High resolution wheat yield mapping using Sentinel-2

Merryn L. Hunt* [a,b], George Alan Blackburn[b], Luis Carrasco[d,e], John W. Redhead[c], Clare S. Rowland[a]

[a] Centre for Ecology and Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster, United Kingdom, LA1 4AP

[b] Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster, United Kingdom, LA1 4AP

[c] Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB

[d] National Institute for Mathematical and Biological Synthesis, 1122 Volunteer Boulevard, University of Tennessee, Knoxville, TN 37996, USA

[e] Department of Ecology and Evolutionary Biology, 569 Dabney Hall, University of Tennessee, Knoxville, TN 37996, USA

* Corresponding author email address: m.hunt3@lancaster.ac.uk

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.

Key Words

Yield estimation; Sentinel-2; Yield mapping; Random Forest regression; Combine harvester

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.

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 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.*, 2017). 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).

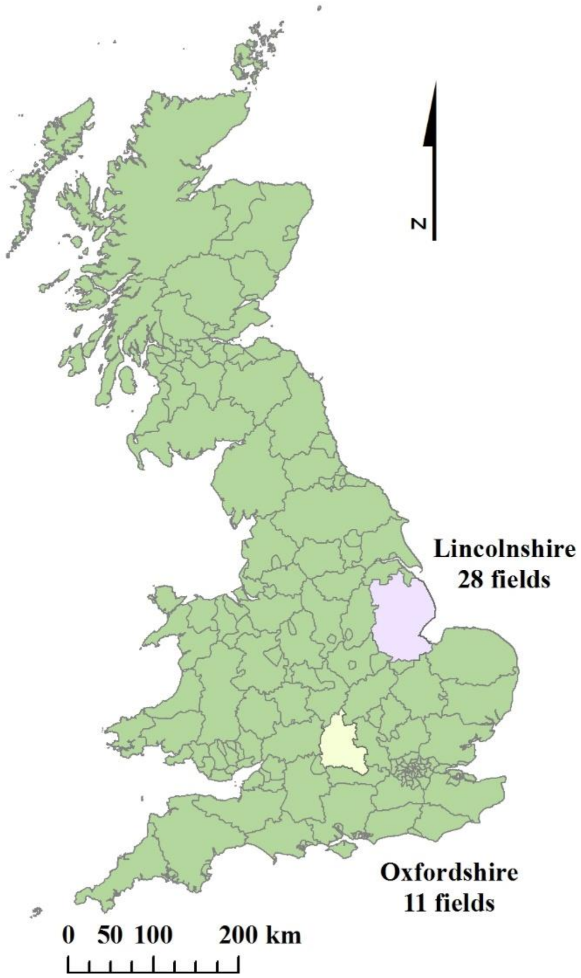

Figure 1: Location of study sites.

3. Data and Methods

Figure 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 3.1-3.3, and the analysis techniques are outlined in sections 3.4-3.6.

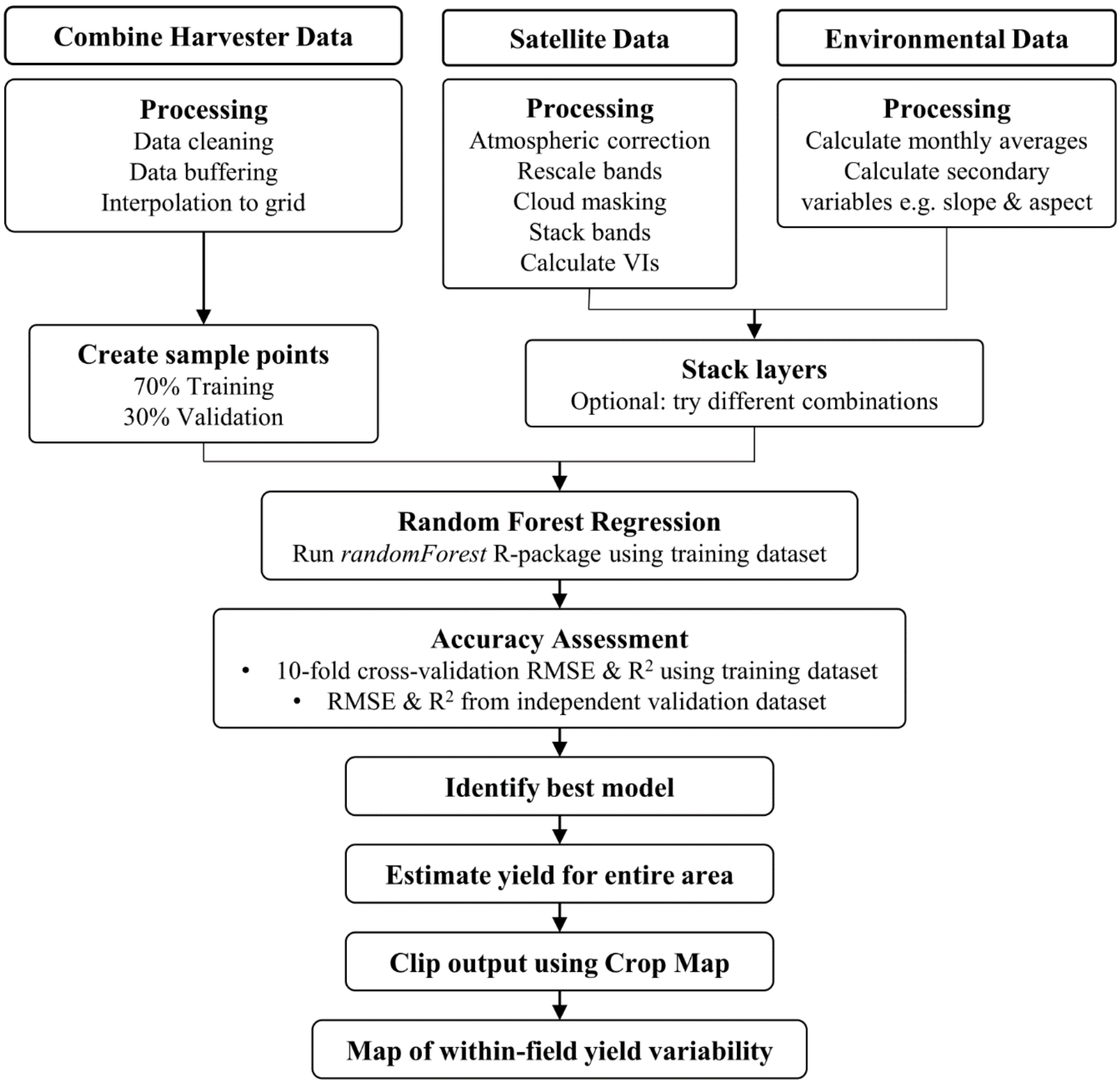

Figure 2: Overview of the method used to estimate yield at high resolution on a landscape scale.

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 6th August and 9th 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 < 30s), accelerating or decelerating (accel. > 0.05 ms^{-2}), or when the speed fell outside the optimum limits to accurately measure the yield (ground speed <2 kmh^{-1} or >8 kmh^{-1}). Finally, data were cleaned on a per field basis removing yield values which fell outside the global mean ± 2.5 sd or the local mean ± 2.5 sd (based on the closest 3 points). A summary of the criteria for data cleaning can be found in figure 3.

Combine Harvester Data cleaning

Criteria for removing points

- No latitude/longitude
- Yield monitor/front attachment switched off
- Turning angle > 0.6 radians for time step < 30s
- Acceleration > 0.05 ms^{-2}
- Ground speed < 2kmh^{-1} or > 8kmh^{-1}
- Outside global mean ± 2.5 sd
- Outside local mean ± 2.5sd

Figure 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 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.

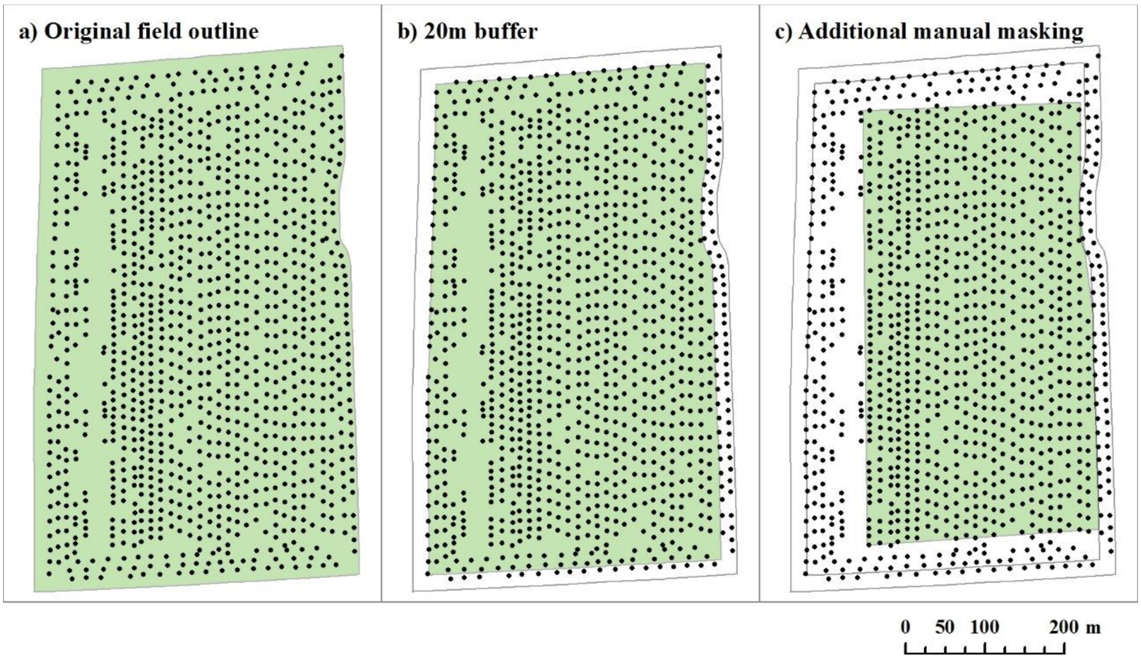

Figure 4: Example yield data points for one field showing a) gaps in the data arising from the data collection and cleaning process and b-c) the stages in the buffering process used to remove these gaps.

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 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 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.

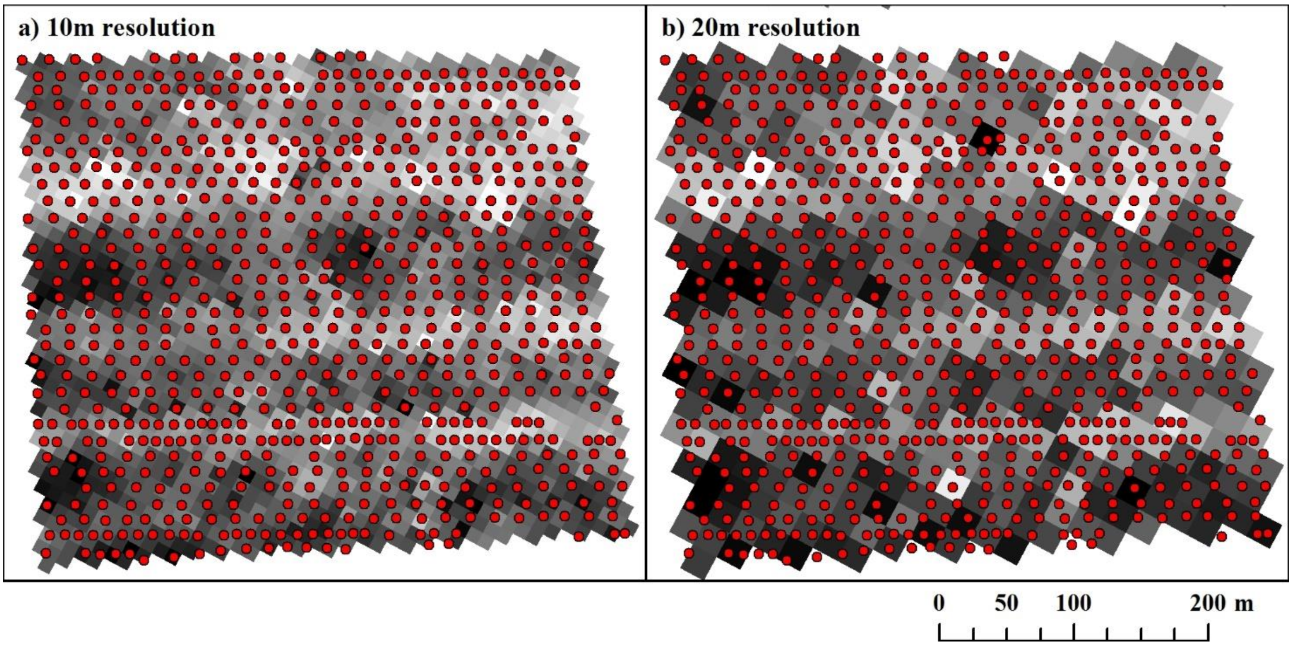

Figure 5: Example of the distribution of yield data points relative to a) 10m and b) 20m resolution interpolated yield data

3.2 Sentinel-2 Data

3.2.1 Sentinel-2 Image Processing

Predominantly cloud-free Sentinel-2 images (Level 1C Top-of-Atmosphere reflectance products; see 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 1. Relatively cloud-free images were available over the growing season for the 29th December 2015, 20th April 2016, 6th June 2016 and 19th July 2016 (table 2). 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 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 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)	29th Dec 2015 20th April 2016 6th June 2016 19th July 2016
Vegetation indices		GCVI, GNDVI, NDVI, SR and WDRVI calculated from Sentinel-2 data	10m	
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	NEXTMap Digital Terrain Model	10m	Created using data collected in 2002 & 2003
	Aspect	Calculated using the NEXTMap DTM	10m	
	Slope		10m	

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 for equations).

Table 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 *et al.*, 2003
Green Normalised Difference Vegetation Index	GNDVI	$GNDVI = \frac{NIR - G}{NIR + G}$	Gitelson *et al.*, 1996
Normalised Difference Vegetation Index	NDVI	$NDVI = \frac{NIR - R}{NIR + R}$	Rouse *et al.*, 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 Environmental Data

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 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.

3.3.2 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 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 Topographic Variables

A 10m resolution digital terrain model (DTM) was obtained from NEXTMap Britain (table 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.

3.4 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; 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). 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 (2018), 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 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 4.

Table 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 2.

Combination	Data layers
Question 1	
S2	Sentinel-2 data
S2_20	Sentinel-2 data resampled to 20m
Question 2	
S2	Sentinel-2
S2_VI	Sentinel-2, VIs
VI	VIs
Question 3	
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
Question 4	
D	Sentinel-2 data December only
A	Sentinel-2 data April only
Jn	Sentinel-2 data June only
J	Sentinel-2 data July only
Question 5a	
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
Question 5b	
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
Question 5c	
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

3.5 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).

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

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 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.

3.6 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.

4. Results

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 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 7; figure 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 6).

Table 5: RMSE and R-squared values calculated from the validation dataset for linear and random forest regressions using vegetation indices calculated for each month.

		Linear Regression		Random Forest Regression	
Month	**VI**	**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 6: RMSE and R-squared values calculated from the validation dataset for random forest regressions using NDVI data accumulated over the growing season.

NDVI	RMSE	RSQ
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

4.2 Random Forest Model Comparison

Validation of the RF models was conducted in two ways, using the 10-fold validation from RF 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 7; figure 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 7: Results of random forest analysis.

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

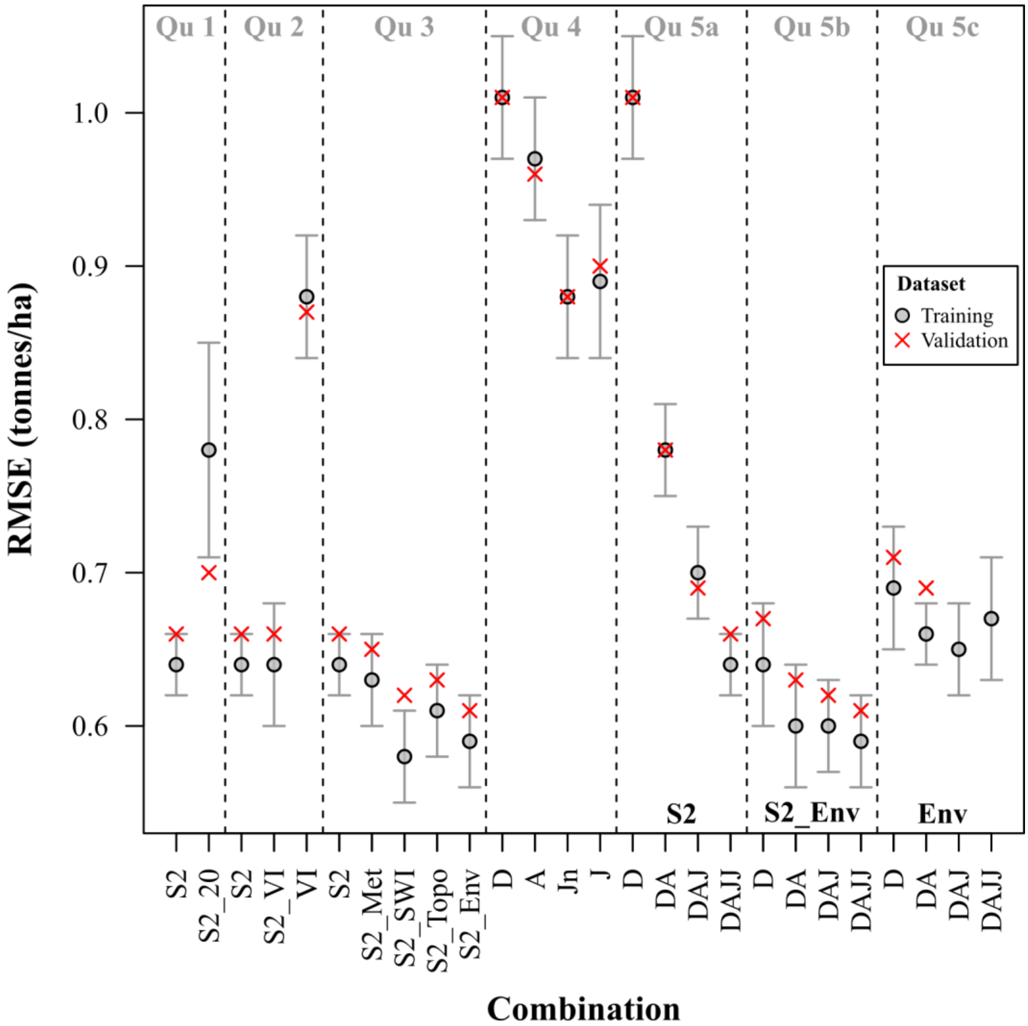

Figure 6: Ten-fold RMSE values from Random Forest analysis calculated using the training dataset and RMSE values from the validation dataset. Error bars produced using the standard deviation in ten-fold RMSE iterations. Specific data values can be found in table 7. For question 5, S2 is the Sentinel-2 only data, S2_Env is the Sentinel-2 and environmental datasets, whilst Env is just the environmental data sets (see table 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 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 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 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 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 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 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 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 ± 0.05 tonnes/ha) than the other environmental variables (RMSE 1.32-1.34 ± 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.

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 7).

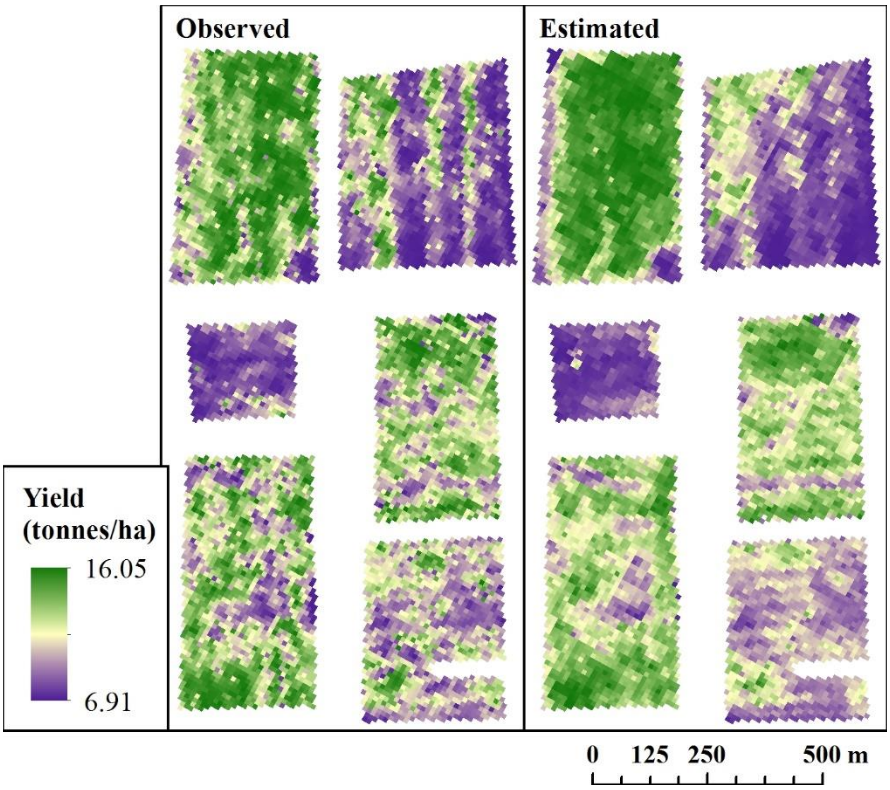

Figure 7: Observed yield interpolated from the combine harvester data (left) and estimated yield from the S2_Env RF model (right) for a selection of fields within the training area.

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 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 8)) and others more complex bimodal distributions (e.g. field 21 (figure 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 8). Additionally, the regression graph confirms the trends shown in the frequency distributions (figure 9).

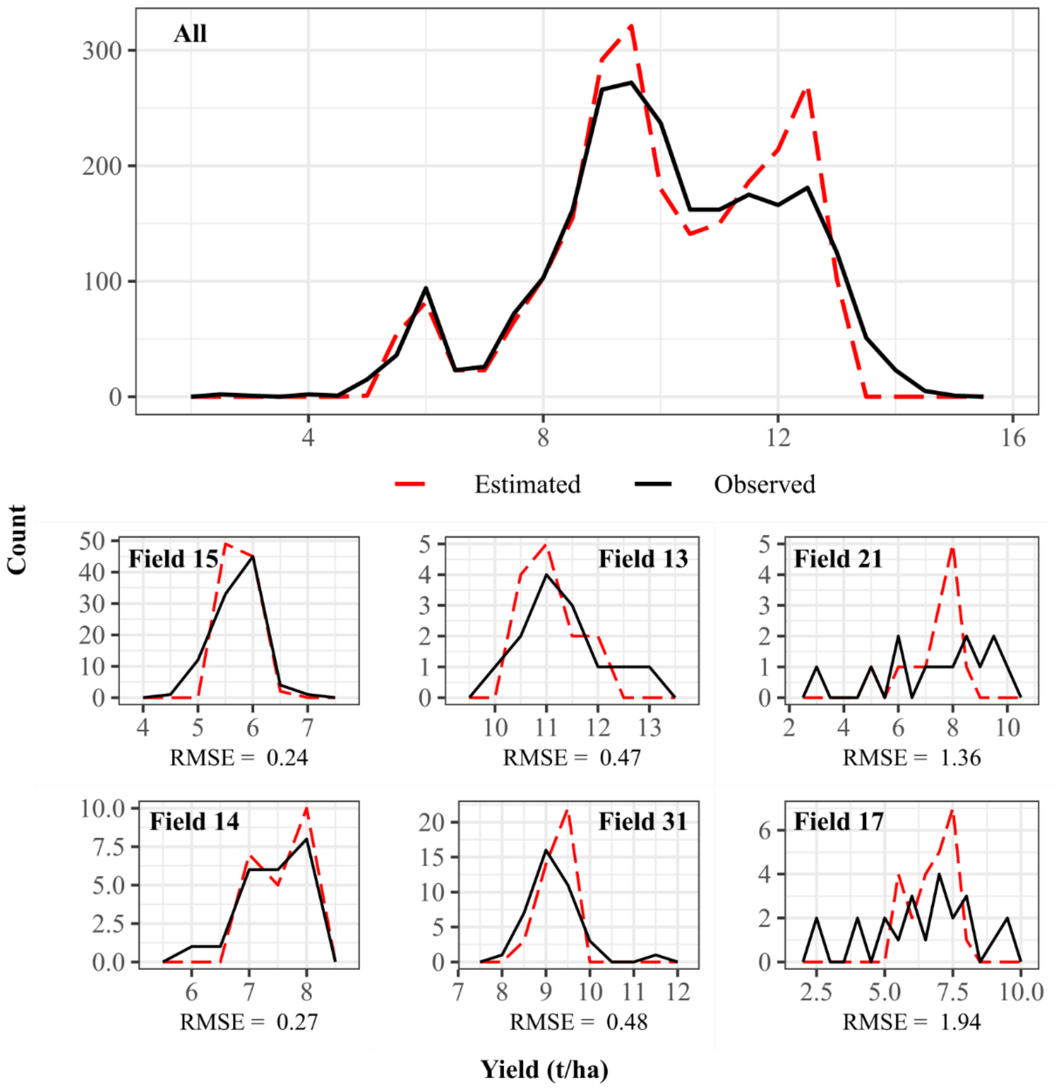

Figure 8: Frequency distributions for observed and estimated yields using the validation data set for the S2_Env model for all fields and a sample of individual fields. Individual fields chosen were those with the two highest (fields 15 and 14), two middle (fields 13 and 31), and two lowest (fields 21 and 17) RMSE values to provide a representative selection.

Table 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		

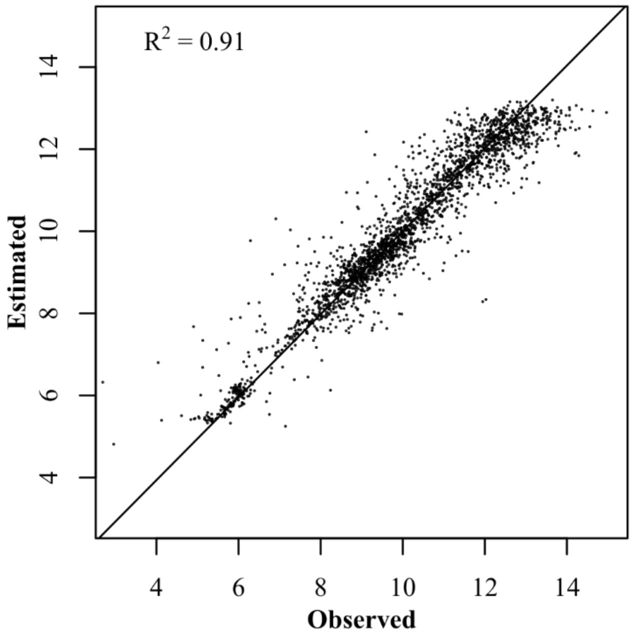

Figure 9: Linear regression between observed and estimated yield for the validation data set from the S2_Env model.

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 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. Extrapolation outside the input data range 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 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).

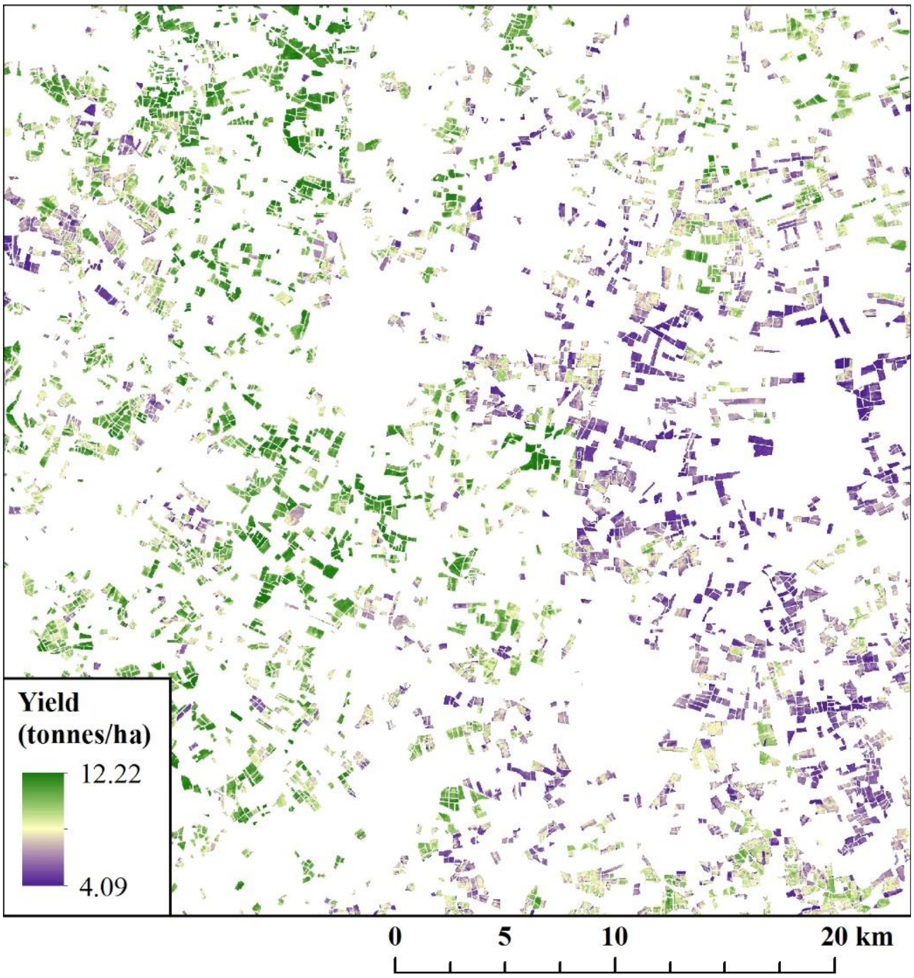

Figure 10: Landscape-scale wheat yield estimation based on S2_Env RF model.

5. Discussion

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.

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.

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 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 interpretation from Sentinel-2 data alone.

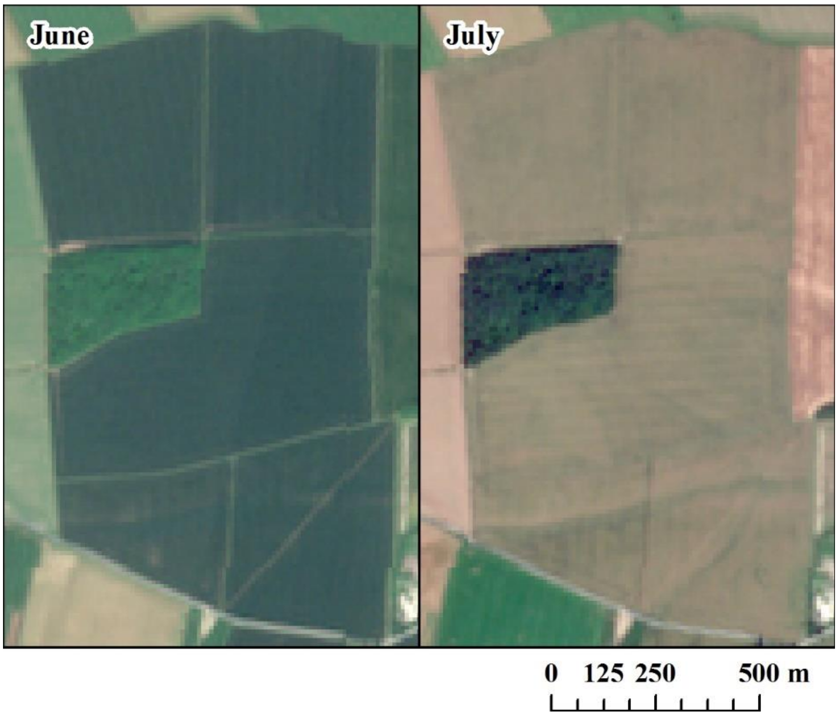

Figure 11: Evidence of crops ripening between successive Sentinel-2 images for June (left) and July (right).

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 estimates (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 predicting 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 2) to other areas, environmental conditions and crop types.

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.

Acknowledgements

We are grateful to the farmers and farming companies who allowed us access to the yield data. MH was funded by Lancaster University through a Lancaster Environment Centre PhD Studentship as part of the Graduate School for the Environment. CR, LC and JR were supported by research programme NE/N018125/1 LTS-M ASSIST - Achieving Sustainable Agricultural Systems, funded by NERC and BBSRC. Land Cover Crops Plus ® was provided by CEH (Land Cover Plus®: Crops © NERC (CEH). © RSAC. © Crown Copyright 2007, License number 10001752). We acknowledge Copernicus for provision of Sentinel data [2015-2016], processed by ESA.

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/publications/2017/january/31/exploiting-yield-maps-and-soil-management-zones.aspx

Anuta, P.E., MacDonald, R.B., 1971. Crop surveys from multiband satellite photography using digital techniques. Remote Sens. Environ. 2, 53–67. doi: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. doi:10.1080/2150704X.2012.744486

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. doi:10.1016/j.rse.2017.04.014

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. doi:10.1016/j.rse.2016.07.030

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. doi:10.1016/0168-1699(95)00049-6

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. doi:10.5721/ItJRS20114335

Breiman, L., 2001. Random Forests 1–33.

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. doi:10.1073/pnas.1616919114

CLAAS, 2018. CLAAS Telematics [WWW Document].

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S. V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161. doi: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. doi:10.1016/j.rse.2018.08.009

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.

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. doi:10.1016/j.rse.2005.03.015

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. doi:10.1016/j.rse.2011.11.026

ESA, 2018. Copernicus Open Access Hub [WWW Document]. URL https://scihub.copernicus.eu/

Everingham, Y., Sexton, J., Skocaj, D., Inman-Bamber, G., 2016. Accurate prediction of sugarcane yield using a random forest algorithm. Agron. Sustain. Dev. 36. doi:10.1007/s13593-016-0364-z

GEO, 2018. Stocktaking Overview of the G20 Global Agricultural Monitoring Initiative.

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. doi: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. doi:10.1029/2002GL016450

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. doi:10.1016/j.rse.2017.06.031

Grisso, R.D., Jasa, P.J., Schroeder, M.A., Wilcox, J.C., 2002. Yield Monitor Accuracy: Successful Farming Magazine Case Study 18, 147–151.

Horning, N., 2018. randomForestPercentCover [WWW Document]. URL https://github.com/nedhorning/RandomForestForRemoteSensing/tree/master/Downloads/randomForestPercentCover

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. doi:10.1016/j.agrformet.2015.02.001

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. doi:10.1016/j.rse.2013.07.018

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. doi: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. doi: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. doi: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. doi: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. doi: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. doi:10.1080/01431160903475357

Jurecka, F., Hlavinka, P., Lukas, V., 2016. Crop yield estimation in the field level using vegetation indices 90–95.

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. doi:10.1371/journal.pone.0157166

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. doi:10.7848/ksgpc.2016.34.4.383

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. doi:10.1109/Multi-Temp.2017.8035204

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. doi:10.1016/j.compag.2017.04.006

Liaw, a, Wiener, M., 2002. Classification and Regression by randomForest. R news 2, 18–22. doi:10.1177/154405910408300516

Lobell, D.B., 2013. Field Crops Research The use of satellite data for crop yield gap analysis. F. Crop. Res. 143, 56–64. doi:10.1016/j.fcr.2012.08.008

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. doi: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. doi:10.1016/j.inpa.2015.06.001

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. doi:10.1007/s11119-013-9336-3

Moriondo, M., Maselli, F., Bindi, M., 2007. A simple model of regional wheat yield based on NDVI data. Eur.

J. Agron. 26, 266–274. doi:10.1016/j.eja.2006.10.007

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. 8596, 1–23. doi:10.1080/14498596.2016.1220872

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. doi:10.1016/j.fcr.2012.02.025

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. doi:10.1007/s10021-005-0054-1

Refaeilzadeh, P., Tang, L., Liu, H., 2009. Cross-Validation, in: Liu, L., Özsu, M.T. (Eds.), Encyclopedia of Database Systems. Springer, Boston, MA. doi: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. doi:10.1016/j.jag.2007.11.003

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. doi:citeulike-article-id:12009708

Rowland, C.S., Morton, R.D., Carrasco, L., McShane, G., O'Neil, A.W., Wood, C.M., 2017. Land Cover Map 2015 (vector, GB).

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. doi:10.1007/BF02990002

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.

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. doi:10.3934/geosci.2017.2.163

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B.Ö., Floury, N., Brown, M., Traver, I.N., Deghaye, 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. doi:10.1016/j.rse.2011.05.028

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. doi:10.1016/S0034-4257(99)00036-X

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. doi:10.3390/rs70201461

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. doi: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. doi:10.1016/j.agrformet.2017.06.015

Yang, C., Everitt, J.H., 2002. Relationships between yield monitor data and airborne multidate multispectral digitial 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. doi: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.

List of Figure Captions

Figure 1: Location of study sites

Figure 2: Overview of the method used to estimate yield at high resolution on a landscape scale

Figure 3: Summary of the criteria for data cleaning

Figure 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

Figure 5: Example of the distribution of yield data points relative to a) 10m and b) 20m resolution interpolated yield data

Figure 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 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 4 for more details).

Figure 7: Observed yield interpolated from the combine harvester data (left) and estimated yield from the S2_Env RF model (right) for a selection of fields within the training area

Figure 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

Figure 9: Linear regression between observed and estimated yield for the validation data set from the S2_Env model

Figure 10: Landscape-scale wheat yield estimation based on S2_Env RF model

Figure 11: Evidence of crops ripening between successive Sentinel-2 images for June (left) and July (right)