans *et al.*, 2005; Kerr and Ostrovsky, 2003; Tebbs *et al.*, 2017).

#### **5.2.4.1 Habitat Heterogeneity**

The UK Land Cover Map 2015 (LCM2015; Rowland *et al.*, 2017b) was used to derive a range of measures of habitat heterogeneity within each 1km square in the UK using FRAGSTATS v4 (McGarigal *et al.*, 2012). The 26 LCM subclasses of the LCM2015 were aggregated into a smaller set of broader land cover classes (details of the groupings can be found in table 5.3) to characterise the main habitat types in the study area. The various FRAGSTATS metrics were calculated for the arable, broadleaved, coniferous, grassland and semi-natural broader land cover classes; the urban, water and coastal classes were excluded as they are less connected with agricultural practices. Details of all 17 class-level area, edge, shape and aggregation metrics calculated for each land cover class can be found in table 5.4.

*Table 5.3: Broad land cover classes used in this study and the corresponding original LCM2015 subclasses.*

<b>LC class</b>	<b>LCM Subclasses</b>
Arable	Arable cereals, arable horticulture, non-rotational horticulture
Broad-leaved	Broad-leaved/mixed woodland
Coniferous	Coniferous woodland
Grassland	Improved grassland, set-aside grassland, neutral grassland, calcareous
Semi-natural	Acid grassland, bracken, dense and open dwarf shrub heath, fen, marsh, swamp, bogs (deep peat), montane habitats, inland bare ground, saltmarsh
Urban	Continuous urban, suburban/rural developed
Water	Water (inland)
Coast	Supra-littoral rock, supra-littoral sediment, littoral rock, littoral sediment
Sea	Sea/estuary

Table 5.4: Details of the habitat structure metrics derived from LCM2015 using FRAGSTATS. Metric descriptions based on McGarigal (2015).

Variable	Abbreviation (Unit)	Description
<b>Area &amp; Edge Metrics</b>		
Area	AREA <sub>AM/CV/MN</sub> (ha)	Area of each patch comprising a landscape mosaic given as mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class.
Percentage of landscape	PLAND (%)	Percentage of the landscape comprised of a particular patch type
Edge Density	ED (m/ha)	Edge length of a particular patch type standardised to a per unit area basis.
Radius of Gyration	GYRATE <sub>AM/CV/MN</sub> (m)	Measure of patch extent given as mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class.
Largest Patch Index	LPI (%)	Quantifies the percentage of the total landscape area comprised by the largest patch.
Total Edge	TE (m)	Absolute measure of total edge length of a particular patch type.
<b>Shape Metrics</b>		
Related Circumscribing Circle	CIRCLE <sub>AM/CV/MN</sub>	Measure of overall patch elongation using the ratio of patch area to the ratio of the smallest circumscribing circle given as mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class.
Contiguity Index	CONTIG <sub>AM/CV/MN</sub>	Measure of spatial connectedness/contiguity of cells within a grid-cell given as the mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class.
Fractal Dimension Index	FRAC <sub>AM/CV/MN</sub>	Measure of shape complexity given as a mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class.
Perimeter-Area Ratio	PARA <sub>AM/CV/MN</sub>	Ratio of patch perimeter to area given as mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class, providing a measure of shape complexity.
Shape Index	SHAPE <sub>AM/CV/MN</sub>	Measures the complexity of patch shape compared to a standard shape (square) of the same size. Values are given as mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class.
<b>Aggregation Metrics</b>		
Patch Cohesion	COHESION	Provides a measure of the physical connectedness of the corresponding patch types.
Landscape Division Index	DIVISION (proportion)	Probability that 2 randomly chosen pixels in the landscape are not situated in the same undissected patch of the corresponding patch type.
Effective Mesh Size	MESH (ha)	Quantifies habitat fragmentation based on the probability that two randomly chosen points in the region under interest are located in the same non-fragmented patch (Jaeger, 2000). The probability is multiplied by the total area of the landscape unit.
Number of Patches	NP	Number of patches of a particular patch type
Patch Density	PD (number of patches per 100 ha)	Number of patches of the corresponding patch type standardised on a per unit area basis
Splitting Index	SPLIT	SPLIT is 1 when the landscape consists of a single patch, increasing in value as the focal patch type is increasingly reduced in area and subdivided into smaller patches.

#### **5.2.4.2 Habitat Productivity**

Google Earth Engine (Gorelick *et al.*, 2017) was used to calculate the NDVI-based habitat productivity metrics using data from Landsat 8 Surface Reflectance Tier 1 collection (Wulder *et al.*, 2019). Monthly greenest pixel composites were produced for March to September using images from 2014 to 2016 to minimise cloud, cloud shadow and haze not removed by the cloud masking (Van Leeuwen *et al.*, 1999). The monthly composites were created by taking the highest NDVI value for each pixel in all images obtained from 2014 to 2016 for each month. Images from March to September were used as this forms the main vegetation growing season in the UK. From these monthly composites individual NDVI metrics were calculated for the area covered by each land cover class within each 1km square in GB, for each month and for the main vegetation growing period (March-September) as a whole; the land cover classes derived from LCM2015 (table 5.3) were used to identify the areas of different land cover classes within each square. The metrics chosen were mean, standard deviation, coefficient of variation, minimum, maximum, range, median, 20<sup>th</sup> and 80<sup>th</sup> percentiles, interquartile range and sum (growing season only) of the NDVI values. These metrics were only calculated for the arable, broadleaved, coniferous, grassland and semi-natural land cover classes. An example of the Google Earth Engine script used to extract the NDVI metrics can be found in Appendix 2.

#### **5.2.5 Random Forest Regression**

Random Forest (RF) regression (Breiman, 2001) was used to produce the maps of wheat yield and farmland bird richness using the data described in sections 5.2.1-5.2.4. RF is a machine learning algorithm that can be used to estimate a continuous response variable using regression analysis. In this study the *randomForest* R package (Liaw and Wiener, 2002) was used to build models with 1000 trees. Further details on Random Forest and its implementation can be found in Appendix 2. The decision to use RF was made based on its ability to handle complex interactions between variables and non-linear responses (Breiman, 2001; A. M. Prasad *et al.*, 2006). RF has been

used in previous studies to successfully estimate both bird species richness (Carrasco *et al.*, 2018; Hunt *et al.*, 2020, Chapter 4) and wheat yield (Hunt *et al.*, 2019a, Chapter 3) from remotely-sensed variables.

The performance of the models built were assessed using  $R^2$  and RMSE values calculated in two different ways: (i) internal validation carried out by the *randomForest* package, and (ii) separate 10-fold cross-validation using the training dataset. For the wheat yield model, an additional validation was carried out using the separate validation dataset.

To produce the wheat yield map, fields containing wheat were identified using the 2016 *Land Cover Plus*®: *Crops* map. Mixed boundary pixels were removed from the dataset by buffering in the field boundaries in the crop map by 20m. Farmland bird richness was estimated for all squares in the study area covering agricultural areas. The wheat yield and farmland bird richness maps were created at resolutions of 10m and 1km respectively, according to the differing resolutions of the training data. To allow further analysis of these two datasets the average wheat yield per 1km square was calculated, to match the farmland bird richness map, producing two spatially equivalent datasets. The focus of this study is the sustainable intensification of arable agriculture, therefore only areas under this type of agriculture or areas potentially influenced by it were included. As such, only squares which contained at least 20% crop cover were included in the subsequent analysis to exclude squares which did not contain much arable land; these squares were identified using the 2016 *Land Cover Plus*®: *Crops* map.

### 5.2.6 Gap Calculation

The aim of this study is to assess the relative performance of arable areas within the landscape based on wheat yield as a measure of agricultural intensity and farmland bird richness as a proxy for environmental quality. The goal of sustainable intensification is to maximise yield without harming the environment. As such, it is important to know what is possible for a specific location. Potential yield and potential bird richness will be affected by environmental conditions, as well as

environmental management. Directly comparing wheat yield and farmland bird richness for different areas without taking into account differences in environmental conditions could be misleading.

One way to account for this underlying variability is through gap analysis. The idea of a yield gap has been widely used in agricultural intensity studies (Lobell *et al.*, 2009; Van Ittersum and Rabbinge, 1997). Yield potential can be defined as the potential yield of a crop when grown under favourable conditions without growth limitations (Lobell *et al.*, 2009). Various methods exist to determine the yield potential including crop model simulation, field experiments and taking the maximum yield of a sample of farmers in a region of interest (Lobell *et al.*, 2009). In this study, the gap concept is applied to both wheat yield and farmland bird richness, comparing the yield and richness values at a specific location to potential values. The aim is to normalise the differences in yield and richness, due to the variable environmental conditions across the study area, so that the remaining variability is due to management. Thus, enabling direct comparison of different locations across the study area.

In this study, the potential wheat yield and farmland bird richness for each 1km square were calculated by spatially stratifying the study area according to the ITE Land Classification of Great Britain (Bunce *et al.*, 1996); a map showing the distribution of the ITE land classes can be found in figure A8 in Appendix 5. The ITE Land Classification splits Great Britain into environmental strata based on 75 environmental variables including climatic data, topographic data and geology data. The environmental strata provide regions with similar characteristics and environmental qualities. This makes the ITE land classes good spatial units to calculate potential against, as they account for the environmental variability in the study area better than simply using a moving window or assuming a maximum value for the entire study area. The potential value for a given 1km square was taken to be the maximum value found across all squares falling within the same ITE land class. The gap was then calculated as the difference between the potential and actual value for each square. This method was applied to both the wheat yield and farmland bird richness datasets.



### **5.2.7 Performance Feature Space**

To assess the performance of arable agriculture, the wheat yield and farmland bird richness gap values for each 1km square were plotted against one another to create a feature space in which relative performance could be plotted. A colour blending approach was adopted to create a continuous performance colour scale by overlaying two separate graphs, the first with points assigned colour based on their yield gap value, the second with colour assigned based on the richness gap value. Overlaying these two graphs creates a continuous colour scale of relative agricultural performance. To explore how farm performance varies in space, a map was created by overlaying the 1km yield gap and richness gap maps, using the same colour schemes as were used to create the feature space.

## **5.3 Results**

### **5.3.1 Wheat Yield and Farmland Bird Richness Gap Data**

The RF models created to estimate wheat yield and farmland bird richness performed relatively well (table 5.5). Wheat yield was estimated with an  $R^2$  value of 0.91 and an RMSE of 0.60 from the cross-validation, while farmland bird richness was estimated with a cross-validation  $R^2$  value of 0.54 and an RMSE of 2.22 (table 5.5). The outputs showed that in the study area, wheat yield ranges from 6.13 to 11.93 tonnes/ha and farmland bird richness ranges from 2.91 to 15.73. The wheat yield map (figure 5.3a) reveals a cluster of high yielding squares to the north east of the image, which is associated with the high quality agricultural land of the Lincolnshire fenlands. The yield values across the rest of the map are more variable and less clustered. In terms of farmland bird richness (figure 5.3c), the northeast also performs well, while clusters of lower richness can be seen in the southeast and northwest, with a more varied picture across the rest of the region.

*Table 5.5: RMSE and R-squared values for the wheat yield and farmland bird richness Random Forest models calculated from (i) the training dataset, (ii) 10-fold cross-validation with the training data (values in the brackets represent the standard deviation for these cross-validation values), and (iii) the validation dataset (wheat yield only).*

<b>Estimated Variable</b>	<b>Training data</b>		<b>Cross-Validation</b>		<b>Validation data</b>	
	<b>RMSE</b>	<b>R<sup>2</sup></b>	<b>RMSE</b>	<b>R<sup>2</sup></b>	<b>RMSE</b>	<b>R<sup>2</sup></b>
Wheat yield	0.61	0.91	0.60 (0.04)	0.91 (0.01)	0.62	0.90
Farmland Bird Richness	2.26	0.56	2.22 (0.1)	0.54 (0.04)	--	--

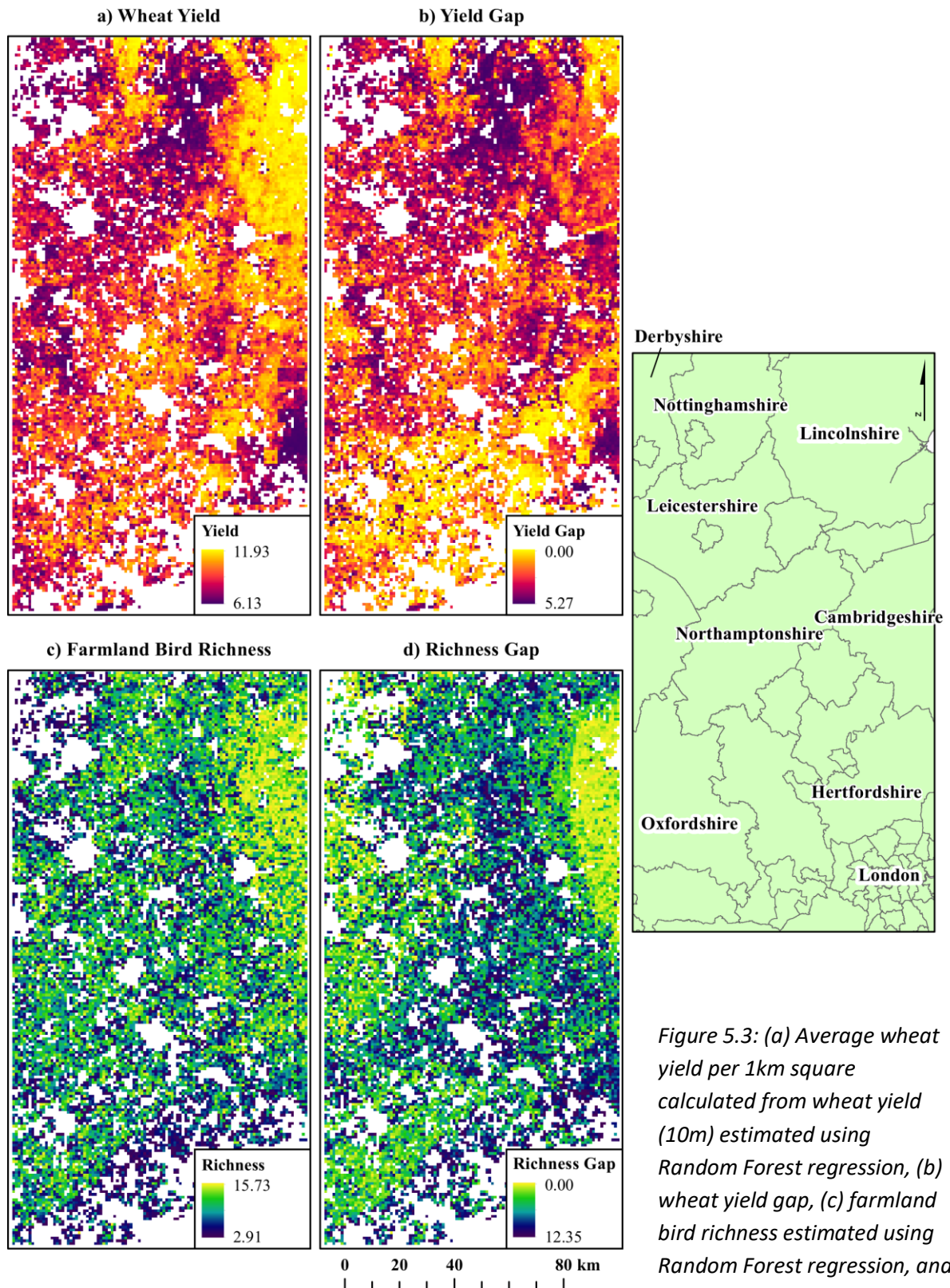


Figure 5.3: (a) Average wheat yield per 1km square calculated from wheat yield (10m) estimated using Random Forest regression, (b) wheat yield gap, (c) farmland bird richness estimated using Random Forest regression, and (d) farmland bird richness gap.

Gap values represent the difference between the potential values and the actual values, with potential values based on the maximum value within in each ITE land class. The colour scale for the gap maps has been inverted to aid comparison as, when considering performance, the ideal scenario is high actual yield/richness values, but low gap values. The white areas represent the 1km squares that contain less than 20% arable land. A map of counties within the study area is provided for spatial context.

To minimise the variability due to location and maximise the variability due to management practices, actual yield and richness values were transformed to gap values. These gap values represent the difference between what is attainable in a given area (potential values) and what is observed (actual values). The potential values were based on the maximum values in each ITE land class, ranging from 8.66 to 11.93 tonnes/ha for wheat yield and 6.84 to 15.73 for farmland bird richness. When considering gap values, it is important to remember that low values indicate better performance, as they suggest areas are closer to reaching their full potential.

The resulting yield and richness gaps indicate different spatial distributions, in terms of performance, compared to the actual yield and richness values. The area of high performance for yield in the north east of the region becomes less well defined when looking at the yield gap, while areas in the south perform better (figure 5.3b). The loss of spatial pattern is to be expected when we normalise for environmental conditions, whilst the variability due to the management practices is retained. Looking at the richness gap reveals a more clearly defined area of high performance in the north east and an improved performance to the north west, compared to the actual farmland bird richness (figure 5.3d). However, generally the richness gap varies unsystematically across the landscape. The yield gap values range from 0 to 5.27 tonnes/ha across the landscape, while richness gap values range from 0 to 12.35.

### **5.3.2 Performance Feature Space and Map**

The performance feature space created in this study is a novel feature space formed by plotting the wheat yield gap against the farmland bird richness gap. A continuous colour scale, representing relative performance, was created by assigning points two colours, one based on the yield gap value and one on the richness gap value (figure 5.4). From the spread of points within the feature space, it is clear that most arable areas have the potential for some improvement in terms of both richness and yield to equal the best performing arable areas in the landscape. However, it is

possible to make some broad distinctions between points, in terms of whether improving richness or yield should be prioritised, based on their relative position within the feature space.

In general terms, points in the lower left section of the feature space represent arable areas with the largest yield and richness gap, indicating there is room for improvement in terms of both yield (intensity) and richness (environmental quality). Points that fall in the upper left section represent arable areas with a small richness gap, but large yield gap, indicating these areas need to focus mostly on improving yield, rather than richness. Points located in the lower right section, represent arable areas with large richness gap, but small yield gap, indicating these areas need to focus mostly on improving richness. And finally, points in the upper right section represent arable areas which are performing best (small yield and richness gaps), indicating there is little or no room for improvement, relative to the other arable areas in the landscape.

These broad classifications are given to demonstrate how this feature space should be interpreted, but, as the feature space represents a continuous scale of agricultural performance, for the majority of arable areas there appears to be the potential to improve both richness and yield. It is important to remember that this feature space provides a relative, not absolute, measure of performance. As such it should not be interpreted as indicating that certain locations are definitively sustainable or unsustainable, rather it provides an indicator of the relative performance of arable areas compared to the landscape as a whole.

Future work could explore the benefits (to interpretation) of splitting the data into more distinct categories (e.g. quadrants) to aid farmers in their interpretation of the feature space and therefore help them identify where they should be focusing their efforts to maximise agricultural performance in terms of both intensity and environmental quality.

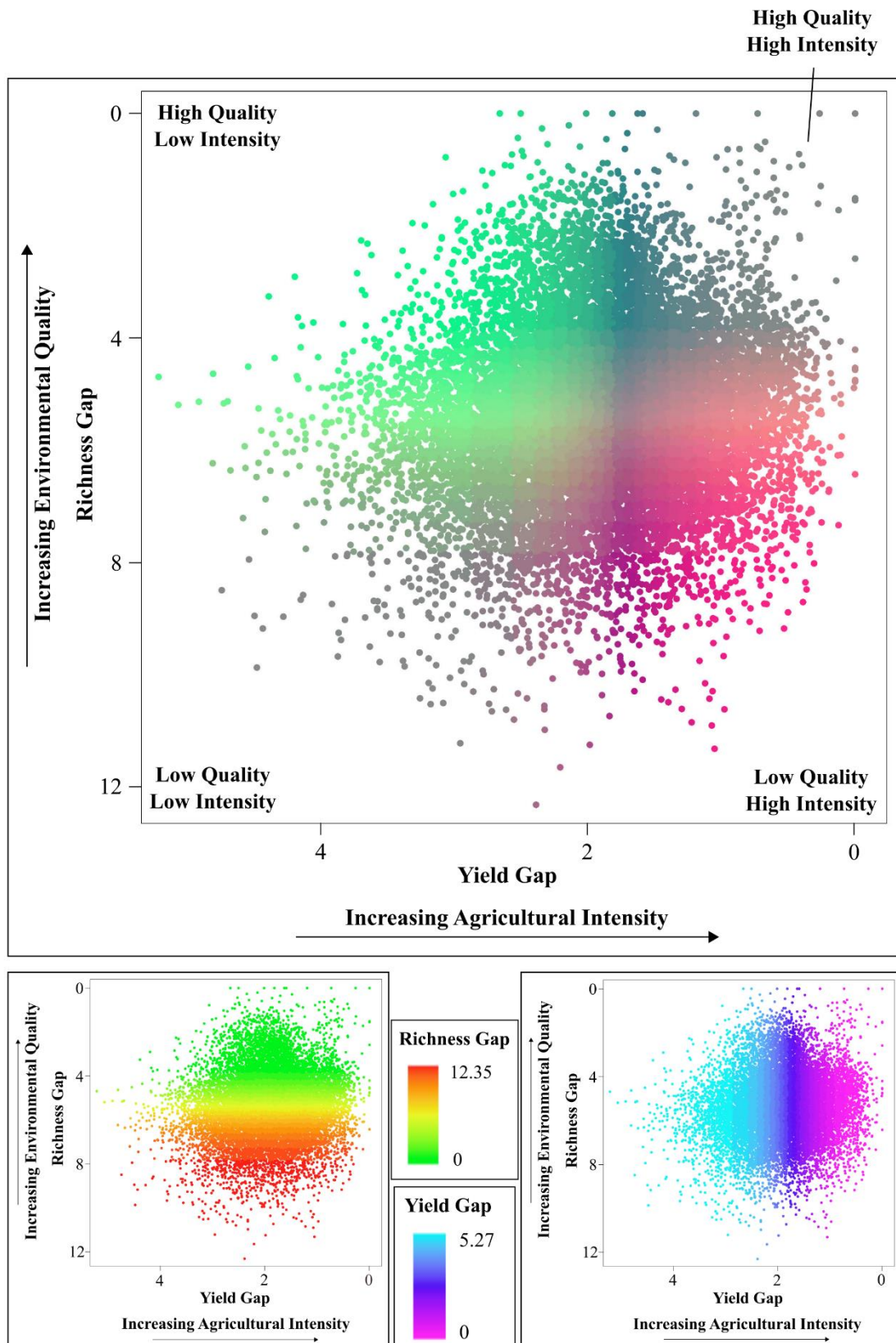


Figure 5.4: Performance feature space created by plotting values for the yield and richness gap for each 1km square. Points have been coloured with two different colour schemes one for yield and one for richness gap (shown in the bottom graphs) to produce a continuous scale to provide an indication of the variable performance level within the feature space.

To explore how farm performance varies in space, a map was created by overlaying the 1km yield gap and richness gap maps, using the same colour schemes as in figure 5.4. From this map, a number of spatial patterns were revealed (figure 5.5). Two main clusters of low performance (pink) squares can be seen, one located to the north, in the western part of Lincolnshire, the other in the southeast in Hertfordshire; these correspond to the squares with relatively large richness and yield gaps. Squares with large richness gaps, but small yield gaps (grey/green squares) are less clearly clustered, but can be found mostly in the southeast, north of London. Squares with small richness gap, but a large yield gap (blue/brown squares), can be found in two main clusters, the largest in the east around Cambridgeshire, with a smaller cluster in the west/northwest. Finally, there are two main clusters of high performance squares (pale green squares), those with small yield and richness gaps. These clusters are located in the northeast, east of Lincolnshire, and in the southwest around Oxfordshire and Buckinghamshire. While it is possible to identify these broad patterns of performance, there is still a lot of localised variability in over- or under-performance. This highlights how this methodology could be used to identify specific areas requiring further investigation or intervention to ensure they reach their potential.



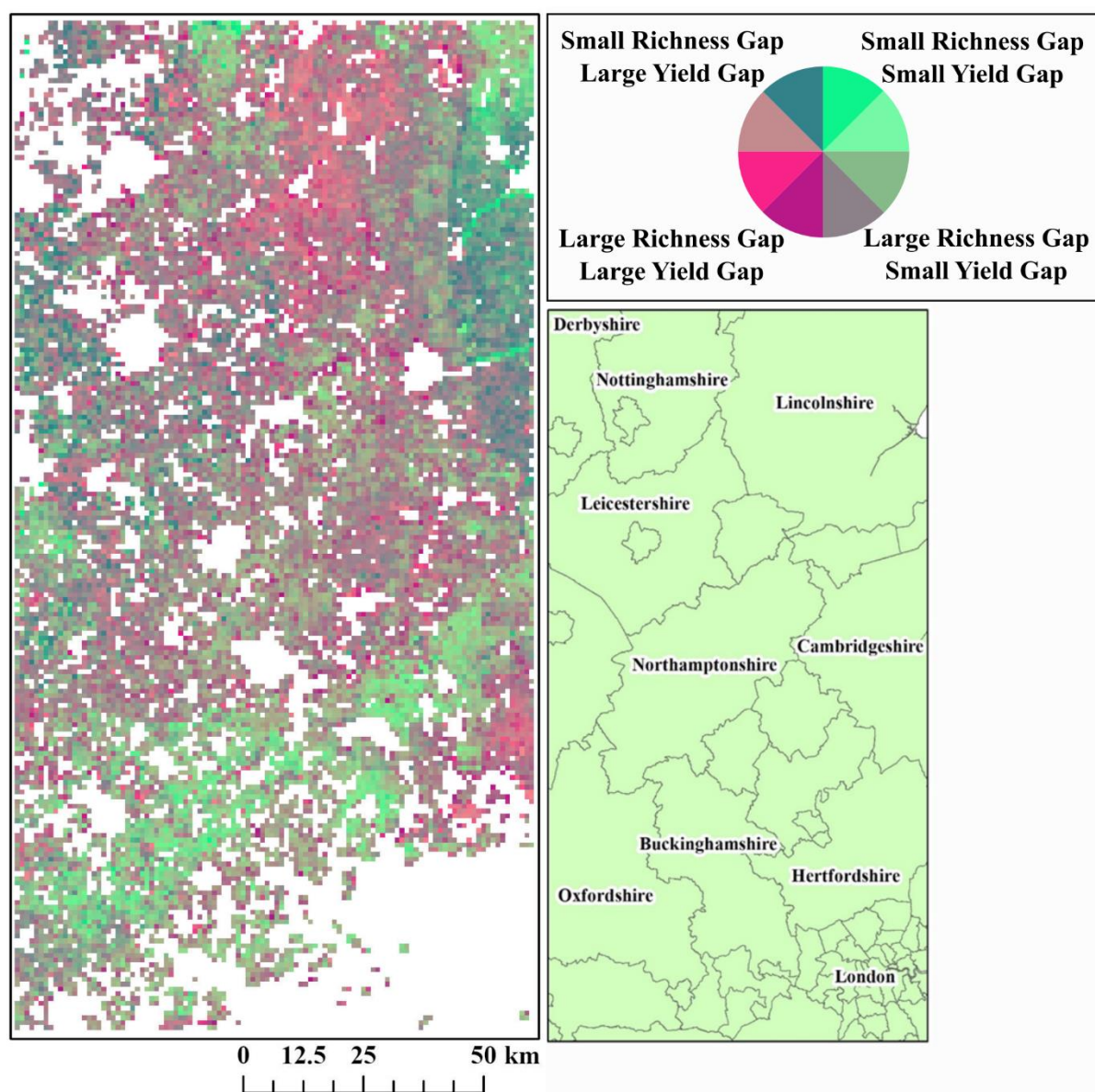


Figure 5.5: Map indicating the relative performance of 1km squares within the landscape. This map was created by overlaying the 1km yield gap and richness gap maps using the same colour schemes as shown in figure 5.4. The legend provided shows 8 distinct colours to make it easier to distinguish squares with different performance properties, in reality this is a continuous scale. White squares are those that contain less than 20% arable land. A map of counties within the study area is provided for spatial context.

## 5.4 Discussion

Being able to efficiently monitor the sustainable intensification of agriculture is critical for ensuring future food demands are met without detrimental environmental impacts, however suitable methods need to be developed. Reliance on traditional data sources such as field and farm



surveys limits the spatial and temporal resolution and coverage of current assessment attempts, affecting their reliability. However, as this study demonstrates, there is significant potential for using satellite data to enhance these assessments. Studies conducted using field data typically focus on a number of individual farms for which data are available (e.g. Elliott *et al.*, 2013; Firbank *et al.*, 2013). The availability of satellite data for all locations, recorded at regular intervals at a range of resolutions and spatial scales, means that farm-scale assessments can be scaled up to the landscape-scale. This allows us, as this study demonstrates, to assess agricultural performance not just for individual farms but for entire landscapes. Combining satellite data with appropriate *in situ* data as demonstrated would allow agricultural performance to be assessed anywhere in the world.

The novel feature space and map presented in this paper provides an effective way of assessing the relative performance of arable areas within a landscape. Using this feature space it is possible to identify arable areas where there appears to be potential for improvement in terms of intensity of the agriculture (i.e. increasing yield), the quality of the surrounding environment, or both. In terms of sustainable intensification, the goal would be for all arable areas to move towards the upper right section of the feature space, where the maximum potential yield and farmland bird richness for each location are achieved. The ability to assess performance of individual arable areas and the landscape in general means that a feature space such as this could provide useful information to multiple stakeholders to ensure progress towards sustainable intensification. Individual farmers, for example, could use this feature space to identify potential areas for improvement to help them to determine where investments and changes to practices should be targeted. It could also allow them to see how their performance compares to that of their neighbours and the wider landscape. Crucially, because we use yield values and bird richness values, farmers could identify where their farms fall in the feature spaces presented above without needed to use EO data, using their own observations of bird richness and their own yield measurements.

Researchers, in collaboration with other stakeholders, could use this feature space to assist in identifying the most effective means of improving performance in specific locations. The feature space could be used to identify the best and worst performing farms in an area to develop best practice guidelines or to understand the impact of different management practices or agri-environment schemes. It could also help to identify the areas/countries where there is the biggest room for improvement both in terms of agricultural production and environmental quality, therefore guiding national and international policies/strategies.

The feature space could also prove useful to governments, aiding them in managing farming subsidy schemes. The new farming subsidy bill for England, for example, has an emphasis on delivering environmentally friendly outcomes, such as habitat enhancement, as well as cultivating land (Stokstad, 2020). A feature space such as the one presented here could help the government to monitor the outcomes of practices implemented by individual farmers, ensuring that farms producing the biggest environmental benefits and meeting requirements of these schemes are rewarded appropriately. The success of specific management techniques, in terms of environmental benefits, will be location dependent. It is therefore important for governments to be able to assess the outcome of the techniques implemented to make sure they are actually providing a benefit, rather than assuming a 'one size fits all' approach to farming.

At present the current feature space assesses performance for 2016. However, intensification is a process rather than an end state, hence it must be monitored over time. EO data could make a significant contribution in this respect, with the availability of regular, repeat satellite imagery allowing change over time to be assessed at various temporal scales. The ability to monitor the performance trajectory of individual farms and countries as a whole will help to ensure we are progressing towards sustainable intensification. The availability of long-term data (e.g. 30-40 years from Landsat data archive) will also allow the establishment of a baseline against which changes over the long- and short-term can be assessed. This will make it easier to identify the impact farming

has had in the past and therefore determine which management techniques should be phased out and which should be maintained/prioritised.

It is important to note that in its current state, the feature space presented here allows a relative, rather than absolute, assessment of performance to be made. While this undoubtedly provides useful information, to ensure we are moving towards sustainable intensification we need to be able to say definitively whether different areas/farms are sustainable or unsustainable. To be able to determine where agricultural performance sits in terms of true sustainability, additional information is required to calibrate the performance feature space to provide an absolute assessment of sustainable intensification. Calibration will require a global data collection campaign alongside consultation/collaboration with international agricultural and environmental experts to determine locally appropriate sustainability criteria and thresholds. Data collection would be required on multiple scales with intensive data collection on a small number of farms monitoring a large number of indicators, and extensive data collection with a smaller number of measurements conducted on a large number of farms. The ground data required to train satellite data-based models will be of interest to multiple research areas, not simply sustainable intensification. Soil carbon, for example, is of considerable interest to climate change scientists because of its role in minimising the impact of climate change. This property, as an indicator of soil quality, would also prove useful in indicating the impact of agricultural management on environmental quality. Some of this data may already exist, so part of the challenge would involve discovering and integrating existing data. Collection of *in situ* data for these indicators could therefore be conducted in collaboration with other schemes focussed on tackling the big environmental issues affecting humanity.

Part of this calibration/data collection process will require consideration of the types of indicators which should be monitored, and the spatial and temporal scales over which these vary and should therefore be assessed. The present study utilises one indicator of agricultural intensity

and one indicator of environmental quality. However, these issues are multi-faceted and many potential indicators exist. From satellite data it is possible to derive multiple indicators of both agricultural intensity and environmental quality, with the latter covering aspects of vegetation and ecosystem health, soil and water quality, and biodiversity (see Hunt *et al.* (2019b, Chapter 2) for a review). Inclusion of multiple satellite-derived indicators could allow more in depth, reliable assessments to be made.

Future work should explore the application of the method presented here to produce performance feature spaces for other areas. This is feasible because multiple countries, such as the United States and France, conduct national surveys similar to that of the British Breeding Bird Survey. Within developed countries the use of yield monitor-equipped combine harvesters are also common. The global coverage and consistent nature of satellite data (e.g. image resolution, data quality) mean that with access to these training datasets, performance feature spaces could be created. Such work would allow us to assess the efficacy of this approach in other countries, with different environmental conditions and agricultural management approaches. Meeting future food demand is a global issue. Being able to assess agricultural performance globally, on multiple scales and compare between countries is therefore crucial to allow us to identify the areas locally and globally where there is the most potential for improvement in terms of both agricultural production and environmental quality. This ability to monitor agricultural performance across multiple scales and diverse agricultural locations, at various temporal resolutions, could facilitate the creation of an operational sustainable intensification monitoring system combining satellite and *in situ* data.

## **5.5 Conclusion**

This study presents an innovative new feature space that can be used to assess the relative performance of arable areas on a landscape-scale, created by combining two novel satellite-derived indicators of agricultural intensity and environmental quality. The creation of this feature space demonstrates how integrating satellite data with *in situ* data can efficiently scale up field

observations to provide performance assessments over large areas which could not be obtained with field data alone. This feature space can be used to assess the relative performance of individual farms relative to their surrounding landscape and identify locations where there is most room for improvement in terms of both agricultural intensity and environmental quality. This information will prove useful to individual farmers, governments and researchers alike as we seek to progress towards a more sustainable form of intensive agriculture.

## **6. Research Outcomes**

Before summarising the key research contributions from this thesis, it is important to note that this research would not have been possible even a few years ago. Advances in cloud computing platforms facilitated much more efficient data access and data processing. Extraction of the habitat productivity measures for every 1km square within Great Britain was carried out using Google Earth Engine (Gorelick *et al.*, 2017). In the past, such a process would have been incredibly time consuming, with the need to download and process each image individually. This would have restricted the area and timescale that could be covered, and the number of metrics that could have been derived, thus limiting the accuracy with which bird richness could be estimated. With Google Earth Engine, however, all processing can be conducted using an internet browser and cloud-processing, meaning only the results need to be downloaded, significantly expanding the potential of the area, timescale and variables that can be explored.

Additionally, the move to free access to all Landsat data (Wulder *et al.*, 2012) meant that sufficient data was available to allow the explanatory variables to be derived for the whole of Great Britain for multiple months across multiple years. The launch of the Sentinel-2 satellites (Drusch *et al.*, 2012), starting in 2015, also played a key role, providing data of high resolution to assess within-field yield variability. Without these freely available sources of data, the bird richness and wheat yield data products could not have been created to allow agricultural performance to be assessed on a landscape scale. With EO data increasing exponentially and cloud computer power increasing, it is becoming more realistic to do regular large-scale land surface monitoring (Woodcock *et al.*, 2020) opening up the potential to monitor sustainable intensification globally. The approach developed in this thesis provides conceptual and analytical foundations for doing this.

### **6.1 Key Contributions of Thesis**

While there are many useful outcomes of this research project, there are a number of key contributions that should be highlighted. Firstly, the literature review presented in chapter 2 is the

first paper to consider the use of EO for monitoring sustainable intensification, demonstrating the wide variety of measures of agricultural intensity and environmental quality that could be derived from satellite data. Bringing together this information in one place creates a useful reference article for researchers of multiple disciplines, not only those interested specifically in sustainable intensification. For example, the potential to monitor variables such as soil organic carbon using EO would be of interest to climate change scientists, while variables such as species richness would be of interest to ecologists. This chapter presents a general overview of the capabilities and benefits of remote sensing, making it accessible to both remote sensing scientists and non-remote sensing scientists alike. It also presents a set of recommendations, including the development of a comprehensive new set of Essential Sustainable Intensification Variables, for the creation of an operational EO-based assessment system which could facilitate global monitoring of sustainable intensification. Creating the ESIVs forms a key step in the creation of this system, ensuring the major dimensions of agricultural and environmental change are captured, allowing sustainable intensification to be monitored.

Chapter 3 demonstrates how high resolution yield data from a small number of fields can be scaled up using satellite data to provide an assessment of within-field yield variability on a landscape scale. While previous efforts had been made to assess within-field yield variability using satellite data (e.g. Burke and Lobell, 2017; Lambert *et al.*, 2017), these studies typically lacked field data of sufficient resolution to be able to assess the accuracy of the maps produced. With its inclusion of combine harvester data, this work demonstrates that yield variability can be estimated with a high degree of accuracy using both Sentinel-2 data on its own and by combining it with environmental data. The use of Random Forest regression in this study demonstrated the value of this user friendly approach to model building. With Random Forest, all potentially important variables can be added and the algorithm assesses the order and way in which variables should be used, dealing internally with the complex interactions and relationships between variables. This reduces the input required by the user, helping to make it a more accessible and transferable approach.

One criticism of Random Forest, and similar algorithms, is the black box nature of the modelling approach, which makes it hard to understand exactly what the model is doing. Chapter 4, however, goes beyond simple estimation to begin to investigate the models created by Random Forest. This highlights how using variable importance and feature contribution analysis techniques allows us to gain an understanding of the strength and nature of the relationships between individual explanatory variables and the response variable. In this case, the work demonstrated the non-linearity of the relationships between key habitat productivity and heterogeneity variables, and bird species richness and diversity. Being able to break down these models allows better interpretation of what is going on in the model, providing information that can be applied in the real world. For example, details of the nature of the variables impacting bird species richness can be used to identify the environmental elements that are most important for promoting high richness, which can inform management practices. Such analysis techniques help to turn these black box models into more open, interpretable and useful white box models.

The culmination of this work, the EO-based assessment of relative agricultural performance, is the first demonstration of this application of EO. Being able to monitor agricultural performance is crucial for ensuring that attempts to sustainably intensify agriculture are successful. Reliance solely on traditional data collection techniques, such as farm surveys, to provide data, results in performance assessments of limited spatial and temporal coverage. With long-term, consistent data coverage across the globe, there is the potential for EO to play a key role in monitoring agricultural intensification building on international efforts such as GeoGLAM. Chapter 5 provides proof of this potential, demonstrating how EO can monitor relative agricultural performance over large areas. More work is needed to create a system to routinely assess sustainable intensification, requiring, for example, monitoring to be conducted over multiple years to track changes. However, the research presented in this thesis represents the first important step in the process of creating an operational EO-based sustainable intensification monitoring system.



## **6.2 Estimation Uncertainty**

Based on the  $R^2$  and RMSE values (chapters 3-5), wheat yield and farmland bird richness are predicted with relatively high accuracy. However, there are of course a number of sources of uncertainty which must be considered. With the yield data there are uncertainties relating to the way in which the data are collected; these include harvest lag time, GPS positional accuracy, and the use of different equipment between fields. Harvest lag time is the delay between the crop being harvested and it actually being measured at the yield monitor. This delay means that the yield measurement recorded may not match the actual harvest position. The length of this delay varies depending on the specific combine harvester model, yield monitor and GPS receiver used. Various estimates of this delay time have been made such as 6-18s (Ping and Dobermann, 2005) and 8-24s (Griffin *et al.*, 2007). Many manufacturers recommend assuming a delay time of 12s (Lyle *et al.*, 2014) when correcting for positional errors. However, the appropriateness of this assumption will vary depending on the equipment used. Hence, whilst harvest time lag can be corrected to some extent, as was done in chapter 3, some uncertainty remains.

The GPS positional accuracy attainable with agricultural grade receivers has been determined by several studies to vary between 1 to 3m (Lyle *et al.*, 2014). In some cases, positional errors may affect all points, leading to measurements offset from the field boundary; this can be solved with by simply applying a positional offset to the whole dataset. In other cases, errors may apply to only a small number of points, which can manifest itself as a false representation of the actual harvest path. The correct harvest path can be identified, for example through use of simple geometrical adjustments, but this relies on the assumption that certain points are correctly positioned, while others are erroneous (Lyle *et al.*, 2014). Therefore, while corrections can be applied to address positional accuracy, there will still be some uncertainty in the resulting datasets.

Due to the large geographical area covered by the sample fields, different equipment will have been used for the yield data collection. As such, the accuracy of the yield monitor and GPS are

likely to differ, even between the same make and model. In this work, when applying corrections to the data, the same criteria were applied to each field. The potential variation in the accuracy of the data collection sensors, means that the appropriateness of the thresholds used for correction may differ slightly between fields. This means that the corrected data presented for some fields may be more reliable than for others. With information on the specific sensors used for data collection, it may be possible to tailor the corrections applied to each field to improve accuracy. However, in this case this information was not available, so this potential source of uncertainty must be taken into consideration when comparing the yield values between individual pixels.

Similarly, the model does not explicitly account for differences in management practices, the start of the growing season or growth condition across the wider landscape. It is assumed that these are, to some extent, captured in the training data, but at present independent data is not available to assess this assumption.

As well as the potential uncertainties that exist in the response variable training data (yield and bird data), there are also some potential sources of uncertainty within the input feature variable data sets. Within the satellite data, for example, there are various potential sources of uncertainty such as possible sensor glitches and atmospheric disturbance (e.g. haze) within the images. Using an empirical model will account for some of this uncertainty, as sources of noise such as haze in the satellite image will be taken into consideration when building the model. However, it will not account for all of the uncertainty.

Another source of uncertainty comes from the accuracy of the land cover and crop maps. Misclassification of the land cover contained within each pixel would lead to inaccuracies in the derivation of the habitat heterogeneity variables, affecting the accuracy of the random forest models. Errors in crop type identification could lead to the yield estimation model being applied to fields which do not contain wheat. As different crops have different characteristics and growth

patterns, applying a model built using wheat yield data to other crops would likely lead to inaccurate yield measurements.

The Met Office climate data used for yield estimation was derived from a set of daily observations of temperature and rainfall interpolated to a uniform 5km grid (Met Office, 2017). The reliability of these data will depend on the accuracy of the interpolation. While the interpolation process is designed to account for effects such as latitude, longitude, altitude and coastal influence, there are likely to be some inaccuracies which will be propagated through the model.

In addition, there is clearly a discrepancy between the resolution at which the climate data is available (5km) and the resolution at which yield is being estimated (10m). The comparably low resolution of the climate data will mean that the data will not be able to fully reflect the micro-climate variations affecting crop growth in different locations. The inability to accurately determine small-scale climate variations will obviously introduce some uncertainty into the yield estimation process. In the absence of higher resolution data that could possibly reduce uncertainty, the Met Office data still appears to contribute valuable information to the yield estimation model, as demonstrated in chapter 3.

Availability of suitable data may not only introduce uncertainty into the modelling process, it also affects our ability to evaluate this uncertainty. While the final yield map was validated using data not used to train the Random Forest model, it would have been beneficial to perform further validation using data from an independent source to provide an independent assessment of the accuracy of the map. However, at present there are no suitable datasets available to carry out such a validation. Defra, for example, does collect yield data, however the spatial aspect of this dataset is insufficient for validation, with yield values assigned to postcodes rather than specific fields. High resolution spatially defined data does exist, however the commercially sensitive nature of this data means that it is not made readily available by the farmers collecting it.

The bird richness/diversity models were unable to explain around 30% of the variation, which may seem high. However, when interpreting this performance it is important to consider the wider research context of these results. Within ecological and biological studies, randomness and noise in data can lead to a significant reduction in the amount of variance that can be explained (Møller and Jennions, 2002). When attempting to predict bird richness/diversity, we are dealing with living organisms that are affected by a vast array of biotic and abiotic factors. The complexity and dynamic nature of these interactions significantly reduce the amount of variance that it is possible to explain. Møller and Jennions (2002), for example, found that from 43 published ecological studies the mean amount of variance explained ( $r^2$ ) was 2.51-5.42%. As such while the bird richness/diversity models do not explain as much variance as might typically be expected from a model that is performing well, in terms of ecological modelling, these models are performing remarkably well.

While these bird richness/diversity models do appear to be performing well, there are of course various sources of uncertainty which must be taken into consideration. One source of uncertainty within the BBS data results from the process of field data collection. The BBS involves volunteers conducting bird counts on two occasions during the survey period (Harris *et al.*, 2019). The number of each species recorded on each day will be affected by elements other than whether birds are actually present. For example, bird numbers will be affected by the weather conditions on the survey day and the time of day that the count is carried out. Additionally, the experience of the volunteer may affect their ability to correctly identify and record different species (Eglington *et al.*, 2010). As such, the data within the BBS, while an incredibly valuable and carefully managed data set, is likely to contain a relatively high level of noise. The relatively large uncertainty in this data is likely to limit the extent to which a model will be able to explain the variance in the data, limiting the maximum achievable  $R^2$  value. It is in this context that the values of  $R^2$  between 0.64 and 0.72 achieved here should be viewed.

Additionally, the relationship between the explanatory variables and individual species is likely to differ, based on the differing habitat requirements of these species. In chapter 4, all farmland birds were grouped to provide an estimate of species richness. As a result, while the observed richness in two squares may be the same, the species making up this richness value may differ. Different species will require different elements in the landscape to support their presence. As such, while two squares may have the same richness, the landscape elements observed may be different. Therefore, attempting to directly estimate species richness may make it harder for Random Forest to accurately determine the relationships between the explanatory variables and farmland bird richness, leading to uncertainty in the estimated values. To address this issue, future work could use satellite data to model the presence of individual species within each square, building species-specific Random Forest models to ensure the presence of all species is estimated for each 1km square with the highest possible accuracy. From this presence data, richness could then be calculated to provide the required biodiversity measure to give an indication of wider environmental quality.

### **6.3 Performance Assessment Scale**

If we are to meet future food demands, we need to maximise the agricultural performance of all fields and farms. To do this, we need to be able to accurately assess the performance of individual fields and farms. This thesis demonstrates that it is possible to estimate the relative performance of arable areas at 1km resolution. However, as each 1km square may contain multiple fields/farms, these squares represent the cumulative performance of these fields. If fields within these squares are managed in different ways this will make it harder to determine the relative performance of each individual field. To improve the accuracy of field-/farm-scale performance assessments, we therefore need to be able to produce a higher resolution performance map. As this thesis demonstrates, EO can be used to assess yield at a relatively high resolution (10m). Bird richness, however, can only be estimated at a resolution of 1km, due to the resolution of the training

data. Therefore, to provide a higher resolution estimate of overall agricultural performance, the resolution at which environmental quality is estimated would need to be improved.

One option would be to further explore the relationships between bird richness and the satellite-derived measures of habitat productivity and heterogeneity. In this thesis a preliminary assessment of variable importance and feature contribution of the different explanatory variables was made. Further exploration of this may allow a better understanding of the factors affecting species richness and the scales over which they are relevant. With this information, it may be possible to focus on the factors affecting richness, rather than richness itself, to provide an indicator of environmental quality within an area. In other words, by identifying the characteristics that make an area suitable for birds, we can obtain an indication of the characteristics that reflect high or low environmental quality. This may allow environmental quality to be assessed at higher resolutions, being no longer constrained by the resolution of the richness data, providing greater flexibility to explore the impacts of specific farms on the environment. Part of this work could involve exploring the relationships for individual species, rather than groups of species, as the exact nature of relationships are likely to vary between species.

Another option to address the issue of scale would be to explore the possibility of utilising alternative measures of environmental quality. Bird richness was used in this study as the indicator of environmental quality largely due to the availability of the BBS data for multiple years. Bird richness is recognised as an important indicator of wider biodiversity (Furness and Greenwood, 2013), and therefore environmental quality, hence this decision is justified. However, as chapter 2 demonstrates, there are multiple potential EO-based indicators of environmental quality, such as soil carbon content and vegetation health. The nature of these indicators, with training data collected at finer scales than the bird data, mean that they lend themselves to being assessed at a higher resolution. With access to suitable training and validation data, the methods presented in this thesis (i.e. Random Forest regression) could be used to create a range of environmental quality

indicators at a higher resolution. This would subsequently allow environmental quality, and therefore agricultural performance, to be assessed at a higher resolution, providing more accurate assessments of individual farm and field performances.

#### **6.4 Realising the Potential of EO to Assess Sustainable Intensification**

This thesis demonstrates how EO can be used to provide an assessment of relative agricultural performance on a landscape scale. While at this stage this does not translate into an actual assessment of the sustainability of agricultural intensification, this work forms an important step on the road to delivering an operational EO-based monitoring system. Moving towards the creation of such a system will require a number of steps. These include:

- ***Performance Map Validation.*** While the data sets used to create the performance feature space (i.e. farmland bird richness and wheat yield) can be validated using the BBS data and combine harvester yield data, it is currently not possible to validate the final performance feature space or map. Validation will require additional ground data to assess the agricultural performance in the landscape. This ground data must include multiple indicators, as agricultural intensity and environmental quality are both multi-faceted elements, as highlighted in chapter 2 of this thesis. Information regarding specific management practices applied to individual fields and farms will be required to allow an assessment as to how the performance map relates to what is actually happening on the ground. This need for ground data emphasises the fact the satellite data alone will not provide a solution to the issue of monitoring agricultural performance, and therefore monitoring sustainable intensification. Rather, satellite data will form one element of a larger monitoring system including intensive and extensive *in situ* data collection to ensure reliable assessments over diverse landscapes can be made on a global scale.
- ***Application of the method to the whole of GB.*** In this thesis, agricultural performance is assessed for an area covering 20,000km<sup>2</sup>. This area was selected largely due to the coverage

offered by the 2016 combine harvester yield data at the time this study was conducted.

Since this time, additional data for other years has become available covering a larger area.

As such, future work should explore the potential for estimating wheat yield for the entirety of Great Britain. Since bird richness has already been estimated for this entire area (chapter 4), scaling up the yield data would allow agricultural performance to be assessed for the whole of Great Britain.

- ***Assessment of agricultural performance over time.*** As intensification is a process rather than an end state, assessments of agricultural performance must be conducted over time. Therefore, an important next step will be to apply the methods presented here to other time periods to track change over time. In GB, the high resolution combine harvester data is available from 2015 onwards, while the BBS has been going since 1994. While this means it will not be possible to assess the impact of past changes in management practices, it will allow changes going forward to be monitored. This will help to ensure current practices are having the desired impacts in terms of both agricultural intensity and environmental quality. Also, applying the methods to other years will help to assess the temporal transferability of these methods and assess their effectiveness for tracking change.
- ***Application of the method in other countries.*** While the methods presented in this thesis were successful in GB, future work must explore their suitability in other areas, with different landscape characteristics and agricultural management practices. To be able to meet future food demands, we need to sustainably intensifying agriculture globally, not just in GB, hence we must ensure that we are able to reliably assess progress in all places.
- ***Identification of the most suitable indicators.*** In this thesis, wheat yield and farmland bird richness are used as indicators of agricultural intensity and environmental quality respectively. These variables were chosen partly due to the availability of suitable *in situ* data for training and validation. As chapter 2 demonstrates there are, however, a wide range of potential indicators that can be derived from EO that could be used to assess agricultural



performance, and therefore sustainable intensification. Consultation with experts from diverse disciplines is required to ensure the most appropriate and widely measurable indicators are chosen. This consultation could result in the creation of a set of Essential Sustainable Intensification Variables (ESIVs) as suggested in chapter 2, to ensure a consistent and efficient approach to monitoring sustainable intensification can be adopted globally.

## **6.5 Summary**

The aim of this thesis was to explore the role of EO in assessing the sustainable intensification of agriculture. The use of satellite data to scale up *in situ* data demonstrates how EO can be used to estimate within-field yield variability and bird richness using an accessible and transferable modelling approach. By combining these two datasets, a novel performance feature space was created demonstrating how EO can be used alongside *in situ* data to provide an assessment of relative agricultural performance on a landscape scale. This information would allow the identification of the best and worst performing farms, helping to identify the best management practices in different areas to inform future decisions. The demonstration of these methods to assess agricultural performance constitutes an important first step in the creation of an operational EO-based monitoring system to assess sustainable intensification and ensure we are able to meet future food demands.

## **References**

- AHDB, 2018. Wheat Growth Guide [WWW Document]. URL <https://cereals.ahdb.org.uk/media/185687/g66-wheat-growth-guide.pdf>
- AHDB, 2016. Exploiting yield maps and soil management zones [WWW Document]. URL <https://cereals.ahdb.org.uk/media/1182974/pr565-final-project-report.pdf>
- Agriculture For Impact (2013). Sustainable Intensification: A new paradigm for african agriculture. A 2013 Montpellier Panel Report. <https://ag4impact.org/publications/montpellier-panel-report2013/>
- Allahyari, M.S., Masouleh, Z.D., Koundinya, V., 2016. Implementing Minkowski fuzzy screening, entropy, and aggregation methods for selecting agricultural sustainability indicators. *Agroecol. Sustain. Food Syst.* 40, 277–294. <https://doi.org/10.1080/21683565.2015.1133467>
- Andrew, M.E., Wulder, M.A., Nelson, T.A., 2014. Potential contributions of remote sensing to ecosystem service assessments. *Prog. Phys. Geogr.* 38, 1–26. <https://doi.org/10.1177/0309133314528942>
- Anuta, P.E., MacDonald, R.B., 1971. Crop surveys from multiband satellite photography using digital techniques. *Remote Sens. Environ.* 2, 53–67. [https://doi.org/10.1016/0034-4257\(71\)90077-0](https://doi.org/10.1016/0034-4257(71)90077-0)
- Armitage, R.P., Ramirez, F.A., Danson, F.M., Ebenezer, Y., 2013. Probability of cloud-free observation conditions across Great Britain estimated using MODIS cloud mask. *Remote Sens. Lett.* 4, 427–435. <https://doi.org/10.1080/2150704X.2012.744486>
- Aue, B., Diekötter, T., Gottschalk, T.K., Wolters, V., Hotes, S., 2014. How High Nature Value (HNV) farmland is related to bird diversity in agro-ecosystems - Towards a versatile tool for biodiversity monitoring and conservation planning. *Agric. Ecosyst. Environ.* 194, 58–64. <https://doi.org/10.1016/j.agee.2014.04.012>
- Azzari, G., Jain, M., Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sens. Environ.* 202, 129–141. <https://doi.org/10.1016/j.rse.2017.04.014>
- Balbi, S., del Prado, A., Gallejones, P., Geevan, C.P., Pardo, G., Pérez-Miñana, E., Manrique, R., Hernandez-Santiago, C., Villa, F., 2015. Modeling trade-offs among ecosystem services in agricultural production systems. *Environ. Model. Softw.* 72, 314–326. <https://doi.org/10.1016/j.envsoft.2014.12.017>
- Barnes, A.P., Thomson, S.G., 2014. Measuring progress towards sustainable intensification: How far can secondary data go? *Ecol. Indic.* 36, 213–220. <https://doi.org/10.1016/j.ecolind.2013.07.001>
- Barnes, E.M., Baker, M.G., 2000. Multispectral data for mapping soil texture: possibilities and limitations. *Appl. Eng. Agric.* 16, 731–741. <https://doi.org/10.13031/2013.5370>
- Battude, M., Al Bitar, A., Morin, D., Cros, J., Huc, M., Marais Sicre, C., Le Dantec, V., Demarez, V., 2016. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sens. Environ.* 184, 668–681. <https://doi.org/10.1016/j.rse.2016.07.030>
- Bauer-Marschallinger, B., Paulik, C., Hochstätter, S., Mistelbauer, T., Modanesi, S., Ciabatta, L., Massari, C., Brocca, L., Wagner, W., 2018. Soil moisture from fusion of scatterometer and SAR: Closing the scale gap with temporal filtering. *Remote Sens.* 10, 1–26. <https://doi.org/10.3390/rs10071030>

- Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* 114, 1312–1323. <https://doi.org/10.1016/j.rse.2010.01.010>
- Benton, T.G., Vickery, J.A., Wilson, J.D., 2003. Farmland biodiversity: Is habitat heterogeneity the key? *Trends Ecol. Evol.* 18, 182–188. [https://doi.org/10.1016/S0169-5347\(03\)00011-9](https://doi.org/10.1016/S0169-5347(03)00011-9)
- Binder, C.R., Feola, G., Steinberger, J.K., 2010. Considering the normative, systemic and procedural dimensions in indicator-based sustainability assessments in agriculture. *Environ. Impact Assess. Rev.* 30, 71–81. <https://doi.org/10.1016/j.eiar.2009.06.002>
- Binder, C.R., Wiek, A., 2006. The role of transdisciplinary processes in sustainability assessment of agricultural systems, in: Häni, F.J., Pintér, L., Herren, H.R. (Eds.), *Sustainable Agriculture: From Principles to Common Practice. Proceedings and Outputs of the First Symposium of the International Forum on Assessing Sustainability in Agriculture (INFASA)*, March 16, 2006, Bern, Switzerland. International Institute for Sustainable Development, pp. 33–48.
- Bingfang, W., 2006. Introduction of China Crop Watch system with remote sensing, in: *ISPRS Archives. XXXVI-8/W48: Remote Sensing Support to Crop Yield Forecast and Area Estimates*. Stresa, Italy, pp. 15–18.
- Biradar, C.M., Xiao, X., 2011. Quantifying the area and spatial distribution of double-and triple-cropping croplands in India with multi-temporal MODIS imagery in 2005. *Int. J. Remote Sens.* 32, 367–386. <https://doi.org/10.1080/01431160903464179>
- Birrell, S.J., Sudduth, K.A., Borgelt, S.C., 1996. Comparison of sensors and techniques for crop yield mapping. *Comput. Electron. Agric.* 14, 215–233. [https://doi.org/10.1016/0168-1699\(95\)00049-6](https://doi.org/10.1016/0168-1699(95)00049-6)
- Boelman, N.T., Stieglitz, M., Rueth, H.M., Sommerkorn, M., Griffin, K.L., Shaver, G.R., Gamon, J.A., 2003. Response of NDVI, biomass, and ecosystem gas exchange to long-term warming and fertilization in wet sedge tundra. *Oecologia* 135, 414–421. <https://doi.org/10.1007/s00442-003-1198-3>
- Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., Zemp, M., 2014. The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. *Bull. Am. Meteorol. Soc.* 95, 1431–1443. <https://doi.org/10.1175/bams-d-13-00047.1>
- Bommarco, R., Kleijn, D., Potts, S.G., 2013a. Ecological intensification: Harnessing ecosystem services for food security. *Trends Ecol. Evol.* 28, 230–238. <https://doi.org/10.1016/j.tree.2012.10.012>
- Bommarco, R., Kleijn, D., Potts, S.G., 2013b. Ecological intensification : harnessing ecosystem services for food security. *Trends Ecol. Evol.* 28, 230–238. <https://doi.org/10.1016/j.tree.2012.10.012>
- Bontemps, S., Arias, M., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Matton, N., Morin, D., Popescu, R., Rabaute, T., Savinaud, M., Sepulcre, G., Valero, S., Ahmad, I., Bégué, A., Wu, B., de Aballeyra, D., Diarra, A., Dupuy, S., French, A., Akhtar, I. ul H., Kussul, N., Lebourgeois, V., Page, M. Le, Newby, T., Savin, I., Verón, S.R., Koetz, B., Defourny, P., 2015. Building a data set over 12 globally distributed sites to support the development of agriculture monitoring applications with Sentinel-2. *Remote Sens.* 7, 16062–16090. <https://doi.org/10.3390/rs71215815>
- Bonthoux, S., Lefèvre, S., Herrault, P.-A., Sheeren, D., 2018. Spatial and Temporal Dependency of NDVI Satellite Imagery in Predicting Bird Diversity over France. *Remote Sens.* 10, 1136. <https://doi.org/10.3390/rs10071136>

- Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: The US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto Int.* 26, 341–358. <https://doi.org/10.1080/10106049.2011.562309>
- Boschetti, M., National, I., Stroppiana, D., National, I., Confalonieri, R., 2014. Estimation of rice production at regional scale with a Light Use Efficiency model and MODIS time series  
Estimation of rice production at regional scale with a Light Use Efficiency model and MODIS time series. <https://doi.org/10.5721/ItJRS20114335>
- Box, E.O., Holben, B.N., Kalb, V., 1989. Accuracy of the AVHRR vegetation index as a predictor of biomass, primary productivity and net CO<sub>2</sub> flux. *Vegetatio* 80, 71–89.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32.
- Buckwell, A., Nordang Uhre, A., Williams, A., Poláková, J., H Blum, W.E., Schiefer, J., Lair, G.J., Heissenhuber, A., Schießl, P., Krämer, C., Haber, W., 2014. The Sustainable Intensification of European Agriculture. A Review Sponsored by the RISE Foundation.
- Bunce, R.G.H., Barr, C.J., Gillespie, M.K., Howard, D.C., 1996. The ITE Land classification: Providing an environmental stratification of Great Britain. *Environ. Monit. Assess.* 39, 39–46. <https://doi.org/10.1007/BF00396134>
- Bunce, R.G.H., Howard, D.C., Clarke, R.T., Lane, A.M.J., 1991. ITE Land Classification: classification of all 1km squares in GB.
- Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc. Natl. Acad. Sci.* 114, 2189–2194. <https://doi.org/10.1073/pnas.1616919114>
- Butchart, S.H.M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.-F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.C., Watson, R., 2010. Global Biodiversity : Indicators of Recent Declines. *Science* (80-. ). 328, 1164–1169.
- Butler, S.J., Freckleton, R.P., Renwick, A.R., Norris, K., 2012. An objective, niche-based approach to indicator species selection. *Methods Ecol. Evol.* 3, 317–326. <https://doi.org/10.1111/j.2041-210X.2011.00173.x>
- Campbell, B.M., Thornton, P., Zougmore, R., van Asten, P., Lipper, L., 2014. Sustainable intensification: What is its role in climate smart agriculture? *Curr. Opin. Environ. Sustain.* 8, 39–43. <https://doi.org/10.1016/j.cosust.2014.07.002>
- Carrasco, L., Norton, L., Henrys, P., Siriwardena, G.M., Rhodes, C.J., Rowland, C.S., Morton, D., 2018. Habitat diversity and structure regulate British bird richness : Implications of non-linear relationships for conservation. *Biol. Conserv.* 226, 256–263. <https://doi.org/10.1016/j.biocon.2018.08.010>
- Carrasco, L., O'Neil, A.W., Daniel Morton, R., Rowland, C.S., 2019. Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. *Remote Sens.* 11, 288. <https://doi.org/10.3390/rs11030288>
- Cassman, K.G., Grassini, P., 2020. A global perspective on sustainable intensification research. *Nat. Sustain.* 3, 262–268. <https://doi.org/10.1038/s41893-020-0507-8>

- Caviglia, O.P., Andrade, F.H., 2010. Sustainable Intensification of Agriculture in the Argentinean Pampas: Capture and Use Efficiency of Environmental Resources. *Am. J. Plant Sci. Biotechnol.* 3, 1–8.
- Chamberlain, D.E., Fuller, R.J., Bunce, R.G.H., Duckworth, J.C., Shrubbs, M., 2000. Changes in the abundance of farmland birds in relation to the timing of agricultural intensification in England and Wales. *J. Appl. Ecol.* 37, 771–788.
- CLAAS, 2018. CLAAS Telematics [WWW Document]. URL <https://www.claas-telematics.com/>
- Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J., Skakun, S. V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens. Environ.* 219, 145–161. <https://doi.org/10.1016/j.rse.2018.09.002>
- Coluzzi, R., Imbrenda, V., Lanfredi, M., Simoniello, T., 2018. A first assessment of the Sentinel-2 Level 1-C cloud mask product to support informed surface analyses. *Remote Sens. Environ.* 217, 426–443. <https://doi.org/10.1016/j.rse.2018.08.009>
- Coops, N.C., Waring, R.H., Wulder, M.A., Pidgeon, A.M., Radeloff, V.C., Manne, L., Journal, S., May, N., 2009a. Bird diversity: a predictable function of satellite-derived estimates of seasonal variation in canopy light absorbance across the United States. *J. Biogeogr.* 36, 905–918. <https://doi.org/10.1111/j.1365>
- Coops, N.C., Wulder, M.A., Iwanicka, D., 2009b. Exploring the relative importance of satellite-derived descriptors of production, topography and land cover for predicting breeding bird species richness over Ontario, Canada. *Remote Sens. Environ.* 113, 668–679. <https://doi.org/10.1016/j.rse.2008.11.012>
- Costanza, R., Mageau, M., 1999. What is a healthy ecosystem? *Aquat. Ecol.* 33, 105–115. <https://doi.org/10.1023/A:1009930313242>
- Crossman, N.D., Burkhard, B., Nedkov, S., Willemsen, L., Petz, K., Palomo, I., Drakou, E.G., Martín-Lopez, B., McPhearson, T., Boyanova, K., Alkemade, R., Egoh, B., Dunbar, M.B., Maes, J., 2013. A blueprint for mapping and modelling ecosystem services. *Ecosyst. Serv.* 4, 4–14. <https://doi.org/10.1016/j.ecoser.2013.02.001>
- Crowder, D.W., Jabbour, R., 2014. Relationships between biodiversity and biological control in agroecosystems: Current status and future challenges. *Biol. Control* 75, 8–17. <https://doi.org/10.1016/j.biocontrol.2013.10.010>
- Čuček, L., Klemeš, J.J., Kravanja, Z., 2012. A review of footprint analysis tools for monitoring impacts on sustainability. *J. Clean. Prod.* 34, 9–20. <https://doi.org/10.1016/j.jclepro.2012.02.036>
- Dantsis, T., Douma, C., Giourga, C., Loumou, A., Polychronaki, E.A., 2010. A methodological approach to assess and compare the sustainability level of agricultural plant production systems. *Ecol. Indic.* 10, 256–263. <https://doi.org/10.1016/j.ecolind.2009.05.007>
- Davis, F.W., Seo, C., Zielinski, W.J., 2007. Regional variation in home-range-scale habitat models for fisher (*Martes pennanti*) in California. *Ecol. Appl.* 17, 2195–2213. <https://doi.org/10.1890/06-1484.1>
- DEFRA, 2019. Agriculture in the United Kingdom 2018 [WWW Document]. URL [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/848641/AUK\\_2018\\_09jul19a.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/848641/AUK_2018_09jul19a.pdf)
- Dietrich, J.P., Schmitz, C., Müller, C., Fader, M., Lotze-Campen, H., Popp, A., 2012. Measuring agricultural land-use intensity - A global analysis using a model-assisted approach. *Ecol. Modell.*

- 232, 109–118. <https://doi.org/10.1016/j.ecolmodel.2012.03.002>
- Diker, K., Heermann, D.F., Brodahl, M.K., Collins, F., 2004. Frequency Analysis of Yield for Delineating Yield Response Zones. *Precis. Agric.* 5, 435–444.
- Dillon, E.J., Hennessy, T., Buckley, C., Donnellan, T., Hanrahan, K., Moran, B., Ryan, M., 2016. Measuring progress in agricultural sustainability to support policy-making. *Int. J. Agric. Sustain.* 14, 31–44. <https://doi.org/10.1080/14735903.2015.1012413>
- Dillon, E.J., Hennessy, T., Hynes, S., 2010. Assessing the sustainability of Irish agriculture. *Int. J. Agric. Sustain.* 8, 131–147. <https://doi.org/10.3763/ijas.2009.0044>
- Donald, P.F., Green, R.E., Heath, M.F., 2001. Agricultural intensification and the collapse of Europe's farmland bird populations. *Proc. R. Soc. B* 268, 25–29. <https://doi.org/10.1098/rspb.2000.1325>
- Donald, P.F., Green, R.E., Heath, M.F., 2001. Agricultural intensification and the collapse of Europe's farmland bird populations. *Proc. R. Soc. B Biol. Sci.* 268, 25–29. <https://doi.org/10.1098/rspb.2000.1325>
- Doraiswamy, P.C., Sinclair, T.R., Hollinger, S., Akhmedov, B., Stern, A., Prueger, J., 2005. Application of MODIS derived parameters for regional crop yield assessment. *Remote Sens. Environ.* 97, 192–202. <https://doi.org/10.1016/j.rse.2005.03.015>
- Doran, J.W., Zeiss, M.R., 2000. Soil health and sustainability: managing the biotic component of soil quality. *Appl. Soil Ecol.* 15, 3–11. [https://doi.org/10.1016/S0929-1393\(00\)00067-6](https://doi.org/10.1016/S0929-1393(00)00067-6)
- Draeger, W., Benson, A.S., 1972. Application of ERTS-1 imagery to agricultural resource evaluation, in: *International Symposium on Remote Sensing of Environment*, Michigan, 8th, 1972, Proceedings: Ann Arbor, Environmental Research Institute of Michigan, v. 2. pp. 1467–1470.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>
- Duro, D.C., Girard, J., King, D.J., Fahrig, L., Mitchell, S., Lindsay, K., Tischendorf, L., 2014. Predicting species diversity in agricultural environments using Landsat TM imagery. *Remote Sens. Environ.* 144, 214–225. <https://doi.org/10.1016/j.rse.2014.01.001>
- Eaton, M., Noble, D., 2018. UK Biodiversity Indicators 2018. Technical background document: The wild bird indicator for the UK and England.
- Eaton, M.A., Noble, D.G., 2019. UK Biodiversity Indicators Technical paper : the wild bird indicator for the UK and England 1–17.
- Eglington, S.M., Davis, S.E., Joys, A.C., Chamberlain, D.E., Noble, D.G., 2010. The effect of observer experience on english Breeding Bird Survey population trends. *Bird Study* 57, 129–141. <https://doi.org/10.1080/00063650903440648>
- Egorov, A.V., Roy, D.P., Zhang, H.K., Li, Z., Yan, L., Huang, H., 2019. Landsat 4, 5 and 7 (1982 to 2017) Analysis Ready Data (ARD) observation coverage over the conterminous United States and implications for terrestrial monitoring. *Remote Sens.* 11, 447. <https://doi.org/10.3390/rs11040447>
- Elliott, J., Firbank, L.G., Drake, B., Cao, Y., Gooday, R., 2013. Exploring the Concept of Sustainable Intensification. ADAS/Firbank, LUPG Commissioned Report.
- ESA, 2018. Copernicus Open Access Hub [WWW Document]. URL <https://scihub.copernicus.eu/>

- Estel, S., Kuemmerle, T., Levers, C., Baumann, M., Hostert, P., 2016. Mapping cropland-use intensity across Europe using MODIS NDVI time series. *Environ. Res. Lett.* 11. <https://doi.org/10.1088/1748-9326/11/2/024015>
- European Commission, 2016. Monitoring Agricultural Resources (MARS) [WWW Document]. URL <https://ec.europa.eu/jrc/en/mars> (accessed 4.9.17).
- Evans, K.L., Greenwood, J.J.D., Gaston, K.J., 2005. Dissecting the species-energy relationship. *Proc. R. Soc. B Biol. Sci.* 272, 2155–2163. <https://doi.org/10.1098/rspb.2005.3209>
- Everingham, Y., Sexton, J., Skocaj, D., Inman-Bamber, G., 2016. Accurate prediction of sugarcane yield using a random forest algorithm. *Agron. Sustain. Dev.* 36. <https://doi.org/10.1007/s13593-016-0364-z>
- FAO, 2014. SAFA Sustainability Assessment of Food and Agriculture Systems: Guidelines (Version 2.2.40) [WWW Document]. URL <http://www.fao.org/3/a-i4113e.pdf>
- FAO, 2013. SAFA (Sustainability Assessment of Food and Agriculture Systems) Indicators.
- FAO, 2009. How to Feed the World in 2050. [http://www.fao.org/fileadmin/templates/wsfs/docs/expert\\_paper/How\\_to\\_Feed\\_the\\_World\\_in\\_2050.pdf](http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf).
- FAO, 2004. The ethics of sustainable agricultural intensification [WWW Document]. FAO Ethics Ser. 3. URL <http://www.fao.org/3/j0902e/j0902e00.htm>
- Farrell, A., Hart, M., 1998. What does sustainability really mean?: The search for useful indicators. *Environ. Sci. Policy Sustain. Dev.* 40, 4–31. <https://doi.org/10.1080/00139159809605096>
- Feng, X., Fu, B., Yang, X., Lü, Y., 2010. Remote sensing of ecosystem services: An opportunity for spatially explicit assessment. *Chinese Geogr. Sci.* 20, 522–535. <https://doi.org/10.1007/s11769-010-0428-y>
- Ferencz, C., Bognár, P., Lichtenberger, J., Hamar, D., Tarcsai, G., Timár, G., Molnár, G., Pásztor, S.Z., Steinbach, P., Székely, B., Ferencz, O.E., Ferencz-Árkos, I., 2004. Crop yield estimation by satellite remote sensing. *Int. J. Remote Sens.* 25, 4113–4149. <https://doi.org/10.1080/01431160410001698870>
- Firbank, L.G., Elliott, J., Drake, B., Cao, Y., Gooday, R., 2013. Evidence of sustainable intensification among British farms. *Agric. Ecosyst. Environ.* 173, 58–65. <https://doi.org/10.1016/j.agee.2013.04.010>
- Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global Consequences of Land Use. *Science* (80-. ). 309, 570–574. <https://doi.org/10.1126/science.1111772>
- Footy, G.M., 2005. Mapping the richness and composition of British breeding birds from coarse spatial resolution satellite sensor imagery. *Int. J. Remote Sens.* 26, 3943–3956. <https://doi.org/10.1080/01431160500165716>
- Footy, G.M., 2004. Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. *Glob. Ecol. Biogeogr.* 13, 315–320. <https://doi.org/10.1111/j.1466-822X.2004.00097.x>
- Fuller, R.J., Hinsley, S.A., Swetnam, R.D., 2004. The relevance of non-farmland habitats, uncropped areas and habitat diversity to the conservation of farmland birds. *Ibis (Lond. 1859)*. 146, 22–31. <https://doi.org/10.1111/j.1474-919X.2004.00357.x>

- Fuller, R.J., Smith, K.W., Grice, P. V., Currie, F.A., Quine, C.P., 2007. Habitat change and woodland birds in Britain: Implications for management and future research. *Ibis* (Lond. 1859). 149, 261–268. <https://doi.org/10.1111/j.1474-919X.2007.00775.x>
- Fuller, R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., Thomson, A.G., Cox, R., Brown, N.J., Clarke, R.T., Rothery, P., Gerard, F.F., 2002. Countryside Survey 2000 Module 7 Land Cover Map 2000. Final Report.
- Furness, R.W., Greenwood, J.J.D., 2013. *Birds as Monitors of Environmental Change*. Springer Science & Business Media.
- Galli, A., Wiedmann, T., Ercin, E., Knoblauch, D., Ewing, B., Giljum, S., 2012. Integrating Ecological, Carbon and Water footprint into a “footprint Family” of indicators: Definition and role in tracking human pressure on the planet. *Ecol. Indic.* 16, 100–112. <https://doi.org/10.1016/j.ecolind.2011.06.017>
- Garnett, T., Appleby, M.C., Balmford, A., Bateman, I.J., Benton, T.G., Bloomer, P., Burlingame, B., Dawkins, M., Dolan, L., Fraser, D., Herrero, M., Hoffman, I., Smith, P., Thornton, P.K., Toulmin, C., Vermeulen, S.J., Godfray, H.C.J., 2013. Sustainable intensification in agriculture: premises and policies. *Science* (80-. ). 341, 33–34. <https://doi.org/10.1126/science.1234485>
- Garnett, T., Godfray, H.C.J., 2012. Sustainable intensification in agriculture: Navigating a course through competing food system priorities. A report on a workshop. Oxford, UK: Food Climate Research Network and the Oxford Martin Programme on the Future of Food, University of Oxford.
- GEO, 2019. GEOGLAM [WWW Document]. URL <http://www.geoglam.org/index.php/en/> (accessed 4.9.17).
- GEO, 2018. Stocktaking Overview of the G20 Global Agricultural Monitoring Initiative.
- Gevaert, C.M., García-Haro, F.J., 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sens. Environ.* 156, 34–44. <https://doi.org/10.1016/j.rse.2014.09.012>
- Gillespie, T.W., Foody, G.M., Rocchini, D., Giorgi, A.P., Saatchi, S., 2008. Measuring and modelling biodiversity from space. *Prog. Phys. Geogr.* 32, 203–221. <https://doi.org/10.1177/0309133308093606>
- Gitelson, A.A., 2004. Wide dynamic range vegetation index for remote quantification of crop biophysical characteristics. *J. Plant Physiol.* 161, 165–173.
- 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 Sens. Environ.* 58, 289–298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)
- Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett.* 30, n/a-n/a. <https://doi.org/10.1029/2002GL016450>
- Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food Security: The Challenge of Feeding 9 Billion People. *Science* (80-. ). 327, 812–818. <https://doi.org/10.1126/science.1185383>
- Godfray, H.C.J., Garnett, T., 2014. Food security and sustainable intensification. *Philos. Trans. R. Soc. B* 369, 2012073. <https://doi.org/10.1098/rstb.2012.0273>
- Goetz, S., Steinberg, D., Dubayah, R., Blair, B., 2007. Laser remote sensing of canopy habitat



- heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sens. Environ.* 108, 254–263. <https://doi.org/10.1016/j.rse.2006.11.016>
- Gómez-Limón, J.A., Sanchez-Fernandez, G., 2010. Empirical evaluation of agricultural sustainability using composite indicators. *Ecol. Econ.* 69, 1062–1075. <https://doi.org/10.1016/j.ecolecon.2009.11.027>
- Gons, H.J., Auer, M.T., Effler, S.W., 2008. MERIS satellite chlorophyll mapping of oligotrophic and eutrophic waters in the Laurentian Great Lakes. *Remote Sens. Environ.* 112, 4098–4106. <https://doi.org/10.1016/j.rse.2007.06.029>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Grassini, P., Eskridge, K.M., Cassman, K.G., 2013. Distinguishing between yield advances and yield plateaus in historical crop production trends. *Nat. Commun.* 4, 2918. <https://doi.org/10.1038/ncomms3918>
- Griffin, T., Brown, J., Lowenberg-DeBoer, J., 2007. Yield Monitor Data Analysis Protocol: A Primer in the Management and Analysis of Precision Agriculture Data. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.2891888>
- Griffiths, G.H., Lee, J., 2000. Landscape pattern and species richness; regional scale analysis from remote sensing. *Int. J. Remote Sens.* 21, 2685–2704. <https://doi.org/10.1080/01431160050110232>
- Griffiths, P., van der Linden, S., Kuemmerle, T., Hostert, P., 2013. A Pixel-Based Landsat Compositing Algorithm for Large Area Land Cover Mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 6, 2088–2101. <https://doi.org/10.1109/JSTARS.2012.2228167>
- Grisso, R.D., Jasa, P.J., Schroeder, M.A., Wilcox, J.C., 2002. Yield Monitor Accuracy: Successful Farming Magazine Case Study. *Appl. Eng. Agric.* 18, 147–151. <https://doi.org/doi:10.13031/2013.7775>
- Gustafson, E.J., 1998. Quantifying Landscape Spatial Pattern: What Is the State of the Art? *Ecosystems* 1, 143–156. <https://doi.org/10.1007/s100219900011>
- Häni, F., Braga, F., Stämpfli, A., Keller, T., Fischer, M., Porsche, H., 2003. RISE, a tool for holistic sustainability assessment at the farm level. *Int. Food Agribus. Manag. Rev.* 6, 78–90.
- Häni, F.J., Stämpfli, A., Gerber, T., Porsche, H., Thalmann, C., Studer, C., 2006. RISE: A tool for improving sustainability in agriculture - a case study with tea farms in southern India, in: Häni, F.J., Pintér, L., Herren, H.R. (Eds.), *Sustainable Agriculture: From Principles to Common Practice. Proceedings and Outputs of the First Symposium of the International Forum on Assessing Sustainability in Agriculture (INFASA)*, March 16, 2006, Bern, Switzerland. International Institute for Sustainable Development, pp. 121–148.
- Harris, S.J., Massimino, D., Eaton, M.A., Gillings, S., Noble, D.G., Balmer, D.E., Pearce-Higgins, J.W., Woodcock, P., 2019. The Breeding Bird Survey 2018, BTO Research Report 717.
- Harris, S.J., Massimino, D., Gillins, S.E., Eaton, M.A., Noble, D.G., Balmer, D.E., Procter, D., Pearce-Higgins, J.W., 2017. The Breeding Bird Survey 2016. Thetford.
- Hein, L., 2014. Biophysical Modelling and Analysis of Ecosystem Services in an Ecosystem Accounting Context DRAFT. <http://img.teebweb.org/wp-content/uploads/2017/01/ANCA-Tech-Guid-9.pdf>.
- Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2018. Disturbance-Informed

- Annual Land Cover Classification Maps of Canada's Forested Ecosystems for a 29-Year Landsat Time Series. *Can. J. Remote Sens.* 44, 67–87. <https://doi.org/10.1080/07038992.2018.1437719>
- Herzog, F., Steiner, B., Bailey, D., Baudry, J., Billeter, R., Bukáček, R., De Blust, G., De Cock, R., Dirksen, J., Dormann, C.F., De Filippi, R., Frossard, E., Liira, J., Schmidt, T., Stöckli, R., Thenail, C., Van Wingerden, W., Bugter, R., 2006. Assessing the intensity of temperate European agriculture at the landscape scale. *Eur. J. Agron.* 24, 165–181. <https://doi.org/10.1016/j.eja.2005.07.006>
- Heywood, V., 1995. *Global Biodiversity Assessment*. Cambridge University Press, Cambridge.
- Hinsley, S.A., Bellamy, P.E., 2000. The influence of hedge structure, management and landscape context on the value of hedgerows to birds: A review. *J. Environ. Manage.* 60, 33–49. <https://doi.org/10.1006/jema.2000.0360>
- Hobi, M.L., Dubinin, M., Graham, C.H., Coops, N.C., Clayton, M.K., Pidgeon, A.M., Radeloff, V.C., 2017. A comparison of Dynamic Habitat Indices derived from different MODIS products as predictors of avian species richness. *Remote Sens. Environ.* 195, 142–152. <https://doi.org/10.1016/j.rse.2017.04.018>
- Hollmann, R., Merchant, C.J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., de Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., van Roozendaal, M., Wagner, W., 2013. The ESA Climate Change Initiative: Satellite Data Records for Essential Climate Variables. *Bull. Am. Meteorol. Soc.* 94, 1541–1552. <https://doi.org/10.1175/bams-d-11-00254.1>
- Honnay, O., Piessens, K., Van Landuyt, W., Hermy, M., Gulinck, H., 2003. Satellite based land use and landscape complexity indices as predictors for regional plant species diversity. *Landsc. Urban Plan.* 63, 241–250. [https://doi.org/10.1016/S0169-2046\(02\)00194-9](https://doi.org/10.1016/S0169-2046(02)00194-9)
- Hooper, D.U., Adair, E.C., Cardinale, B.J., Byrnes, J.E.K., Hungate, B.A., Matulich, K.L., Gonzalez, A., Duffy, J.E., Gamfeldt, L., Connor, M.I., 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* 486, 105–108. <https://doi.org/10.1038/nature11118>
- Horning, N., 2014. randomForestPercentCover [WWW Document]. URL [https://bitbucket.org/rsbiodiv/randomforestpercentcover/src/6217badf2fbebcb9c3e319df2a81faa20b85fa2d/rf\\_percentCover.R?at=master&fileviewer=file-view-default](https://bitbucket.org/rsbiodiv/randomforestpercentcover/src/6217badf2fbebcb9c3e319df2a81faa20b85fa2d/rf_percentCover.R?at=master&fileviewer=file-view-default)
- Horton, M.L., Heilman, J.L., 1973. Crop identification using ERTS imagery, in: *NASA Goddard Space Flight Center, Symposium on Significant Results Obtained from ERTS-1*, New Carrollton, Maryland, March 1973, Proceedings, NASA SP-327, v.1, Section A. pp. 27–33.
- Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., Su, W., Zhang, X., Zhu, D., Wu, W., 2015. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agric. For. Meteorol.* 204, 106–121. <https://doi.org/10.1016/j.agrformet.2015.02.001>
- Huete, A., Didan, K., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213. <https://doi.org/10.3390/rs4103201>
- Huete, A.R., 1989. Soil influences in remotely sensed vegetation-canopy spectra, in: Asrar, G. (Ed.), *Theory and Applications of Optical Remote Sensing*. Wiley, New York, pp. 107–141.
- Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., Rowland, C.S., 2019a. High resolution wheat yield mapping using Sentinel-2. *Remote Sens. Environ.* 233, 111410. <https://doi.org/10.1016/j.rse.2019.111410>

- Hunt, M.L., Blackburn, G.A., Rowland, C.S., 2019b. Monitoring the Sustainable Intensification of Arable Agriculture : the Potential Role of Earth Observation. *Int. J. Appl. Earth Obs. Geoinf.* 81, 125–136. <https://doi.org/10.1016/j.jag.2019.05.013>
- Hunt, M.L., Blackburn, G.A., Siriwardena, G.M., Carrasco, L., Rowland, C.S., 2020. Satellite-derived environmental heterogeneity and productivity as indicators of bird diversity. Yet to be Submitt. Publ.
- Hurlbert, Haskell, 2003. The Effect of Energy and Seasonality on Avian Species Richness and Community Composition. *Am. Nat.* 161, 83. <https://doi.org/10.2307/3078884>
- Ines, A.V.M., Das, N.N., Hansen, J.W., Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sens. Environ.* 138, 149–164. <https://doi.org/10.1016/j.rse.2013.07.018>
- Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., Rodes, I., 2017. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sens.* 9, 95. <https://doi.org/10.3390/rs9010095>
- Intermap-Technologies, 2009. NEXTMap British Digital Terrain (DTM) Model Data by Intermap [WWW Document]. NERC Earth Obs. Data Centre, 2018. URL <https://catalogue.ceda.ac.uk/uuid/998a28d8a5ed4564863a0daa0f731e8d>
- Ishwaran, H., Kogalur, U.B., Gorodeski, E.Z., Minn, A.J., Lauer, M.S., 2010. High-dimensional variable selection for survival data. *J. Am. Stat. Assoc.* 105, 205–217. <https://doi.org/10.1198/jasa.2009.tm08622>
- Jackson, T.J., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., Doriaswamy, P., Hunt, E.R., 2004. Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. *Remote Sens. Environ.* 92, 475–482. <https://doi.org/10.1016/j.rse.2003.10.021>
- Jain, M., Mondal, P., DeFries, R.S., Small, C., Galford, G.L., 2013. Mapping cropping intensity of smallholder farms: A comparison of methods using multiple sensors. *Remote Sens. Environ.* 134, 210–223. <https://doi.org/10.1016/j.rse.2013.02.029>
- Jain, M., Srivastava, A.K., Balwinder-Singh, Joon, R.K., McDonald, A., Royal, K., Lisaius, M.C., Lobell, D.B., 2016. Mapping smallholder wheat yields and sowing dates using micro-satellite data. *Remote Sens.* 8, 1–18. <https://doi.org/10.3390/rs8100860>
- Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.M., Gerber, J.S., Reddy, V.R., Kim, S.H., 2016. Random forests for global and regional crop yield predictions. *PLoS One* 11, 1–15. <https://doi.org/10.1371/journal.pone.0156571>
- Jin, Z., Azzari, G., Burke, M., Aston, S., Lobell, D.B., 2017a. Mapping smallholder yield heterogeneity at multiple scales in eastern Africa. *Remote Sens.* 9. <https://doi.org/10.3390/rs9090931>
- Jin, Z., Azzari, G., Lobell, D.B., 2017b. Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches. *Agric. For. Meteorol.* 247, 207–220. <https://doi.org/10.1016/j.agrformet.2017.08.001>
- Jordan, C.F., 1969. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* 50, 663–666. <https://doi.org/10.2307/1936256>
- Ju, W., Gao, P., Zhou, Y., Chen, J.M., Chen, S., Li, X., 2010. Prediction of summer grain crop yield with a process-based ecosystem model and remote sensing data for the northern area of the Jiangsu Province, China. *Int. J. Remote Sens.* 31, 1573–1587.

<https://doi.org/10.1080/01431160903475357>

- Jurecka, F., Hlavinka, P., Lukas, V., 2016. Crop yield estimation in the field level using vegetation indices. *MendelNet* 2016 90–95.
- Kandasamy, S., Baret, F., Verger, A., Neveux, P., Weiss, M., 2013. A comparison of methods for smoothing and gap filling time series of remote sensing observations - application to MODIS LAI products. *Biogeosciences* 10, 4055–4071. <https://doi.org/10.5194/bg-10-4055-2013>
- Kayad, A.G., Al-Gaadi, K.A., Tola, E., Madugundu, R., Zeyada, A.M., Kalaitzidis, C., 2016. Assessing the spatial variability of alfalfa yield using satellite imagery and ground-based data. *PLoS One* 11, 1–15. <https://doi.org/10.1371/journal.pone.0157166>
- Kerr, J.T., Cihlar, J., 2003. Land use and cover with intensity of agriculture for Canada from satellite and census data. *Glob. Ecol. Biogeogr.* 12, 161–172. <https://doi.org/10.1046/j.1466-822X.2003.00017.x>
- Kerr, J.T., Ostrovsky, M., 2003. From space to species: Ecological applications for remote sensing. *Trends Ecol. Evol.* 18, 299–305. [https://doi.org/10.1016/S0169-5347\(03\)00071-5](https://doi.org/10.1016/S0169-5347(03)00071-5)
- Kim, N., Lee, Y.-W., 2016. Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* 34, 383–390. <https://doi.org/10.7848/ksgpc.2016.34.4.383>
- Kissling, W.D., Ahumada, J.A., Bowser, A., Fernandez, M., Fernández, N., García, E.A., Guralnick, R.P., Isaac, N.J.B., Kelling, S., Los, W., McRae, L., Mihoub, J.B., Obst, M., Santamaria, M., Skidmore, A.K., Williams, K.J., Agosti, D., Amariles, D., Arvanitidis, C., Bastin, L., De Leo, F., Egloff, W., Elith, J., Hobern, D., Martin, D., Pereira, H.M., Pesole, G., Peterseil, J., Saarenmaa, H., Schigel, D., Schmeller, D.S., Segata, N., Turak, E., Uhlir, P.F., Wee, B., Hardisty, A.R., 2018. Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biol. Rev.* 93, 600–625. <https://doi.org/10.1111/brev.12359>
- Kuemmerle, T., Erb, K., Meyfroidt, P., Müller, D., Verburg, P.H., Estel, S., Haberl, H., Hostert, P., Jepsen, M.R., Kastner, T., Levers, C., Lindner, M., Plutzer, C., Verkerk, P.J., van der Zanden, E.H., Reenberg, A., 2013. Challenges and opportunities in mapping land use intensity globally. *Curr. Opin. Environ. Sustain.* 5, 484–493. <https://doi.org/10.1016/j.cosust.2013.06.002>
- Kuenzer, C., Ottinger, M., Wegmann, M., Guo, H., Wang, C., Zhang, J., Dech, S., Wikelski, M., 2014. Earth observation satellite sensors for biodiversity monitoring: potentials and bottlenecks. *Int. J. Remote Sens.* 35, 6599–6647. <https://doi.org/10.1080/01431161.2014.964349>
- Lakhankar, T., Ghedira, H., Temimi, M., Azar, A.E., Khanbilvardi, R., 2009. Effect of land cover heterogeneity on soil moisture retrieval using active microwave remote sensing data. *Remote Sens.* 1, 80–91. <https://doi.org/10.3390/rs1020080>
- Lambert, M.J., Blaes, X., Traore, P.S., Defourny, P., 2017. Estimate yield at parcel level from S2 time serie in sub-Saharan smallholder farming systems. 2017 9th Int. Work. Anal. Multitemporal Remote Sens. Images, MultiTemp 2017. <https://doi.org/10.1109/Multi-Temp.2017.8035204>
- Lampkin, N.H., Pearce, B.D., Leake, A.R., Creissen, H., Gerrard, C.L., Girling, R., Lloyd, S., Padel, S., Smith, J., Smith, L.G., Vieweger, A., Wolfe, M.S., 2015. The Role of Agroecology in Sustainable Intensification. Report for the Land Use Policy Group. Organic Research Centre, Elm Farm and Game & Wildlife Conservation Trust.
- Lasne, Y., Paillou, P., Ruffié, G., Serradilla, C., Freeman, A., Farr, T., McDonald, K., Chapman, B., 2008. Effect of Salinity on the Dielectric Properties of Geological Materials: Implication for Soil Moisture Detection by Means of Remote Sensing. *IEEE Trans. Geosci. Remote Sens.* 6, 1674–

1688. <https://doi.org/10.1109/TGRS.2008.916220>
- Li, J., Roy, D.P., 2017. A global analysis of Sentinel-2A, Sentinel-2B and Landsat-8 data revisit intervals and implications for terrestrial monitoring. *Remote Sens.* 9, 902. <https://doi.org/10.3390/rs9090902>
- Li, L., Friedl, M.A., Xin, Q., Gray, J., Pan, Y., Frolking, S., 2014. Mapping crop cycles in China using MODIS-EVI time series. *Remote Sens.* 6, 2473–2493. <https://doi.org/10.3390/rs6032473>
- Li, Z., Guo, X., 2012. Detecting climate effects on vegetation in northern mixed rairie using NOAA AVHRR 1-km time-series NDVI data. *Remote Sens.* 4, 120–134. <https://doi.org/10.3390/rs4010120>
- Li, Z., Xu, D., Guo, X., 2014. Remote sensing of ecosystem health: opportunities, challenges, and future perspectives. *Sensors* 14, 21117–21139. <https://doi.org/10.3390/s141121117>
- Liaqat, M.U., Cheema, M.J.M., Huang, W., Mahmood, T., Zaman, M., Khan, M.M., 2017. Evaluation of MODIS and Landsat multiband vegetation indices used for wheat yield estimation in irrigated Indus Basin. *Comput. Electron. Agric.* 138, 39–47. <https://doi.org/10.1016/j.compag.2017.04.006>
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. *R news* 2, 18–22. <https://doi.org/10.1177/154405910408300516>
- Link, W.A., Sauer, J.R., 1997. Estimation of population trajectories from count data. *Biometrics* 53, 63–72.
- Lobell, D.B., 2013. Field Crops Research The use of satellite data for crop yield gap analysis. *F. Crop. Res.* 143, 56–64. <https://doi.org/10.1016/j.fcr.2012.08.008>
- Lobell, D.B., Cassman, K.G., Field, C.B., 2009. Crop Yield Gaps: Their Importance, Magnitudes, and Causes. *Annu. Rev. Environ. Resour.* 34, 179–204. <https://doi.org/10.1146/annurev.enviro.041008.093740>
- Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. *Remote Sens. Environ.* 164, 324–333. <https://doi.org/10.1016/j.rse.2015.04.021>
- Lopresti, M.F., Di Bella, C.M., Degioanni, A.J., 2015. Relationship between MODIS-NDVI data and wheat yield: A case study in Northern Buenos Aires province, Argentina. *Inf. Process. Agric.* 2, 73–84. <https://doi.org/10.1016/j.inpa.2015.06.001>
- Loveland, T.R., Reed, B.C., Ohlen, D.O., Brown, J.F., Zhu, Z., Yang, L., Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sens.* 21, 1303–1330. <https://doi.org/10.1080/014311600210191>
- Luoto, M., Toivonen, T., Heikkinen, R.K., 2002. Prediction of total and rare plant species richness in agricultural landscapes from satellite images and topographic data. *Landsc. Ecol.* 17, 195–217. <https://doi.org/10.1023/A:1020288509837>
- Luoto, M., Virkkala, R., Heikkinen, R.K., Rainio, K., 2004. Predicting Bird Species Richness Using Remote Sensing in Boreal Agricultural-Forest Mosaics. *Ecol. Appl.* 14, 1946–1962.
- Lyle, G., Bryan, B.A., Ostendorf, B., 2014. Post-processing methods to eliminate erroneous grain yield measurements: Review and directions for future development. *Precis. Agric.* 15, 377–402. <https://doi.org/10.1007/s11119-013-9336-3>
- Mairota, P., Cafarelli, B., Didham, R.K., Lovergine, F.P., Lucas, R.M., Nagendra, H., Rocchini, D., Tarantino, C., 2015. Challenges and opportunities in harnessing satellite remote-sensing for

- biodiversity monitoring. *Ecol. Inform.* 30, 207–214.  
<https://doi.org/10.1016/j.ecoinf.2015.08.006>
- Martinez, J.M., Guyot, J.L., Cochonneau, G., 2008. Monitoring of surface water quality in large rivers with satellite imagery - Application to the Amazon basin.  
[www.iwra.org/congress/resource/abs649\\_article.pdf](http://www.iwra.org/congress/resource/abs649_article.pdf).
- Matson, P.A., Parton, W.J., Power, A.G., Swift, M.J., 1997. Agricultural intensification and ecosystem properties. *Science* (80-. ). 277, 504–509. <https://doi.org/10.1126/science.277.5325.504>
- Mcgarigal, K., 2015. Fragstats Help. [https://doi.org/10.1016/S0022-3913\(12\)00047-9](https://doi.org/10.1016/S0022-3913(12)00047-9)
- McGarigal, K., Cushman, S.A., Ene, E., 2012. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following website:  
<http://www.umass.edu/landeco/research/fragsta>.
- Meeus, J.H.A., 1993. The transformation of agricultural landscapes in Western Europe. *Sci. Total Environ.* 129, 171–190. [https://doi.org/10.1016/0048-9697\(93\)90169-7](https://doi.org/10.1016/0048-9697(93)90169-7)
- Met-Office, 2017. UKCP09: Met Office gridded land surface climate observations - daily temperature and precipitation at 5km resolution [WWW Document]. URL  
<https://catalogue.ceda.ac.uk/uuid/319b3f878c7d4cbfbdb356e19d8061d6>
- Mi, C., Huettmann, F., Guo, Y., Han, X., Wen, L., 2017. Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* 5, e2849. <https://doi.org/10.7717/peerj.2849>
- Mingwei, Z., Qingbo, Z., Zhongxin, C., Jia, L., Yong, Z., Chongfa, C., 2008. Crop discrimination in Northern China with double cropping systems using Fourier analysis of time-series MODIS data. *Int. J. Appl. Earth Obs. Geoinf.* 10, 476–485. <https://doi.org/10.1016/j.jag.2007.11.002>
- Mitchell, G., May, A., McDonald, A., 1995. PICABUE: a methodological framework for the development of indicators of sustainable development. *Int.J. Sustain. Dev. World Ecol* 2, 104–123. <https://doi.org/10.1080/13504509509469893>
- Møller, A.P., Jennions, M.D., 2002. How much variance can be explained by ecologists and evolutionary biologists? *Oecologia* 132, 492–500. <https://doi.org/10.1007/s00442-002-0952-2>
- Morelli, F., Jerzak, L., Tryjanowski, P., 2014. Birds as useful indicators of high nature value (HNV) farmland in Central Italy. *Ecol. Indic.* 38, 236–242.  
<https://doi.org/10.1016/j.ecolind.2013.11.016>
- Moriondo, M., Maselli, F., Bindi, M., 2007. A simple model of regional wheat yield based on NDVI data. *Eur. J. Agron.* 26, 266–274. <https://doi.org/10.1016/j.eja.2006.10.007>
- Mueller, N., Lewis, A., Roberts, D., Ring, S., Melrose, R., Sixsmith, J., Lymburner, L., McIntyre, A., Tan, P., Curnow, S., Ip, A., 2016. Water observations from space: Mapping surface water from 25years of Landsat imagery across Australia. *Remote Sens. Environ.* 174, 341–352.  
<https://doi.org/10.1016/j.rse.2015.11.003>
- Musumba, M., Palm, C., Grabowski, P., Snapp, S.S., 2017. A Framework for Selecting and Analyzing Indicators of Sustainable intensification.
- Myneni, R.B., Nemani, R.R., Running, S.W., 1997. Estimation of global leaf area index and absorbed PAR using radiative transfer models. *IEEE Trans. Geosci. Remote Sens.* 35, 1380–1393.  
<https://doi.org/10.1109/36.649788>

- Nagendra, H., 2001. Using remote sensing to assess biodiversity. *Int. J. Remote Sens.* 22, 2377–2400. <https://doi.org/10.1080/01431160117096>
- Newton, I., 2004. The recent declines of farmland bird populations in Britain: An appraisal of causal factors and conservation actions. *Ibis (Lond. 1859)*. 146, 579–600. <https://doi.org/10.1111/j.1474-919X.2004.00375.x>
- Niedertscheider, M., Kastner, T., Fetzl, T., Haberl, H., Kroisleitner, C., Plutzer, C., Erb, K.-H., 2016. Mapping and analysing cropland use intensity from a NPP perspective. *Environ. Res. Lett.* 11. <https://doi.org/10.1088/1748-9326/11/1/014008>
- Nigam, R., Vyas, S.S., Bhattacharya, B.K., Oza, M.P., 2017. Retrieval of regional LAI over agricultural land from an Indian geostationary satellite and its application for crop yield estimation. *J. Spat. Sci.* 62, 103–125. <https://doi.org/10.1080/14498596.2016.1220872>
- OneSoil, 2018. OneSoil [WWW Document]. URL <https://onesoil.ai/en/> (accessed 4.8.19).
- Pacheco, A., McNairn, H., 2010. Evaluating multispectral remote sensing and spectral unmixing analysis for crop residue mapping. *Remote Sens. Environ.* 114, 2219–2228. <https://doi.org/10.1016/j.rse.2010.04.024>
- Padilla, F.L.M., Maas, S.J., González-Dugo, M.P., Mansilla, F., Rajan, N., Gavilán, P., Domínguez, J., 2012. Monitoring regional wheat yield in Southern Spain using the GRAMI model and satellite imagery. *F. Crop. Res.* 130, 145–154. <https://doi.org/10.1016/j.fcr.2012.02.025>
- Palczewska, A., Palczewski, J., Robinson, R.M., Neagu, D., 2014. Interpreting random forest classification models using a feature contribution method, in: Bouabana-Tebibel, T., Rubin, S.H. (Eds.), *Integration of Reusable Systems*. Springer, pp. 193–218.
- Parihar, J.S., Justice, C., Soares, J., Leo, O., Kosuth, P., Jarvis, I., Williams, D., Bingfang, W., Latham, J., Becker-Reshef, I., 2012. GEO-GLAM: A GEOSS-G20 initiative on Global Agricultural Monitoring. *Proc. 39th COSPAR Sci. Assem. Mysore, India, 14-22 July 2012*.
- Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith, D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S., McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential biodiversity variables. *Science (80-. )*. 339, 277–278. <https://doi.org/10.1126/science.1229931>
- Petit, S., Griffiths, L., S. Smart, S., M. Smith, G., C. Stuart, R., M. Wright, S., 2004. Effects of area and isolation of woodland patches on herbaceous plant species richness across Great Britain. *Landsc. Ecol.* 19, 463–472. <https://doi.org/10.1023/B:LAND.0000036141.30359.53>
- Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H., Turner, W., 2014. Satellite remote sensing for applied ecologists: Opportunities and challenges. *J. Appl. Ecol.* 51, 839–848. <https://doi.org/10.1111/1365-2664.12261>
- Pettorelli, N., Wegmann, M., Skidmore, A., Mùcher, S., Dawson, T.P., Fernandez, M., Lucas, R., Schaepman, M.E., Wang, T., O'Connor, B., Jongman, R.H.G., Kempeneers, P., Sonnenschein, R., Leidner, A.K., Böhm, M., He, K.S., Nagendra, H., Dubois, G., Fatoyinbo, T., Hansen, M.C., Paganini, M., de Klerk, H.M., Asner, G.P., Kerr, J.T., Estes, A.B., Schmeller, D.S., Heiden, U., Rocchini, D., Pereira, H.M., Turak, E., Fernandez, N., Lausch, A., Cho, M.A., Alcaraz-Segura, D., McGeoch, M.A., Turner, W., Mueller, A., St-Louis, V., Penner, J., Vihervaara, P., Belward, A., Reyers, B., Geller, G.N., 2016. Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. *Remote Sens. Ecol. Conserv.* 2, 122–

131. <https://doi.org/10.1002/rse2.15>
- Phillips, L.B., Hansen, A.J., Flather, C.H., 2008. Evaluating the species energy relationship with the newest measures of ecosystem energy: NDVI versus MODIS primary production. *Remote Sens. Environ.* 112, 3538–3549. <https://doi.org/10.1016/j.rse.2008.04.012>
- Phillips, L.B., Hansen, A.J., Flather, C.H., Robinson-Cox, J., 2010. Applying species — energy theory to conservation : a case study for North American birds. *Ecol. Appl.* 20, 2007–2023.
- Ping, J.L., Dobermann, A., 2005. Processing of Yield Map Data. *Precis. Agric.* 6, 193–212. <https://doi.org/10.1007/s11119-005-1035-2>
- Power, A.G., 2010. Ecosystem services and agriculture: tradeoffs and synergies. *Philos. Trans. R. Soc. B* 365, 2959–2971. <https://doi.org/10.1098/rstb.2010.0143>
- Prasad, A.K., Chai, L., Singh, R.P., Kafatos, M., 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. *Int. J. Appl. Earth Obs. Geoinf.* 8, 26–33. <https://doi.org/10.1016/j.jag.2005.06.002>
- Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199. <https://doi.org/10.1007/s10021-005-0054-1>
- Pretty, J., 2008. Agricultural sustainability: concepts, principles and evidence. *Philos. Trans. R. Soc. B* 363, 447–465. <https://doi.org/10.1098/rstb.2007.2163>
- Pretty, J., Benton, T.G., Bharucha, Z.P., Dicks, L. V., Flora, C.B., Godfray, H.C.J., Goulson, D., Hartley, S., Lampkin, N., Morris, C., Pierzynski, G., Prasad, P.V.V., Reganold, J., Rockström, J., Smith, P., Thorne, P., Wratten, S., 2018. Global assessment of agricultural redesign for sustainable intensification. *Nat. Sustain.* 1, 441–446. <https://doi.org/10.1038/s41893-018-0114-0>
- Pretty, J., Bharucha, Z.P., 2014. Sustainable intensification in agricultural systems. *Ann. Bot.* 114, 1571–1596. <https://doi.org/10.1093/aob/mcu205>
- Pretty, J., Toulmin, C., Williams, S., 2011. Sustainable intensification in African agriculture. *Int. J. Agric. Sustain.* 9, 5–24. <https://doi.org/10.3763/ijas.2010.0583>
- Rasul, G., Thapa, G.B., 2004. Sustainability of ecological and conventional agricultural systems in Bangladesh: An assessment based on environmental, economic and social perspectives. *Agric. Syst.* 79, 327–351. [https://doi.org/10.1016/S0308-521X\(03\)00090-8](https://doi.org/10.1016/S0308-521X(03)00090-8)
- Refaeilzadeh, P., Tang, L., Liu, H., 2009. Cross-Validation, in: Liu, L., Özsu, M.T. (Eds.), *Encyclopedia of Database Systems*. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-39940-9\\_565](https://doi.org/10.1007/978-0-387-39940-9_565)
- Ren, J., Chen, Z., Zhou, Q., Tang, H., 2008. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *Int. J. Appl. Earth Obs. Geoinf.* 10, 403–413. <https://doi.org/10.1016/j.jag.2007.11.003>
- Rhodes, C.J., Henrys, P., Siriwardena, G.M., Whittingham, M.J., Norton, L.R., 2015. The relative value of field survey and remote sensing for biodiversity assessment. *Methods Ecol. Evol.* 6, 772–781. <https://doi.org/10.1111/2041-210X.12385>
- Rigby, D., Woodhouse, P., Young, T., Burton, M., 2001. Constructing a farm level indicator of sustainable agricultural practice. *Ecol. Econ.* 39, 463–478. [https://doi.org/10.1016/S0921-8009\(01\)00245-2](https://doi.org/10.1016/S0921-8009(01)00245-2)
- Robinson, R.A., Wilson, J.D., Crick, H.Q.P., 2001. The importance of arable habitat for farmland birds



- in grassland landscapes. *J. Appl. Ecol.* 38, 1059–1069.
- Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S., Kark, S., Levin, N., Lucas, K., Luoto, M., Nagendra, H., Oldeland, J., Ricotta, C., Southworth, J., Neteler, M., 2010. Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecol. Inform.* 5, 318–329. <https://doi.org/10.1016/j.ecoinf.2010.06.001>
- Rodrigues, G.S., Rodrigues, I.A., Buschinelli, C.C. de A., de Barros, I., 2010. Integrated farm sustainability assessment for the environmental management of rural activities. *Environ. Impact Assess. Rev.* 30, 229–239. <https://doi.org/10.1016/j.eiar.2009.10.002>
- Rouse, J.W., Hass, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the great plains with ERTS. *Third Earth Resour. Technol. Satell. Symp.* 1, 309–317. <https://doi.org/citeulike-article-id:12009708>
- Rowland, C.S., Morton, R.D., Carrasco, L., McShane, G., O’Neil, A.W., Wood, C.M., 2017a. Land Cover Map 2015 (vector, GB).
- Rowland, C.S., Morton, R.D., Carrasco, L., McShane, G., O’Neil, A.W., Wood, C.M., 2017b. Land Cover Map 2015 (25m raster, GB).
- Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M., Loveland, T.R., Vermote, E., Zhang, C., 2010. Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sens. Environ.* 114, 35–49. <https://doi.org/10.1016/j.rse.2009.08.011>
- Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* 145, 154–172. <https://doi.org/10.1016/j.rse.2014.02.001>
- Roy, R., Chan, N.W., 2012. An assessment of agricultural sustainability indicators in Bangladesh: Review and synthesis. *Environmentalist* 32, 99–110. <https://doi.org/10.1007/s10669-011-9364-3>
- Schut, M., van Asten, P., Okafor, C., Hicintuka, C., Mapatano, S., Nabahungu, N.L., Kagabo, D., Muchunguzi, P., Njukwe, E., Dongsop-Nguezet, P.M., Sartas, M., Vanlauwe, B., 2016. Sustainable intensification of agricultural systems in the Central African Highlands: The need for insitutional innovation. *Agric. Syst.* 145, 165–176. <https://doi.org/10.1016/j.agry.2016.03.005>
- Sehgal, V.K., Sastri, C.V.S., Kalra, N., Dadhwal, V.K., 2005. Farm-level yield mapping for Precision Crop Management by linking remote sensing inputs and a crop simulation model. *J. Indian Soc. Remote Sens.* 33, 131–136. <https://doi.org/10.1007/BF02990002>
- Senf, C., Leitão, P.J., Pflugmacher, D., van der Linden, S., Hostert, P., 2015. Mapping land cover in complex Mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. *Remote Sens. Environ.* 156, 527–536. <https://doi.org/10.1016/j.rse.2014.10.018>
- Seto, K.C., Fleishman, E., Fay, J.P., Betrus, C.J., 2004. Linking spatial patterns of bird and butterfly species richness with Landsat TM derived NDVI. *Int. J. Remote Sens.* 25, 4309–4324. <https://doi.org/10.1080/0143116042000192358>
- Shanahan, J.F., Schepers, J.S., Francis, D.D., Varvel, G.E., Wilhelm, W.W., Tringe, J.M., Schlemmer,

- M.R., Major, D.J., 2001. Use of Remote-Sensing Imagery to Estimate Corn Grain Yield. *Agron. J.* 93, 583–589.
- Shao, Y., Lunetta, R.S., Wheeler, B., Liames, J.S., Campbell, J.B., 2016. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. *Remote Sens. Environ.* 174, 258–265. <https://doi.org/10.1016/j.rse.2015.12.023>
- Shoshany, M., Goldshleger, N., Chudnovsky, A., 2013. Monitoring of agricultural soil degradation by remote-sensing methods: a review. *Int. J. Remote Sens.* 34, 6152–6181. <https://doi.org/10.1080/01431161.2013.793872>
- Shriar, A.J., 2000. Agricultural intensity and its measurement in frontier regions. *Agrofor. Syst.* 49, 301–318. <https://doi.org/10.1023/A:1006316131781>
- Siebrecht, N., 2020. Sustainable agriculture and its implementation gap - Overcoming obstacles to implementation. *Sustain.* 12. <https://doi.org/10.3390/su12093853>
- Singh, R., Semwal, D.P., Rai, A., Chhikara, R.S., 2002. Small area estimation of crop yield using remote sensing satellite data. *Int. J. Remote Sens.* 23, 49–56. <https://doi.org/10.1080/01431160010014756>
- Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, A.K., 2009. An overview of sustainability assessment methodologies. *Ecol. Indic.* 9, 189–212. <https://doi.org/10.1016/j.ecolind.2008.05.011>
- Skakun, S., Vermote, E., Roger, J.-C., Franch, B., 2017. Combined Use of Landsat-8 and Sentinel-2A Images for Winter Crop Mapping and Winter Wheat Yield Assessment at Regional Scale. *AIMS Geosci.* 3, 163–186. <https://doi.org/10.3934/geosci.2017.2.163>
- Smith, A., Snapp, S., Chikowo, R., Thorne, P., Bekunda, M., Glover, J., 2017. Measuring sustainable intensification in smallholder agroecosystems: A review. *Glob. Food Sec.* 12, 127–138. <https://doi.org/10.1016/j.gfs.2016.11.002>
- Snapp, S.S., Grabowski, P., Chikowo, R., Smith, A., Anders, E., Sirrine, D., Chimonyo, V., Bekunda, M., 2018. Maize yield and profitability tradeoffs with social, human and environmental performance: Is sustainable intensification feasible? *Agric. Syst.* 162, 77–88. <https://doi.org/10.1016/j.agsy.2018.01.012>
- Stein, A., Gerstner, K., Kreft, H., 2014. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecol. Lett.* 17, 866–880. <https://doi.org/10.1111/ele.12277>
- Stein, A., Riley, J., Halberg, N., 2001. Issues of scale for environmental indicators. *Agric. Ecosyst. Environ.* 87, 215–232. [https://doi.org/10.1016/S0167-8809\(01\)00280-8](https://doi.org/10.1016/S0167-8809(01)00280-8)
- Stoate, C., Boatman, N.D., Borralho, R.J., Carvalho, C.R., de Snoo, G.R., Eden, P., 2001. Ecological impacts of arable intensification in Europe. *J. Environ. Manage.* 63, 337–365. <https://doi.org/10.1006/jema.2001.0473>
- Stoate, C., Boatman, N.D., Borralho, R.J., Carvalho, C.R., De Snoo, G.R., Eden, P., 2001. Ecological impacts of arable intensification in Europe. *J. Environ. Manage.* 63, 337–365. <https://doi.org/10.1006/jema.2001.0473>
- Stokstad, E., 2020. United Kingdom to embark on “agricultural revolution” in break from EU farm subsidies [WWW Document]. URL <https://www.sciencemag.org/news/2020/01/united-kingdom-embark-agricultural-revolution-break-eu-farm-subsidies>
- Sulistioadi, Y.B., Tseng, K.H., Shum, C.K., Hidayat, H., Sumaryono, M., Suhardiman, A., Setiawan, F., Sunarso, S., 2015. Satellite radar altimetry for monitoring small rivers and lakes in Indonesia.

- Hydrol. Earth Syst. Sci. 19, 341–359. <https://doi.org/10.5194/hess-19-341-2015>
- Sullivan, M.J.P., Pearce-Higgins, J.W., Newson, S.E., Scholefield, P., Brereton, T., Oliver, T.H., 2017. A national-scale model of linear features improves predictions of farmland biodiversity. *J. Appl. Ecol.* 54, 1776–1784. <https://doi.org/10.1111/1365-2664.12912>
- Tan, B., Morisette, J.T., Wolfe, R.E., Gao, F., Ederer, G.A., Nightingale, J., Pedelty, J.A., 2011. An Enhanced TIMESAT Algorithm for Estimating Vegetation Phenology Metrics From MODIS Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 4, 361–371. <https://doi.org/10.1109/JSTARS.2010.2075916>
- Tebbs, E., Rowland, C.S., Smart, S., Maskell, L., Norton, L., 2017. Regional-Scale High Spatial Resolution Mapping of Aboveground Net Primary Productivity (ANPP) from Field Survey and Landsat Data: A Case Study for the Country of Wales. *Remote Sens.* 9, 801. <https://doi.org/10.3390/rs9080801>
- Teillard, F., Allaire, G., Cahuzac, E., Léger, F., Maigné, E., Tichit, M., 2012. A novel method for mapping agricultural intensity reveals its spatial aggregation: Implications for conservation policies. *Agric. Ecosyst. Environ.* 149, 135–143. <https://doi.org/10.1016/j.agee.2011.12.018>
- Teluguntla, P., Thenkabail, P., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K., Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Remote Sens.* 144, 325–340. <https://doi.org/10.1016/j.isprsjprs.2018.07.017>
- Temme, A.J.A.M., Verburg, P.H., 2011. Mapping and modelling of changes in agricultural intensity in Europe. *Agric. Ecosyst. Environ.* 140, 46–56. <https://doi.org/10.1016/j.agee.2010.11.010>
- Terraube, J., Archaux, F., Deconchat, M., van Halder, I., Jactel, H., Barbaro, L., 2016. Forest edges have high conservation value for bird communities in mosaic landscapes. *Ecol. Evol.* 6, 5178–5189. <https://doi.org/10.1002/ece3.2273>
- Thieme, A., Yadav, S., Oddo, P.C., Fitz, J.M., McCartney, S., King, L.A., Keppler, J., McCarty, G.W., Hively, W.D., 2020. Using NASA Earth observations and Google Earth Engine to map winter cover crop conservation performance in the Chesapeake Bay watershed. *Remote Sens. Environ.* 248, 111943. <https://doi.org/10.1016/j.rse.2020.111943>
- Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. *PNAS* 108, 20260–20264. <https://doi.org/10.1073/pnas.1116437108>
- Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R., Polasky, S., 2002. Agricultural sustainability and intensive production practices. *Nature* 418, 671–677. <https://doi.org/10.1038/nature01014>
- Tilman, D., Fargione, J., Wolff, B., D’Antonio, C., Dobson, A., Howarth, R., Schindler, D., Schlesinger, W.H., Simberloff, D., Swackhamer, D., 2001. Forecasting agriculturally driven global environmental change. *Science* (80-. ). 292, 281–284. <https://doi.org/10.1126/science.1057544>
- Torres, R., Snoeijs, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B.Ö., Floury, N., Brown, M., Traver, I.N., Deghayes, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L’Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., Rostan, F., 2012. GMES Sentinel-1 mission. *Remote Sens. Environ.* 120, 9–24. <https://doi.org/10.1016/j.rse.2011.05.028>
- Trombetti, M., Riaño, D., Rubio, M.A., Cheng, Y.B., Ustin, S.L., 2008. Multi-temporal vegetation canopy water content retrieval and interpretation using artificial neural networks for the continental USA. *Remote Sens. Environ.* 112, 203–215. <https://doi.org/10.1016/j.rse.2007.04.013>

- Tscharntke, T., Klein, A.M., Kruess, A., Steffan-Dewenter, I., Thies, C., 2005. Landscape perspectives on agricultural intensification and biodiversity - Ecosystem service management. *Ecol. Lett.* 8, 857–874. <https://doi.org/10.1111/j.1461-0248.2005.00782.x>
- Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, A.K., Szantoi, Z., Buchanan, G., Dech, S., Dwyer, J., Herold, M., Koh, L.P., Leimgruber, P., Taubenboeck, H., Wegmann, M., Wikelski, M., Woodcock, C., 2015. Free and open-access satellite data are key to biodiversity conservation. *Biol. Conserv.* 182, 173–176. <https://doi.org/10.1016/j.biocon.2014.11.048>
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* 18, 306–314. [https://doi.org/10.1016/S0169-5347\(03\)00070-3](https://doi.org/10.1016/S0169-5347(03)00070-3)
- USDA FAS, 2019. GLAM - Global Agricultural Monitoring [WWW Document]. URL <https://www.pecad.fas.usda.gov/glam.htm> (accessed 4.9.17).
- Van Cauwenbergh, N., Biala, K., Biielders, C., Brouckaert, V., Franchois, L., Garcia Ciudad, V., Hermy, M., Mathijs, E., Muys, B., Reijnders, J., Sauvenier, X., Valckx, J., Vanclooster, M., Van der Veken, B., Wauters, E., Peeters, A., 2007. SAFE-A hierarchical framework for assessing the sustainability of agricultural systems. *Agric. Ecosyst. Environ.* 120, 229–242. <https://doi.org/10.1016/j.agee.2006.09.006>
- Van Ittersum, M.K., Rabbinge, R., 1997. Concepts in production ecology for analysis and quantification of agricultural input-output combinations. *F. Crop. Res.* 52, 197–208.
- Van Leeuwen, W.J.D., Huete, A.R., Laing, T.W., 1999. MODIS vegetation index compositing approach: A prototype with AVHRR data. *Remote Sens. Environ.* 69, 264–280. [https://doi.org/10.1016/S0034-4257\(99\)00022-X](https://doi.org/10.1016/S0034-4257(99)00022-X)
- Van Passel, S., Meul, M., 2012. Multilevel and multi-user sustainability assessment of farming systems. *Environ. Impact Assess. Rev.* 32, 170–180. <https://doi.org/10.1016/j.eiar.2011.08.005>
- Verger, A., Baret, F., Weiss, M., Kandasamy, S., Vermote, E., 2013. The CACAO method for smoothing, gap filling, and characterizing seasonal anomalies in satellite time series. *IEEE Trans. Geosci. Remote Sens.* 51, 1963–1972. <https://doi.org/10.1109/TGRS.2012.2228653>
- Vickery, J.A., Tallowin, J.R., Feber, R.E., Asteraki, E.J., Atkinson, P.W., Fuller, R.J., Brown, V.K., 2001. The management of lowland neutral grasslands in Britain: Effects of agricultural practices on birds and their food resources. *J. Appl. Ecol.* 38, 647–664. <https://doi.org/10.1046/j.1365-2664.2001.00626.x>
- Vogelmann, J.E., Helder, D., Morfitt, R., Choate, M.J., Merchant, J.W., Bulley, H., 2001. Effects of Landsat 5 Thematic Mapper and Landsat 7 Enhanced Thematic Mapper plus radiometric and geometric calibrations and corrections on landscape characterization. *Remote Sens. Environ.* 78, 55–70. [https://doi.org/10.1016/S0034-4257\(01\)00249-8](https://doi.org/10.1016/S0034-4257(01)00249-8)
- Vuolo, F., Ng, W.-T., Atzberger, C., 2017. Smoothing and gap-filling of high resolution multi-spectral time series: Example of Landsat data. *Int. J. Appl. Earth Obs. Geoinf.* 57, 202–213. <https://doi.org/10.1016/j.jag.2016.12.012>
- Wagner, W., Lemoine, G., Rott, H., 1999. A method for estimating soil moisture from ERS Scatterometer and soil data. *Remote Sens. Environ.* 70, 191–207. [https://doi.org/10.1016/S0034-4257\(99\)00036-X](https://doi.org/10.1016/S0034-4257(99)00036-X)
- Wagner, W., Naeimi, V., Scipal, K., Jeu, R., Martínez-Fernández, J., 2007. Soil moisture from operational meteorological satellites. *Hydrogeol. J.* 15, 121–131. <https://doi.org/10.1007/s10040-006-0104-6>

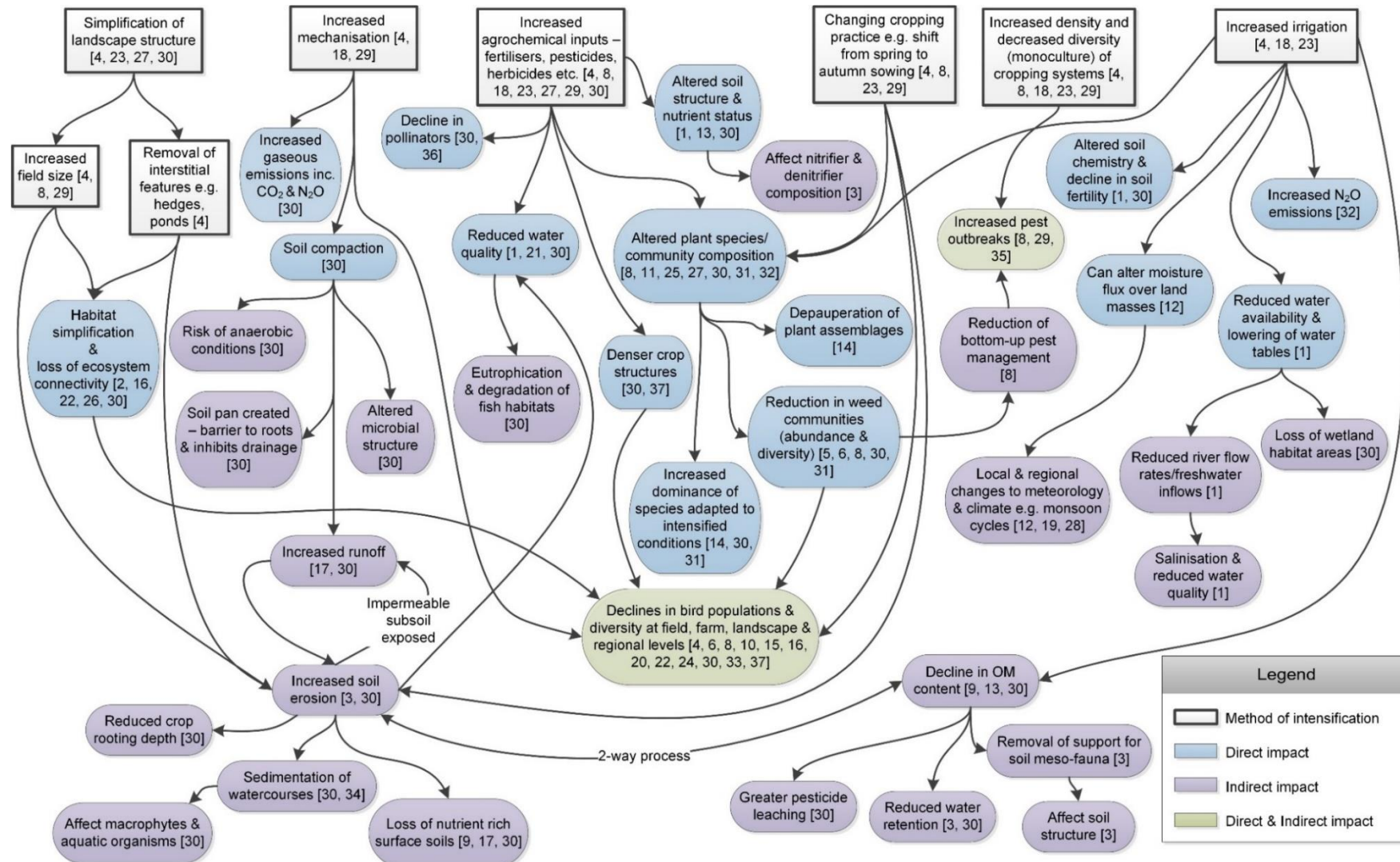
- Wang, K., Franklin, S.E., Guo, X., Cattet, M., 2010. Remote sensing of ecology, biodiversity and conservation: A review from the perspective of remote sensing specialists. *Sensors* 10, 9647–9667. <https://doi.org/10.3390/s101109647>
- Wang, Q., Atkinson, P.M., 2018. Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sens. Environ.* 204, 31–42. <https://doi.org/10.1016/j.rse.2017.10.046>
- Weiers, S., Bock, M., Wissen, M., Rossner, G., 2004. Mapping and indicator approaches for the assessment of habitats at different scales using remote sensing and GIS methods. *Landsc. Urban Plan.* 67, 43–65. [https://doi.org/10.1016/S0169-2046\(03\)00028-8](https://doi.org/10.1016/S0169-2046(03)00028-8)
- Welling, S.H., Refsgaard, H.H.F., Brockhoff, P.B., Clemmensen, L.K.H., 2016. Forest Floor Visualisations of Random Forest. ArXiv e-prints.
- Whelan, C.J., Wenny, D.G., Marquis, R.J., 2008. Ecosystem services provided by bats. *Ann. N. Y. Acad. Sci.* 1134, 25–60. <https://doi.org/10.1196/annals.1439.003>
- Whitcraft, A.K., Becker-Reshef, I., Justice, C.O., 2015. A framework for defining spatially explicit earth observation requirements for a global agricultural monitoring initiative (GEOGLAM). *Remote Sens.* 7, 1461–1481. <https://doi.org/10.3390/rs70201461>
- Wiek, A., Binder, C., 2005. Solution spaces for decision-making - A sustainability assessment tool for city-regions. *Environ. Impact Assess. Rev.* 25, 589–608. <https://doi.org/10.1016/j.eiar.2004.09.009>
- Wilson, A.M., Fuller, R.J., 2002. Bird Populations and Environmental Change: Countryside Survey 2000 Module 5. BTO Res. Rep.
- Woodcock, C.E., Loveland, T.R., Herold, M., Bauer, M.E., 2020. Transitioning from change detection to monitoring with remote sensing: A paradigm shift. *Remote Sens. Environ.* 238, 111558. <https://doi.org/10.1016/j.rse.2019.111558>
- Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., Allen, R.G., Anderson, M.C., Belward, A.S., Cohen, W.B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J.D., Hostert, P., Hughes, M.J., Huntington, J., Johnson, D.M., Kennedy, R., Kilic, A., Li, Z., Lyburner, L., McCorkel, J., Pahlevan, N., Scambos, T.A., Schaaf, C., Schott, J.R., Sheng, Y., Storey, J., Vermote, E., Vogelmann, J., White, J.C., Wynne, R.H., Zhu, Z., 2019. Current status of Landsat program, science, and applications. *Remote Sens. Environ.* 225, 127–147. <https://doi.org/10.1016/j.rse.2019.02.015>
- Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* 122, 2–10. <https://doi.org/10.1016/j.rse.2012.01.010>
- Xie, Y., Wang, P., Bai, X., Khan, J., Zhang, S., Li, L., Wang, L., 2017. Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using Landsat imagery and the CERES-Wheat model. *Agric. For. Meteorol.* 246, 194–206. <https://doi.org/10.1016/j.agrformet.2017.06.015>
- Xiong, J., Thenkabail, P.S., Tilton, J.C., Gumma, M.K., Teluguntla, P., Oliphant, A., Congalton, R.G., Yadav, K., Gorelick, N., 2017. Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on google earth engine. *Remote Sens.* 9, 1065. <https://doi.org/10.3390/rs9101065>
- Yan, H., Xiao, X., Huang, H., Liu, J., Chen, J., Bai, X., 2014. Multiple Cropping Intensity in China Derived from Agro-meteorological Observations and MODIS Data. *Chinese Geogr. Sci.* 24, 205–219. <https://doi.org/10.1007/s11769-013-0637-2>

- Yang, C., Everitt, J.H., 2002. Relationships between yield monitor data and airborne multiband multispectral digital imagery for grain sorghum. *Precis. Agric.* 3, 373–388.
- Yang, C., Everitt, J.H., Bradford, J.M., 2009. Evaluating high resolution SPOT 5 satellite imagery to estimate crop yield. *Precis. Agric.* 10, 292–303. <https://doi.org/10.1016/j.compag.2010.12.012>
- Yang, C., Everitt, J.H., Bradford, J.M., Escobar, D.E., 2000. Mapping grain sorghum growth and yield variations using airborne multispectral digital imagery. *Trans. ASAE* 43, 1927–1938.
- Zahm, F., Viaux, P., Vilain, L., Girardin, P., Mouchet, C., 2008. Assessing Farm Sustainability with the IDEA Method -from the Concept of Agriculture Sustainability to Case Studies on Farms. *Sustain. Dev.* 16. <https://doi.org/10.1002/sd.380>
- Zhang, G., Lu, Y., 2012. Bias-corrected random forests in regression. *J. Appl. Stat.* 39, 151–160. <https://doi.org/10.1080/02664763.2011.578621>
- Zhang, Y., Atkinson, P.M., Li, X., Ling, F., Wang, Q., Du, Y., 2017. Learning-Based Spatial-Temporal Superresolution Mapping of Forest Cover with MODIS Images. *IEEE Trans. Geosci. Remote Sens.* 55, 600–614. <https://doi.org/10.1109/TGRS.2016.2613140>
- Zhen, L., Routray, J.K., Zoebisch, M.A., Chen, G., Xie, G., Cheng, S., 2005. Three dimensions of sustainability of farming practices in the North China Plain: A case study from Ningjin County of Shandong Province, PR China. *Agric. Ecosyst. Environ.* 105, 507–522. <https://doi.org/10.1016/j.agee.2004.07.012>

## **Appendices**

### **Appendix 1: Supplementary Material for Chapter 2**

**Figure A1: Some of the key environmental impacts of various mechanisms of agricultural intensification with references. Reference numbers correspond to the list on pages 157-159.**





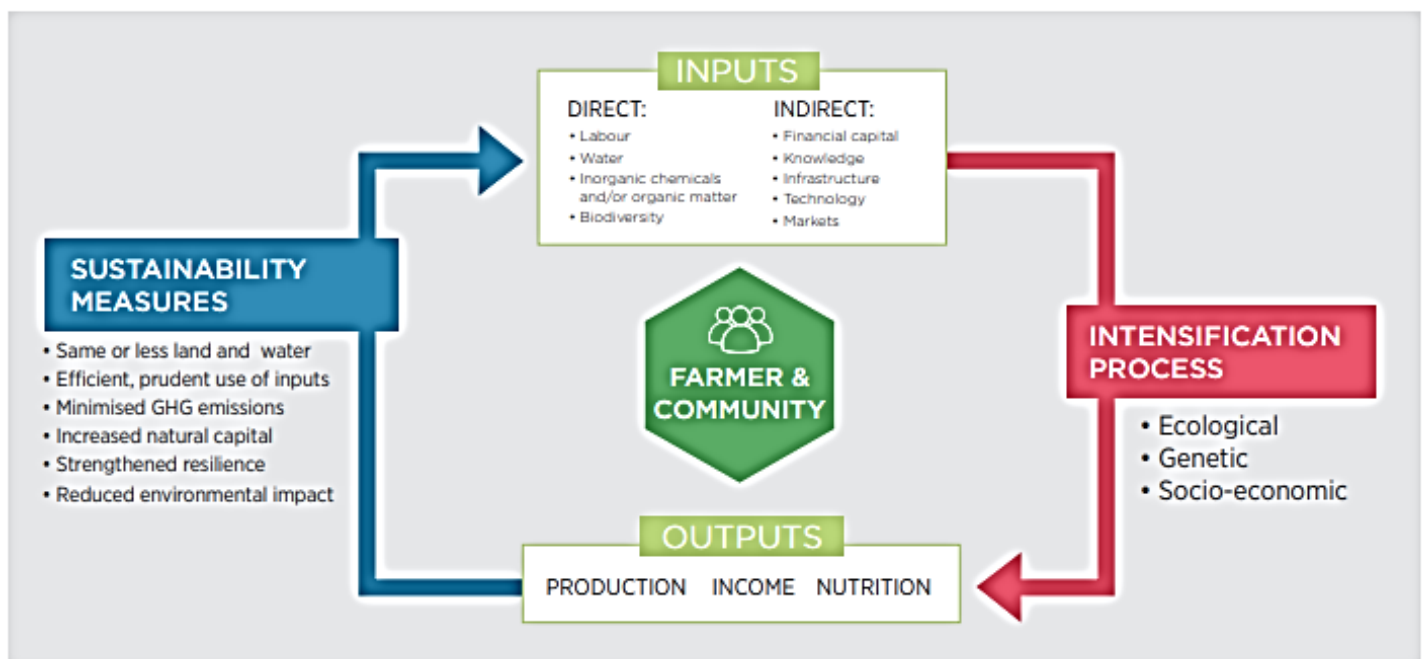
## References for figure A1

1. Alauddin, M., Quiggin, J., 2008. Agricultural intensification, irrigation and the environment in South Asia: Issues and policy options. *Ecol. Econ.* 65, 111–124. doi:10.1016/j.ecolecon.2007.06.004
2. Baum, K.A., Haynes, K.J., Dilleuth, F.P., Cronin, J.T., 2004. The matrix enhances the effectiveness of corridors and stepping stones. *Reports Ecol.* 85, 2671–2676. doi:10.1890/04-0500
3. Benckiser, G., 1997. Organic inputs and soil metabolism, in: Benckiser, G. (Ed.), *Fauna in Soil Ecosystems: Recycling Processes, Nutrient Fluxes and Agricultural Production*. Dekker, New York, pp. 7–62.
4. Benton, T.G., Vickery, J.A., Wilson, J.D., 2003. Farmland biodiversity: Is habitat heterogeneity the key? *Trends Ecol. Evol.* 18, 182–188. doi:10.1016/S0169-5347(03)00011-9
5. Blažek, P., Lepš, J., 2015. Victims of agricultural intensification: Mowing date affects *Rhinanthus* spp. regeneration and fruit ripening. *Agric. Ecosyst. Environ.* 211, 10–16. doi:10.1016/j.agee.2015.04.022
6. Campbell, L.H., Avery, M.I., Donald, P., Evans, A.D., Green, R.E., Wilson, J.D., 1997. A review of the indirect effects of pesticides on birds. *JNCC Report No. 277*. Peterborough.
7. Ciccolini, V., Bonari, E., Ercoli, L., Pellegrino, E., 2016. Phylogenetic and multivariate analyses to determine the effect of agricultural land-use intensification and soil physico-chemical properties on N-cycling microbial communities in drained Mediterranean peaty soils. *Biol. Fertil. Soils* 52, 811–824. doi:10.1007/s00374-016-1121-9
8. Crowder, D.W., Jabbour, R., 2014. Relationships between biodiversity and biological control in agroecosystems: Current status and future challenges. *Biol. Control* 75, 8–17. doi:10.1016/j.biocontrol.2013.10.010
9. De Jong, S.M., Paracchini, M.L., Bertolo, F., Folving, S., Megier, J., De Roo, A.P.J., 1999. Regional assessment of soil erosion using the distributed model SEMMED and remotely sensed data. *Catena* 37, 291–308. doi:10.1016/S0341-8162(99)00038-7
10. Donald, P.F., Green, R.E., Heath, M.F., 2001. Agricultural intensification and the collapse of Europe's farmland bird populations. *Proc. R. Soc. B* 268, 25–29. doi:10.1098/rspb.2000.1325
11. Dorrough, J., Scroggie, M.P., 2008. Plant responses to agricultural intensification. *J. Appl. Ecol.* 45, 1274–1283. doi:10.1111/j.1365-2664.2008.01501.x
12. Douglas, E.M., Beltrán-Przekurat, A., Niyogi, D., Pielke, R.A., Vörösmarty, C.J., 2009. The impact of agricultural intensification and irrigation on land-atmosphere interactions and Indian monsoon precipitation - A mesoscale modeling perspective. *Glob. Planet. Change* 67, 117–128. doi:10.1016/j.gloplacha.2008.12.007
13. Edwards, C.A., 1984. Changes in agricultural practice and their impact on soil organisms, in: Jenkins, D. (Ed.), *Agriculture and the Environment. Proceedings of ITE Symposium no.13 Held at Monks Wood Experimental Station on 28-29 February and 1 March 1984*. pp. 56–65.
14. Firbank, L.G., Petit, S., Smart, S., Blain, A., Fuller, R.J., 2008. Assessing the impacts of agricultural intensification on biodiversity: a British perspective. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 777–787. doi:10.1098/rstb.2007.2183
15. Franke, A.C., Lotz, L.A.P., Van Der Burg, W.J., Van Overbeek, L., 2009. The role of arable weed seeds for agroecosystem functioning. *Weed Res.* 49, 131–141. doi:10.1111/j.1365-3180.2009.00692.x
16. Gabriel, D., Sait, S.M., Hodgson, J.A., Schmutz, U., Kunin, W.E., Benton, T.G., 2010. Scale matters: The impact of organic farming on biodiversity at different spatial scales. *Ecol. Lett.* 13, 858–869. doi:10.1111/j.1461-0248.2010.01481.x

17. Ganasri, B.P., Ramesh, H., 2016. Assessment of soil erosion by RUSLE model using remote sensing and GIS - A case study of Nethravathi Basin. *Geosci. Front.* 7, 953–961. doi:10.1016/j.gsf.2015.10.007
18. González-Estébanez, F.J., García-Tejero, S., Mateo-Tomás, P., Olea, P.P., 2011. Effects of irrigation and landscape heterogeneity on butterfly diversity in Mediterranean farmlands. *Agric. Ecosyst. Environ.* 144, 262–270. doi:10.1016/j.agee.2011.09.002
19. Gordon, L.J., Steffen, W., Jonsson, B.F., Folke, C., Falkenmark, M., Johannessen, A., 2005. Human modification of global water vapor flows from the land surface. *Proc. Natl. Acad. Sci.* 102, 7612–7617. doi:10.1073/pnas.0500208102
20. Guerrero, I., Morales, M.B., Oñate, J.J., Geiger, F., Berendse, F., Snoo, G. de, Eggers, S., Pärt, T., Bengtsson, J., Clement, L.W., Weisser, W.W., Olszewski, A., Ceryngier, P., Hawro, V., Liira, J., Aavik, T., Fischer, C., Flohre, A., Thies, C., Tschardt, T., 2012. Response of ground-nesting farmland birds to agricultural intensification across Europe: Landscape and field level management factors. *Biol. Conserv.* 152, 74–80. doi:10.1016/j.biocon.2012.04.001
21. Herzig, A., Dymond, J., Ausseil, A.-G., 2016. Exploring limits and trade-offs of irrigation and agricultural intensification in the Ruamahanga catchment, New Zealand. *New Zeal. J. Agric. Res.* 59, 216–234. doi:10.1080/00288233.2016.1183685
22. Jeliakov, A., Mimet, A., Chargé, R., Jiguet, F., Devictor, V., Chiron, F., 2016. Impacts of agricultural intensification on bird communities: New insights from a multi-level and multi-facet approach of biodiversity. *Agric. Ecosyst. Environ.* 216, 9–22. doi:10.1016/j.agee.2015.09.017
23. José-María, L., Armengot, L., Blanco-Moreno, J.M., Bassa, M., Sans, F.X., 2010. Effects of agricultural intensification on plant diversity in Mediterranean dryland cereal fields. *J. Appl. Ecol.* 47, 832–840. doi:10.1111/j.1365-2664.2010.01822.x
24. Karp, D.S., Rominger, A.J., Zook, J., Ranganathan, J., Ehrlich, P.R., Daily, G.C., 2012. Intensive agriculture erodes  $\beta$ -diversity at large scales. *Ecol. Lett.* 15, 963–970. doi:10.1111/j.1461-0248.2012.01815.x
25. Kleijn, D., Van Der Voort, L.A.C., 1997. Conservation headlands for rare arable weeds: The effects of fertilizer application and light penetration on plant growth. *Biol. Conserv.* 81, 57–67. doi:10.1016/S0006-3207(96)00153-X
26. Maron, M., Fitzsimons, J.A., 2007. Agricultural intensification and loss of matrix habitat over 23 years in the West Wimmera, south-eastern Australia. *Biol. Conserv.* 135, 587–593. doi:10.1016/j.biocon.2006.10.051
27. Meeus, J.H.A., 1993. The transformation of agricultural landscapes in Western Europe. *Sci. Total Environ.* 129, 171–190.
28. Pielke, R.A., 2001. Influence of the spatial distribution of vegetation and soils on the prediction of cumulus convective rainfall. *Rev. Geophys.* 39, 151–177. doi:10.1029/1999RG000072
29. Roubos, C.R., Rodriguez-Saona, C., Isaacs, R., 2014. Mitigating the effects of insecticides on arthropod biological control at field and landscape scales. *Biol. Control* 75, 28–39. doi:10.1016/j.biocontrol.2014.01.006
30. Stoate, C., Boatman, N.D., Borralho, R.J., Carvalho, C.R., de Snoo, G.R., Eden, P., 2001. Ecological impacts of arable intensification in Europe. *J. Environ. Manage.* 63, 337–365. doi:10.1006/jema.2001.0473
31. Storkey, J., Meyer, S., Still, K.S., Leuschner, C., 2012. The impact of agricultural intensification and land-use change on the European arable flora. *Proc. R. Soc. B Biol. Sci.* 279, 1421–1429. doi:10.1098/rspb.2011.1686
32. Tilman, D., Fargione, J., Wolff, B., D'Antonio, C., Dobson, A., Howarth, R., Schindler, D., Schlesinger, W.H., Simberloff, D., Swackhamer, D., 2001. Forecasting agriculturally driven global environmental change. *Science* (80-. ). 292, 281–284. doi:10.1126/science.1057544

33. Tscharntke, T., Klein, A.M., Kruess, A., Steffan-Dewenter, I., Thies, C., 2005. Landscape perspectives on agricultural intensification and biodiversity - Ecosystem service management. *Ecol. Lett.* 8, 857–874. doi:10.1111/j.1461-0248.2005.00782.x
34. Van Oost, K., Govers, G., Desmet, P., 2000. Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. *Landsc. Ecol.* 15, 577–589. doi:10.1023/A:1008198215674
35. Welch, K.D., Harwood, J.D., 2014. Temporal dynamics of natural enemy-pest interactions in a changing environment. *Biol. Control* 75, 18–27. doi:10.1016/j.biocontrol.2014.01.004
36. Williams, P.H., 1982. The distribution and decline of British bumble bees (*Bombus Latr.*). *J. Apic. Res.* 21, 236–245. doi:10.1080/00218839.1982.11100549
37. Wilson, J.D., Evans, J., Browne, S.J., King, J.R., 1997. Territory Distribution and Breeding Success of Skylarks *Alauda arvensis* on Organic and Intensive Farmland in Southern England. *J. Appl. Ecol.* 34, 1462–1478.

**Figure A2: Theoretical model of Sustainable Intensification (Agriculture For Impact, 2013).**



**Table A1: Examples of data sources and indicators used by various authors to investigate agricultural intensity.**

Study Area	Data sources	Indicator(s)
Beijing mountainous region (Zhang and Li, 2016)	• Statistical Yearbook 2013	• Emergy analysis of agricultural inputs
Northern Spain (Armengot et al., 2011)	• Interviews	• Mean annual inputs of exogenous nitrogen • Weed control intensity • Cereal ratio • Crop diversity • Seed origin
Germany (Egorov et al., 2014)	• Yearly interviews with farmers and land-owners	• Land Use Intensity Index (LUI): summation of fertilization level, mowing frequency and grazing intensity
France (Teillard et al., 2012)	• French FADN (Farm Accountancy Data Network) • Datasets from agricultural social security, CAP declarations, national bovine identification • Topo-climatic data	• Input cost/ha
India (Biradar and Xiao, 2011)	• MODIS (EVI, NDVI, LSWI) • Government agricultural census data • Field ground-truth data inc. crop types and cropping pattern	• Cropping intensity
India (Jain et al., 2013)	• Landsat 5 TM & Landsat 7 ETM+ • MODIS (EVI) • Ground-truth data: Landsat, Quickbird, Worldview & Google Earth Imagery	• Cropping intensity • Multi-cropping
India (Singh et al., 2002)	• Crop cutting experiments • Crop yield estimation surveys • IRS-1B LISS-II	• Crop yield
Northern China (Mingwei et al., 2008)	• MODIS (NDVI)	• Crop acreage
China (Yan et al., 2014)	• Agricultural meteorological stations – crop calendar/crop phenological data • MODIS (EVI) • National LU/LC dataset	• Multi-cropping
China (L. Li et al., 2014)	• MODIS (EVI) • National survey data	• Cropping intensity
China (Xie et al., 2014)	• Secondary agricultural statistics e.g. China Rural Statistical Yearbook	• Emergy analysis of inputs to arable land per ha
United States (Johnson, 2013)	• Cropland Data Layer LC classifications (derived from Landsat TM by Agricultural Statistics Service) • NASS Census of Agriculture & June Acreage Survey	• Area of annually tilled cropland
US Central Great Plains (Wardlow and Egbert, 2008)	• MODIS (NDVI)	• Cropping area
Canada (Kerr and Cihlar, 2003)	• Canadian Census of Agriculture • SPOT 4 VEG	• Agricultural pollution

Table A1 continued

Study area	Data sources	Indicator(s)
Belgium, France, the Netherlands and Switzerland ( <i>Le Féon et al., 2010</i> )	• Standardised interviews with farmers	• Global Intensity Index: based on nitrogen input, livestock density and pesticide input
Europe (Donald et al., 2001)	• FAOSTAT database	• PCA analysis based on political & economic differences
Europe ( <i>Herzog et al., 2006</i> )	• Interviews • Geo-referenced aerial photographs • European Fourier-Adjusted & Interpolated NDVI dataset (Stockli & Vidale, 2004)	• Nitrogen output • Density of livestock units • Number of pesticide applications
European Union ( <i>Reidsma et al., 2006</i> )	• FADN survey	• Input costs • Irrigation use • Organic or not?
European Union ( <i>Temme and Verburg, 2011</i> )	• Agricultural statistics & census data • CORINE land cover map • Land Use/Cover Area frame statistical Survey (LUCAS) dataset	• Total nitrogen input
Europe & Turkey ( <i>Estel et al., 2016</i> )	• MODIS NDVI (Terra & Aqua satellites) • GlobCorine LC Map • Annual fallow/active crop maps	• Cropping frequency • Multi-cropping • Fallow cycles • Crop duration ratio
Russia ( <i>de Beurs and Ioffe, 2014</i> )	• Landsat 5 • MODIS • All-Russia Agricultural Census • Field Observations	• Cropping intensity
Asia ( <i>Gray et al., 2014</i> )	• MODIS (EVI)	• Multi-cropping
World ( <i>Johnston et al., 2011</i> )	• Global census data from FAOSTAT database – M3 cropland datasets	• Yield gap analysis
World ( <i>Niedertscheider et al., 2016</i> )	• Earthstat gridded maps of crop yields & crop area	• Human appropriation of net primary productivity (HANPP)
World ( <i>Potter et al., 2010</i> )	• National level fertiliser data – based on questionnaires • Global maps of harvested Area (from Monfreda et al. 2008) • FAO Gridded Livestock of the World maps	• Fertiliser inputs of N & P
World ( <i>Siebert et al., 2010</i> )	• MIRCA2000 dataset – monthly growing areas of 26 irrigated & rain-fed crop classes	• Cropping intensity • Crop duration ratio • Extent of fallow land
World ( <i>Thenkabail et al., 2009</i> )	• AVHRR • SPOT • JERS-1 • CRU rainfall time series (1961-2000) • Global Elevation dataset • Global Tree Cover data • Google Earth • Groundtruth data	• Irrigated area

**Table A2: Example EO-based methods used by researchers to assess agricultural intensity.**

Indicator	Example Methods
Crop yield	<p><i>NB: A brief review of crop yield estimation techniques can be found in the introduction to Doraiswamy et al. (2005, 2004, 2003) and Kasampalis et al. (2018) provide an overview of crop growth models.</i></p> <ul style="list-style-type: none"> <li>• Empirical regression-based modelling linking satellite-derived data (e.g. NDVI) to detailed official crop statistics (Becker-Reshef et al., 2010; Lobell et al., 2013; Salazar et al., 2007) or ground survey data (Ferencz et al., 2004; Ren et al., 2008; Yang et al., 2006); supplementary data may be included as additional explanatory variables (Balaghi et al., 2008; Prasad et al., 2006)</li> <li>• Estimate yield using regression models based on seasonal growth profiles from satellite-derived VIs (Kalubarme et al., 2003; Lai et al., 2018; Nagy et al., 2018; Son et al., 2014)</li> <li>• Crop yield simulation models incorporating satellite-derived data as either direct inputs or for calibration (Doraiswamy et al., 2005; Lobell et al., 2013; Moriondo et al., 2007)</li> <li>• Derive crop biomass using the Monteith light use efficiency approach (Awad, 2019; Leblon et al., 1991; Liu et al., 2010; Monteith, 1972; Morel et al., 2014; Pan et al., 2009; Patel et al., 2006)</li> </ul>
Cropping area	<p><i>NB: Gallego (2004) provides a review of some common EO-based land cover area estimation techniques.</i></p> <ul style="list-style-type: none"> <li>• Pixel counting &amp; sub-pixel analysis (spectral unmixing, linear mixing models, mixture modelling) applied to classified satellite images; ground data used as an auxiliary/validation tool (Gallego, 2004; Gallego et al., 2014; Gumma et al., 2014; Vibhute and Gawali, 2013)</li> <li>• Regression analysis combining satellite-derived information with an accurate sample (e.g. ground survey data) (Gallego, 2004; Gallego et al., 2014; Vibhute and Gawali, 2013)</li> <li>• Derive a cropland probability layer using a combination of classified images and satellite-derived land surface phenology metrics (de Beurs and Ioffe, 2014)</li> <li>• SAR time-series decomposition (Canisius et al., 2018; Ponnurangam and Rao, 2018; Xu et al., 2019); SAR data may be integrated with multispectral data to assist crop classification (Gao et al., 2018; Shuai et al., 2019)</li> </ul>
Cropping intensity – number of years a field is sown with crops and actually reaches harvest (de Beurs and Ioffe, 2014)	Jain et al. (2013) provide a comparison of different methods to map cropping intensity. Methods included are:
Cropping frequency – number of years a pixel was cropped over an observation period (Estel et al., 2016)	<ul style="list-style-type: none"> <li>• NDVI threshold method – define threshold for cropped land for a particular season based on satellite-derived NDVI, training data and regression tree analysis; use to classify pixels as cropped or uncropped agriculture for all seasons of interest</li> <li>• EVI peak method: (1) define threshold based on satellite-derived EVI and training data for cropped and non-cropped areas; (2) identify peaks in EVI time series; if peak exceeds threshold then classify it as cropped agriculture</li> <li>• Hierarchical training technique (using EVI): (1) define the percent of each pixel cropped using higher resolution ground-truth imagery; (2) use this to calibrate EVI to quantify the percent of each pixel that was cropped in each season</li> <li>• Apply a series of simple decision rules to satellite-derived phenology metrics (phenology model) to distinguish cropped pixels from fallow lands. <i>Example rules in de Beurs &amp; Ioffe (2014).</i></li> </ul>

Table A2 continued

Indicator	Example Methods
Multi-cropping – number of harvests within a single year (i.e. growing season) (Estel <i>et al.</i> , 2016)	<ul style="list-style-type: none"> <li>• Use Temporal Mixture Analysis of end-member phenologies to determine whether a pixel is single, double or triple cropped (Jain <i>et al.</i>, 2013)</li> <li>• Use a satellite-based phenology algorithm (e.g. Biradar and Xiao, 2011) to delineate the number of cropping cycles in a year</li> <li>• Apply time-series segmentation/iterative moving-window methodology to (smoothed) EVI time series to identify greening and browning phases and therefore cropping cycles; use to determine number of cropping cycles per year (Gray <i>et al.</i>, 2014)</li> </ul>
Cropping intensity – number of cropping cycles per year (L. Li <i>et al.</i> , 2014)	<ul style="list-style-type: none"> <li>• Use TIMESAT computer software to count the number of vegetation peaks in NDVI per growing season (Jönsson and Eklundh, 2004 in L. Li <i>et al.</i>, 2014; Z. Li <i>et al.</i>, 2014)</li> <li>• Determine the number of growth cycles in a year by counting the number of peaks (using thresholding techniques) on a crop growth curve based on satellite derived VI (e.g. EVI) (Yan <i>et al.</i>, 2014)</li> </ul>
Crop duration ratio	Ratio of the time period (during the growing season) for which a pixel was cropped and the total length of the growing season (Estel <i>et al.</i> , 2016)
Fallow cycles – recurring periods of fallow cropland	To identify fallow cycles: (1) Map active/fallow farmland based on NDVI time series (Estel <i>et al.</i> , 2015); (2) Filter time series for 'chain segments' i.e. certain number of consecutive fallow years; (3) count chain occurrence per pixel across entire time series; (4) summarise all chain segments using a weighting scheme (see Estel <i>et al.</i> , 2016 for details).

**Table A3: Example EO-based methods used by researchers to assess vegetation health.**

Indicator	Example EO-based Methods
<i>Crop condition</i>	<p><i>NB: A review of remote sensing methods for assessing crop condition can be found in Vibhute and Gawali (2013).</i></p> <ul style="list-style-type: none"> <li>• Vegetation indices e.g. NDVI, NDWI, SAVI etc. – assume the higher the indices, the better the crop condition (Ali and Pelkey, 2013; Vibhute and Gawali, 2013)</li> <li>• Same-period comparing – compare EO-derived data (e.g. NDVI, LAI, VCI) of a specific period with data from period in history to determine areas of deterioration, no-change &amp; improvement (Vibhute and Gawali, 2013; Wu <i>et al.</i>, 2015, 2014; Zhang <i>et al.</i>, 2014)</li> <li>• Crop growth profile monitoring (Jiang <i>et al.</i>, 2003; Vibhute and Gawali, 2013; Wu <i>et al.</i>, 2015, 2014; Zhang <i>et al.</i>, 2014)</li> </ul>
<b>Plant Trait Mapping</b>	<p><i>NB: Homolova et al. (2013) and Andrew et al. (2014) provide reviews of remote sensing techniques for mapping various plant traits.</i></p>
<i>Biophysical traits inc. biomass, fAPAR, photosynthetic capacity</i>	<ul style="list-style-type: none"> <li>• Empirical models (e.g. simple linear regression) relating limited field trait observations to EO-derived data such as vegetation indices (e.g. NDVI, EVI) &amp; classified images (Baret and Guyot, 1991; Chen <i>et al.</i>, 2010; Homolová <i>et al.</i>, 2013; Jackson <i>et al.</i>, 2004; Karnieli <i>et al.</i>, 2013; Sakowska <i>et al.</i>, 2016; Schino <i>et al.</i>, 2003; Sibanda <i>et al.</i>, 2015; Turner <i>et al.</i>, 1999)</li> <li>• Radiative Transfer Models (RTM) (Homolová <i>et al.</i>, 2013; Myneni <i>et al.</i>, 1997)</li> <li>• Estimate fAPAR using Neural Networks (Baghdadi <i>et al.</i>, 2016)</li> <li>• Hyperspectral methods such as partial least squares regression (Hansen and Schjoerring, 2003)</li> </ul>
<i>Structural traits inc. crop/canopy height, leaf area index (LAI), biomass, canopy morphology</i>	<p><i>NB: The introduction to Atzberger (2010) provides a brief overview of methods used to estimate LAI.</i></p> <ul style="list-style-type: none"> <li>• Empirical models using spectral data, VIs or image texture metrics (Andrew <i>et al.</i>, 2014; Baret and Guyot, 1991; Clevers <i>et al.</i>, 2017; Delegido <i>et al.</i>, 2011; Z. Li <i>et al.</i>, 2014; Wulder <i>et al.</i>, 2004)</li> <li>• RTM-based approaches e.g. SAIL (Atzberger, 2010; Doraiswamy <i>et al.</i>, 2004; Frampton <i>et al.</i>, 2013; Homolová <i>et al.</i>, 2013; Jackson <i>et al.</i>, 2004; Myneni <i>et al.</i>, 1997; Verhoef, 1984)</li> <li>• Estimate using correlation between surface properties and backscatter from active sensors (Andrew <i>et al.</i>, 2014; Z. Li <i>et al.</i>, 2014) e.g. LiDAR (Drake <i>et al.</i>, 2002; van Leeuwen and Nieuwenhuis, 2010) &amp; Radar (Brisco and Brown, 1998; Kasischke <i>et al.</i>, 1997)</li> </ul>
<i>Biochemical traits inc. chlorophyll (Ch) &amp; water content, nitrogen (N) &amp; phosphorous (P) status</i>	<ul style="list-style-type: none"> <li>• Empirical methods using vegetation indices (VIs) (Andrew <i>et al.</i>, 2014; Khanna <i>et al.</i>, 2007; Sakowska <i>et al.</i>, 2016) for example: Double-peak Canopy Nitrogen Index (DCNI) (Chen <i>et al.</i>, 2010); Modified triangle vegetation index 2 (Bagheri <i>et al.</i>, 2013); NDVI (Cheng <i>et al.</i>, 2008); Narrowband Green NDVI (NGNDVI) (Bausch and Khosla, 2010); Normalised Difference Water Index (NDWI) (Gao, 1996); Short-wave Infrared Water Stress Index (SIWSI) (Briant <i>et al.</i>, 2010; Fensholt and Sandholt, 2003); Triangular Greenness Index (TGI) (Hunt <i>et al.</i>, 2013); MERIS Terrestrial Chlorophyll Index (MTCI) (Dash and Curran, 2007)</li> <li>• RTM inversion e.g. REGFLEC (Andrew <i>et al.</i>, 2014; Boegh <i>et al.</i>, 2013; Frampton <i>et al.</i>, 2013; Homolová <i>et al.</i>, 2013; Houborg and Boegh, 2008; Jackson <i>et al.</i>, 2004; Trombetti <i>et al.</i>, 2008)</li> <li>• Estimate chlorophyll content using the red edge position (REP) (Z. Li <i>et al.</i>, 2014)</li> </ul>



**Table A4: Example EO-based methods used by researchers to assess soil quality and soil erosion/protection.**

Indicator	Example EO-based Methods
<b>Soil Quality</b>	<i>NB: Shoshany et al. (2013) provide a review of EO methods for monitoring agricultural soil degradation</i>
<i>Soil carbon (C) &amp; organic matter (OM)</i>	<ul style="list-style-type: none"> <li>• Empirical modelling (e.g. partial least squares regression, random forest) (Andrew et al., 2014; Castaldi et al., 2019, 2016; Gholizadeh et al., 2018; Stevens et al., 2010)</li> <li>• Quantify using particular absorption features in the VIS-NIR-SWIR region (Ben Dor et al., 1999 in Shoshany et al., 2013) or the degree of concavity of the reflectance spectrum in VIS wavelengths (Andrew et al., 2014; Palacios-Orueta and Ustin, 1998)</li> </ul>
<i>Crop residue/conservation tillage density</i>	<ul style="list-style-type: none"> <li>• Map organic residue (or non-photosynthetic vegetation) cover using spectral unmixing approaches (Andrew et al., 2014; Pacheco and McNairn, 2010)</li> <li>• Spectral indices designed for detecting crop residues include: Normalised Difference Tillage Index (NDTI) and Normalised Difference Senescent Vegetation Index (NDSVI) (Daughtry et al., 2006, 2005)</li> <li>• Map crop residue using a multiband reflectance algorithm e.g. Crop Residue Index Multiband (CRIM) (Biard and Baret, 1997)</li> </ul>
<i>Nitrogen (N) status/availability</i>	<ul style="list-style-type: none"> <li>• Assess based on two premises: (1) N mineralisation &amp; subsequent availability to growing crop will be proportional to OM content i.e. darker soil implies high soil nitrate levels (Scharf et al., 2002); (2) N stress increases canopy reflectance over all visible wavelengths (Beatty et al., 2000 in Scharf et al., 2002) – indices combining VIS &amp; NIR regions may maximise sensitivity to N stress (Eitel et al., 2011; Scharf et al., 2002; Tilling et al., 2007)</li> <li>• Assess nitrogen status using surface indicators of subsurface nutrient conditions using multispectral and hyperspectral techniques (Shoshany et al., 2013)</li> </ul>
<i>Soil salinity</i>	<p><i>NB: A review of the potentials and constraints of remote sensing-based soil salinity mapping can be found in Metternicht &amp; Zinck (2003).</i></p> <ul style="list-style-type: none"> <li>• Machine learning and regression-based models e.g. Multilayer Perception Neural Networks, Artificial Neural Networks, Gaussian Processes, Partial Least Square Regression, Support Vector Regression and Random Forest (Farifteh et al., 2007; Hoa et al., 2019; Taghadosi et al., 2019)</li> <li>• Categorical mapping of regions of differing soil salinity (e.g. high, medium, low) using hyperspectral satellite data and image classification (e.g. minimum distance, maximum likelihood) and spectral unmixing techniques (Ghosh et al., 2012; Hamzeh et al., 2016)</li> <li>• Quantitative mapping of soil salinity using indices derived from hyperspectral data and regression techniques (e.g. partial least squares regression, linear regression) (Bai et al., 2018; Hamzeh et al., 2013; Kumar et al., 2015; Mashimbye et al., 2012; Qian et al., 2019; Weng et al., 2008)</li> <li>• Distinguish ‘normal’ soil from moderately or severely salt-affected soils using brightness approach (Koshal 2010 in Shoshany et al., 2013)</li> <li>• Use spectral indices including Salinity Index (SI), Normalised Differential Salinity Index (NDSI) &amp; Brightness Index (BI) (Asfaw et al., 2018; Dehni and Lounis, 2012; Khan et al., 2005; Shoshany et al., 2013)</li> <li>• Detect salinisation-related surface roughness features (e.g. crusting) using variation in radar backscattering &amp; InSAR coherence signals (C, P &amp; R wavelengths) (Metternicht and Zinck, 2003; Shoshany et al., 2013; Taylor et al., 1996)</li> </ul>
<i>Soil moisture (SM) content</i>	<p><i>NB: Srivastava (2017) provides a review of satellite-based methods for monitoring soil moisture, while Petropoulos et al. (2018) provide an overview of the state of the art of EO techniques to derive operational estimates of soil moisture.</i></p> <ul style="list-style-type: none"> <li>• Empirical &amp; semi-empirical models relating backscattering coefficient to soil water content (&amp; soil surface roughness) (Attarzadeh et al., 2018; Bao et al., 2018; Bousbih et al., 2018; Dubois et al., 1995; Genis et al., 2013; Hajj et al., 2017; Hosseini et al., 2015; Huang et al., 2019; Zhang et al., 2017)</li> </ul>

Table A4 continued

Indicator	Example EO-based Methods
Soil moisture (SM) content continued	<ul style="list-style-type: none"> <li>• RTM-based approaches relating soil dielectric constant to soil moisture (Babiet <i>et al.</i>, 2018; Dubois <i>et al.</i>, 1995; Hosseini <i>et al.</i>, 2015; Jackson, 2002; Wagner <i>et al.</i>, 2007)</li> <li>• Spectral indices e.g. Normalised Multiband Drought Index (NMDI) (Shoshany <i>et al.</i>, 2013; Wang and Qu, 2009)</li> <li>• Retrieval from thermal data using apparent thermal inertia (ATI) (Shoshany <i>et al.</i>, 2013; Verstraeten <i>et al.</i>, 2006; Wang and Qu, 2009)</li> <li>• Map surface roughness &amp; SM in sparsely vegetated landscapes using a multi-angle (radar-based) approach &amp; an Integral Equation Model retrieval algorithm (Rahman <i>et al.</i>, 2008)</li> </ul>
<b>Soil erosion/ protection</b>	NB: Vrielling (2006) provide a review of satellite-based techniques for assessing erosion
Vegetation cover	<ul style="list-style-type: none"> <li>• Assess degree of protection based on amount of vegetation cover e.g. percentage ground cover, LAI (Cyr <i>et al.</i>, 1995; Dwivedi <i>et al.</i>, 1997; Fadul <i>et al.</i>, 1999; Metternicht and Zinck, 1998; Wang <i>et al.</i>, 2013)</li> </ul>
Erosion feature detection	<ul style="list-style-type: none"> <li>• Visual interpretation of high resolution images (Dwivedi <i>et al.</i>, 1997; Fadul <i>et al.</i>, 1999; Wang <i>et al.</i>, 2013)</li> <li>• Estimate metric dimensions &amp; volume of individual patches of sheet, rill &amp; gully erosion &amp; densities (Metternicht and Zinck, 1998; Shoshany <i>et al.</i>, 2013)</li> <li>• Potential to use InSAR multi-temporal interferometric coherence change technique (Shoshany <i>et al.</i>, 2013)</li> </ul>
Erosion modelling	<ul style="list-style-type: none"> <li>• Erosion Potential Index (EPI) (Shoshany <i>et al.</i>, 2013)</li> <li>• Integrate EO-derived data into soil loss/erosion models (Cyr <i>et al.</i>, 1995) such as USLE (Universal Soil Loss Equation), ANSWERS (Areal Non-Point Source Watershed Environment Response Simulation), SEMMED (Soil Erosion Model for Mediterranean Regions) (De Jong, 1994; De Jong <i>et al.</i>, 1999; Ganasri and Ramesh, 2016; Shoshany <i>et al.</i>, 2013). EO derived-data includes: land use/land cover map (Baban and Yusof, 2001; De Jong <i>et al.</i>, 1999; Ganasri and Ramesh, 2016; Sharma and Singh, 1995), interception or total vegetation cover (De Jong, 1994), and soil parameter data (Baban and Yusof, 2001; De Jong <i>et al.</i>, 1999; Sharma and Singh, 1995)</li> </ul>

**Table A5: Example EO-based methods used by researchers to assess water quality and water availability.**

Indicator	Example EO-based Methods
<b>Water Quality</b>	<i>NB: Reviews of various EO-based techniques for assessing various water quality parameters can be found in Gholizadeh et al. (2016a, 2016b); Dornhofer &amp; Oppelt (2016) and Chang et al. (2015)</i>
<i>Water Quality Indices</i>	<ul style="list-style-type: none"> <li>Water Quality indices derived from different combinations of spectral bands (Vignolo et al., 2006; Wen and Yang, 2010)</li> </ul>
<i>Physical water quality parameters: total suspended solids (TSS), turbidity, suspended sediment concentration (SSC), chlorophyll concentration, temperature &amp; water clarity</i>	<ul style="list-style-type: none"> <li>Empirical (simple or multiple regression) modelling – relate field data (or water quality indices derived from field data) to satellite data (e.g. band ratios) to estimate water quality parameters (Blix et al., 2018; Carpenter and Carpenter, 1983; Chen et al., 2007; Ha et al., 2017; Hu et al., 2004; Kloiber et al., 2002; Lavery et al., 1993; Liu et al., 2017; Pereira-Sandoval et al., 2018; Ritchie and Cooper, 2001; Sòria-perpinyà et al., 2019)</li> <li>Spectral unmixing-based approach – end-member spectra related to physical water quality parameters such as SSC (Martinez et al., 2008)</li> </ul>
<i>Chemical water quality parameters: concentration of total nitrogen (TN), NO<sub>3</sub>-N (nitrate as nitrogen) &amp; total phosphorous (TP)</i>	<ul style="list-style-type: none"> <li>Empirical (regression) modelling – relate field data to satellite data to estimate water quality parameters (Chen and Quan, 2012; Wu et al., 2010)</li> <li>Use neural network modelling (e.g. back-propagation neural network model) to establish a retrieval model for concentrations of TN &amp; TP on the basis of satellite data (Xiao et al., 2015)</li> </ul>
<i>Water quality proxy</i>	<ul style="list-style-type: none"> <li>Assess health of vegetation alongside water bodies as a proxy for water quality, using vegetation indices (e.g. NDVI, EVI) (Trivero et al., 2013)</li> <li>Identification and mapping of submergent aquatic vegetation using image interpretation and classification techniques (Ackleson and Klemas, 1987; Dogan et al., 2009; Wolter et al., 2005; Yang, 2005)</li> </ul>
<b>Water Availability</b>	
<i>Water body area &amp; configuration</i>	<ul style="list-style-type: none"> <li>Detect/classify water bodies using optical data (NIR &amp; SWIR regions) or spectral indices (e.g. NDVI &amp; NDWI) (Andrew et al., 2014; Frazier and Page, 2000; Mueller et al., 2016; Smith, 1997; Tulbure and Broich, 2013)</li> <li>WiPE water body classification algorithm (Ngoc et al., 2019)</li> <li>Determine water body area using pixel counting &amp; vector-based GIS methods (Verpoorter et al., 2012)</li> <li>Quantify the spatial configuration of water bodies (e.g. number of water bodies, mean surface water body area) based on classified satellite images using FRAGSTATS software (v4) (McGarigal et al., 2012; Tulbure and Broich, 2013)</li> </ul> <p><i>NB: Detection of water bodies may be enhanced through use of techniques such as Principal Component Analysis (PCA) (Verpoorter et al., 2012)</i></p>
<i>Water level &amp; volume</i>	<ul style="list-style-type: none"> <li>Estimate water level using satellite altimetry (Guo et al., 2009; Koblinsky et al., 1993; Michailovsky et al., 2012; Smith, 1997; Sulistioadi et al., 2015)</li> <li>Use satellite-derived LC data as an input into models to estimate the volume of water yield available for consumptive purposes (Crossman et al., 2013)</li> </ul>
<i>Water use efficiency &amp; crop water stress</i>	<ul style="list-style-type: none"> <li>Use satellite-derived data including VIs, NIR &amp; TIR data as inputs into evapotranspiration models such as SEBAL (Surface Energy Balance for Land) (Bastiaanssen, 2000; Bastiaanssen et al., 1998; Mutiga et al., 2010), METRIC (Allen et al., 2007) and ALEXI/DisALEXI (Anderson et al., 2011) to predict actual evapotranspiration as an indicator of crop water stress and whether water is being used as intended (Gonzalez-Dugo et al., 2009; Mutiga et al., 2010)</li> <li>Use TIR satellite data to calculate Evaporative Stress Index (ESI) to detect drought conditions and to infer crop health (Anderson and Kustas, 2008)</li> </ul>

**Table A6: Example EO-based methods used by researchers to assess biodiversity.**

Indicator	Example EO-based Methods
	<i>NB: EO-based techniques for monitoring biodiversity have been review by a number of authors including Gillespie et al. (2008), Kerr and Ostrovsky (2003), Kuenzer et al. (2014), Mairota et al. (2015), Nagendra (2001), Rocchini et al. (2010), Turner et al. (2003) and Wang et al. (2010)</i>
<i>Direct mapping of individuals and associations</i>	<ul style="list-style-type: none"> <li>• Map individual plants or associations of single species by applying pixel- or object-based classification procedures to high spatial resolution data (Ban, 2003; Clark <i>et al.</i>, 2001; Crossman <i>et al.</i>, 2013; Feng <i>et al.</i>, 2010; Gillespie <i>et al.</i>, 2008; Lauver, 1997; Nagendra and Gadgil, 1999; Turner <i>et al.</i>, 2003; Vibhute and Gawali, 2013). Species differentiation aided by differences in size, shape and vertical structure of canopies in active RS (e.g. LiDAR) or hyperspectral data (Andrew <i>et al.</i>, 2014)</li> <li>• Extract unique multi-temporal signature for different crops from VIs (e.g. NDVI &amp; EVI) (Wardlow <i>et al.</i>, 2007)</li> <li>• Harmonic (Fourier) analysis of NDVI time series (Jakubauskas <i>et al.</i>, 2002; Mingwei <i>et al.</i>, 2008)</li> <li>• Classify vegetation types using visual and digital interpretation of false colour composites and SAR images based on derived characteristics including size, shape and texture (Blaes <i>et al.</i>, 2005; Ravan <i>et al.</i>, 1995)</li> </ul>
<i>Plant (and animal) species diversity</i>	<ul style="list-style-type: none"> <li>• Assess species diversity and distribution patterns by examining direct relationships between EO-derived spectral radiance values and species distribution patterns recorded from field observations (Feng <i>et al.</i>, 2010; Nagendra, 2001)</li> <li>• Use satellite-based land use and landscape complexity indices (e.g. patch shape indices) to predict regional plant species diversity (Honnay <i>et al.</i>, 2003)</li> <li>• Predict distribution or probability of occurrence of individual species and species assemblages using multiple regression analysis and EO-based data such as land cover maps (Jennings, 2000; Kerr <i>et al.</i>, 2001; Kerr and Ostrovsky, 2003; Luoto <i>et al.</i>, 2002a; Saveraid <i>et al.</i>, 2001); supplementary material such as climate and topography data may be incorporated (Cumming, 2000; Nagendra, 2001)</li> </ul>
<i>Habitat suitability</i>	<ul style="list-style-type: none"> <li>• Model distribution and abundance of single species using detailed information about known habitat requirements and EO-derived land cover, habitat maps and landscape metrics (Amici <i>et al.</i>, 2010; Feng <i>et al.</i>, 2010; Kerr <i>et al.</i>, 2001; Kerr and Ostrovsky, 2003; Z. Li <i>et al.</i>, 2014; Luoto <i>et al.</i>, 2002b; Mairota <i>et al.</i>, 2015; Nagendra, 2001; Weiers <i>et al.</i>, 2004)</li> <li>• Habitat suitability parameters include: spectral and textural indexes (Muñoz and Felicísimo, 2004; Stickler and Southworth, 2008); canopy cover (Davis <i>et al.</i>, 2007); NPP (Meynard and Quinn, 2007); existence of suitable water bodies (Weiers <i>et al.</i>, 2004); and hedgerow networks (Vannier <i>et al.</i>, 2011)</li> </ul>
<i>Landscape structure</i>	<ul style="list-style-type: none"> <li>• Derive quantitative measures of landscape structure (e.g. composition, isolation and complexity) from land cover classifications (Gustafson, 1998; Kuenzer <i>et al.</i>, 2014; Luoto <i>et al.</i>, 2002a; Rocchini <i>et al.</i>, 2010)</li> <li>• Landscape metrics can be computed by software products including FRAGSTATS (v4) using raster or vector data (McGarigal <i>et al.</i>, 2012)</li> <li>• Landscape diversity may be represented using diversity indices combining <i>richness</i> (number of classes present) and <i>evenness</i> (distribution of area among classes) (Gustafson, 1998). Examples include Shannon's and Simpson's diversity indices (Shannon and Weaver, 1948 and Simpson, 1949 in Gustafson, 1998).</li> <li>• Use image classification to map landscape connectivity elements (e.g. hedgerows) (Vannier <i>et al.</i>, 2011)</li> <li>• Quantify landscape fragmentation using pattern indices (Saura, 2004) such as number of patches and mean patch size (Turner and Ruscher, 1988 in Saura, 2004) and patch cohesion index (Schumaker, 1996)</li> </ul>

Table A6 continued

Indicator	Example EO-based Methods
<i>Species richness</i>	<ul style="list-style-type: none"> <li>• Species-Energy Theory – species richness is proportional to NPP, derived from e.g. NDVI (Currie, 1991; Kerr and Ostrovsky, 2003; Z. Li <i>et al.</i>, 2014; Nagendra, 2001)</li> <li>• Spectral Variation Hypothesis – assume higher variation in spectra implies higher habitat heterogeneity, allowing coexistence of more species and consequently higher species richness (Diamond, 1988 in Fairbanks and McGwire, 2004; Z. Li <i>et al.</i>, 2014; Palmer <i>et al.</i>, 2002; Rocchini <i>et al.</i>, 2007)</li> <li>• Estimate spatial variation in species richness based on NDVI variability, vegetation classification map &amp; multiple regression analysis (Bawa <i>et al.</i>, 2002; Bino <i>et al.</i>, 2008; Bonthoux <i>et al.</i>, 2018; Carrasco <i>et al.</i>, 2018; Fairbanks and McGwire, 2004; Gould, 2000)</li> </ul>
<i>Invasive species</i>	<p><i>NB: Bradley et al. (2014) provide a review of remote sensing-based techniques for detecting invasive species</i></p> <ul style="list-style-type: none"> <li>• Identify invasive species using visual interpretation, pixel-based &amp; object-based classification, &amp; spectral mixing/unmixing approaches (Huang and Asner, 2009; Z. Li <i>et al.</i>, 2014; Walsh <i>et al.</i>, 2008)</li> <li>• Map vegetation species from spectral and textural data using image classification techniques (e.g. maximum likelihood classification) (Kimothe and Dasari, 2010; Laba <i>et al.</i>, 2010, 2008; Mirik and Ansley, 2012), neural networks (Fuller, 2005) and principal component analysis (Tsai <i>et al.</i>, 2007)</li> <li>• Texture-augmented image analysis (Tsai and Chou, 2006)</li> <li>• Maximum Entropy Model (Evangelista <i>et al.</i>, 2009)</li> </ul>

**Table A7: Example EO-based methods used by researchers to assess ecosystem health.**

Indicator		Example EO-based Methods
<i>NB: EO-based techniques for monitoring ecosystem health have been reviewed by a number of authors including Andrew et al. (2014), Feng et al. (2010) and Z. Li et al. (2014)</i>		
Vigour	Net Primary Productivity (NPP) & Gross Primary Productivity (GPP)	<ul style="list-style-type: none"> <li>• Model based on Light Use Efficiency (LUE) Concept (Feng et al., 2010; Monteith, 1972 and Prince 1991 in Z. Li et al., 2014; Ruimy and Saugier, 1994)</li> <li>• Statistical empirical model of GPP or NPP &amp; a vegetation indices such as NDVI or EVI (Feng et al., 2010; Z. Li et al., 2014; Olofsson et al., 2008; Xu et al., 2012)</li> <li>• Estimate GPP based on photosynthetic capacity quantified using satellite-based leaf Ch content estimates e.g. from <math>CI_{green}</math> index (Gitelson et al., 2008; Houborg et al., 2013)</li> </ul>
	Fractional cover of green vegetation, non-photosynthetic vegetation (NPV) & bare soil	<ul style="list-style-type: none"> <li>• Spectral mixing approach (Asner and Heidebrecht, 2002; Gill and Phinn, 2009; Gitelson, 2013; Z. Li et al., 2014; Pacheco and McNairn, 2010)</li> <li>• Empirical model of fractional vegetation cover &amp; vegetation indices (Carlson and Ripley, 1997; Gitelson, 2013; Guerschman et al., 2009; Z. Li et al., 2014; Wang et al., 2018)</li> <li>• Estimate using a neural network based on NIR &amp; red reflectances (Baret et al., 1995; Gitelson, 2013)</li> </ul>
	Biochemical properties inc. N, P & chlorophyll	<ul style="list-style-type: none"> <li>• Empirical modelling (based on biochemical spectra features) inc. simple linear regression, partial least-squares regression (PLSR) &amp; stepwise linear regression (SMLR) (Homolová et al., 2013; Z. Li et al., 2014)</li> </ul>
Organisation	Species richness & biodiversity	See table A6
	Structural traits	See table A3
Resilience		<ul style="list-style-type: none"> <li>• Assessed based on a ratio of a given ES health indicator, e.g. aboveground biomass, measured pre- &amp; post-disturbance (Z. Li et al., 2014)</li> <li>• Vegetation indices e.g. NDVI time series frequently used to assess/monitor variation in vegetation health &amp; deviation from normal conditions over time or in response to specific disturbances (Z. Li et al., 2014) such as climate change (Li and Guo, 2012), wildfires (Díaz-Delgado et al., 2002) &amp; grazing intensity (Numata et al., 2007)</li> </ul>
Ecosystem Services as a Proxy for Ecosystem Health		<ul style="list-style-type: none"> <li>• Use EO data (e.g. land cover) as an input for Ecosystem Services Models e.g. InVEST, ARIES, SolVES, GUMBO to assess the ability of an ecosystem to provide various ESS</li> <li>• Indirect modelling techniques include deriving empirical models of ESS or their providers based on spatial environmental covariates, and using maps of biophysical drivers of ESS supply to parameterise mechanistic models (Andrew et al., 2014).</li> </ul>

## References for Tables A1-A7

- Ackleson, S.G., Klemas, V., 1987. Remote sensing of submerged aquatic vegetation in lower chesapeake bay: A comparison of Landsat MSS to TM imagery. *Remote Sens. Environ.* 22, 235–248. [https://doi.org/10.1016/0034-4257\(87\)90060-5](https://doi.org/10.1016/0034-4257(87)90060-5)
- Ali, R., Pelkey, N., 2013. Satellite images indicate vegetation degradation due to invasive herbivores in the Andaman Islands. *Curr. Sci.* 105, 209–214.
- Allen, R.G., Tasumi, M., Morse, A., Trezza, R., Wright, J.L., Bastiaanssen, W., Kramber, W., Lorite, I., Robison, C.W., 2007. Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC) —Applications. *J. Irrig. Drain. Eng.* 133, 395–406. [https://doi.org/10.1061/\(ASCE\)0733-9437\(2007\)133:4\(395\)](https://doi.org/10.1061/(ASCE)0733-9437(2007)133:4(395))
- Amici, V., Geri, F., Battisti, C., 2010. An integrated method to create habitat suitability models for fragmented landscapes. *J. Nat. Conserv.* 18, 215–223. <https://doi.org/10.1016/j.jnc.2009.10.002>
- Anderson, M., Kustas, W., 2008. Thermal remote sensing of drought and evapotranspiration. *Eos (Washington, DC)*. 89, 233–240. <https://doi.org/10.1029/2008EO260001>
- Anderson, M.C., Kustas, W.P., Norman, J.M., Hain, C.R., Mecikalski, J.R., Schultz, L., González-Dugo, M.P., Cammalleri, C., D'Urso, G., Pimstein, A., Gao, F., 2011. Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrol. Earth Syst. Sci.* 15, 223–239. <https://doi.org/10.5194/hess-15-223-2011>
- Andrew, M.E., Wulder, M.A., Nelson, T.A., 2014. Potential contributions of remote sensing to ecosystem service assessments. *Prog. Phys. Geogr.* 38, 1–26. <https://doi.org/10.1177/0309133314528942>
- Armengot, L., José-María, L., Blanco-Moreno, J.M., Bassa, M., Chamorro, L., Xavier Sans, F., 2011. A novel index of land use intensity for organic and conventional farming of Mediterranean cereal fields. *Agron. Sustain. Dev.* 31, 699–707. <https://doi.org/10.1007/s13593-011-0042-0>
- Asfaw, E., Suryabagavan, K. V., Argaw, M., 2018. Soil salinity modeling and mapping using remote sensing and GIS: The case of Wonji sugar cane irrigation farm, Ethiopia. *J. Saudi Soc. Agric. Sci.* 17, 250–258. <https://doi.org/10.1016/j.jssas.2016.05.003>
- Asner, G.P., Heidebrecht, K.B., 2002. Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: comparing multispectral and hyperspectral observations. *Int. J. Remote Sens.* 23, 3939–3958. <https://doi.org/10.1080/01431160110115960>
- Attarzadeh, R., Amini, J., Notarnicola, C., Greifeneder, F., 2018. Synergetic Use of Sentinel-1 and Sentinel-2 Data for Soil Moisture Mapping at Plot Scale. *Remote Sens.* 10, 1285. <https://doi.org/10.3390/rs10081285>
- Atzberger, C., 2010. Inverting the PROSAIL canopy reflectance model using neural nets trained on streamlined databases. *J. Spectr. Imaging* 1, 1–13. <https://doi.org/10.1255/jsi.2010.a2>
- Awad, M., 2019. Toward Precision in Crop Yield Estimation Using Remote Sensing and Optimization Techniques. *Agriculture* 9, 54. <https://doi.org/10.3390/agriculture9030054>
- Baban, S.M.J., Yusof, K.W., 2001. Modelling soil erosion in tropical environments using remote sensing and geographical information systems. *Hydrol. Sci. -Journal-des Sci. Hydrol.* 46, 191–198. <https://doi.org/10.1080/02626660109492815>
- Bablet, A., Vu, P.V.H., Jacquemoud, S., Viallefont-Robinet, F., Fabre, S., Briottet, X., Sadeghi, M., Whiting, M.L., Baret, F., Tian, J., 2018. MARMIT: A multilayer radiative transfer model of soil reflectance to estimate surface soil moisture content in the solar domain (400–2500 nm). *Remote Sens. Environ.* 217, 1–17. <https://doi.org/10.1016/j.rse.2018.07.031>

- Baghdadi, N., Hajj, M. El, Zribi, M., Fayad, I., 2016. Coupling SAR C-band and optical data for soil moisture and leaf area index retrieval over irrigated grasslands. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 9, 1229–1243. <https://doi.org/10.1109/JSTARS.2015.2464698>
- Bagheri, N., Ahmadi, H., Kazem Alavipanah, S., Omid, M., 2013. Multispectral remote sensing for site-specific nitrogen fertilizer management. *Pesqui. Agropecuária Bras.* 48, 1394–1401. <https://doi.org/10.1590/S0100-204X2013001000011>
- Bai, L., Wang, C., Zang, S., Wu, C., Luo, J., Wu, Y., 2018. Mapping soil alkalinity and salinity in northern songnen plain, China with the hj-1 hyperspectral imager data and partial least squares regression. *Sensors* 18, 3855. <https://doi.org/10.3390/s18113855>
- Balaghi, R., Tychon, B., Eerens, H., Jlibene, M., 2008. Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. *Int. J. Appl. Earth Obs. Geoinf.* 10, 438–452. <https://doi.org/10.1016/j.jag.2006.12.001>
- Ban, Y., 2003. Synergy of multitemporal ERS-1 SAR and Landsat TM data for classification of agricultural crops. *Can. J. Remote Sens.* 29, 518–526. <https://doi.org/10.5589/m03-014>
- Bao, Y., Lin, L., Wu, S., Kwal Deng, K.A., Petropoulos, G.P., 2018. Surface soil moisture retrievals over partially vegetated areas from the synergy of Sentinel-1 and Landsat 8 data using a modified water-cloud model. *Int. J. Appl. Earth Obs. Geoinf.* 72, 76–85. <https://doi.org/10.1016/j.jag.2018.05.026>
- Baret, F., Clevers, J.G.P.W., Steven, M.D., 1995. The robustness of canopy gap fraction estimates from red and near-infrared reflectances: A comparison of approaches. *Remote Sens. Environ.* 54, 141–151. [https://doi.org/10.1016/0034-4257\(95\)00136-0](https://doi.org/10.1016/0034-4257(95)00136-0)
- Baret, F., Guyot, G., 1991. Potentials and Limits of Vegetation Indices for LAI and APAR Assessment. *Remote Sens. Environ.* 35, 161–173. [https://doi.org/10.1016/0034-4257\(91\)90009-U](https://doi.org/10.1016/0034-4257(91)90009-U)
- Bastiaanssen, W.G.M., 2000. SEBAL-based sensible and latent heat fluxes in the irrigated Gediz Basin, Turkey. *J. Hydrol.* 229, 87–100. [https://doi.org/10.1016/S0022-1694\(99\)00202-4](https://doi.org/10.1016/S0022-1694(99)00202-4)
- Bastiaanssen, W.G.M., Menenti, M., Feddes, R.A., Holtslag, A.A.M., 1998. A remote sensing surface energy balance algorithm for land (SEBAL) 1. Formulation. *J. Hydrol.* 212–213, 198–212. [https://doi.org/10.1016/S0022-1694\(98\)00253-4](https://doi.org/10.1016/S0022-1694(98)00253-4)
- Bausch, W.C., Khosla, R., 2010. QuickBird satellite versus ground-based multi-spectral data for estimating nitrogen status of irrigated maize. *Precis. Agric.* 11, 274–290. <https://doi.org/10.1007/s11119-009-9133-1>
- Bawa, K., Rose, J., Ganeshaiah, K.N., Barve, N., Kiran, M.C., Umashaanker, R., 2002. Assessing biodiversity from space: An example from the Western Ghats, India. *Conserv. Ecol.* 6, 7. <https://doi.org/10.5751/ES-00434-060207>
- Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* 114, 1312–1323. <https://doi.org/10.1016/j.rse.2010.01.010>
- Biard, F., Baret, F., 1997. Crop Residue Estimation Using Multiband Reflectance. *Remote Sens. Environ.* 59, 530–536. [https://doi.org/10.1016/S0034-4257\(96\)00125-3](https://doi.org/10.1016/S0034-4257(96)00125-3)
- Bino, G., Levin, N., Darawshi, S., Van Der Hal, N., Reich-Solomon, a., Kark, S., 2008. Accurate prediction of bird species richness patterns in an urban environment using Landsat-derived NDVI and spectral unmixing. *Int. J. Remote Sens.* 29, 3675–3700. <https://doi.org/10.1080/01431160701772534>
- Biradar, C.M., Xiao, X., 2011. Quantifying the area and spatial distribution of double-and triple-cropping croplands in India with multi-temporal MODIS imagery in 2005. *Int. J. Remote Sens.* 32, 367–386. <https://doi.org/10.1080/01431160903464179>



- Blaes, X., Vanhalle, L., Defourny, P., 2005. Efficiency of crop identification based on optical and SAR image time series. *Remote Sens. Environ.* 96, 352–365. <https://doi.org/10.1016/j.rse.2005.03.010>
- Blix, K., Pálffy, K., Tóth, V.R., Eltoft, T., 2018. Remote sensing of water quality parameters over Lake Balaton by using Sentinel-3 OLCI. *Water* 10, 1428. <https://doi.org/10.3390/w10101428>
- Boegh, E., Houborg, R., Bienkowski, J., Braban, C.F., Dalgaard, T., Van Dijk, N., Dragosits, U., Holmes, E., Magliulo, V., Schelde, K., Di Tommasi, P., Vitale, L., Theobald, M.R., Cellier, P., Sutton, M.A., 2013. Remote sensing of LAI, chlorophyll and leaf nitrogen pools of crop- and grasslands in five European landscapes. *Biogeosciences* 10, 6279–6307. <https://doi.org/10.5194/bg-10-6279-2013>
- Bonthoux, S., Lefèvre, S., Herrault, P.-A., Sheeren, D., 2018. Spatial and Temporal Dependency of NDVI Satellite Imagery in Predicting Bird Diversity over France. *Remote Sens.* 10, 1136. <https://doi.org/10.3390/rs10071136>
- Bousbih, S., Zribi, M., Hajj, M. El, Baghdadi, N., Lili-Chabaane, Z., Gao, Q., Fanise, P., 2018. Soil moisture and irrigation mapping in a semi-arid region, based on the synergetic use of Sentinel-1 and Sentinel-2 data. *Remote Sens.* 10, 1953. <https://doi.org/10.3390/rs10121953>
- Bradley, B.A., 2014. Remote detection of invasive plants: A review of spectral, textural and phenological approaches. *Biol. Invasions* 16, 1411–1425. <https://doi.org/10.1007/s10530-013-0578-9>
- Briant, G., Gond, V., Laurance, S.G.W., 2010. Habitat fragmentation and the desiccation of forest canopies: A case study from eastern Amazonia. *Biol. Conserv.* 143, 2763–2769. <https://doi.org/10.1016/j.biocon.2010.07.024>
- Brisco, B., Brown, R.J., 1998. Agricultural Applications with Radar, in: Henderson, F.M., Lewis, A.J. (Eds.), *Principles and Applications of Imaging Radar - Manual of Remote Sensing Third Edition, Volume 2*. John Wiley & Sons, New York, pp. 381–406.
- Canisius, F., Shang, J., Liu, J., Huang, X., Ma, B., Jiao, X., Geng, X., Kovacs, J.M., Walters, D., 2018. Tracking crop phenological development using multi-temporal polarimetric Radarsat-2 data. *Remote Sens. Environ.* 210, 508–518. <https://doi.org/10.1016/j.rse.2017.07.031>
- Carlson, T.N., Ripley, D.A., 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.* 62, 241–252. [https://doi.org/10.1016/S0034-4257\(97\)00104-1](https://doi.org/10.1016/S0034-4257(97)00104-1)
- Carpenter, D.J., Carpenter, S.M., 1983. Modeling Inland Water Quality Using Landsat Data. *Remote Sens. Environ.* 13, 345–352. [https://doi.org/10.1016/0034-4257\(83\)90035-4](https://doi.org/10.1016/0034-4257(83)90035-4)
- Carrasco, L., Norton, L., Henrys, P., Siriwardena, G.M., Rhodes, C.J., Rowland, C.S., Morton, D., 2018. Habitat diversity and structure regulate British bird richness : Implications of non-linear relationships for conservation. *Biol. Conserv.* 226, 256–263. <https://doi.org/10.1016/j.biocon.2018.08.010>
- Castaldi, F., Hueni, A., Chabrilat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B., 2019. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. *ISPRS J. Photogramm. Remote Sens.* 147, 267–282. <https://doi.org/10.1016/j.isprsjprs.2018.11.026>
- Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S., Casa, R., 2016. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. *Remote Sens. Environ.* 179, 54–65. <https://doi.org/10.1016/j.rse.2016.03.025>
- Chang, N. Bin, Imen, S., Vannah, B., 2015. Remote sensing for monitoring surface water quality status and ecosystem state in relation to the nutrient cycle: A 40-year perspective. *Crit. Rev. Environ. Sci. Technol.* 45, 101–166. <https://doi.org/10.1080/10643389.2013.829981>
- Chen, J., Quan, W., 2012. Using Landsat/TM imagery to estimate nitrogen and phosphorus concentration in Taihu Lake, China. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 5, 273–280. <https://doi.org/10.1109/JSTARS.2011.2174339>

- Chen, P., Haboudane, D., Tremblay, N., Wang, J., Vigneault, P., Li, B., 2010. New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat. *Remote Sens. Environ.* 114, 1987–1997. <https://doi.org/10.1016/j.rse.2010.04.006>
- Chen, Z., Muller-Karger, F.E., Hu, C., 2007. Remote sensing of water clarity in Tampa Bay. *Remote Sens. Environ.* 109, 249–259. <https://doi.org/10.1016/j.rse.2007.01.002>
- Cheng, Y.B., Ustin, S.L., Riaño, D., Vanderbilt, V.C., 2008. Water content estimation from hyperspectral images and MODIS indexes in Southeastern Arizona. *Remote Sens. Environ.* 112, 363–374. <https://doi.org/10.1016/j.rse.2007.01.023>
- Clark, P.E., Seyfried, M.S., Harris, B., 2001. Intermountain plant community classification using Landsat TM and SPOT HRV data. *J. Range Manag.* 54, 152–160. <https://doi.org/10.2307/4003176>
- Clevers, J.G.P.W., Kooistra, L., van den Brande, M.M.M., 2017. Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sens.* 9, 405. <https://doi.org/10.3390/rs9050405>
- Crossman, N.D., Burkhard, B., Nedkov, S., Willemen, L., Petz, K., Palomo, I., Drakou, E.G., Martín-Lopez, B., McPhearson, T., Boyanova, K., Alkemade, R., Egoh, B., Dunbar, M.B., Maes, J., 2013. A blueprint for mapping and modelling ecosystem services. *Ecosyst. Serv.* 4, 4–14. <https://doi.org/10.1016/j.ecoser.2013.02.001>
- Cumming, G.S., 2000. Using habitat models to map diversity: Pan-African species richness of ticks (Acari: Ixodida). *J. Biogeogr.* 27, 425–440. <https://doi.org/10.1046/j.1365-2699.2000.00419.x>
- Currie, D.J., 1991. Energy and Large-Scale Patterns of Animal-and Plant-Species Richness. *Am. Nat.* 137, 27–49. <https://doi.org/10.1086/285144>
- Cyr, L., Bonn, F., Pesant, A., 1995. Vegetation indices derived from remote sensing for an estimation of soil protection against water erosion. *Ecol. Modell.* 79, 277–285. [https://doi.org/10.1016/0304-3800\(94\)00182-H](https://doi.org/10.1016/0304-3800(94)00182-H)
- Dash, J., Curran, P.J., 2007. Evaluation of the MERIS terrestrial chlorophyll index (MTCI). *Adv. Sp. Res.* 39, 100–104. <https://doi.org/10.1016/j.asr.2006.02.034>
- Daughtry, C.S.T., Doraiswamy, P.C., Hunt, E.R., Stern, A.J., McMurtrey Iii, J.E., Prueger, J.H., 2006. Remote sensing of crop residue cover and soil tillage intensity. *Soil Tillage Res.* 91, 101–108. <https://doi.org/10.1016/j.still.2005.11.013>
- Daughtry, C.S.T., Hunt, E.R., Doraiswamy, P.C., McMurtrey, J.E., 2005. Remote sensing the spatial distribution of crop residues. *Agron. J.* 97, 864–871. <https://doi.org/10.2134/agronj2003.0291>
- Davis, F.W., Seo, C., Zielinski, W.J., 2007. Regional variation in home-range-scale habitat models for fisher (*Martes pennanti*) in California. *Ecol. Appl.* 17, 2195–2213. <https://doi.org/10.1890/06-1484.1>
- de Beurs, K.M., Ioffe, G., 2014. Use of Landsat and MODIS data to remotely estimate Russia's sown area. *J. Land Use Sci.* 9, 377–401. <https://doi.org/10.1080/1747423X.2013.798038>
- De Jong, S.M., 1994. Derivation of vegetative variables from a landsat TM image for modelling soil erosion. *Earth Surf. Process. Landforms* 19, 165–178. <https://doi.org/10.1002/esp.3290190207>
- De Jong, S.M., Paracchini, M.L., Bertolo, F., Folving, S., Megier, J., De Roo, A.P.J., 1999. Regional assessment of soil erosion using the distributed model SEMMED and remotely sensed data. *Catena* 37, 291–308. [https://doi.org/10.1016/S0341-8162\(99\)00038-7](https://doi.org/10.1016/S0341-8162(99)00038-7)
- Dehni, A., Lounis, M., 2012. Remote sensing techniques for salt affected soil mapping: Application to the Oran region of Algeria. *Procedia Eng.* 33, 188–198. <https://doi.org/10.1016/j.proeng.2012.01.1193>

- Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors* 11, 7063–7081. <https://doi.org/10.3390/s110707063>
- Díaz-Delgado, R., Lloret, F., Pons, X., Terradas, J., 2002. Satellite evidence of decreasing resilience in Mediterranean plant communities after recurrent wildfires. *Ecology* 83, 2293–2303. [https://doi.org/10.1890/0012-9658\(2002\)083\[2293:SEODRI\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2293:SEODRI]2.0.CO;2)
- Dogan, O.K., Akyurek, Z., Beklioglu, M., 2009. Identification and mapping of submerged plants in a shallow lake using quickbird satellite data. *J. Environ. Manage.* 90, 2138–2143. <https://doi.org/10.1016/j.jenvman.2007.06.022>
- Donald, P.F., Green, R.E., Heath, M.F., 2001. Agricultural intensification and the collapse of Europe's farmland bird populations. *Proc. R. Soc. B* 268, 25–29. <https://doi.org/10.1098/rspb.2000.1325>
- Doraiswamy, P.C., Hatfield, J.L., Jackson, T.J., Akhmedov, B., Prueger, J., Stern, A., 2004. Crop condition and yield simulations using Landsat and MODIS. *Remote Sens. Environ.* 92, 548–559. <https://doi.org/10.1016/j.rse.2004.05.017>
- Doraiswamy, P.C., Moulin, S., Cook, P.W., Stern, A., 2003. Crop Yield Assessment from Remote Sensing. *Photogramm. Eng. Remote Sens.* 69, 665–674. <https://doi.org/10.14358/pers.69.6.665>
- Doraiswamy, P.C., Sinclair, T.R., Hollinger, S., Akhmedov, B., Stern, A., Prueger, J., 2005. Application of MODIS derived parameters for regional crop yield assessment. *Remote Sens. Environ.* 97, 192–202. <https://doi.org/10.1016/j.rse.2005.03.015>
- Dörnhöfer, K., Oppelt, N., 2016. Remote sensing for lake research and monitoring - Recent advances. *Ecol. Indic.* 64, 105–122. <https://doi.org/10.1016/j.ecolind.2015.12.009>
- Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M.A., Chazdon, R.L., Weishampel, J.F., Prince, S., 2002. Estimation of tropical forest structural characteristics, using large-footprint lidar. *Remote Sens. Environ.* 79, 305–319. [https://doi.org/10.1016/S0034-4257\(01\)00281-4](https://doi.org/10.1016/S0034-4257(01)00281-4)
- Dubois, P.C., van Zyl, J., Engman, T., 1995. Measuring Soil Moisture with Imaging Radars. *IEEE Trans. Geosci. Remote Sens.* 33, 915–926. <https://doi.org/10.1109/36.406677>
- Dwivedi, R.S., Sankar, T.R., Venkataratnam, L., Karale, R.L., Gawande, S.P., Rao, K.V.S., Senchaudhary, S., Bhaumik, K.R., Mukharjee, K.K., 1997. The inventory and monitoring of eroded lands using remote sensing data. *Int. J. Remote Sens.* 18, 107–119. <https://doi.org/10.1080/014311697219303>
- Egorov, E., Prati, D., Durka, W., Michalski, S., Fischer, M., Schmitt, B., Blaser, S., Brändle, M., 2014. Does land-use intensification decrease plant phylogenetic diversity in local grasslands? *PLoS One* 9, e103252. <https://doi.org/10.1371/journal.pone.0103252>
- Eitel, J.U.H., Vierling, L.A., Litvak, M.E., Long, D.S., Schulthess, U., Ager, A.A., Krofcheck, D.J., Stoscheck, L., 2011. Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. *Remote Sens. Environ.* 115, 3640–3646. <https://doi.org/10.1016/j.rse.2011.09.002>
- Estel, S., Kuemmerle, T., Alcántara, C., Levers, C., Prishchepov, A., Hostert, P., 2015. Mapping farmland abandonment and recultivation across Europe using MODIS NDVI time series. *Remote Sens. Environ.* 163. <https://doi.org/10.1016/j.rse.2015.03.028>
- Estel, S., Kuemmerle, T., Levers, C., Baumann, M., Hostert, P., 2016. Mapping cropland-use intensity across Europe using MODIS NDVI time series. *Environ. Res. Lett.* 11. <https://doi.org/10.1088/1748-9326/11/2/024015>
- Evangelista, P.H., Stohlgren, T.J., Morissette, J.T., Kumar, S., 2009. Mapping invasive tamarisk (*Tamarix*): A comparison of single-scene and time-series analyses of remotely sensed data. *Remote Sens.* 1, 519–533. <https://doi.org/10.3390/rs1030519>

- Fadul, H.M., Salih, A.A., Ali, I.A., Inanaga, S., 1999. Use of remote sensing to map gully erosion along the Atbara River, Sudan. *JAG* 1, 314–1999. [https://doi.org/10.1016/S0303-2434\(99\)85010-7](https://doi.org/10.1016/S0303-2434(99)85010-7)
- Fairbanks, D.H.K., McGwire, K.C., 2004. Patterns of floristic richness in vegetation communities of California: Regional scale analysis with multi-temporal NDVI. *Glob. Ecol. Biogeogr.* 13, 221–235. <https://doi.org/10.1111/j.1466-822X.2004.00092.x>
- Farifteh, J., Van der Meer, F., Atzberger, C., Carranza, E.J.M., 2007. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sens. Environ.* 110, 59–78. <https://doi.org/10.1016/j.rse.2007.02.005>
- Feng, X., Fu, B., Yang, X., Lü, Y., 2010. Remote sensing of ecosystem services: An opportunity for spatially explicit assessment. *Chinese Geogr. Sci.* 20, 522–535. <https://doi.org/10.1007/s11769-010-0428-y>
- Fensholt, R., Sandholt, I., 2003. Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment. *Remote Sens. Environ.* 87, 111–121. <https://doi.org/10.1016/j.rse.2003.07.002>
- Ferencz, C., Bognár, P., Lichtenberger, J., Hamar, D., Tarcsai, G., Timár, G., Molnár, G., Pásztor, S.Z., Steinbach, P., Székely, B., Ferencz, O.E., Ferencz-Árkos, I., 2004. Crop yield estimation by satellite remote sensing. *Int. J. Remote Sens.* 25, 4113–4149. <https://doi.org/10.1080/01431160410001698870>
- Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J. Photogramm. Remote Sens.* 82, 83–92. <https://doi.org/10.1016/j.isprsjprs.2013.04.007>
- Frazier, P.S., Page, K.J., 2000. Water body detection and delineating with Landsat TM data. *Photogramm. Eng. Remote Sens.* 66, 1461–1467.
- Fuller, D.O., 2005. Remote detection of invasive *Melaleuca* trees (*Melaleuca quinquenervia*) in South Florida with multispectral IKONOS imagery. *Int. J. Remote Sens.* 26, 1057–1063. <https://doi.org/10.1080/01430060512331314119>
- Gallego, F.J., 2004. Remote sensing and land cover area estimation. *Int. J. Remote Sens.* 25, 3019–3047. <https://doi.org/10.1080/01431160310001619607>
- Gallego, F.J., Kussul, N., Skakun, S., Kravchenko, O., Shelestov, A., Kussul, O., 2014. Efficiency assessment of using satellite data for crop area estimation in Ukraine. *Int. J. Appl. Earth Obs. Geoinf.* 29, 22–30. <https://doi.org/10.1016/j.jag.2013.12.013>
- Ganasri, B.P., Ramesh, H., 2016. Assessment of soil erosion by RUSLE model using remote sensing and GIS - A case study of Nethravathi Basin. *Geosci. Front.* 7, 953–961. <https://doi.org/10.1016/j.gsf.2015.10.007>
- Gao, B.C., 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Gao, H., Wang, C., Wang, G., Zhu, J., Tang, Y., Shen, P., Zhu, Z., 2018. A crop classification method integrating GF-3 PolSAR and sentinel-2A optical data in the Dongting lake basin. *Sensors* 18, 3139. <https://doi.org/10.3390/s18093139>
- Genis, A., Vulfson, L., Blumberg, D.G., Sprinstin, M., Kotlyar, A., Freilikher, V., Ben-Asher, J., 2013. Retrieving parameters of bare soil surface roughness and soil water content under arid environment from ERS-1, -2 SAR data. *Int. J. Remote Sens.* 34, 6202–6215. <https://doi.org/10.1080/01431161.2013.793862>
- Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens. Environ.* 218, 89–103. <https://doi.org/10.1016/j.rse.2018.09.015>
- Gholizadeh, M.H., Melesse, A.M., Reddi, L., 2016a. A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors* 16, 1298. <https://doi.org/10.3390/s16081298>

- Gholizadeh, M.H., Melesse, A.M., Reddi, L., 2016b. Spaceborne and airborne sensors in water quality assessment. *Int. J. Remote Sens.* 37, 3143–3180. <https://doi.org/10.1080/01431161.2016.1190477>
- Ghosh, G., Kumar, S., Saha, S.K., 2012. Hyperspectral Satellite Data in Mapping Salt-Affected Soils Using Linear Spectral Unmixing Analysis. *J. Indian Soc. Remote Sens.* 40, 129–136. <https://doi.org/10.1007/s12524-011-0143-x>
- Gill, T.K., Phinn, S.R., 2009. Improvements to ASTER-Derived Fractional Estimates of Bare Ground in a Savanna Rangeland. *IEEE Trans. Geosci. Remote Sens.* 47, 662–670. <https://doi.org/10.1109/TGRS.2008.2004628>
- Gillespie, T.W., Foody, G.M., Rocchini, D., Giorgi, A.P., Saatchi, S., 2008. Measuring and modelling biodiversity from space. *Prog. Phys. Geogr.* 32, 203–221. <https://doi.org/10.1177/0309133308093606>
- Gitelson, A.A., 2013. Remote estimation of crop fractional vegetation cover: the use of noise equivalent as an indicator of performance of vegetation indices. *Int. J. Remote Sens.* 34, 6054–6066. <https://doi.org/10.1080/01431161.2013.793868>
- Gitelson, A.A., Viña, A., Masek, J.G., Verma, S.B., Suyker, A.E., 2008. Synoptic monitoring of gross primary productivity of maize using landsat data. *IEEE Geosci. Remote Sens. Lett.* 5, 133–137. <https://doi.org/10.1109/LGRS.2008.915598>
- Gonzalez-Dugo, M.P., Neale, C.M.U., Mateos, L., Kustas, W.P., Prueger, J.H., Anderson, M.C., Li, F., 2009. A comparison of operational remote sensing-based models for estimating crop evapotranspiration. *Agric. For. Meteorol.* 149, 1843–1853. <https://doi.org/10.1016/j.agrformet.2009.06.012>
- Gould, W., 2000. Remote Sensing of Vegetation, Plant Species Richness, and Regional Biodiversity Hotspots. *Ecol. Appl.* 10, 1861–1870. <https://doi.org/10.2307/2641244>
- Gray, J., Friedl, M., Froking, S., Ramankutty, N., Nelson, A., Gumma, M.K., 2014. Mapping Asian cropping intensity with MODIS. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 7, 3373–3379. <https://doi.org/10.1109/JSTARS.2014.2344630>
- Guerschman, J.P., Hill, M.J., Renzullo, L.J., Barrett, D.J., Marks, A.S., Botha, E.J., 2009. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sens. Environ.* 113, 928–945. <https://doi.org/10.1016/j.rse.2009.01.006>
- Gumma, M.K., Thenkabail, P.S., Maunahan, A., Islam, S., Nelson, A., 2014. Mapping seasonal rice cropland extent and area in the high cropping intensity environment of Bangladesh using MODIS 500m data for the year 2010. *ISPRS J. Photogramm. Remote Sens.* 91, 98–113. <https://doi.org/10.1016/j.isprsjprs.2014.02.007>
- Guo, J., Chang, X., Gao, Y., Sun, J., Hwang, C., 2009. Lake level variations monitored with satellite altimetry waveform retracking. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2, 80–86. <https://doi.org/10.1109/JSTARS.2009.2021673>
- Gustafson, E.J., 1998. Quantifying Landscape Spatial Pattern: What Is the State of the Art? *Ecosystems* 1, 143–156. <https://doi.org/10.1007/s100219900011>
- Ha, N.T.T., Thao, N.T.P., Koike, K., Nhuan, M.T., 2017. Selecting the Best Band Ratio to Estimate Chlorophyll-a Concentration in a Tropical Freshwater Lake Using Sentinel 2A Images from a Case Study of Lake Ba Be (Northern Vietnam). *ISPRS Int. J. Geo-Information* 6, 290. <https://doi.org/10.3390/ijgi6090290>
- Hajj, M. El, Baghdadi, N., Zribi, M., Bazzi, H., 2017. Synergic use of Sentinel-1 and Sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas. *Remote Sens.* 9, 1292. <https://doi.org/10.3390/rs9121292>
- Hamzeh, S., Naseri, A.A., AlaviPanah, S.K., Bartholomeus, H., Herold, M., 2016. Assessing the accuracy of hyperspectral and multispectral satellite imagery for categorical and quantitative mapping of salinity

- stress in sugarcane fields. *Int. J. Appl. Earth Obs. Geoinf.* 52, 412–421. <https://doi.org/10.1016/j.jag.2016.06.024>
- Hamzeh, S., Naseri, A.A., AlaviPanah, S.K., Mojaradi, B., Bartholomeus, H.M., Clevers, J.G.P.W., Behzad, M., 2013. Estimating salinity stress in sugarcane fields with spaceborne hyperspectral vegetation indices. *Int. J. Appl. Earth Obs. Geoinf.* 21, 282–290. <https://doi.org/10.1016/j.jag.2012.07.002>
- Hansen, P.M., Schjoerring, J.K., 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* 86, 542–553. [https://doi.org/10.1016/S0034-4257\(03\)00131-7](https://doi.org/10.1016/S0034-4257(03)00131-7)
- Herzog, F., Steiner, B., Bailey, D., Baudry, J., Billeter, R., Bukáček, R., De Blust, G., De Cock, R., Dirksen, J., Dormann, C.F., De Filippi, R., Frossard, E., Liira, J., Schmidt, T., Stöckli, R., Thenail, C., Van Wingerden, W., Bugter, R., 2006. Assessing the intensity of temperate European agriculture at the landscape scale. *Eur. J. Agron.* 24, 165–181. <https://doi.org/10.1016/j.eja.2005.07.006>
- Hoa, P.V., Giang, N.V., Binh, N.A., Hai, L.V.H., Pham, T.D., Hasanlou, M., Bui, D.T., 2019. Soil salinity mapping using SAR Sentinel-1 data and advanced machine learning algorithms: A case study at Ben Tre Province of the Mekong River Delta (Vietnam). *Remote Sens.* 11, 128. <https://doi.org/10.3390/rs11020128>
- Homolová, L., Malenovský, Z., Clevers, J.G.P.W., García-Santos, G., Schaepman, M.E., 2013. Review of optical-based remote sensing for plant trait mapping. *Ecol. Complex.* 15, 1–16. <https://doi.org/10.1016/j.ecocom.2013.06.003>
- Honnay, O., Piessens, K., Van Landuyt, W., Hermy, M., Gulink, H., 2003. Satellite based land use and landscape complexity indices as predictors for regional plant species diversity. *Landsc. Urban Plan.* 63, 241–250. [https://doi.org/10.1016/S0169-2046\(02\)00194-9](https://doi.org/10.1016/S0169-2046(02)00194-9)
- Hosseini, R., Newlands, N.K., Dean, C.B., Takemura, A., 2015. Statistical modeling of soil moisture, integrating satellite remote-sensing (SAR) and ground-based data. *Remote Sens.* 7, 2752–2780. <https://doi.org/10.3390/rs70302752>
- Houborg, R., Boegh, E., 2008. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sens. Environ.* 112, 186–202. <https://doi.org/10.1016/j.rse.2007.04.012>
- Houborg, R., Cescatti, A., Migliavacca, M., Kustas, W.P., 2013. Satellite retrievals of leaf chlorophyll and photosynthetic capacity for improved modeling of GPP. *Agric. For. Meteorol.* 177, 10–23. <https://doi.org/10.1016/j.agrformet.2013.04.006>
- Hu, C., Chen, Z., Clayton, T.D., Swarzenski, P., Brock, J.C., Muller-Karger, F.E., 2004. Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from Tampa Bay, FL. *Remote Sens. Environ.* 93, 423–441. <https://doi.org/10.1016/j.rse.2004.08.007>
- Huang, C.-Y., Asner, G.P., 2009. Applications of Remote Sensing to Alien Invasive Plant Studies. *Sensors* 9, 4869–4889. <https://doi.org/10.3390/s90604869>
- Huang, S., Ding, J., Zou, J., Liu, B., Zhang, J., Chen, W., 2019. Soil moisture retrieval based on sentinel-1 imagery under sparse vegetation coverage. *Sensors* 19, 589. <https://doi.org/10.3390/s19030589>
- Hunt, E.R., Doraiswamy, P.C., McMurtrey, J.E., Daughtry, C.S.T., Perry, E.M., Akhmedov, B., 2013. A visible band index for remote sensing leaf chlorophyll content at the canopy scale. *Int. J. Appl. Earth Obs. Geoinf.* 21, 103–112. <https://doi.org/10.1016/j.jag.2012.07.020>
- Jackson, T.J., 2002. Remote sensing of soil moisture: Implications for groundwater recharge. *Hydrogeol. J.* 10, 40–51. <https://doi.org/10.1007/s10040-001-0168-2>

- Jackson, T.J., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., Doriaswamy, P., Hunt, E.R., 2004. Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. *Remote Sens. Environ.* 92, 475–482. <https://doi.org/10.1016/j.rse.2003.10.021>
- Jain, M., Mondal, P., DeFries, R.S., Small, C., Galford, G.L., 2013. Mapping cropping intensity of smallholder farms: A comparison of methods using multiple sensors. *Remote Sens. Environ.* 134, 210–223. <https://doi.org/10.1016/j.rse.2013.02.029>
- Jakubauskas, M.E., Legates, D.R., Kastens, J.H., 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Comput. Electron. Agric.* 37, 127–139. [https://doi.org/10.1016/S0168-1699\(02\)00116-3](https://doi.org/10.1016/S0168-1699(02)00116-3)
- Jennings, M.D., 2000. Gap analysis: Concepts, methods, and recent results. *Landsc. Ecol.* 15, 5–20. <https://doi.org/10.1023/A:1008184408300>
- Jiang, D., Wang, N., Yang, X., Wang, J., 2003. Study on the interaction between NDVI profile and the growing status of crops. *Chinese Geogr. Sci.* 13, 62–65. <https://doi.org/10.1007/s11769-003-0086-4>
- Johnson, D.M., 2013. A 2010 map estimate of annually tilled cropland within the conterminous United States. *Agric. Syst.* 114, 95–105. <https://doi.org/10.1016/j.agry.2012.08.004>
- Johnston, M., Licker, R., Foley, J., Holloway, T., Mueller, N.D., Barford, C., Kucharik, C., 2011. Closing the gap: global potential for increasing biofuel production through agricultural intensification. *Environ. Res. Lett.* 6, 1–11. <https://doi.org/10.1088/1748-9326/6/3/034028>
- Kalubarme, M.H., Potdar, M.B., Manjunath, K.R., Mahey, R.K., Siddhu, S.S., 2003. Growth profile based crop yield models: A case study of large area wheat yield modelling and its extendibility using atmospheric corrected NOAA AVHRR data. *Int. J. Remote Sens.* 24, 2037–2054. <https://doi.org/10.1080/01431160210156018>
- Karnieli, A., Bayarjargal, Y., Bayasgalan, M., Mandakh, B., Dugarjav, C., Burgheimer, J., Khudulmur, S., Bazha, S.N., Gunin, P.D., 2013. Do vegetation indices provide a reliable indication of vegetation degradation? A case study in the Mongolian pastures. *Int. J. Remote Sens.* 34, 6243–6262. <https://doi.org/10.1080/01431161.2013.793865>
- Kasampalis, D., Alexandridis, T., Deva, C., Challinor, A., Moshou, D., Zalidis, G., 2018. Contribution of Remote Sensing on Crop Models: A Review. *J. Imaging* 4, 52. <https://doi.org/10.3390/jimaging4040052>
- Kasischke, E.S., Melack, J.M., Dobson, M.C., 1997. The use of imaging radars for ecological applications - A review. *Remote Sens. Environ.* 59, 141–156. [https://doi.org/10.1016/S0034-4257\(96\)00148-4](https://doi.org/10.1016/S0034-4257(96)00148-4)
- Kerr, J.T., Cihlar, J., 2003. Land use and cover with intensity of agriculture for Canada from satellite and census data. *Glob. Ecol. Biogeogr.* 12, 161–172. <https://doi.org/10.1046/j.1466-822X.2003.00017.x>
- Kerr, J.T., Ostrovsky, M., 2003. From space to species: Ecological applications for remote sensing. *Trends Ecol. Evol.* 18, 299–305. [https://doi.org/10.1016/S0169-5347\(03\)00071-5](https://doi.org/10.1016/S0169-5347(03)00071-5)
- Kerr, J.T., Southwood, T.R.E., Cihlar, J., 2001. Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. *PNAS* 98, 11365–11370. <https://doi.org/10.1073/pnas.201398398>
- Khan, N.M., Rastokuev, V. V., Sato, Y., Shiozawa, S., 2005. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agric. Water Manag.* 77, 96–109. <https://doi.org/10.1016/j.agwat.2004.09.038>
- Khanna, S., Palacios-Orueta, A., Whiting, M.L., Ustin, S.L., Riaño, D., Litago, J., 2007. Development of angle indexes for soil moisture estimation, dry matter detection and land-cover discrimination. *Remote Sens. Environ.* 109, 154–165. <https://doi.org/10.1016/j.rse.2006.12.018>

- Kimothi, M.M., Dasari, A., 2010. Methodology to map the spread of an invasive plant (*Lantana camara* L.) in forest ecosystems using Indian remote sensing satellite data. *Int. J. Remote Sens.* 31, 3273–3289. <https://doi.org/10.1080/01431160903121126>
- Kloiber, S.M., Brezonik, P.L., Olmanson, L.G., Bauer, M.E., 2002. A procedure for regional lake water clarity assessment using Landsat multispectral data. *Remote Sens. Environ.* 82, 38–47. [https://doi.org/10.1016/S0034-4257\(02\)00022-6](https://doi.org/10.1016/S0034-4257(02)00022-6)
- Koblinsky, C.J., Clarke, R.T., Brenner, A.C., Frey, H., 1993. Measurement of river level variations with satellite altimetry. *Water Resour. Res.* 29, 1839–1848. <https://doi.org/10.1029/93WR00542>
- Kuenzer, C., Ottinger, M., Wegmann, M., Guo, H., Wang, C., Zhang, J., Dech, S., Wikelski, M., 2014. Earth observation satellite sensors for biodiversity monitoring: potentials and bottlenecks. *Int. J. Remote Sens.* 35, 6599–6647. <https://doi.org/10.1080/01431161.2014.964349>
- Kumar, S., Gautam, G., Saha, S.K., 2015. Hyperspectral remote sensing data derived spectral indices in characterizing salt-affected soils: a case study of Indo-Gangetic plains of India. *Environ. Earth Sci.* 73, 3299–3308. <https://doi.org/10.1007/s12665-014-3613-y>
- Laba, M., Blair, B., Downs, R., Monger, B., Philpot, W., Smith, S., Sullivan, P., Baveye, P.C., 2010. Use of textural measurements to map invasive wetland plants in the Hudson River National Estuarine Research Reserve with IKONOS satellite imagery. *Remote Sens. Environ.* 114, 876–886. <https://doi.org/10.1016/j.rse.2009.12.002>
- Laba, M., Downs, R., Smith, S., Welsh, S., Neider, C., White, S., Richmond, M., Philpot, W., Baveye, P., 2008. Mapping invasive wetland plants in the Hudson River National Estuarine Research Reserve using quickbird satellite imagery. *Remote Sens. Environ.* 112, 286–300. <https://doi.org/10.1016/j.rse.2007.05.003>
- Lai, Y.R., Pringle, M.J., Kopittke, P.M., Menzies, N.W., Orton, T.G., Dang, Y.P., 2018. An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. *Int. J. Appl. Earth Obs. Geoinf.* 72, 99–108. <https://doi.org/10.1016/j.jag.2018.07.013>
- Lauver, C.L., 1997. Mapping species diversity patterns in the Kansas shortgrass region by integrating remote sensing and vegetation analysis. *J. Veg. Sci.* 8, 387–394. <https://doi.org/10.2307/3237328>
- Lavery, P., Pattiaratchi, C., Wyllie, A., Hick, P., 1993. Water Quality Monitoring in Estuarine Waters Using the Landsat Thematic Mapper. *Remote Sens. Environ.* 46, 268–280. [https://doi.org/10.1016/0034-4257\(93\)90047-2](https://doi.org/10.1016/0034-4257(93)90047-2)
- Le Féon, V., Schermann-Legionnet, A., Delettre, Y., Aviron, S., Billeter, R., Bugter, R., Hendrickx, F., Burel, F., 2010. Intensification of agriculture, landscape composition and wild bee communities: A large scale study in four European countries. *Agric. Ecosyst. Environ.* 137, 143–150. <https://doi.org/10.1016/j.agee.2010.01.015>
- Leblon, B., Guerif, M., Baret, F., 1991. The use of remotely sensed data in estimation of PAR use efficiency and biomass production of flooded rice. *Remote Sens. Environ.* 38, 147–158. [https://doi.org/10.1016/0034-4257\(91\)90076-I](https://doi.org/10.1016/0034-4257(91)90076-I)
- Li, L., Friedl, M.A., Xin, Q., Gray, J., Pan, Y., Frohling, S., 2014. Mapping crop cycles in China using MODIS-EVI time series. *Remote Sens.* 6, 2473–2493. <https://doi.org/10.3390/rs6032473>
- Li, Z., Guo, X., 2012. Detecting climate effects on vegetation in northern mixed rairie using NOAA AVHRR 1-km time-series NDVI data. *Remote Sens.* 4, 120–134. <https://doi.org/10.3390/rs4010120>
- Li, Z., Xu, D., Guo, X., 2014. Remote sensing of ecosystem health: opportunities, challenges, and future perspectives. *Sensors* 14, 21117–21139. <https://doi.org/10.3390/s141121117>



- Liu, H., Li, Q., Shi, T., Hu, S., Wu, G., Zhou, Q., 2017. Application of Sentinel 2 MSI Images to Retrieve Suspended Particulate Matter Concentrations in Poyang Lake. *Remote Sens.* 9, 761. <https://doi.org/10.3390/rs9070761>
- Liu, J., Pattey, E., Miller, J.R., McNairn, H., Smith, A., Hu, B., 2010. Estimating crop stresses, aboveground dry biomass and yield of corn using multi-temporal optical data combined with a radiation use efficiency model. *Remote Sens. Environ.* 114, 1167–1177. <https://doi.org/10.1016/j.rse.2010.01.004>
- Lobell, D.B., Baldos, U.L.C., Hertel, T.W., 2013. Climate adaptation as mitigation: the case of agricultural investments. *Environ. Res. Lett.* 8, 1–12. <https://doi.org/10.1088/1748-9326/8/1/015012>
- Luoto, M., Kuussaari, M., Toivonen, T., 2002a. Modelling butterfly distribution based on remote sensing data. *J. Biogeogr.* 29, 1027–1037. <https://doi.org/10.1046/j.1365-2699.2002.00728.x>
- Luoto, M., Toivonen, T., Heikkinen, R.K., 2002b. Prediction of total and rare plant species richness in agricultural landscapes from satellite images and topographic data. *Landsc. Ecol.* 17, 195–217. <https://doi.org/10.1023/A:1020288509837>
- Mairota, P., Cafarelli, B., Didham, R.K., Lovergine, F.P., Lucas, R.M., Nagendra, H., Rocchini, D., Tarantino, C., 2015. Challenges and opportunities in harnessing satellite remote-sensing for biodiversity monitoring. *Ecol. Inform.* 30, 207–214. <https://doi.org/10.1016/j.ecoinf.2015.08.006>
- Martinez, J.M., Guyot, J.L., Cochonneau, G., 2008. Monitoring of surface water quality in large rivers with satellite imagery - Application to the Amazon basin. [www.iwra.org/congress/resource/abs649\\_article.pdf](http://www.iwra.org/congress/resource/abs649_article.pdf).
- Mashimbye, Z.E., Cho, M.A., Nell, J.P., De clerck, W.P., Van niekerk, A., Turner, D.P., 2012. Model-Based Integrated Methods for Quantitative Estimation of Soil Salinity from Hyperspectral Remote Sensing Data: A Case Study of Selected South African Soils. *Pedosphere* 22, 640–649. [https://doi.org/10.1016/S1002-0160\(12\)60049-6](https://doi.org/10.1016/S1002-0160(12)60049-6)
- McGarigal, K., Cushman, S.A., Ene, E., 2012. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical Maps [WWW Document]. URL <http://www.umass.edu/landeco/research/fragstats/fragstats.html> (accessed 4.21.17).
- Metternicht, G.I., Zinck, J.A., 2003. Remote sensing of soil salinity: potentials and constraints. *Remote Sens. Environ.* 85, 1–20. [https://doi.org/10.1016/S0034-4257\(02\)00188-8](https://doi.org/10.1016/S0034-4257(02)00188-8)
- Metternicht, G.I., Zinck, J.A., 1998. Evaluating the information content of JERS-1 SAR and Landsat TM data for discrimination of soil erosion features. *ISPRS J. Photogramm. Remote Sens.* 53, 143–153. [https://doi.org/10.1016/S0924-2716\(98\)00004-5](https://doi.org/10.1016/S0924-2716(98)00004-5)
- Meynard, C.N., Quinn, J.F., 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. *J. Biogeogr.* 34, 1455–1469. <https://doi.org/10.1111/j.1365-2699.2007.01720.x>
- Michailovsky, C.I., McEnnis, S., Berry, P.A.M., Smith, R., Bauer-Gottwein, P., 2012. River monitoring from satellite radar altimetry in the Zambezi River basin. *Hydrol. Earth Syst. Sci.* 16, 2181–2192. <https://doi.org/10.5194/hess-16-2181-2012>
- Mingwei, Z., Qingbo, Z., Zhongxin, C., Jia, L., Yong, Z., Chongfa, C., 2008. Crop discrimination in Northern China with double cropping systems using Fourier analysis of time-series MODIS data. *Int. J. Appl. Earth Obs. Geoinf.* 10, 476–485. <https://doi.org/10.1016/j.jag.2007.11.002>
- Mirik, M., Ansley, R.J., 2012. Utility of Satellite and Aerial Images for Quantification of Canopy Cover and Infilling Rates of the Invasive Woody Species Honey Mesquite (*Prosopis Glandulosa*) on Rangeland. *Remote Sens.* 4, 1947–1962. <https://doi.org/10.3390/rs4071947>

- Monteith, J.L., 1972. Solar Radiation and Productivity in Tropical Ecosystems. *J. Appl. Ecol.* 9, 747–766. <https://doi.org/10.2307/2401901>
- Morel, J., Todoroff, P., Bégué, A., Bury, A., Martiné, J.F., Petit, M., 2014. Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: A case study on reunion island. *Remote Sens.* 6, 6620–6635. <https://doi.org/10.3390/rs6076620>
- Moriondo, M., Maselli, F., Bindi, M., 2007. A simple model of regional wheat yield based on NDVI data. *Eur. J. Agron.* 26, 266–274. <https://doi.org/10.1016/j.eja.2006.10.007>
- Mueller, N., Lewis, A., Roberts, D., Ring, S., Melrose, R., Sixsmith, J., Lymburner, L., McIntyre, A., Tan, P., Curnow, S., Ip, A., 2016. Water observations from space: Mapping surface water from 25years of Landsat imagery across Australia. *Remote Sens. Environ.* 174, 341–352. <https://doi.org/10.1016/j.rse.2015.11.003>
- Muñoz, J., Felicísimo, Á.M., 2004. Comparison of statistical methods commonly used in predictive modelling. *J. Veg. Sci.* 15, 285–292. <https://doi.org/10.1111/j.1654-1103.2004.tb02263.x>
- Mutiga, J.K., Su, Z., Woldai, T., 2010. Using satellite remote sensing to assess evapotranspiration: Case study of the upper Ewaso Ng'iro North Basin, Kenya. *Int. J. Appl. Earth Obs. Geoinf.* 12S, S100–S108. <https://doi.org/10.1016/j.jag.2009.09.012>
- Myneni, R.B., Nemani, R.R., Running, S.W., 1997. Estimation of global leaf area index and absorbed PAR using radiative transfer models. *IEEE Trans. Geosci. Remote Sens.* 35, 1380–1393. <https://doi.org/10.1109/36.649788>
- Nagendra, H., 2001. Using remote sensing to assess biodiversity. *Int. J. Remote Sens.* 22, 2377–2400. <https://doi.org/10.1080/01431160117096>
- Nagendra, H., Gadgil, M., 1999. Satellite Imagery as a Tool for Monitoring Species Diversity: An Assessment. *J. Appl. Ecol.* 36, 388–397. <https://doi.org/10.1046/j.1365-2664.1999.00406.x>
- Nagy, A., Fehér, J., Tamás, J., 2018. Wheat and maize yield forecasting for the Tisza river catchment using MODIS NDVI time series and reported crop statistics. *Comput. Electron. Agric.* 151, 41–49. <https://doi.org/10.1016/j.compag.2018.05.035>
- Ngoc, D.D., Loisel, H., Jamet, C., Vantrepotte, V., Duforêt-Gaurier, L., Minh, C.D., Mangin, A., 2019. Coastal and inland water pixels extraction algorithm (WiPE) from spectral shape analysis and HSV transformation applied to Landsat 8 OLI and Sentinel-2 MSI. *Remote Sens. Environ.* 223, 208–228. <https://doi.org/10.1016/j.rse.2019.01.024>
- Niedertscheider, M., Kastner, T., Fetzl, T., Haberl, H., Kroisleitner, C., Plutzer, C., Erb, K.-H., 2016. Mapping and analysing cropland use intensity from a NPP perspective. *Environ. Res. Lett.* 11. <https://doi.org/10.1088/1748-9326/11/1/014008>
- Numata, I., Roberts, D.A., Chadwick, O.A., Schimel, J., Sampaio, F.R., Leonidas, F.C., Soares, J. V., 2007. Characterization of pasture biophysical properties and the impact of grazing intensity using remotely sensed data. *Remote Sens. Environ.* 109, 314–327. <https://doi.org/10.1016/j.rse.2007.01.013>
- Olofsson, P., Lagergren, F., Lindroth, A., Lindström, J., Klemedtsson, L., Kutsch, W., Eklundh, L., 2008. Towards operational remote sensing of forest carbon balance across Northern Europe. *Biogeosciences* 5, 817–832. <https://doi.org/10.5194/bg-5-817-2008>
- Pacheco, A., McNairn, H., 2010. Evaluating multispectral remote sensing and spectral unmixing analysis for crop residue mapping. *Remote Sens. Environ.* 114, 2219–2228. <https://doi.org/10.1016/j.rse.2010.04.024>
- Palacios-Orueta, A., Ustin, S.L., 1998. Remote sensing of soil properties in the Santa Monica Mountains I. Spectral analysis. *Remote Sens. Environ.* 65, 170–183. [https://doi.org/10.1016/S0034-4257\(98\)00024-8](https://doi.org/10.1016/S0034-4257(98)00024-8)

- Palmer, M.W., Earls, P.G., Hoagland, B.W., White, P.S., Wohlgemuth, T., 2002. Quantitative tools for perfecting species lists. *Environmetrics* 13, 121–137. <https://doi.org/10.1002/env.516>
- Pan, G., Sun, G.J., Li, F.M., 2009. Using QuickBird imagery and a production efficiency model to improve crop yield estimation in the semi-arid hilly Loess Plateau, China. *Environ. Model. Softw.* 24, 510–516. <https://doi.org/10.1016/j.envsoft.2008.09.014>
- Patel, N.R., Bhattacharjee, B., Mohammed, A.J., Tanupriya, B., Saha, S.K., 2006. Remote sensing of regional yield assessment of wheat in Haryana, India. *Int. J. Remote Sens.* 27, 4071–4090. <https://doi.org/10.1080/01431160500377188>
- Pereira-Sandoval, M., Ruiz-Verdu, A., Tenjo, C., Delegido, J., Urrego, P., Pena, R., Vicente, E., Soria, Juan, Soria, Javier, Moreno, J., 2018. Calibration and Validation of Algorithms for the Estimation of Chlorophyll-A Concentration and Secchi Depth in Inland Waters with Sentinel-2. *Limnetica* 38, 471–487. <https://doi.org/10.1109/igarss.2018.8517371>
- Petropoulos, G.P., Srivastava, P.K., Piles, M., Pearson, S., 2018. Earth observation-based operational estimation of soil moisture and evapotranspiration for agricultural crops in support of sustainable water management. *Sustainability* 10, 181. <https://doi.org/10.3390/su10010181>
- Ponnuram, G.G., Rao, Y.S., 2018. The application of compact polarimetric decomposition algorithms to L-band PolSAR data in agricultural areas. *Int. J. Remote Sens.* 39, 8337–8360. <https://doi.org/10.1080/01431161.2018.1488281>
- Potter, P., Ramankutty, N., Bennett, E.M., Donner, S.D., 2010. Characterizing the spatial patterns of global fertilizer application and manure production. *Earth Interact.* 14, 1–22. <https://doi.org/10.1175/2009EI288.1>
- Prasad, A.K., Chai, L., Singh, R.P., Kafatos, M., 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. *Int. J. Appl. Earth Obs. Geoinf.* 8, 26–33. <https://doi.org/10.1016/j.jag.2005.06.002>
- Qian, T., Tsunekawa, A., Peng, F., Masunaga, T., Wang, T., Li, R., 2019. Derivation of salt content in salinized soil from hyperspectral reflectance data: A case study at Minqin Oasis, Northwest China. *J. Arid Land* 11, 111–122. <https://doi.org/10.1007/s40333-019-0091-9>
- Rahman, M.M., Moran, M.S., Thoma, D.P., Bryant, R., Holifield Collins, C.D., Jackson, T., Orr, B.J., Tischler, M., 2008. Mapping surface roughness and soil moisture using multi-angle radar imagery without ancillary data. *Remote Sens. Environ.* 112, 391–402. <https://doi.org/10.1016/j.rse.2006.10.026>
- Ravan, S.A., Roy, P.S., Sharma, C.M., 1995. Space remote sensing for spatial vegetation characterization. *J. Biosci* 20, 427–438.
- Reidsma, P., Tekelenburg, T., Van Den Berg, M., Alkemade, R., 2006. Impacts of land-use change on biodiversity: An assessment of agricultural biodiversity in the European Union. *Agric. Ecosyst. Environ.* 114, 86–102. <https://doi.org/10.1016/j.agee.2005.11.026>
- Ren, J., Chen, Z., Zhou, Q., Tang, H., 2008. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *Int. J. Appl. Earth Obs. Geoinf.* 10, 403–413. <https://doi.org/10.1016/j.jag.2007.11.003>
- Ritchie, J.C., Cooper, C.M., 2001. Remote sensing techniques for determining water quality: Applications to TMDLs, in: *TMDL Science Issues Conference (2001)*, Water Environment Federation, Alexandria, VA. pp. 367–374.
- Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S., Kark, S., Levin, N., Lucas, K., Luoto, M., Nagendra, H., Oldeland, J., Ricotta, C., Southworth, J., Neteler, M., 2010. Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecol. Inform.* 5, 318–329. <https://doi.org/10.1016/j.ecoinf.2010.06.001>

- Rocchini, D., Ricotta, C., Chiarucci, A., 2007. Using satellite imagery to assess plant species richness: The role of multispectral systems. *Appl. Veg. Sci.* 10, 325–331. <https://doi.org/10.1111/j.1654-109X.2007.tb00431.x>
- Ruimy, A., Saugier, B., 1994. Methodology for the estimation of terrestrial net primary production from remotely sensed data. *J. Geophys. Res.* 99, 5263–5283. <https://doi.org/10.1029/93JD03221>
- Sakowska, K., Juszczak, R., Gianelle, D., 2016. Remote Sensing of Grassland Biophysical Parameters in the Context of the Sentinel-2 Satellite Mission. *J. Sensors* 2016, 1–16. <https://doi.org/10.1155/2016/4612809>
- Salazar, L., Kogan, F., Roytman, L., 2007. Use of remote sensing data for estimation of winter wheat yield in the United States. *Int. J. Remote Sens.* 28, 3795–3811. <https://doi.org/10.1080/01431160601050395>
- Saura, S., 2004. Effects of remote sensor spatial resolution and data aggregation on selected fragmentation indices. *Landsc. Ecol.* 19, 197–209. <https://doi.org/10.1023/B:LAND.0000021724.60785.65>
- Saveraid, E.H., Debinski, D.M., Kindscher, K., Jakubauskas, M.E., 2001. A comparison of satellite data and landscape variables in predicting bird species occurrences in the Greater Yellowstone Ecosystem, USA. *Landsc. Ecol.* 16, 71–83. <https://doi.org/10.1023/A:1008119219788>
- Scharf, P., Schmidt, J.P., Kitchen, N.R., Sudduth, K.A., 2002. Remote sensing for nitrogen management. *J. Soil Water Conserv.* 57, 518–524.
- Schino, G., Borfecchia, F., De Cecco, L., Dibari, C., Iannetta, M., Martini, S., Pedrotti, F., 2003. Satellite estimate of grass biomass in a mountainous range in central Italy. *Agrofor. Syst.* 59, 157–162. <https://doi.org/10.1023/A:1026308928874>
- Schumaker, N.H., 1996. Using landscape indices to predict habitat connectivity. *Ecology* 77, 1210–1225. <https://doi.org/10.2307/2265590>
- Sharma, K.D., Singh, S., 1995. Satellite remote sensing for soil erosion modelling using the ANSWERS model. *Hydrol. Sci.* 40, 259–272. <https://doi.org/10.1080/02626669509491408>
- Shoshany, M., Goldshleger, N., Chudnovsky, A., 2013. Monitoring of agricultural soil degradation by remote-sensing methods: a review. *Int. J. Remote Sens.* 34, 6152–6181. <https://doi.org/10.1080/01431161.2013.793872>
- Shuai, G., Zhang, J., Basso, B., Pan, Y., Zhu, X., Zhu, S., Liu, H., 2019. Multi-temporal RADARSAT-2 polarimetric SAR for maize mapping supported by segmentations from high-resolution optical image. *Int. J. Appl. Earth Obs. Geoinf.* 74, 1–15. <https://doi.org/10.1016/j.jag.2018.08.021>
- Sibanda, M., Mutanga, O., Rouget, M., 2015. Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS J. Photogramm. Remote Sens.* 110, 55–65. <https://doi.org/10.1016/j.isprsjprs.2015.10.005>
- Siebert, S., Portmann, F.T., Döll, P., 2010. Global patterns of cropland use intensity. *Remote Sens.* 2, 1625–1643. <https://doi.org/10.3390/rs2071625>
- Singh, R., Semwal, D.P., Rai, A., Chhikara, R.S., 2002. Small area estimation of crop yield using remote sensing satellite data. *Int. J. Remote Sens.* 23, 49–56. <https://doi.org/10.1080/01431160010014756>
- Smith, L.C., 1997. Satellite remote sensing of river inundation area, stage, and discharge: a review. *Hydrol. Process.* 11, 1427–1439. [https://doi.org/10.1002/\(SICI\)1099-1085\(199708\)11:10<1427::AID-HYP473>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1099-1085(199708)11:10<1427::AID-HYP473>3.0.CO;2-S)
- Son, N.T., Chen, C.F., Chen, C.R., Minh, V.Q., Trung, N.H., 2014. A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation. *Agric. For. Meteorol.* 197, 52–64. <https://doi.org/10.1016/j.agrformet.2014.06.007>

- Sòria-perpinyà, X., Urrego, P., Pereira-sandoval, M., Ruiz-verdú, A., Peña, R., Soria, J.M., Delegido, J., Vicente, E., Moreno, J., València, U. De, José, C.C., Martínez, B., Martínez, B., 2019. Monitoring the ecological state of a hypertrophic lake ( Albufera of València , Spain ) using multitemporal Sentinel-2 images 38, 457–469. <https://doi.org/10.23818/limn.38.26>
- Srivastava, P.K., 2017. Satellite Soil Moisture: Review of Theory and Applications in Water Resources. *Water Resour. Manag.* 31, 3161–3176. <https://doi.org/10.1007/s11269-017-1722-6>
- Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., van Wesemael, B., 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma* 158, 32–45. <https://doi.org/10.1016/j.geoderma.2009.11.032>
- Stickler, C.M., Southworth, J., 2008. Application of multi-scale spatial and spectral analysis for predicting primate occurrence and habitat associations in Kibale National Park, Uganda. *Remote Sens. Environ.* 112, 2170–2186. <https://doi.org/10.1016/j.rse.2007.10.013>
- Sulistioadi, Y.B., Tseng, K.H., Shum, C.K., Hidayat, H., Sumaryono, M., Suhardiman, A., Setiawan, F., Sunarso, S., 2015. Satellite radar altimetry for monitoring small rivers and lakes in Indonesia. *Hydrol. Earth Syst. Sci.* 19, 341–359. <https://doi.org/10.5194/hess-19-341-2015>
- Taghadosi, M.M., Hasanlou, M., Eftekhari, K., 2019. Soil salinity mapping using dual-polarized SAR Sentinel-1 imagery. *Int. J. Remote Sens.* 40, 237–252. <https://doi.org/10.1080/01431161.2018.1512767>
- Taylor, G.R., Mah, A.H., Kruse, F.A., Kierein-Young, K.S., Hewson, R.D., Bennett, B.A., 1996. Characterization of Saline Soils Using Airborne Radar Imagery. *Remote Sens. Environ.* 57, 127–142. [https://doi.org/10.1016/0034-4257\(95\)00239-1](https://doi.org/10.1016/0034-4257(95)00239-1)
- Teillard, F., Allaire, G., Cahuzac, E., Léger, F., Maigné, E., Tichit, M., 2012. A novel method for mapping agricultural intensity reveals its spatial aggregation: Implications for conservation policies. *Agric. Ecosyst. Environ.* 149, 135–143. <https://doi.org/10.1016/j.agee.2011.12.018>
- Temme, A.J.A.M., Verburg, P.H., 2011. Mapping and modelling of changes in agricultural intensity in Europe. *Agric. Ecosyst. Environ.* 140, 46–56. <https://doi.org/10.1016/j.agee.2010.11.010>
- Thenkabail, P.S., Biradar, C.M., Noojipady, P., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., Gangalakunta, O.R.P., Turrall, H., Cai, X., Vithanage, J., Schull, M. a., Dutta, R., 2009. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. *Int. J. Remote Sens.* 30, 3679–3733. <https://doi.org/10.1080/01431160802698919>
- Tilling, A.K., O’Leary, G.J., Ferwerda, J.G., Jones, S.D., Fitzgerald, G.J., Rodriguez, D., Belford, R., 2007. Remote sensing of nitrogen and water stress in wheat. *F. Crop. Res.* 104, 77–85. <https://doi.org/10.1016/j.fcr.2007.03.023>
- Trivero, P., Borasi, M., Biamino, W., Cavagnero, M., Rinaudo, C., Bonansea, M., Lanfri, S., 2013. River pollution remediation monitored by optical and infrared high-resolution satellite images. *Environ. Monit. Assess.* 185, 7647–7658. <https://doi.org/10.1007/s10661-013-3125-3>
- Trombetti, M., Riaño, D., Rubio, M.A., Cheng, Y.B., Ustin, S.L., 2008. Multi-temporal vegetation canopy water content retrieval and interpretation using artificial neural networks for the continental USA. *Remote Sens. Environ.* 112, 203–215. <https://doi.org/10.1016/j.rse.2007.04.013>
- Tsai, F., Chou, M.J., 2006. Texture augmented analysis of high resolution satellite imagery in detecting invasive plant species. *J. Chinese Inst. Eng.* 29, 581–592. <https://doi.org/10.1080/02533839.2006.9671155>
- Tsai, F., Lin, E.K., Yoshino, K., 2007. Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species. *Int. J. Remote Sens.* 28, 1023–1039. <https://doi.org/10.1080/01431160600887706>

- Tulbure, M.G., Broich, M., 2013. Spatiotemporal dynamic of surface water bodies using Landsat time-series data from 1999 to 2011. *ISPRS J. Photogramm. Remote Sens.* 79, 44–52. <https://doi.org/10.1016/j.isprsjprs.2013.01.010>
- Turner, D.P., Cohen, W.B., Kennedy, R.E., Fassnacht, K.S., Briggs, J.M., 1999. Relationships between Leaf Area Index and Landsat TM Spectral Vegetation Indices across Three Temperate Zone Sites. *Remote Sens. Environ.* 70, 52–68. [https://doi.org/10.1016/S0034-4257\(99\)00057-7](https://doi.org/10.1016/S0034-4257(99)00057-7)
- Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* 18, 306–314. [https://doi.org/10.1016/S0169-5347\(03\)00070-3](https://doi.org/10.1016/S0169-5347(03)00070-3)
- van Leeuwen, M., Nieuwenhuis, M., 2010. Retrieval of forest structural parameters using LiDAR remote sensing. *Eur. J. For. Res.* 129, 749–770. <https://doi.org/10.1007/s10342-010-0381-4>
- Vannier, C., Vasseur, C., Hubert-Moy, L., Baudry, J., 2011. Multiscale ecological assessment of remote sensing images. *Landsc. Ecol.* 26, 1053–1069. <https://doi.org/10.1007/s10980-011-9626-y>
- Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sens. Environ.* 16, 125–141. [https://doi.org/10.1016/0034-4257\(84\)90057-9](https://doi.org/10.1016/0034-4257(84)90057-9)
- Verpoorter, C., Kutser, T., Tranvik, L., 2012. Automated mapping of water bodies using Landsat multispectral data. *Limnol. Oceanogr. Methods* 10, 1037–1050. <https://doi.org/10.4319/lom.2012.10.1037>
- Verstraeten, W.W., Veroustraete, F., Van Der Sande, C.J., Grootaers, I., Feyen, J., 2006. Soil moisture retrieval using thermal inertia, determined with visible and thermal spaceborne data, validated for European forests. *Remote Sens. Environ.* 101, 299–314. <https://doi.org/10.1016/j.rse.2005.12.016>
- Vibhute, A.D., Gawali, B.W., 2013. Analysis and Modeling of Agricultural Land use using Remote Sensing and Geographic Information System: a Review. *Int. J. Eng. Res. Appl.* 3, 81–91.
- Vignolo, A., Pochettino, A., Cicerone, D., 2006. Water quality assessment using remote sensing techniques: Medrano Creek, Argentina. *J. Environ. Manage.* 81, 429–433. <https://doi.org/10.1016/j.jenvman.2005.11.019>
- Vrieling, A., 2006. Satellite remote sensing for water erosion assessment: A review. *Catena* 65, 2–18. <https://doi.org/10.1016/j.catena.2005.10.005>
- Wagner, W., Naeimi, V., Scipal, K., Jeu, R., Martínez-Fernández, J., 2007. Soil moisture from operational meteorological satellites. *Hydrogeol. J.* 15, 121–131. <https://doi.org/10.1007/s10040-006-0104-6>
- Walsh, S.J., McCleary, A.L., Mena, C.F., Shao, Y., Tuttle, J.P., González, A., Atkinson, R., 2008. QuickBird and Hyperion data analysis of an invasive plant species in the Galapagos Islands of Ecuador: Implications for control and land use management. *Remote Sens. Environ.* 112, 1927–1941. <https://doi.org/10.1016/j.rse.2007.06.028>
- Wang, B., Jia, K., Liang, S., Xie, X., Wei, X., Zhao, X., Yao, Y., Zhang, X., 2018. Assessment of Sentinel-2 MSI Spectral Band Reflectances for Estimating Fractional Vegetation Cover. *Remote Sens.* 10, 1927. <https://doi.org/10.3390/rs10121927>
- Wang, K., Franklin, S.E., Guo, X., Cattet, M., 2010. Remote sensing of ecology, biodiversity and conservation: A review from the perspective of remote sensing specialists. *Sensors* 10, 9647–9667. <https://doi.org/10.3390/s101109647>
- Wang, L., Huang, J., Du, Y., Hu, Y., Han, P., 2013. Dynamic assessment of soil erosion risk using landsat TM and HJ satellite data in danjiangkou reservoir area, China. *Remote Sens.* 5, 3826–3848. <https://doi.org/10.3390/rs5083826>
- Wang, L., Qu, J.J., 2009. Satellite remote sensing applications for surface soil moisture monitoring: A review. *Front. Earth Sci. China* 3, 237–247. <https://doi.org/10.1007/s11707-009-0023-7>

- Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. *Remote Sens. Environ.* 112, 1096–1116. <https://doi.org/10.1016/j.rse.2007.07.019>
- Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sens. Environ.* 108, 290–310. <https://doi.org/10.1016/j.rse.2006.11.021>
- Weiers, S., Bock, M., Wissen, M., Rossner, G., 2004. Mapping and indicator approaches for the assessment of habitats at different scales using remote sensing and GIS methods. *Landsc. Urban Plan.* 67, 43–65. [https://doi.org/10.1016/S0169-2046\(03\)00028-8](https://doi.org/10.1016/S0169-2046(03)00028-8)
- Wen, X., Yang, X., 2010. Monitoring of water quality using remote sensing techniques. *Appl. Mech. Mater.* 29–32, 2360–2364. <https://doi.org/10.4028/www.scientific.net/AMM.29-32.2360>
- Weng, Y., Gong, P., Zhu, Z., 2008. Soil salt content estimation in the Yellow River delta with satellite hyperspectral data. *Can. J. Remote Sens.* 34, 259–270. <https://doi.org/10.5589/m08-017>
- Wolter, P.T., Johnston, C.A., Niemi, G.J., 2005. Mapping submergent aquatic vegetation in the US Great Lakes using Quickbird satellite data. *Int. J. Remote Sens.* 26, 5255–5274. <https://doi.org/10.1080/01431160500219208>
- Wu, B., Gommers, R., Zhang, M., Zeng, H., Yan, N., Zou, W., Zheng, Y., Zhang, N., Chang, S., Xing, Q., van Heijden, A., 2015. Global crop monitoring: A satellite-based hierarchical approach. *Remote Sens.* 7, 3907–3933. <https://doi.org/10.3390/rs70403907>
- Wu, B., Meng, J., Li, Q., Yan, N., Du, X., Zhang, M., 2014. Remote sensing-based global crop monitoring: Experiences with China's CropWatch system. *Int. J. Digit. Earth* 7, 113–137. <https://doi.org/10.1080/17538947.2013.821185>
- Wu, C., Wu, J., Qi, J., Zhang, L., Huang, H., Lou, L., Chen, Y., 2010. Empirical estimation of total phosphorus concentration in the mainstream of the Qiantang River in China using Landsat TM data. *Int. J. Remote Sens.* 31, 2309–2324. <https://doi.org/10.1080/01431160902973873>
- Wulder, M.A., Hall, R.J., Coops, Nicholas, C., Franklin, S.E., 2004. High spatial resolution remotely sensed data for ecosystem characterisation. *Bioscience* 54, 511–521.
- Xiao, X., Jian, X., Xiongfei, W., Chengfang, H., Xuejun, C., Zhaohui, W., Dengzhong, Z., 2015. Evaluation method of water quality for river based on multi-spectral remote sensing data, in: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. <https://doi.org/10.5194/isprsarchives-XL-7-W3-1517-2015>
- Xie, H., Zou, J., Jiang, H., Zhang, N., Choi, Y., 2014. Spatiotemporal pattern and driving forces of arable land-use intensity in China: Toward sustainable land management using emergy analysis. *Sustainability* 6, 3504–3520. <https://doi.org/10.3390/su6063504>
- Xu, C., Li, Y., Hu, J., Yang, X., Sheng, S., Liu, M., 2012. Evaluating the difference between the normalized difference vegetation index and net primary productivity as the indicators of vegetation vigor assessment at landscape scale. *Environ. Monit. Assess.* 184, 1275–1286. <https://doi.org/10.1007/s10661-011-2039-1>
- Xu, L., Zhang, H., Wang, C., Zhang, B., Liu, M., 2019. Crop classification based on temporal information using Sentinel-1 SAR time-series data. *Remote Sens.* 11, 53. <https://doi.org/10.3390/rs11010053>
- Yan, H., Xiao, X., Huang, H., Liu, J., Chen, J., Bai, X., 2014. Multiple Cropping Intensity in China Derived from Agro-meteorological Observations and MODIS Data. *Chinese Geogr. Sci.* 24, 205–219. <https://doi.org/10.1007/s11769-013-0637-2>

- Yang, C., Everitt, J.H., Bradford, J.M., 2006. Comparison of QuickBird satellite imagery and airborne imagery for mapping grain sorghum yield patterns. *Precis. Agric.* 7, 33–44. <https://doi.org/10.1007/s11119-005-6788-0>
- Yang, X., 2005. Remote sensing and GIS applications for estuarine ecosystem analysis: An overview. *Int. J. Remote Sens.* 26, 5347–5356. <https://doi.org/10.1080/01431160500219406>
- Zhang, M., Wu, B., Yu, M., Zou, W., Zheng, Y., 2014. Crop condition assessment with adjusted NDVI using the uncropped arable land ratio. *Remote Sens.* 6, 5774–5794. <https://doi.org/10.3390/rs6065774>
- Zhang, W., Li, H., 2016. Characterizing and Assessing the Agricultural Land Use Intensity of the Beijing Mountainous Region. *Sustainability* 8, 1–18. <https://doi.org/10.3390/su8111180>
- Zhang, Y., Atkinson, P.M., Li, X., Ling, F., Wang, Q., Du, Y., 2017. Learning-Based Spatial-Temporal Superresolution Mapping of Forest Cover with MODIS Images. *IEEE Trans. Geosci. Remote Sens.* 55, 600–614. <https://doi.org/10.1109/TGRS.2016.2613140>



## **Appendix 2: A basic introduction to Random Forest and its implementation**

### **Random Forest Regression**

Random Forest is a supervised learning algorithm that uses an ensemble learning method for both classification and regression analysis (Boehmke and Greenwell, 2020; Breiman, 2001; Liaw and Wiener, 2002; A. M. Prasad et al., 2006). Random Forest builds a large number of individual decision trees that are merged together to provide a more accurate and stable prediction.

As a supervised machine learning model, Random Forest learns to map input variables (input features) to outputs (target feature) during the training phase of model building. During this phase, the model learns any relationships between the input features and target feature. Bootstrap aggregation, or bagging, is used to add an element of randomness during training to prevent overfitting. Bagging allows each tree to draw a random sample from the original training data set using replacement, resulting in different trees.

Random Forest adds additional randomness into the tree-growing process. While growing a decision tree, Random Forests perform *split-variable randomisation*, so each time a split is to be performed, the search for the split variable is limited to a random subset of the original target features. Of this subset, the feature chosen to split the data is determined using the *best split* approach; the feature chosen at each splitting node is the one that minimises the Gini impurity (in the case of classification) and the Sum of Squared Error (in the case of regression). This ensures that the ensemble model, once all trees are combined, makes fair use of all potentially predictive features, and does not rely too heavily on any individual feature.

The individual trees within Random Forest are therefore not only trained using different sets of data, but also using different input features to make decisions. The result is a series of uncorrelated trees that, when combined, produce a more accurate prediction than that of any individual tree.

Prediction is performed using the trained Random Forest algorithm by passing the same input features used for training through the rules of each of the randomly created decision trees. Once each tree is fully grown, the results from the individual trees are aggregated through averaging for regression and using majority vote for classification. Combining these multiple decision trees creates a single ensemble model which outperforms any of the individual decision trees.

In this project, the Random Forest algorithm was implemented using the *randomForest* package available within *R*. The *randomForestSRC* package was used alongside the *forestFloor* package in subsequent analysis to unpick the relationships between individual input features and the target features (see chapter 4). An example of the script used to estimate crop yield (chapters 3 and 5) using Random Forest is given below, with the code in black/blue and comments in green. The same basic process was used to build the Random Forest models to estimate bird diversity in chapters 4 and 5.

## Example RStudio Script to run Random Forest Regression Analysis

### Step 1: Set input and output file names and locations

```
# Name and path of the observed wheat yield data
outFile <- '...'          # CSV file containing the wheat yield for each
                           # point in the training dataset

# Name and path of the input image that will be used for predictions
inPredImage <- '...'      # TIFF file comprising the stacked raster bands
                           # of the input feature variables (e.g. Sentinel-2
                           # bands, precipitation data)

# Name and path of the output GeoTiff predicted image
outImage <- '...'
```

### Step 2: Load training data and extract corresponding input feature variables for each point

```
# Read training data
pointTable <- read.csv(outFile, header=TRUE)
# Identify which columns contain XY coordinates
xy <- SpatialPoints(pointTable[,1:2])      # The xy coordinates are
                                           # needed to extract the input
                                           # feature variables for each
                                           # yield data point

# Identify the column containing the response variable (yield data)
response <- as.numeric(pointTable[,3])

# Load the image containing the input feature variables
satImage <- stack(inPredImage)

# Set no data value
for (b in 1:nlayers(satImage)) {NAvalue(satImage@layers[[b]]) <- nd}

# Extract values for input feature variable values for the corresponding
# yield data points
trainvals <- cbind(response, extract(satImage, xy))      # The resulting
                                                         # database
                                                         # contains the
                                                         # data used to
                                                         # train Random
                                                         # Forest

# Remove NA values from the training data
trainvals_no_na <- na.omit(trainvals)
```

### **Step 3: Build the Random Forest model**

```
# Run Random Forest
randfor <- randomForest(response ~ . , data = trainvals_no_na,
importance=TRUE)
# Using "response ~." uses all available input feature variables
to build the Random Forest model. To include only specific
variables an equation must be define e.g. "response ~ variable_1 +
variable_2 +...+ variable_n"
```

### **Step 4: Predict yield**

```
# Use the randomForest model to estimate yield across the full study
area
predict(satImage, randfor, filename = outImage, progress = 'text',
format = 'GTiff', datatype = 'FLT4S', type = 'response', overwrite=TRUE)
# "satImage" is the stacked raster containing the input feature
variables; "randfor" is the Random Forest model built in the
previous step; and "outImage" identifies the name and location of
the prediction image
```

### **Step 5: Accuracy Assessment**

#### **Internal validation**

```
# Print models OOB error
print(paste("Variance explained by the RF model =", randfor$rsq[500],
sep= ""))
```

#### **External Validation**

```
# Create empty data frame for rmse and rsquared values
rmse <- ()
rsquared <- ()
# Perform 10-fold cross-validation
for (i in 1:10){
  # Set size of sample for validation
  val_samp <- sample(c(1:nrow(trainvals_no_na)),
round(nrow(trainvals_no_na)/10))
  # Build Random Forest model (excluding the data for validation)
  randfor <- randomForest(response ~. , data=trainvals_no_na[-
val_samp,])
  # Use randfor model to predict yield values for the validation
data sample
  pred <- predict(randfor, trainvals_no_na[val_samp,])
  # Identify the observed values for validation data sample
  obs <- trainvals_no_na[val_samp,1]
  # Calculate RMSE value and add it to the data frame
  rmse <- c(rmse,sqrt(mean((pred-obs)^2)))
  # Calculate r-squared value and add it to the data frame
  rsquared <- c(rsquared,cor(pred,obs)^2)
}
```

```
# Calculate the average rmse and rsquared values for the 10 iterations
rmseMEAN <- mean(rmse)
rsquaredMEAN <- mean(rsquared)

# Print mean values
print(paste("Ten-fold cross validation RMSE mean =", rmseMEAN))
print(paste("Ten-fold cross validation r-squared mean =", rsquaredMEAN))
```

### ***Random Forest Input Feature Variables Preparation***

The input feature variables used to build the various Random Forest models used throughout this thesis were extracted from satellite data and environmental data (e.g. temperature and precipitation) using a variety of software depending on the nature of the variables being extracted and the scale/coverage required.

Input feature variables used to estimate crop yield (chapter 3) consisted of Sentinel-2 satellite data and a series of environmental data (e.g. precipitation and temperature data). For the most part these data required only a limited amount of basic pre-processing (clipping to the study area, stacking layers into single raster file, resampling to 10m resolution, etc.) which was conducted using ArcMap10. In addition to this, a series of vegetation indices were derived from the Sentinel-2 data using a combination of Erdas Imagine and ArcMap10.

The habitat productivity and heterogeneity input feature variables used to estimate bird diversity (chapter 4 & chapter 5) required more extensive processing prior to being used to build the Random Forest models. The habitat heterogeneity metrics were derived using FRAGSTATS, a computer software program that is designed to compute a variety of landscape metrics for categorical map patterns (McGarigal et al., 2012b). FRAGSTATS was used to derive a series of class-level metrics from LCM2000 (Fuller et al., 2002) and LCM2015 (Rowland et al., 2017) for chapters 4 and 5 respectively, using a sampling strategy of uniform 1km tiles designed to line up with the Countryside Survey and BBS data. Details of the metrics derived can be found in table A10 (Appendix 4).

The habitat productivity variables were derived in Google Earth Engine (GEE), a cloud-based platform that provides access to high-performance computing resources allowing planetary-scale geospatial analysis (Gorelick et al., 2017). Without GEE calculating these metrics would not have been possible due to the volume of data involved and the processing power required. An example of the script used to calculate and extract the various NDVI metrics from Landsat 5 and Landsat 7 data, for analysis in chapter 4, can be seen below. The same basic process was used to extract data from Landsat 8 for the analysis in chapter 5, with some slight alterations to account for differences between the sensors.

**Example Google Earth Engine script to extract Habitat Productivity Variables (NDVI metrics) for each broad land class within each Countryside Survey square. Green text shows comments, other colours of text show the javascript code used by GEE.**

**Step 1: Create collection of cloud free Landsat 7 images for each month**

```
// Load Landsat 7 ImageCollection.
var L7collection = ee.ImageCollection('LANDSAT/LE07/C01/T1_SR')
// Filter collection to obtain all images from 1999-2002
.filter(ee.Filter.calendarRange(1999,2002,'year'))
// Filter collection using to obtain only images for GB
.filterBounds(geometry) // "geometry" must be digitised prior to
                           running script
// Map function over the image collection to mask out cloud
.map(function(img) {
  var mask = img.select(['pixel_qa']).bitwiseAnd(32).eq(0);
  return img.updateMask(mask)
//Add NDVI to each image in the collection
.addBands(img.normalizedDifference(['B4', 'B3']))
// Add image capture time to each image

.addBands(img.metadata('system:time_start').subtract(1167609600000).divide(
86400000));
});

// Separate collection by month and reproject each image to British
National Grid projection
var L7mar = L7collection.filter(ee.Filter.calendarRange(3,3,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L7apr = L7collection.filter(ee.Filter.calendarRange(4,4,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L7may = L7collection.filter(ee.Filter.calendarRange(5,5,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L7jun = L7collection.filter(ee.Filter.calendarRange(6,6,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L7jul = L7collection.filter(ee.Filter.calendarRange(7,7,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L7aug = L7collection.filter(ee.Filter.calendarRange(8,8,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L7sep = L7collection.filter(ee.Filter.calendarRange(9,9,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
```

## Step 2: Create Maximum Value Composite for each month from Landsat 7 images

```
// Create a greenest pixel composite for each month
var L7mar_gPC = L7mar.qualityMosaic('nd');
var L7apr_gPC = L7apr.qualityMosaic('nd');
var L7may_gPC = L7may.qualityMosaic('nd');
var L7jun_gPC = L7jun.qualityMosaic('nd');
var L7jul_gPC = L7jul.qualityMosaic('nd');
var L7aug_gPC = L7aug.qualityMosaic('nd');
var L7sep_gPC = L7sep.qualityMosaic('nd');

// Define RED band for NDVI calculation for each greenest pixel composite
var L7mar_red = L7mar_gPC.select('B3');
var L7apr_red = L7apr_gPC.select('B3');
var L7may_red = L7may_gPC.select('B3');
var L7jun_red = L7jun_gPC.select('B3');
var L7jul_red = L7jul_gPC.select('B3');
var L7aug_red = L7aug_gPC.select('B3');
var L7sep_red = L7sep_gPC.select('B3');

// Define NIR band for NDVI calculation for each greenest pixel composite
var L7mar_nir = L7mar_gPC.select('B4');
var L7apr_nir = L7apr_gPC.select('B4');
var L7may_nir = L7may_gPC.select('B4');
var L7jun_nir = L7jun_gPC.select('B4');
var L7jul_nir = L7jul_gPC.select('B4');
var L7aug_nir = L7aug_gPC.select('B4');
var L7sep_nir = L7sep_gPC.select('B4');

// Calculate NDVI band for each greenest pixel composite
var L7mar_ndvi =
L7mar_nir.subtract(L7mar_red).divide(L7mar_nir.add(L7mar_red));
var L7apr_ndvi =
L7apr_nir.subtract(L7apr_red).divide(L7apr_nir.add(L7apr_red));
var L7may_ndvi =
L7may_nir.subtract(L7may_red).divide(L7may_nir.add(L7may_red));
var L7jun_ndvi =
L7jun_nir.subtract(L7jun_red).divide(L7jun_nir.add(L7jun_red));
var L7jul_ndvi =
L7jul_nir.subtract(L7jul_red).divide(L7jul_nir.add(L7jul_red));
var L7aug_ndvi =
L7aug_nir.subtract(L7aug_red).divide(L7aug_nir.add(L7aug_red));
var L7sep_ndvi =
L7sep_nir.subtract(L7sep_red).divide(L7sep_nir.add(L7sep_red));

// Rename NDVI band
var L7mar_ndvi = L7mar_ndvi.select('B4').rename('ndvi');
var L7apr_ndvi = L7apr_ndvi.select('B4').rename('ndvi');
var L7may_ndvi = L7may_ndvi.select('B4').rename('ndvi');
var L7jun_ndvi = L7jun_ndvi.select('B4').rename('ndvi');
var L7jul_ndvi = L7jul_ndvi.select('B4').rename('ndvi');
var L7aug_ndvi = L7aug_ndvi.select('B4').rename('ndvi');
var L7sep_ndvi = L7sep_ndvi.select('B4').rename('ndvi');
```

### Step 3: Create collection of cloud free Landsat 5 images for each month

```
// Load Landsat 5 ImageCollection.
var L5collection = ee.ImageCollection('LANDSAT/LT05/C01/T1_SR')
// Filter collection to obtain all images from 1999-2002
.filter(ee.Filter.calendarRange(1999,2002,'year'))
// Filter collection using to obtain only images for GB
.filterBounds(geometry) // "geometry" must be digitised prior to
                          running script

// Map function over the image collection to mask out cloud
.map(function(img) {
  var mask = img.select(['pixel_qa']).bitwiseAnd(32).eq(0);
  return img.updateMask(mask)
//Add NDVI to each image in the collection
.addBands(img.normalizedDifference(['B4', 'B3']))
// Add image capture time to each image

.addBands(img.metadata('system:time_start').subtract(1167609600000).divide(
86400000));
});

// Separate collection by month and reproject each image to British
National Grid projection
var L5mar = L5collection.filter(ee.Filter.calendarRange(3,3,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L5apr = L5collection.filter(ee.Filter.calendarRange(4,4,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L5may = L5collection.filter(ee.Filter.calendarRange(5,5,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L5jun = L5collection.filter(ee.Filter.calendarRange(6,6,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L5jul = L5collection.filter(ee.Filter.calendarRange(7,7,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L5aug = L5collection.filter(ee.Filter.calendarRange(8,8,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
var L5sep = L5collection.filter(ee.Filter.calendarRange(9,9,'month'))
  .map(function(img) {
    return img.reproject('EPSG:4326',null,30);
  });
```

### Step 4: Create Maximum Value Composite for each month from Landsat 7 images

```
// Create a greenest pixel composite for each month
var L5mar_gPC = L5mar.qualityMosaic('nd');
var L5apr_gPC = L5apr.qualityMosaic('nd');
var L5may_gPC = L5may.qualityMosaic('nd');
var L5jun_gPC = L5jun.qualityMosaic('nd');
var L5jul_gPC = L5jul.qualityMosaic('nd');
var L5aug_gPC = L5aug.qualityMosaic('nd');
var L5sep_gPC = L5sep.qualityMosaic('nd');
```

```

// Define RED band for NDVI calculation for each greenest pixel composite
var L5mar_red = L5mar_gPC.select('B3');
var L5apr_red = L5apr_gPC.select('B3');
var L5may_red = L5may_gPC.select('B3');
var L5jun_red = L5jun_gPC.select('B3');
var L5jul_red = L5jul_gPC.select('B3');
var L5aug_red = L5aug_gPC.select('B3');
var L5sep_red = L5sep_gPC.select('B3');

// Define NIR band for NDVI calculation for each greenest pixel composite
var L5mar_nir = L5mar_gPC.select('B4');
var L5apr_nir = L5apr_gPC.select('B4');
var L5may_nir = L5may_gPC.select('B4');
var L5jun_nir = L5jun_gPC.select('B4');
var L5jul_nir = L5jul_gPC.select('B4');
var L5aug_nir = L5aug_gPC.select('B4');
var L5sep_nir = L5sep_gPC.select('B4');

// Calculate NDVI band for each greenest pixel composite
var L5mar_ndvi =
L5mar_nir.subtract(L5mar_red).divide(L5mar_nir.add(L5mar_red));
var L5apr_ndvi =
L5apr_nir.subtract(L5apr_red).divide(L5apr_nir.add(L5apr_red));
var L5may_ndvi =
L5may_nir.subtract(L5may_red).divide(L5may_nir.add(L5may_red));
var L5jun_ndvi =
L5jun_nir.subtract(L5jun_red).divide(L5jun_nir.add(L5jun_red));
var L5jul_ndvi =
L5jul_nir.subtract(L5jul_red).divide(L5jul_nir.add(L5jul_red));
var L5aug_ndvi =
L5aug_nir.subtract(L5aug_red).divide(L5aug_nir.add(L5aug_red));
var L5sep_ndvi =
L5sep_nir.subtract(L5sep_red).divide(L5sep_nir.add(L5sep_red));

// Rename NDVI band
var L5mar_ndvi = L5mar_ndvi.select('B4').rename('ndvi');
var L5apr_ndvi = L5apr_ndvi.select('B4').rename('ndvi');
var L5may_ndvi = L5may_ndvi.select('B4').rename('ndvi');
var L5jun_ndvi = L5jun_ndvi.select('B4').rename('ndvi');
var L5jul_ndvi = L5jul_ndvi.select('B4').rename('ndvi');
var L5aug_ndvi = L5aug_ndvi.select('B4').rename('ndvi');
var L5sep_ndvi = L5sep_ndvi.select('B4').rename('ndvi');

// Join greenest pixel composites from L5 & L7 to form single image
collection for each month
var mar_ndvi = ee.ImageCollection.fromImages([L5mar_ndvi,L7mar_ndvi]);
var apr_ndvi = ee.ImageCollection.fromImages([L5apr_ndvi,L7apr_ndvi]);
var may_ndvi = ee.ImageCollection.fromImages([L5may_ndvi,L7may_ndvi]);
var jun_ndvi = ee.ImageCollection.fromImages([L5jun_ndvi,L7jun_ndvi]);
var jul_ndvi = ee.ImageCollection.fromImages([L5jul_ndvi,L7jul_ndvi]);
var aug_ndvi = ee.ImageCollection.fromImages([L5aug_ndvi,L7aug_ndvi]);
var sep_ndvi = ee.ImageCollection.fromImages([L5sep_ndvi,L7sep_ndvi]);

// Function to mask out NDVI values above 1 & below -1 (i.e. unrealistic
values)
var maskNDVI = function(image){
  var ndvi_gt1_mask = image.select('ndvi').lte(1);
  var ndvi_lt1_mask = image.select('ndvi').gte(-1);
  var ndvi_gt1 = image.select('ndvi').updateMask(ndvi_gt1_mask);
  var ndvi_lt1 = ndvi_gt1.select('ndvi').updateMask(ndvi_lt1_mask);
  return ndvi_lt1;
};

```



```

// Map maskNDVI function over the image collections to mask out odd values
var mar_ndvimasked = mar_ndvi.map(maskNDVI);
var apr_ndvimasked = apr_ndvi.map(maskNDVI);
var may_ndvimasked = may_ndvi.map(maskNDVI);
var jun_ndvimasked = jun_ndvi.map(maskNDVI);
var jul_ndvimasked = jul_ndvi.map(maskNDVI);
var aug_ndvimasked = aug_ndvi.map(maskNDVI);
var sep_ndvimasked = sep_ndvi.map(maskNDVI);

// Create Maximum Value Composite (MVC) for each month by taking the
highest NDVI value for each pixel between Landsat 5 and Landsat 7
var mar_ndvimax = mar_ndvimasked.reduce(ee.Reducer.max());
var apr_ndvimax = apr_ndvimasked.reduce(ee.Reducer.max());
var may_ndvimax = may_ndvimasked.reduce(ee.Reducer.max());
var jun_ndvimax = jun_ndvimasked.reduce(ee.Reducer.max());
var jul_ndvimax = jul_ndvimasked.reduce(ee.Reducer.max());
var aug_ndvimax = aug_ndvimasked.reduce(ee.Reducer.max());
var sep_ndvimax = sep_ndvimasked.reduce(ee.Reducer.max());

// Rename the MVC band
var mar_ndvimax = mar_ndvimax.select('ndvi_max').rename('ndvi');
var apr_ndvimax = apr_ndvimax.select('ndvi_max').rename('ndvi');
var may_ndvimax = may_ndvimax.select('ndvi_max').rename('ndvi');
var jun_ndvimax = jun_ndvimax.select('ndvi_max').rename('ndvi');
var jul_ndvimax = jul_ndvimax.select('ndvi_max').rename('ndvi');
var aug_ndvimax = aug_ndvimax.select('ndvi_max').rename('ndvi');
var sep_ndvimax = sep_ndvimax.select('ndvi_max').rename('ndvi');

// Create single collection of all the MVCs
var ndviCollection = ee.ImageCollection.fromImages([mar_ndvimax,
apr_ndvimax, may_ndvimax, jun_ndvimax, jul_ndvimax, aug_ndvimax,
sep_ndvimax]);

```

**Step 5: Calculate and export the average of each NDVI metric for each land cover type within each 1km Countryside Survey square. The method for arable land is provided below as an example, this process was repeated for each broad land cover class.**

```

// Create arable landcover mask
var arable_mask = LCMArableWGS.select('b1').eq(1); // "LCMArableWGS" is a
single band raster image where arable land has a value of 1, non-arable
land has a value of 0

// Apply arable mask to each NDVI metric image to mask out all pixels which
do not contain arable land
var ar_mar = mar_ndvimax.select('ndvi').updateMask(arable_mask);
var ar_apr = apr_ndvimax.select('ndvi').updateMask(arable_mask);
var ar_may = may_ndvimax.select('ndvi').updateMask(arable_mask);
var ar_jun = jun_ndvimax.select('ndvi').updateMask(arable_mask);
var ar_jul = jul_ndvimax.select('ndvi').updateMask(arable_mask);
var ar_aug = aug_ndvimax.select('ndvi').updateMask(arable_mask);
var ar_sep = sep_ndvimax.select('ndvi').updateMask(arable_mask);

// Calculate the average of each NDVI metric (e.g. minimum NDVI) per 1km CS
square
// "ar_mean" is provided as an example, repeat process substituting in each
NDVI metric

```

```

var arable = ar_mar.reduceRegions({           // Replace "ar_mar" with the month
  collection: CS_bird_squaresWGS,           // Shapefile containing outlines
  reducer: ee.Reducer.mean(),               // of 1km Countryside Survey squares
  scale: 30                                // Change reducer depending on
})                                           // metric of interest
// Creates a feature collection with the average value of the specified
// NDVI metric for each CS square, which can be exported as a CSV file

//Export 'arable' table to Google Drive
Export.table.toDrive({
  collection: arable,                       // Table to export
  description: 'ar_mean_mar',               // File name
  fileFormat: 'CSV',                       // Export as CSV file
  folder: 'NDVI_per_month',                // Set Google Drive folder to export file
})                                           // into

// Repeat this process for each month and NDVI metric combination

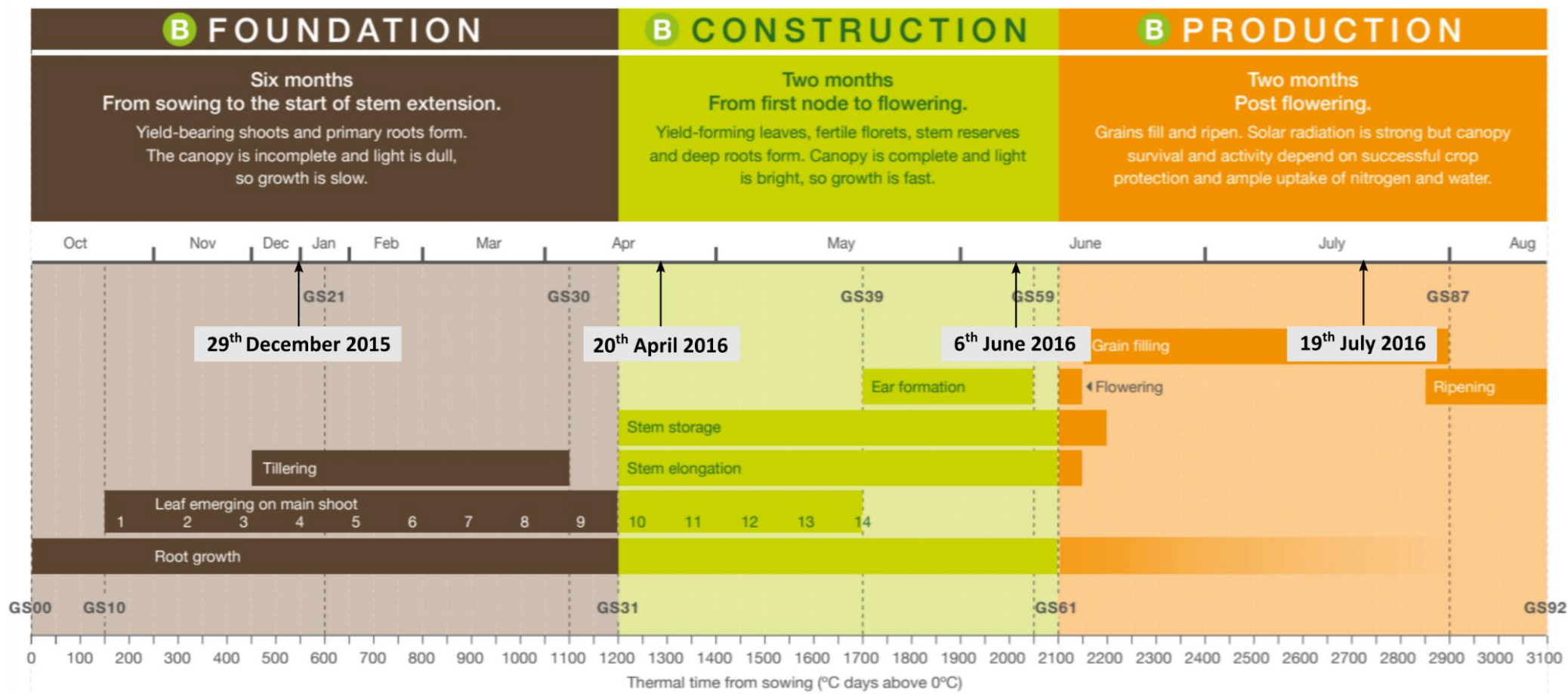
```

## References for Appendix 2

- Boehmke, B., Greenwell, B., 2020. Hands On Machine Learning with R. Taylor & Francis.
- Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.
- Fuller, R.M., Smith, G.M., Sanderson, J.M., Hill, R.A., Thomson, A.G., Cox, R., Brown, N.J., Clarke, R.T., Rothery, P., Gerard, F.F., 2002. Countryside Survey 2000 Module 7 Land Cover Map 2000. Final Report.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R news 2, 18–22. <https://doi.org/10.1177/154405910408300516>
- McGarigal, K., Cushman, S.A., Ene, E., 2012. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following website: <http://www.umass.edu/landeco/research/fragsta>.
- Prasad, A.M., Iversen, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems 9, 181–199. <https://doi.org/10.1007/s10021-005-0054-1>
- Rowland, C.S., Morton, R.D., Carrasco, L., McShane, G., O’Neil, A.W., Wood, C.M., 2017. Land Cover Map 2015 (25m raster, GB).

### Appendix 3: Supplementary Material for Chapter 3

Figure A3: Timing of the Sentinel-2 images used for yield estimation relative to the key wheat growth stages. Wheat growth stages taken from (AHDB, 2018).



#### Appendix 4: Supplementary Material for Chapter 4

**Table A8: Farmland and woodland bird species groupings based on the BTO/JNCC/RSPB wild bird indicators for the UK and England (Eaton & Noble, 2018). Italicised text indicates species for which no data was available in the CS2000 data.**

Farmland Birds Species	Woodland Birds Species	
Corn Bunting	Blackbird	Long-tailed Tit
Goldfinch	Blackcap	Marsh Tit
Greenfinch	Blue Tit	Nightingale
Grey partridge	Bullfinch	Nuthatch
Jackdaw	<i>Capercaillie</i>	<i>Pied Flycatcher</i>
Kestrel	Chaffinch	Redstart
Lapwing	Chiffchaff	Robin
Linnet	Coal Tit	Siskin
Reed Bunting	<i>Common Crossbill</i>	Song Thrush
Rook	Dunnock	Sparrowhawk
Skylark	Garden Warbler	Spotted
Starling	Goldcrest	Flycatcher
Stock Dove	Great Tit	Tawny Owl
Tree Sparrow	Green Woodpecker	Treecreeper
Turtle Dove	Great Spotted	Tree Pipit
Yellow Wagtail	Woodpecker	Willow Tit
Yellowhammer	Jay	Willow Warbler
Whitethroat	Lesser Redpoll	Wood Warbler
Woodpigeon	Lesser Spotted	Wren
	Woodpecker	
	Lesser Whitethroat	

**Table A9: Broader land cover class groups and the original LCM2000 subclasses.**

LC group	LCM Subclass
Arable	Arable cereals, arable horticulture, non-rotational horticulture
Broad-leaved	Broad-leaved/mixed woodland
Coniferous	Coniferous woodland
Grassland	Improved grassland, set-aside grassland, neutral grassland, calcareous
Semi-natural	Acid grassland, bracken, dense and open dwarf shrub heath, fen, marsh, swamp, bogs (deep peat), montane habitats, inland bare ground, saltmarsh
Urban	Continuous urban, suburban/rural developed
Water	Water (inland)
Coast	Supra-littoral rock, supra-littoral sediment, littoral rock, littoral sediment
Sea	Sea/estuary

**Table A10: Details of the habitat structure metrics derived from LCM2000 using FRAGSTATS. Information based on the McGarigal (2015).**

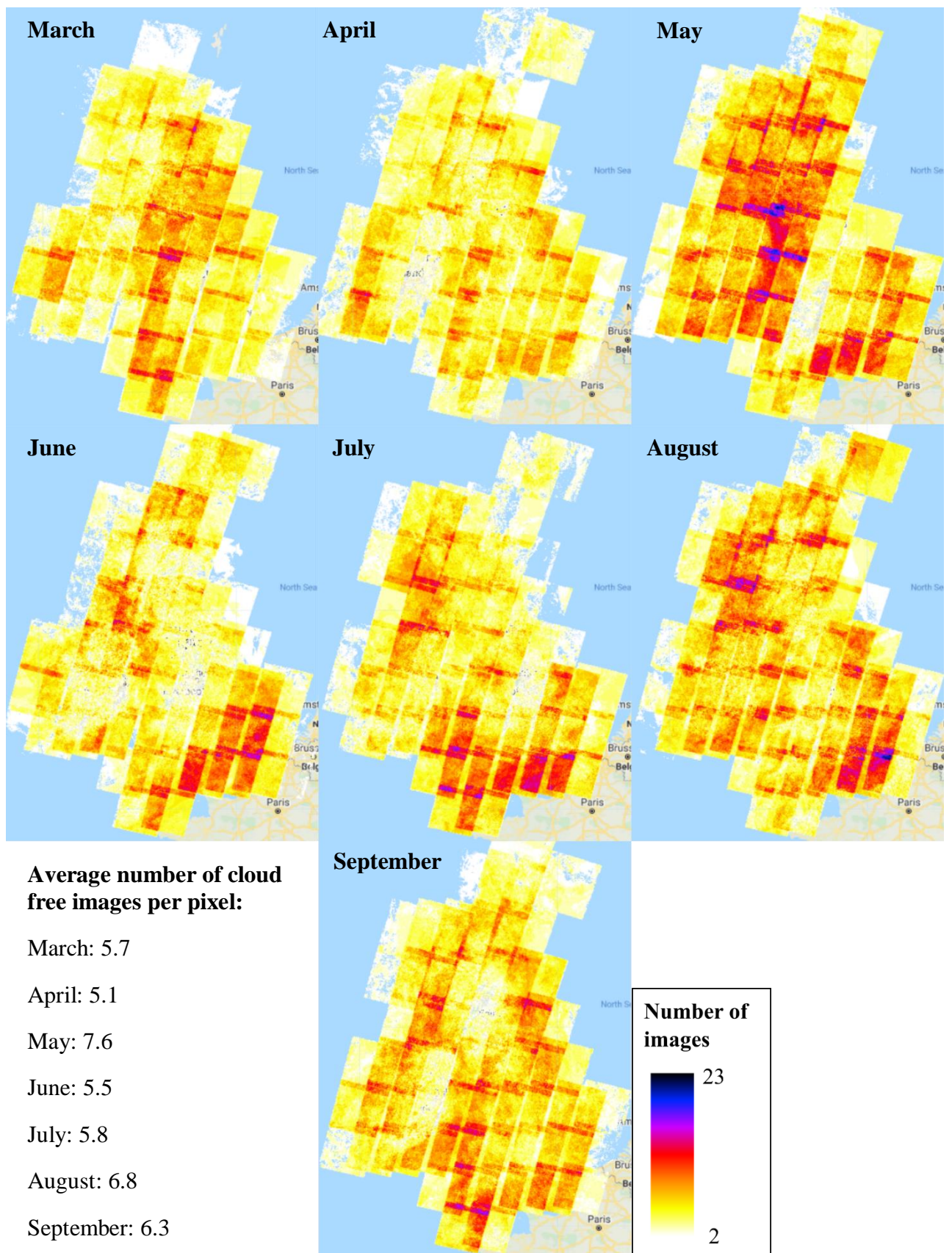
Variable	Abbreviation (Unit)	Description
<i>Area &amp; Edge Metrics</i>		
Area	AREA <sub>AM/CV/MN</sub> (ha)	Area of each patch comprising a landscape mosaic given as mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class. Measured in hectares (ha).
Percentage of landscape	PLAND (%)	Percentage of the landscape comprised of a particular patch type
Edge Density	ED (m/ha)	Edge length of a particular patch type standardised to a per unit area basis. Measured in metres per hectare (m/ha).
Radius of Gyration	GYRATE <sub>AM/CV/MN</sub> (m)	Measure of patch extent given as mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class. GYRATE increases as the patch increases in extent. Measured in metres (m)
Largest Patch Index	LPI (%)	Quantifies the percentage of the total landscape area comprised by the largest patch. The LPI approaches 0 when the largest patch of the corresponding type is increasingly small. An LPI value of 100 indicates the entire landscape is made up of a single patch of the corresponding patch type. Measured in percent (%).
Total Edge	TE (m)	Absolute measure of total edge length of a particular patch type. Measured in metres (m)
<i>Shape Metrics</i>		
Related Circumscribing Circle	CIRCLE <sub>AM/CV/MN</sub>	Measure of overall patch elongation using the ratio of patch area to the ratio of the smallest circumscribing circle given as mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class. Highly convoluted but narrow patches give a low index value, while narrow and elongated patches have a high index value.
Contiguity Index	CONTIG <sub>AM/CV/MN</sub>	Measure of spatial connectedness/contiguity of cells within a grid-cell given as the mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class. An index value of 0 represents a one-pixel patch, increasing to a limit of 1 as connectedness increases.
Fractal Dimension Index	FRAC <sub>AM/CV/MN</sub>	Measure of shape complexity which approaches 1 for shapes with simple perimeters (e.g. squares) and 2 for shapes with highly convoluted, plane-filling perimeters. Values are given as a mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class.
Perimeter-Area Ratio	PARA <sub>AM/CV/MN</sub>	Ratio of patch perimeter to area given as mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class, providing a measure of shape complexity.
Shape Index	SHAPE <sub>AM/CV/MN</sub>	Measures the complexity of patch shape compared to a standard shape (square) of the same size. Values are given as mean (MN), coefficient of variation (CV) or area-weighted mean (AM) per class. Higher values indicate more complex patch shapes.

Table A10 continued

Variable	Abbreviation (Unit)	Description
<i>Aggregation Metrics</i>		
Patch Cohesion	COHESION	Provides a measure of the physical connectedness of the corresponding patch types. COHESION approaches 0 as the proportion of the focal class decreases and becomes increasingly subdivided, and therefore less physically connected.
Landscape Division Index	DIVISION (proportion)	Probability that 2 randomly chosen pixels in the landscape are not situated in the same undissected patch of the corresponding patch type. DIVISION is 0 when the landscape consists of a single patch, approaching 1 as the proportion of the landscape comprised of the focal patch type decrease and those patches decrease in size.
Effective Mesh Size	MESH (ha)	Size of patches one gets when dividing the total landscape into patches of equal size in such a way that the new configuration leads to the same degree of landscape division as obtained by for the observed cumulative area distribution. Quantifies habitat fragmentation based on the probability that two randomly chosen points in the region under interest are located in the same non-fragmented patch (Jaeger, 2000). The probability is multiplied by the total area of the landscape unit. The more barriers (e.g. roads, railroads) in the landscape, the lower the probability that the two locations will be located in the same patch, and the lower the effective mesh size
Number of Patches	NP	Number of patches of a particular patch type
Patch Density	PD (number of patches per 100 ha)	Number of patches of the corresponding patch type standardised on a per unit area basis
Splitting Index	SPLIT	Number of patches one gets when dividing the total landscape into patches of equal size in such a way that the new configuration leads to the same degree of landscape division as obtained by for the observed cumulative area distribution. SPLIT is 1 when the landscape consists of a single patch, increasing in value as the focal patch type is increasingly reduced in area and subdivided into smaller patches.



**Figure A4: Number of cloud-free images available for each pixel for each month.**



**Figure A5: Monthly Maximum NDVI Value Composites created by merging Landsat 5 and Landsat 7 to produce near-complete coverage for the whole of GB.**





**Table A11: R-squared values for Random Forest models built using individual habitat productivity and heterogeneity variables. For each model, the variable of interest was calculated for each land cover class separately. For example, the model for March NDVI Mean contains the individual mean March NDVIs calculated for the arable, broadleaved woodland, coniferous woodland, grassland and semi-natural land cover classes.**

Habitat Productivity Variables	R-squared value			
	Farmland birds		Woodland birds	
	Richness	Diversity	Richness	Diversity
<b><i>Growing season (March-September)</i></b>				
NDVI Sum	0.61	0.59	0.59	0.70
NDVI Coefficient of variation	0.61	0.56	0.55	0.67
NDVI Mean	0.62	0.59	0.61	0.73
NDVI Median	0.58	0.57	0.60	0.72
NDVI Minimum	0.58	0.55	0.57	0.68
NDVI Standard deviation	0.61	0.55	0.54	0.65
NDVI Maximum	0.61	0.56	0.58	0.71
NDVI Range	0.60	0.55	0.55	0.66
NDVI 20 <sup>th</sup> percentile	0.62	0.59	0.60	0.72
NDVI 80 <sup>th</sup> percentile	0.60	0.56	0.58	0.70
NDVI Interquartile range	0.61	0.56	0.56	0.66
<b><i>March</i></b>				
NDVI Mean	0.60	0.58	0.58	0.69
NDVI Median	0.60	0.57	0.57	0.67
NDVI Maximum	0.65	0.61	0.58	0.71
NDVI Minimum	0.58	0.54	0.53	0.65
NDVI Range	0.65	0.58	0.59	0.68
NDVI Standard Deviation	0.63	0.57	0.56	0.67
NDVI 20 <sup>th</sup> percentile	0.58	0.55	0.57	0.68
NDVI 80 <sup>th</sup> percentile	0.62	0.58	0.60	0.71
NDVI Interquartile range	0.63	0.57	0.58	0.70
All NDVI metrics	0.63	0.61	0.63	0.73
<b><i>April</i></b>				
NDVI Mean	0.60	0.57	0.58	0.69
NDVI Median	0.60	0.57	0.57	0.67
NDVI Maximum	0.65	0.62	0.60	0.72
NDVI Minimum	0.59	0.54	0.53	0.64
NDVI Range	0.66	0.59	0.60	0.71
NDVI Standard Deviation	0.64	0.57	0.61	0.71
NDVI 20 <sup>th</sup> percentile	0.58	0.55	0.58	0.66
NDVI 80 <sup>th</sup> percentile	0.63	0.60	0.58	0.70
NDVI Interquartile range	0.63	0.58	0.58	0.69
All NDVI metrics	0.66	0.62	0.65	0.73

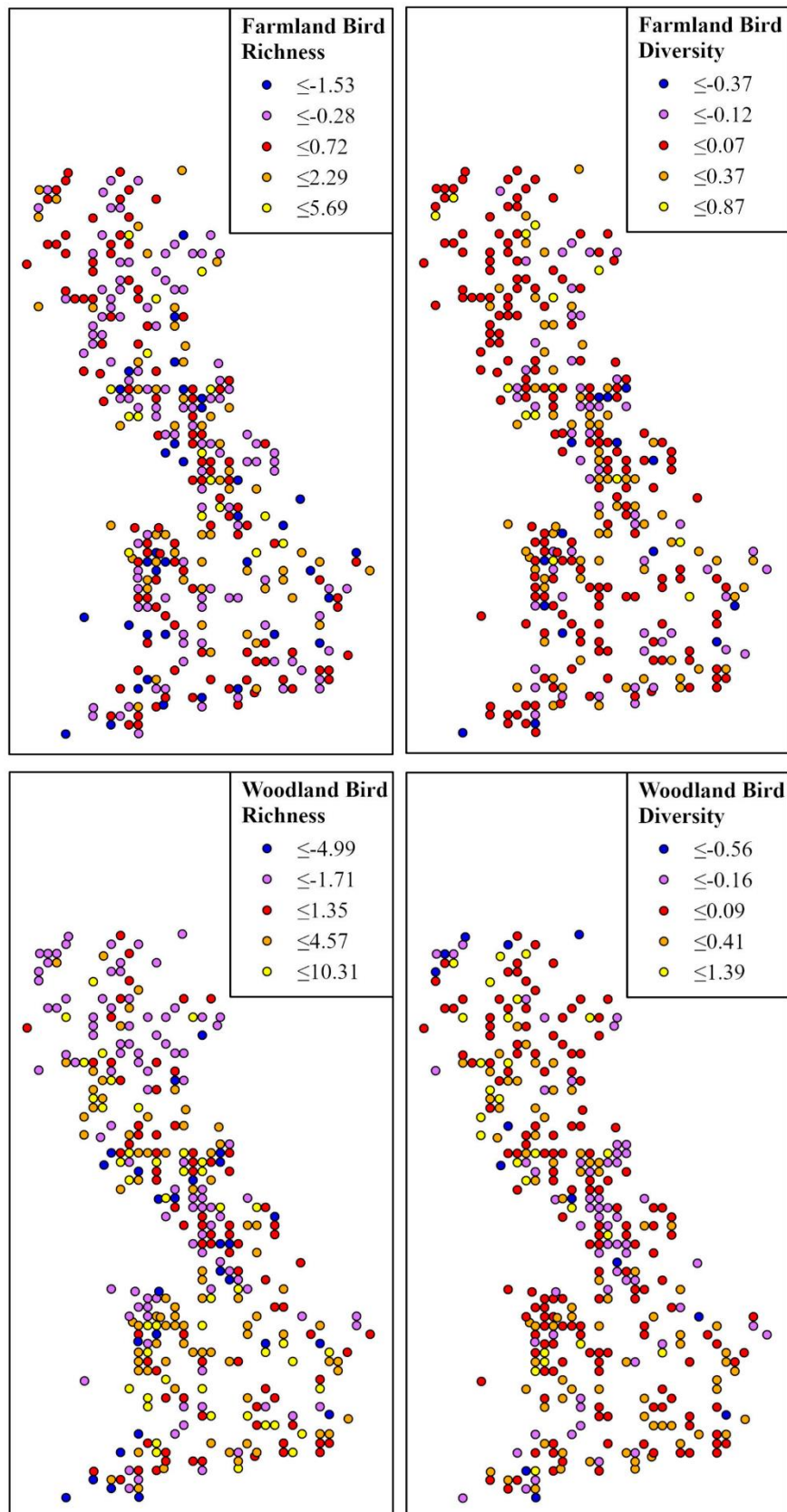
Table A11 continued

Habitat Productivity Variables	R-squared value			
	Farmland birds		Woodland birds	
	Richness	Diversity	Richness	Diversity
<b>May</b>				
NDVI Mean	0.60	0.56	0.54	0.65
NDVI Median	0.60	0.56	0.55	0.66
NDVI Maximum	0.65	0.60	0.55	0.66
NDVI Minimum	0.58	0.53	0.52	0.62
NDVI Range	0.63	0.58	0.57	0.67
NDVI Standard Deviation	0.61	0.57	0.55	0.66
NDVI 20 <sup>th</sup> percentile	0.60	0.56	0.53	0.64
NDVI 80 <sup>th</sup> percentile	0.61	0.57	0.55	0.66
NDVI Interquartile range	0.57	0.55	0.55	0.65
All NDVI metrics	0.64	0.61	0.57	0.68
<b>June</b>				
NDVI Mean	0.57	0.51	0.55	0.61
NDVI Median	0.58	0.52	0.57	0.62
NDVI Maximum	0.63	0.56	0.59	0.65
NDVI Minimum	0.57	0.52	0.53	0.62
NDVI Range	0.61	0.54	0.53	0.62
NDVI Standard Deviation	0.58	0.53	0.50	0.60
NDVI 20 <sup>th</sup> percentile	0.57	0.54	0.56	0.66
NDVI 80 <sup>th</sup> percentile	0.61	0.59	0.60	0.67
NDVI Interquartile range	0.57	0.55	0.51	0.60
All NDVI metrics	0.63	0.61	0.61	0.70
<b>July</b>				
NDVI Mean	0.57	0.52	0.54	0.65
NDVI Median	0.57	0.52	0.55	0.66
NDVI Maximum	0.63	0.57	0.60	0.69
NDVI Minimum	0.58	0.54	0.54	0.64
NDVI Range	0.65	0.60	0.56	0.65
NDVI Standard Deviation	0.62	0.58	0.55	0.65
NDVI 20 <sup>th</sup> percentile	0.57	0.53	0.56	0.66
NDVI 80 <sup>th</sup> percentile	0.59	0.54	0.59	0.68
NDVI Interquartile range	0.62	0.57	0.54	0.64
All NDVI metrics	0.64	0.59	0.61	0.68
<b>August</b>				
NDVI Mean	0.61	0.55	0.60	0.68
NDVI Median	0.61	0.54	0.60	0.68
NDVI Maximum	0.60	0.54	0.62	0.70
NDVI Minimum	0.62	0.55	0.53	0.65
NDVI Range	0.65	0.58	0.57	0.66
NDVI Standard Deviation	0.63	0.56	0.55	0.65
NDVI 20 <sup>th</sup> percentile	0.62	0.55	0.57	0.66
NDVI 80 <sup>th</sup> percentile	0.61	0.56	0.59	0.69
NDVI Interquartile range	0.63	0.56	0.56	0.65
All NDVI metrics	0.66	0.60	0.65	0.71

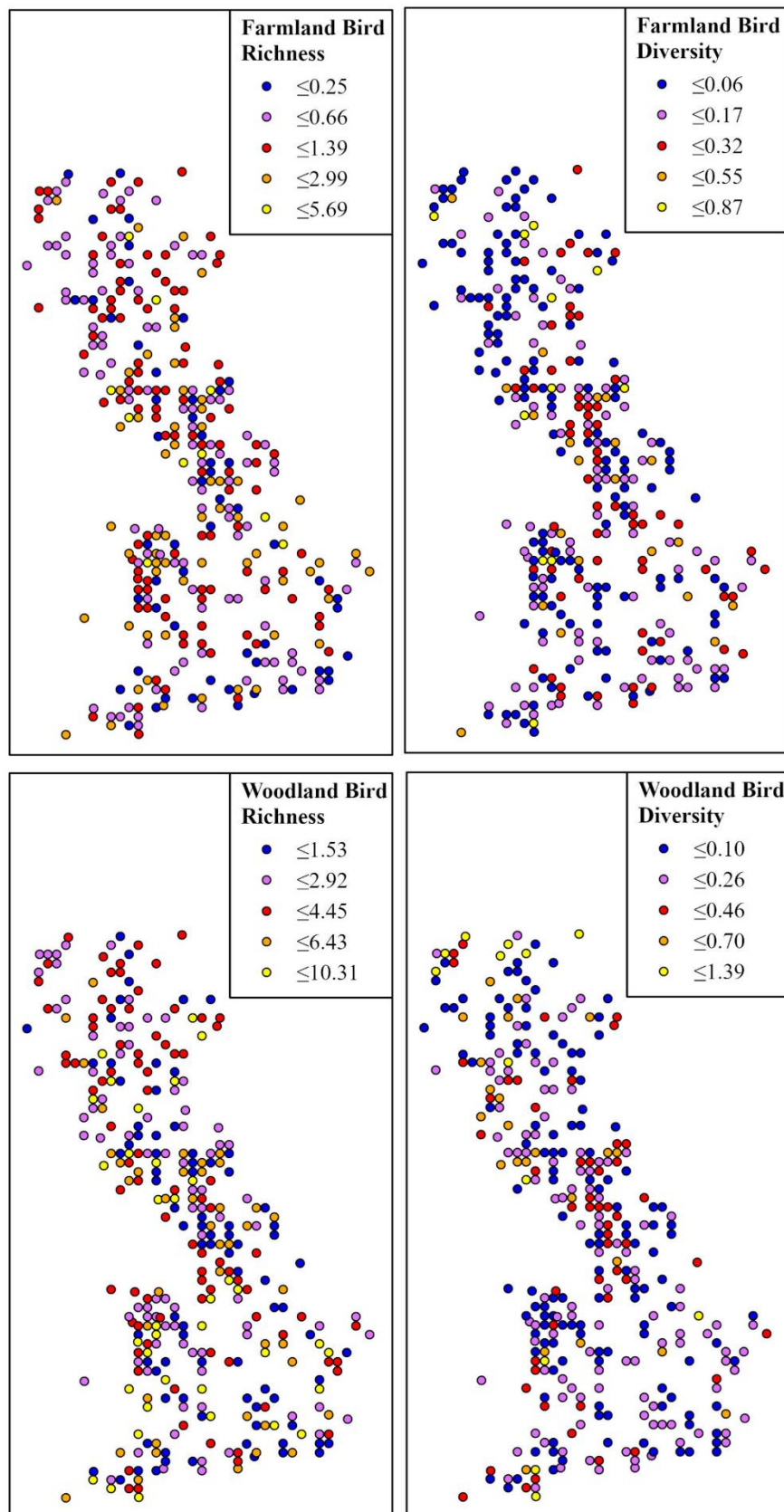
Table A11 continued

Habitat Productivity Variables	R-squared value			
	Farmland birds		Woodland birds	
	Richness	Diversity	Richness	Diversity
<b>September</b>				
NDVI Mean	0.64	0.60	0.61	0.71
NDVI Median	0.64	0.60	0.61	0.71
NDVI Maximum	0.63	0.58	0.64	0.72
NDVI Minimum	0.63	0.60	0.55	0.66
NDVI Range	0.66	0.61	0.60	0.68
NDVI Standard Deviation	0.64	0.59	0.56	0.66
NDVI 20 <sup>th</sup> percentile	0.63	0.57	0.60	0.71
NDVI 80 <sup>th</sup> percentile	0.64	0.60	0.62	0.72
NDVI Interquartile range	0.63	0.57	0.58	0.67
All NDVI metrics	0.68	0.64	0.68	0.74
<b>Habitat Heterogeneity Variables</b>				
AREA_AM	0.70	0.61	0.64	0.72
AREA_MN	0.68	0.61	0.64	0.71
PLAND	0.71	0.63	0.67	0.73
ED	0.71	0.62	0.65	0.70
GYRATE_AM	0.70	0.62	0.63	0.72
GYRATE_MN	0.69	0.61	0.60	0.70
LPI	0.70	0.62	0.65	0.72
TE	0.71	0.62	0.65	0.71
CIRCLE_AM	0.60	0.56	0.58	0.69
CIRCLE_MN	0.61	0.55	0.55	0.66
CONTIG_AM	0.69	0.62	0.62	0.71
CONTIG_MN	0.65	0.60	0.58	0.67
FRAC_AM	0.62	0.57	0.57	0.68
FRAC_MN	0.61	0.54	0.57	0.68
PARA_AM	0.66	0.59	0.61	0.70
PARA_MN	0.62	0.58	0.56	0.67
SHAPE_AM	0.64	0.57	0.58	0.68
SHAPE_MN	0.65	0.55	0.59	0.68
COHESION	0.69	0.62	0.65	0.72
DIVISION	0.64	0.57	0.62	0.69
MESH	0.70	0.62	0.66	0.72
NP	0.60	0.53	0.57	0.66
PD	0.60	0.54	0.57	0.67
SPLIT	0.67	0.59	0.63	0.70

Figure A6: Difference between the observed and predicted values for farmland bird and woodland bird richness and diversity.



**Figure A7: Absolute difference between the observed and predicted values for farmland bird and woodland bird richness and diversity.**



## Appendix 5: Supplementary Material for Chapter 5

**Figure A8: ITE Land Classification of Great Britain (2007).**

