CRANFIELD UNIVERSITY

NIKOLAOS-CHRISTOS VAVLAS

DERIVING CROP PRODUCTIVITY INDICATORS FROM SATELLITE SYNTHETIC APERTURE RADAR TO ASSESS WHEAT PRODUCTION AT FIELD-SCALE

SCHOOL OF WATER, ENERGY AND ENVIRONMENT

PhD

Academic Year 2021

Supervisors: T. W. Waine, G. M. Richter (Rothamsted Research)
Associate Supervisors: P. J. Burgess and J. Meersmans

September 2021

CRANFIELD UNIVERSITY

SCHOOL OF WATER, ENERGY AND ENVIRONMENT

PhD

Academic Year 2021

Nikolaos Christos Vavlas

Deriving Crop Productivity Indicators From Satellite Synthetic Aperture Radar To Assess Wheat Production At Field-scale

Supervisors: T. W. Waine, G. M. Richter (Rothamsted Research)
Associate Supervisors: P. J. Burgess and J. Meersmans
September 2021

ABSTRACT

The deployment of high-revisit satellite-based radar sensors raises the question of whether the data collected can provide quantitative information to improve agricultural productivity. This thesis aims to develop and test mathematical algorithms to describe the dynamic backscatter of high-resolution Synthetic Aperture Radar (Sentinel-1) in order to describe the development and productivity of wheat at field-scale. A time series of the backscatter ratio (VH/VV), collected over a cropping season, could be characterised by a growth and a senescence logistic curve and related to critical phases of crop development. The curve parameters, referred to as Crop Productivity Indicators (CPIs), compared well with the crop production for three years at the Rothamsted experimental farm. The combination of different parameters (e.g. midpoints of the two curves) helped to define CPIs, such as duration, that significantly ($r = 0.61$, $p = 0.05$) correlated with measured yields. Field observations were used to understand the wheat evolution by sampling canopy characteristics across the seasons. The correlation between the samples and the CPIs showed that structural changes, like biomass increase, influence the CPIs during the growth phase, and that declining plant water content was correlated with VH/VV values during maturation. The methodology was upscaled to other farms in Hertfordshire and Norfolk. The ANOVA identified significant effects ($p<0.001$) of farm management, year (weather conditions) and the interaction between soil type and year on the selected CPIs. Multilinear regression models between yields and selected CPIs displayed promising predictive power ($R^2= 0.5$) across different farms in the same year. However, these models could not explain yield differences within high-yielding farms across seasons because of the dominant effect of weather patterns on the CPIs in each year. The potential impact of the research includes estimation of yield across the landscape, phenology monitoring and indication biophysical parameters. Future work on SAR-derived CPIs should focus on improving the correlations with biophysical properties, applying of the methodology in other crops, with different soils and climates.

Keywords: Crop development, Indicators, remote sensing, Sentinel-1, wheat, temporal curves, upscale

ACKNOWLEDGEMENTS

First, I like to thank all the practitioners for the crucial yield and management information (2017- 2020) provided by collaborating farms (Dominic Swan, Catalyst Farming; George Badger, Cereal Rural) under the project “Advancing EO Applications in Agriculture” (Innovate-UK Project No: 102682) and St Pauls Warden Estate, East Hall, under the research programme NE/N018125/1 LTS-M ASSIST—Achieving Sustainable Agricultural Systems. I also want to acknowledge the funding through the Soil Agricultural Research Innovation Accelerator (AgRIA) PhD programme, between Cranfield University and Rothamsted Research.

I would like to thank all my supervisors G. M. Richter as Rothamsted’s lead scientist of the Innovate-UK project, and T. W. Waine, J. Meersmans and P. J. Burgess for their guidance and continuous discussions about the project as well as their encouragement and motivation. Thanks for the amazing four years of collaboration and discussion that I will never forget, and I can’t wait for future discussions and collaborations!

I also thank Dr Giacomo Fontanelli, for his support in the beginning of my project, giving numerous insights into the workings of Synthetic Aperture Radar (SAR), discussions and his support in the field work and encouragement and Stephen Hayward for setting up the server-based database within the Innovate-UK project mentioned above. I also thank Dr Iain Cameron (Environment Systems Ltd.; Aberystwyth) for providing pre-processed satellite images during and beyond the Innovate-UK funded project and having interesting discussions about satellite data. I am grateful to the Rothamsted statistic team and especially Andrew Mead for the insightful discussions and Lab Manager Chris Hall for his support in the lab. It was very nice to also help and collaborate with visiting students in the “Richter” team: Nico Bosatta, Pierre Laureau, Robin Comte in different years that assisted me in the fieldwork and also my fellow PhD students Xavier Albano, Timo Breure, Jonah Prout, Patricia Ortega-Ramos and Hannah McGrath for helping mentally but also with fieldwork whenever I needed help. Also, numerous other PhD students and other people that I have fun to be around all these years.

Last but not least, I would like to thank my parents, Angelis and Maria, for their unconditional support, my wife Eleni Pyloudi (getting married during this period) and now our baby Sofia-Olga Vavlas for the sacrifice they have taken so I could finish my PhD thesis with long working days and short weekends, and they are always my strength.

TABLE OF CONTENTS

ABSTRACT i
ACKNOWLEDGEMENTS iii
LIST OF FIGURES vii
LIST OF TABLES xi
LIST OF ABBREVIATIONS AND SYMBOLS 1
1 Introduction 3
1.1 Importance of spaceborne remote sensing in agriculture 3
1.2 The potential of Synthetic Aperture Radar (SAR) 4
1.3 Use of SAR in agriculture 6
1.3.1 Soil moisture content detection with radar 6
1.3.2 Radar response to vegetation 7
1.3.3 Potential applications of time series monitoring 9
1.4 Hypothesis and objectives 10
1.5 How to read the thesis 12
2 Deriving wheat crop productivity indicators using Sentinel-1 time series 14
2.1 Introduction 15
2.2 Materials and Methods 17
2.2.1 Study area 17
2.2.2 Site ground measurements of wheat development and yield 18
2.2.3 SAR data to obtain VH/VV time series in each wheat field 19
2.2.4 SAR pre-processing for field specific VH/VV time series 21
2.2.5 Temporal analysis and wheat development definition 21
2.2.6 Logistic curve fitting 22
2.2.7 Parameter optimisation 25
2.2.8 Automatic curve extraction and correlation analysis 25
2.3 Results 25
2.3.1 Annual analysis of VH/VV ratio curve parameters, 2017 to 2019 25
2.3.2 Relationship between SAR-derived parameters and Yield 31
2.4 Discussion 33
2.4.1 Technical and statistical process, the goodness of fit, robustness, uncertainty 34
2.4.2 Environmental and biophysical understanding 35
2.4.3 Opportunities for agronomic management and modelling 36
2.5 Conclusions 37
3 Biological interpretation of satellite radar data for Assessing crop development and yield Formation 38
3.1 Introduction 39
3.2 Materials and Methods 41
3.2.1 Study site 41
3.2.2 Field data and observations 42
3.2.3 Satellite data and time series 47
3.3 Results 48
3.3.1 Variability of field measurements 48

3.3.2 SAR temporal development for the different polarisations and their response based on key crop development stages 51
3.3.3 Detection of annual differences in vegetation dynamics and yield formation based on the VH/VV ratio 53
3.3.4 Dynamics of VH/VV-ratio time series on fields from the same year with different productivity 55
3.3.5 Relationship between CPIs and biophysical variables 57
3.4 Discussion 64
3.4.1 Temporal variation of SAR profiles and field observations 64
3.4.2 CPIs connection to biophysical properties of wheat 66
3.4.3 Sampling approach 67
3.5 Conclusions 71
4 Effect of management and pedoclimatic conditions on SAR derived crop productivity indicators 72
4.1 Introduction 73
4.2 Materials and methods 74
4.2.1 Study sites 74
4.2.2 Selected CPIs for the upscale 79
4.2.3 Quality assurance for the CPI outlier definition and elimination 81
4.2.4 Sensitivity of CPIs 82
4.2.5 Analysis of variance (ANOVA) for CPIs related to the biophysical environment and crop management 82
4.2.6 Multiple linear regression analysis (MLR) 83
4.3 Results 86
4.3.1 Relationship between SAR-derived CPIs and yield 86
4.3.2 Variation of key CPIs across locations 90
4.3.3 Analysis of variance for the selected CPIs 91
4.3.4 Multiple linear regression (MLR) to estimate wheat yield 92
4.4 Discussion 100
4.4.1 Landscape and management factors effects on CPIs 100
4.4.2 Combination of the CPIs to increase yield prediction 101
4.4.3 Applications of SAR-derived CPIs 102
4.5 Conclusions 104
5 SYNTHESIS 105
5.1 Deriving quantitative CPIs from SAR-backscatter dynamics 106
5.2 CPIs and biophysical properties of field and plot scale 109
5.3 Potential benefits of regional monitoring and agricultural applications 111
5.4 Research Impact 113
5.5 Recommendations for future work 115
5.6 Conclusions 118
REFERENCES 119
APPENDICES 135

LIST OF FIGURES

Figure 1-1 Different scattering mechanisms based on the surface structure (European Space Agency, 2007) 5

Figure 1-2 Representation of vertical (i.e. VV) (a), horizontal (i.e. HH) (b) and cross-polarisation (i.e. VH) (c) of the electric field vector of an electromagnetic wave. (European Space Agency, 2007) 5

Figure 1-3 the effect of the crop stage in the field that determines the surface and the dominant scattering mechanism. Graphic courtesy of Dr Giacomo Fontanelli; Poster "Satellites in Agriculture" at the 175 Anniversary Exhibition Rothamsted Research." 8

Figure 1-4 Structure of the thesis in terms of objectives and outputs of the different chapters deriving Crop Productivity Indicators (CPI) from high-frequency backscatter curves of Synthetic Aperture Radar (SAR-Cross Pol) and assigning statistics (ANOVA & Multi-Linear Regression) – Place "*" Published in Remote Sensing (Vavlas et al., 2020) 12

Figure 2-1 Google Earth image of Rothamsted Research experimental farm, with wheat field boundaries from 2017, 2018 and 2019 indicated using red lines 17

Figure 2-2 Total monthly rainfall (mm) and mean air temperatures (°C) on the primary axis (left) and maximum potential soil moisture deficit (PSMD; mm) on the secondary axis (right) at Rothamsted farm from Sep 2016 to August 2019..... 18

Figure 2-3 Flow diagram explaining the steps for the temporal analysis of SAR time series, including the data and process 20

Figure 2-4 a) Mean VH/VV ratio from Sentinel-1 orbit 132 of wheat field (for the Swayers_4 field in 2018) together with a smoothed curve of the mean (with the date of the maximum point indicated) and b) first (f'x) and second (f''x) derivative (bottom) of the smoothed field-averaged line. The date of the mid-point of the growth period, and the date of the start, mid-point and minimum value of the maturation period are indicated as vertical dotted lines. The secondary y axis in the first graph displays the VH/VV ratio on a decibel scale for reference..... 22

Figure 2-5 Mean SAR VH/VV ratio for a wheat field (Swayers_4 in 2018) with 228 pixels (1.3 ha), between January and September 2018, showing the standard deviation for each date (n = 41), and displaying 12 parameters as calculated using a smoothed line and two logistic curves (see Table 2-2). All calculations were completed using a linear scale and the secondary y axis displays the decibel scale for reference. 23

Figure 2-6 VH/VV ratio time series (blue dots) with the smoothed curve (dotted green line) and two logistic curves fitted (black line) for nine fields during 2017, orbit 132. The vertical dotted lines indicate the dates of G_midP (green), TZmax (red), inflorescence (purple), and S_midP (yellow) as determined from the VH/VV backscatter time series analysis 26

Figure 2-7 VH/VV ratio time series (blue dots) with the smoothed curve (green dashed line) and two logistic curves fitted (black line) for five fields during 2018, orbit 132. The vertical dotted lines indicate the dates of G_midP (green), TZmax (red), inflorescence (purple), and S_midP (yellow) as determined from the VH/VV backscatter time series analysis 27

Figure 2-8 VH/VV ratio time series (blue dots) with the smoothed curve (green dashed line) and two logistic curves fitted (black line) for four fields in 2019, orbit 132. The vertical dotted lines indicate the dates of G_midP (green), TZmax (red), inflorescence (purple), and S_midP (yellow) as determined from the VH/VV backscatter time series analysis 27

Figure 2-9 Correlation matrix with 2017 (green circles), 2018 (red squares) and 2019 (blue diamond) using the 17 fields. r values are colour-coded green when there was a 95% level of significance for a two-tailed test. 32

Figure 2-10 Scatter plot showing the relationship between Yield and a) Duration and b) the timing of the maximum value of VH/VV (TZmax). The r and best fit linear line for each relationship, across the three seasons, is shown 33

Figure 3-1 Google Earth image of Rothamsted Research experimental farm, with wheat field boundaries from 2018 (red) and 2019 (Blue) 41

Figure 3-2 Sample collection from Sawyers_3 field creating a W shape across the field with the total samples (blue) from different dates. The centroid (red) is calculated by the repeated blue dots assigned in each point (red). This centroid provides the area with a radius of 15 m creating plots (orange) to extract the SAR polarisations for the individual points in the field. The field is buffed 10 m to reduce the effect of the field boundaries in SAR 45

Figure 3-3 Key growth stages (GS) observed in the fields considered in this study across the growing season and the three phases, foundation (red), construction (green) and production (blue) phase. 46

Figure 3-4 Mean values and associated variability of a) the LAI at the booting stage and b) the tiller density at the maturation phase in selected fields across two seasons: 2018 (Sawyers 2, Sawyers_4, Stackyard and West Barnfield) and 2019 (Drapers, Whitehall, Gknott 2 and 3, Sawyers_3, and West Barnfield) using the five sample points per field. 49

Figure 3-5 Mean values and associated variability of dry biomass distribution between different components at a) stem elongation in April and b) before harvest across two seasons (2018, 2019) using the five sample points per field. 50

Figure 3-6 SAR backscatter polarisations of a) VH/VV, b) VV, c) VH) and field measurements of d) height, e) plant water content (PWC), and f) above ground biomass (AGB) against growth stages on the BBCH scale for the West Barnfield field for two periods: 2017-18 and 2018-19. 52

Figure 3-7 images at the ripening stage of wheat between the two seasons of the same field (West Barnfield) displaying the difference in colour, development stage and ear maturation level as 2018 matured a week earlier, is more yellow and ear is more inclined to the ground than 2019 53

Figure 3-8 a) Means of SAR backscatter polarisation ratio from Sentinel-1 for 2018 (**red**) and 2019 (**blue**) together with the smoothed lines in West Barnfield and b) field measurements for AGB (▲), c) Canopy Height (o) and d) PWC (×) for the experimental in relation to the wheat phenological development characterised below Table 3-3 - 54

Figure 3-9 Field data and temporal evolution of SAR using the average of two wheat fields in 2018, West Barnfield (blue) and Sawyers 4 (red), with a significant

difference in biomass. Field measurements for b) above ground biomass (AGB), c) canopy height, and d) plant water content together with the development stages (vertical lines), Temporal VH/VV ratio SAR smoothed with the development stages (down) 56

Figure 3-10 CPI calculation using the two logistic curve fitting for 2019 harvest season. Every column represents the same field with its five sample positions and the parenthesis displays the number of pixel used to average the VH/VV ratio..... 60

Figure 3-11 The relationship between the SAR VH/VV ratio and the canopy height for three wheat development periods: stem elongation, booting and ear emergence, and grain filling. The x and o symbols display two different fields in 2018. 65

Figure 3-12 Display of the curve signature similarity of field average value of VH/VV (hatched line) smoothed curve with five individual plots smoothed lines (P6 to P9) in the same wheat field. 69

Figure 3-13 Yield comparison between the combine harvest from the machinery as field average in t/ha (x-axis) and the hand harvest in each sample point by five points per field (y-axis) as well as the 1:1 line (black) for the two different years (2018-red, 2019-blue) 70

Figure 4-1 Google Earth images from the four farms that display fields available for wheat cultivation in each area. Rothamsted Farm, Hertfordshire (a), East Hall, Hertfordshire (b), Hydehall, Norfolk (c) and Salle Farm, Norfolk (d) 75

Figure 4-2 Monthly mean temperature (a) and monthly total precipitation (b) for Rothamsted (Roth) in Hertfordshire and Salle in Norfolk..... 78

Figure 4-3 Seasonal variations of the total precipitation across the three growing seasons in the two regions (a) Rothamsted – AL5 and (b) Reepham/Salle – NR10 using the local weather stations in the farms. 79

Figure 4-4 Presentation of important CPIs derived from the previous chapters that were used for the statistical analysis in the different farms. 80

Figure 4-5 Examples of poor curve fitting that the IQR application on the senescence CPIs flagged up the fields (red circles) that were then excluded from the analysis in Hydehall farm. 81

Figure 4-6 Correlation matrix for the 2017-2020 season using the two Norfolk farms displaying the correlation among the CPIs. The correlation values (r) are colour-coded green when there was a 95% level of significance for a two-tailed test 88

Figure 4-7 Correlation matrix for the 2017 season across the four farms displaying the correlation among the CPIs. The correlation values (r) are colour-coded green when there was a 95% level of significance for a two-tailed test 89

Figure 4-8 Actual (calibration) vs predicted values of yield (t/ha) for the four different farms in the 2016-17 growing season using the best model yield =f (Gmid, Smid, Gmax, Smax) based on the AIC criterion. The black line represents the 1:1 line and the grey is the regression with R^2=0.5. The 95% confidence limits are highlighted as faded regions. 97

Figure 4-9 Yield prediction for the four different years in Norfolk based on the 2017 four farm regression model (CS2) and comparison with the 1:1 line (black). The

regression line is displayed as blue and the 95% confidence limits are highlighted as faded regions. 98

Figure 4-10 Comparison of predicted yields against actual yields for the Norfolk region using the 2017 growing season as calibration (cyan 1:1 line) for the multiple linear regression model and the rest of the years for comparison (black lines 1:1) as well as the regression lines for each year (grey lines). The 95% confidence limits are highlighted as faded regions. model yield=f(Gmid,Smid,Gmax,Smax) 99

Figure 5-1 Example of the VH/VV temporal curves of the same field with the VH/VV ratio field average points and standard deviation, based on four different orbits (8, 81 for descending and 59, 132 for ascending) passing from the same farm in Norfolk displaying differences during the vegetation period. 107

Figure 5-2 Example of using NDVI as the base to define areas with similar characteristics that can be examined using the SAR temporal analysis. The black box shows the time of the optical image acquisition and the smoothed lines display the SAR VH/VV ratio temporal evolution on the three zones in 2018. 116

LIST OF TABLES

Table 1-1 Microwave bands description commonly used by satellite sensors. https://earthdata.nasa.gov/learn/backgrounders/what-is-sar ; highlighting C-Band of Sentinel-1 4

Table 2-1 Wheat field names and number of 10 m x 10 m pixels per field on Rothamsted farm 18

Table 2-2 Definitions of the VH/VV ratio curve parameters and the anticipated associated crop development stage (BBCH or Zadoks scale) 24

Table 2-3 Values of derived parameters (linear scale), yield data, and the root mean standard error of the growth (RMSEG) and maturation (RMSES) stages, for each of the winter wheat fields in 2017, 2018 and 2019 29

Table 2-4 Mean and standard deviation values of 12 derived parameters (linear scale) and the wheat yield for each season of the analysis 30

Table 3-1 Field names and management information from the Rothamsted experimental farm; *Nitram fertiliser with 34.5% N; 42

Table 3-2 Description of the field observation types the frequency and replicates as well as the instruments used during the fieldwork 44

Table 3-3 Key crop development stages of winter wheat of “West Barnfield” site in the two growing seasons and the final yield provided at harvest 55

Table 3-4 Comparisons of the two scales (field and plot) between the CPIs derived from a) 2018 using the mean values of VH/VV for five fields and of VH/VV values for five plots within five fields, and b) 2019 using the mean values of VH/VV for six fields and of VH/VV for five plots (15 m radius circle, approximate 7 pixels) within the six fields , by displaying the mean, standard deviation (SD) and coefficient of variation (CV). The yield measurement for the field was derived from the combine harvester and for the within field samples, the yields was derived from the hand harvesting of individual plants before harvest. N is number of observations and A is the mean area (ha) of the sampling unit. 58

Table 3-5 Levels of significant difference: $0.05<p<0.1(^{o})$, $0.01<p<0.05(^{*})$, $0.001<p<0.01$ $(^{**})$ and $p<0.001$ $(^{***})$ between field and plot scale in each year and between the two years on the same scale 58

Table 3-6 Correlation coefficient (r) between CPIs and field observations (50 points in total) using both growing seasons (2018 & 2019), significant r (P level>0.05) coloured green and red. details of each indicator definition can be found in Table 2-2. 63

Table 4-1 Total number of wheat fields provided for each season and farm before quality assurance. (* no yield data available); proposed boxes mark the subsample in the analysis (two case studies) 75

Table 4-2 Different groups of residual soil N effect (https://www.fas.scot/downloads/tn731-nitrogen-recommendations-for-cereals-oilseed-rape-and-potatoes/) 76

Table 4-3 Number of fields for each farm based on the residual soil N group 76

Table 4-4 Number of fields per farm and soil characteristics (Cranfield University, 2021) across the four selected farms with wheat as a crop. 77

Table 4-5 Soil properties (average) describing the dominant soil series under arable management in each of the four farms (Cranfield University, 2021) 77

Table 4-6 Selected CPIs definitions that displayed contribution towards biophysical properties of wheat taken from previous chapters 80

Table 4-7 Information about the different factors considered in the ANOVA based on the data available from the selected farms. 83

Table 4-8 Mean and standard deviation (SD) of the important CPIs from the four different farms during 2017 to compare the regional effects on CPIs* and yield data.) TZmax = booting stage, G_midP = midpoint of growth, S_midP = midpoint of senescence, G_max = biomass proxy at the end of the construction phase, Smax = water proxy at the start of maturation (see Table 4-6) 90

Table 4-9 ANOVA results using three harvest years (2017-2019) for the four selected farms, testing the influence of annual meteorology (Year), management (Farm), residual nitrogen (Pr_crop_N), Soil type and the interactions of soil season and nitrogen residual and soil on the important CPIs at four different levels of significance: $0.05<p<0.1(^{o})$, $0.01<p<0.05(^{*})$, $0.001<p<0.01$ $(^{**})$ and $p<0.001$ $(^{***})$ (number of samples =210) 92

Table 4-10 Multiple linear regression results for the three main models at four different levels of significance: $0.05<p<0.1(^{o})$, $0.01<p<0.05(^{*})$, $0.001<p<0.01$ $(^{**})$ and $p<0.001$ $(^{***})$ for the Norfolk farms in years 2017-2020 93

Table 4-11 Results of multiple linear regression for the three main models at four different levels of significance: 0.05<p<0.1(o), $0.01<p<0.05(^{*})$, $0.001<p<0.01$ $(^{**})$ and $p<0.001$ $(^{***})$ for the four farms during the 2017 growing season (CS2). 94

Table 4-12 Adjusted coefficient of determination R^2_{adj} and AIC of the multiple linear regressions of selected variables relative to wheat yield across four farms in one season and across four seasons in two farms in Norfolk using the important CPIs (Duration, TZmax=booting stage, Gmid=midpoint of growth, Smid=midpoint of senescence, Gmax=biomass proxy at the end of the construction phase, Smax=water proxy at the start of maturation). In bold the best model in each case 96

Table 4-13 RMSE and Pbias for the different years in Norfolk farm based on re-calibration of the individual year (different rows) of the best-selected formula. The calibration data is shown in italic. 100

LIST OF ABBREVIATIONS AND SYMBOLS

AGB	Above Ground Biomass
AIC	Akaike information criterion
ANOVA	Analysis of variance
CPI	Crop Productivity Indicator
CV	Coefficient of variation
D_base	Tillering backscatter
D_max	Structure change (Inflorescence)
DM	Dry Matter
DOY	Day of the year
Duration	Duration of “full” vegetation to maturation
EO	Earth Observation
ESA	European Space Agency
G_base	Baseline value for the growth stage
G_max	Max value for growth stage
G_midP	Time of midpoint of growth period (t_0, G)
G_steep	Steepness of logistic curve for growth period
IQR	Inter quartile range
LAI	Leaf Area Index
MLR	Multi linear regression
PWC	Plant water content
RMSE	Root Mean Square Error
RS	Remote sensing
S_base	Baseline value at the end of the season
S_max	Value at the start of grain filling
S_midP	Time of midpoint of maturation (t_0, S)
S_steep	Steepness of logistic curve for maturation period
SAR	Synthetic Aperture Radar
SMC	Soil moisture content
TDR	Time domain reflectometry
TZmax	Time of maximum point
UAV	Unmanned aerial vehicles
γ°	Backscatter (gamma nought)
σ°	Backscatter (sigma nought)

1 INTRODUCTION

1.1 Importance of spaceborne remote sensing in agriculture

The increase of the global population and rising demand for food creates a threat to food security (FAO, 2011). New strategies are necessary to deal with two important challenges to sustainable agricultural development: (i) closing the yield gap (difference between attainable and actual production) (Rayfuse and Weisfelt, 2012) and (ii) avoiding long term soil degradation (Giannakis et al., 2017). One tool that can enhance our understanding of the impact of agricultural practices and assist in developing early warning systems of low yield is the use of remote sensing (RS) to monitor crops (Jin et al., 2018; Veloso et al., 2017). Remote sensing is the science and the art of acquiring information about an object by observing it from a distance (Fischer, Hemphill and Kover, 1976). Sensors attached to different platforms, such as handheld devices, Unmanned Aerial Vehicles (UAVs), and aeroplanes or satellites, can acquire data remotely (Fischer, Hemphill and Kover, 1976). Satellite data have the potential to support farming decisions by gathering and displaying environmental variables across the landscape (Earl et al., 2003). The close connection between mainly optical RS and precision farming is possible because crops show physical responses to environmental variables, like water or nitrogen stress (Blaes et al., 2006; Casanova, Judge and Miyoung Jang, 2007; van Emmerik et al., 2015), influenced by time and location (Vicente-Guijalba, Martinez-Marin and Lopez-Sanchez, 2014).

The agriculture sector can benefit from spaceborne RS data using different types of satellite-derived products for different applications (Kansakar and Hossain, 2016; Steele-Dunne et al., 2017; De Sy et al., 2012). Optical data are commonly used for vegetation analysis, as the spectral properties of a leaf are determined in large part, by its chemical composition, such as photosynthetically active pigments (e.g. chlorophyll) (Behrenfeld and Falkowski, 1997; Pinker and Laszlo, 1992). As leaves senesce, green chlorophyll pigments decrease, and other pigments dominate, these differences are then exploited for species discrimination or crop monitoring, but in areas characterised by high cloud cover, radar (microwave) based observations have the ability to penetrate the clouds and, due to the nature of the signal, are unaffected by day/night acquisition (Bush and Ulaby, 1978). Synthetic Aperture Radar (SAR) can contribute to the crop monitoring framework, although its potential has not been fully established (Friesen, Steele-Dunne and van de Giesen, 2012; Paget, Long and Madsen, 2016; Steele-Dunne et al., 2017). Considering that radar is sensitive to the surface structure and the dielectric properties, it can be used as a quantification tool for the canopy properties and moisture. Crop identification and phenological development can be identified by high temporal frequency SAR backscatter data (Bargiel, 2017) and parameters such as, for

example, the growth duration, an important factor for yield formation (Veloso et al., 2017). The research is particularly timely because spaceborne data have become easily accessible from the European Space Agency (ESA) with the Sentinel satellites, (Aschbacher and Milagro-Pérez, 2012), which has led to an increase of interest in the use of RS applications in agriculture.

1.2 The potential of Synthetic Aperture Radar (SAR)

SAR is an active sensor operating in the microwave spectrum domain (1 mm < λ < 1 m), able to produce an image of a target. Radar and SAR are active devices that provide their energy source and illumination: they transmit an electromagnetic wave towards a target and then measure the energy scattered back. The bands are different wavelengths of SAR designated as K, X, C, L, and P. Table 1-1 notes how the wavelength, frequency and different applications can vary with eight bands.

Table 1-1 Microwave bands description commonly used by satellite sensors. https://earthdata.nasa.gov/learn/backgrounders/what-is-sar ; highlighting C-Band of Sentinel-1

Band	Wavelength	Frequency	Typical Application
Ka	1.1–0.8 cm	27–40 GHz	Rarely used for SAR (airport surveillance)
K	1.7–1.1 cm	18–27 GHz	rarely used (H_2O absorption)
Ku	2.4–1.7 cm	12–18 GHz	rarely used for SAR (satellite altimetry)
X	3.8–2.4 cm	8–12 GHz	urban monitoring, ice and snow, little penetration into vegetation cover; fast coherence decay in vegetated areas
C	**7.5–3.8 cm**	**4–8 GHz**	**SAR Workhorse (global mapping; change detection; monitoring of areas with low to moderate penetration; higher coherence); ice, maritime ocean navigation**
S	15–7.5 cm	2–4 GHz	agriculture monitoring (NISAR will carry an S-band channel; expends C-band applications to higher vegetation density)
L	30–15 cm	1–2 GHz	geophysical monitoring; biomass and vegetation mapping; high penetration, InSAR
P	100–30 cm	0.3–1 GHz	Biomass. First p-band spaceborne SAR will be launched ~2020; vegetation mapping.

When an electromagnetic wave emitted by a transmitter arrives on a rough surface, energy will be scattered in all directions with different intensities (Figure 1-1). In the case of monostatic systems (when both the transmitter and receiver antennas are located on the same platform), the measure of energy scattered back toward the receiver is called backscatter (σ°). Imaging

radar can transmit vertical (V) or horizontal (H) electric field vectors and receive either vertical or horizontal signals or both (European Space Agency, 2007).

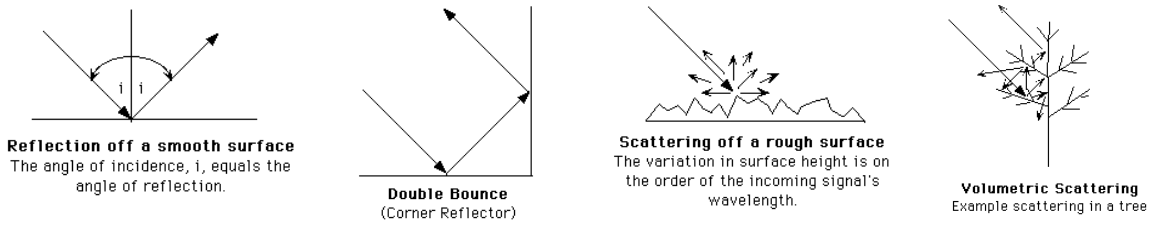


Figure 1-1 Different scattering mechanisms based on the surface structure (European Space Agency, 2007)

In Europe, Sentinel 1 C-band (5.405 GHz) SAR instruments operate in dual polarisation (VV+VH), implemented through one vertical transmit V polarised and two parallel receive vectors for H and V polarisation over the land (Figure 1-2) (European Space Agency, 2007; Østergaard et al., 2011). Thanks to the two satellites orbiting in near-polar, sun-synchronous orbits, it is possible to benefit from the 6-days return time of satellites on the same orbit, leading to a widespread observation of vegetation growth (Torres et al., 2012).

VV: vertical plane transmitted and received (like-polarized)

VH: vertical plane transmitted, and horizontal plane received (cross-polarized)

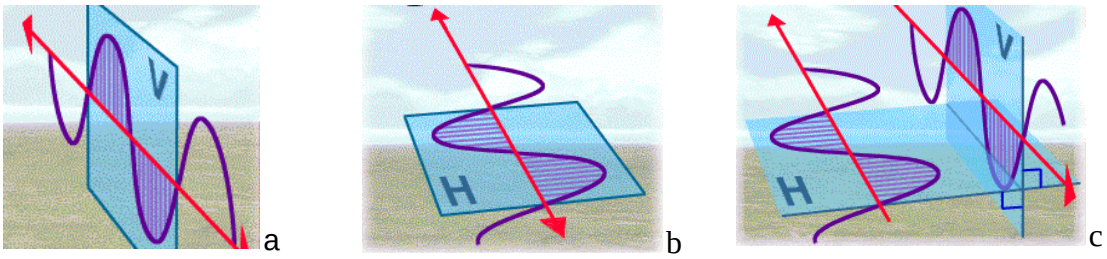


Figure 1-2 Representation of vertical (i.e. VV) (a), horizontal (i.e. HH) (b) and cross-polarisation (i.e. VH) (c) of the electric field vector of an electromagnetic wave with red arrows as direction of transmitting downwards and receiving upwards signal. (European Space Agency, 2007)

The radar backscatter carries the information about the surface of the object. This reflection is driven by radar parameters (wavelength or frequency, Incidence angle) and surface parameters (Dielectric constant, Surface roughness). The high frequency in data acquisition, the improved spatial resolution of Sentinel-1, the ability to penetrate clouds, and the free availability from the ESA are the main reasons these SAR data have been selected as a focus for this thesis.

The physical process responsible for the VV is direct surface reflection and surface or volume scattering. The VH is associated with multiple scattering because of volume scattering or surface roughness and is usually weaker (Figure 1-1). When the electromagnetic wave scatters from a object, the way the wave is polarised can change. The scattering mechanism and the returns from different surfaces vary with the incidence angle of the electromagnetic ray (Brown et al., 2003; Huang et al., 2015; Mladenova et al., 2013).

The amplitude component of the SAR signal is affected by speckle, which is commonly and incorrectly considered a noise that inherently exists in the backscatter and reduces the quality of the radar. Speckle in SAR is caused by coherent processing of the signals from multiple distributed objects, which is generally severe, causing difficulties for image interpretation. Several different methods are used to reduce the effect of speckle, based on different models describing the phenomenon (Kanevsky, 2008). For example, the multiple-look processing averages out the speckle (incoherent average) using several "looks" at the object in a single radar sweep (Franceschetti and Lanari, 1999).

1.3 Use of SAR in agriculture

1.3.1 Soil moisture content detection with radar

Soil moisture content (SMC) is an important variable affecting land surface atmospheric interactions in fields such as hydrology and agronomy. The microwave region is especially attractive for the remote sensing of SMC since the presence of water in soil strongly affects the microwave return, and at the same time, the return is not directly influenced by the chemical and mineralogical properties of the soil material (Ulaby et al., 1978; Lundien, 1971; Edgerton et al., 1971).

The sensitivity of microwave signals to the SMC is dependent on the effect of water on the dielectric constant, as has been established in multiple studies (Ulaby et al., 1986; Wang et al., 1980; Jackson et al., 1997). Scattering of the radar signal from a surface, however, depends on its overall physical nature. For example, on bare agricultural soil, the backscatter mainly depends on soil moisture and surface roughness (Figure 1-3) in addition to the sensor configuration as electromagnetic wavelength, incidence angle and polarisation (Zribi et al., 2007; Baghdadi et al., 2007: Pasolli et al., 2011). The soil moisture spatial variability has been investigated using ground observations and remote sensing measurements. Reynolds (1970) classified the source of moisture variability into static (e.g., soil characteristics and local topography) and dynamic (e.g., weather patterns and vegetation development) components. The water sensitivity of microwave signal is examined for different applications as research

has been done on soil moisture (Alexakis et al., 2017; Choker et al., 2017; Gherboudj et al., 2011) and irrigation applications at the field (Bastiaanssen, Molden and Makin, 2000; Gao et al., 2018) and landscape scale (Petropoulos et al., 2018).

Soil roughness (irregularities of the surfaces) is caused by various factors, such as land management, rock fragments, as well as soil texture and aggregates. Soil surface roughness affects the water retention, infiltration, overland flow and ultimately erosion and sediment transport (Amoah, Amatya and Nnaji, 2013). It is also relevant for measuring SMC using SAR σ° data, as it affects the scattering mechanisms (Alvarez-Mozos et al., 2006; Bousbih et al., 2017; Rahman et al., 2008; Sahebi, Bonn and Gwyn, 2003; Srivastava et al., 2003). It is important for the development of a methodology that aims to describe vegetation dynamics to understand the effect of soil moisture and roughness in the scattering mechanism. It is crucial to consider this effect when low vegetation covers the field's surface, and the signal variation is strongly affected by changes in moisture and soil roughness (Harfenmeister, Spengler and Weltzien, 2019).

1.3.2 Radar response to vegetation

In the microwave region of the electromagnetic spectrum, the wavelength affects the intensity of the penetrating energy and scattering by the vegetation together with the canopy architecture and plant water content (Table 1-1). Backscatter is affected by the orientation, size and shape of vegetation components (leaves, stalks, fruit), the crop dielectric properties (Alexakis et al., 2017; Steele-Dunne et al., 2017; Vreugdenhil et al., 2018), and the cropping characteristics (Harfenmeister, Spengler and Weltzien, 2019; Mattia et al., 2003), such as plant density and biomass. A crop canopy will interact as a group of volume scatterers. At the same time, soil contributes on the surface scattering of the energy that penetrated the crop (Figure 1-1,Figure 1-3). The main scattering mechanisms in vegetation displayed in Figure 1-1 are a) Single bounce, b) double-bounce and c) volume scattering (European Space Agency, 2007).

Vegetation geometry includes both the macrostructure and microstructure of a canopy. The first describes the canopy's height and the number of plants per unit area, and the latter refers to the stem and leaves. Canopy components with similar size or larger that the radar wavelength are expected to affect the scattering in contrast to the smaller in size that attenuate the energy (McNairn et al., 2009a). The physical variables that affect σ° are the structure of the plants, plant density, row orientation and the plant water content (PWC), which is linked to the dielectric constant. PWC changes with crop phenology stage and impacts on the mixture of direct and volumetric backscattering mechanisms. It is expected that the lower PWC and

reduced canopy would allow greater penetration once senescence begins and the soil will interact more with the backscatter (Figure 1-3).

Among the polarisations, cross-polarisations (both HV and VH) have been found to be efficient for crop and soil conditions monitoring and most affected by changes in structure with the crop life cycle (Moran et al., 2012; Inoue et al., 2014). The backscatter ratio VH/VV that is calculated based on the two polarisations could be used to assess the dynamics of crops because the value is relatively insensitive to soil moisture (Fieuzal, Baup and Marais-Sicre, 2013; Khabbazan et al., 2019; Veloso et al., 2017; Vreugdenhil et al., 2018). In the literature so far, there is lack of quantitative analyses using the VH/VV ratio to quantify changes in the crop dynamics. Veloso et al. (2017) displayed the importance of using the VH/VV ratio for biophysical parameter retrieval such as biomass and identifying changes in canopy structure that can be connected with the development stages and need further investigation as it showed great potential. Similar conclusions were derived about the use of the ratio and time series analysis relationship with biomass, canopy height and plant water content in different crop growth stages on wheat, canola, corn and oilseed rape (Canisius et al., 2018; Harfenmeister, Spengler and Weltzien, 2019; Vreugdenhil et al., 2018)

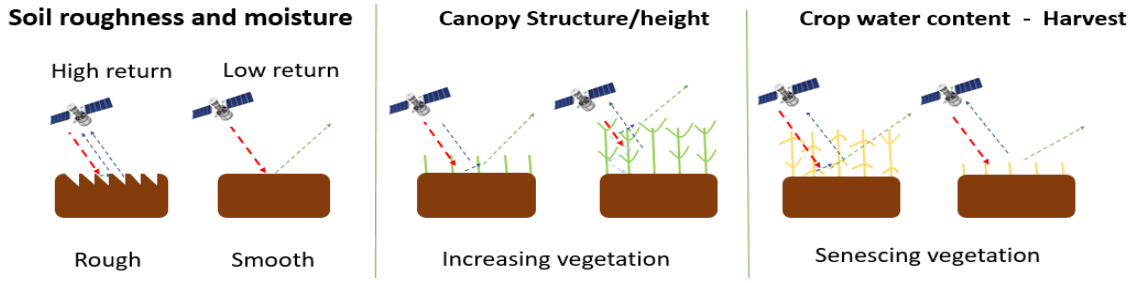


Figure 1-3 the effect of the crop stage in the field that determines the surface and the dominant scattering mechanism. Graphic courtesy of Dr Giacomo Fontanelli; Poster “Satellites in Agriculture” at the 175 Anniversary Exhibition Rothamsted Research.”

SAR can therefore be used to assess the biophysical changes of the agricultural landscape (Ndikumana et al., 2018) through the whole growing season, as it is sensitive to the structure and biomass of the crop (Harfenmeister, Spengler and Weltzien, 2019; Mattia et al., 2003). In combination with the sensitivity of radar to the water in the surface (Alexakis et al., 2017; Steele-Dunne et al., 2017; Vreugdenhil et al., 2018), SAR can be used to quantify the moisture and roughness of the field surface (Betbeder, Fieuzal and Baup, 2016; Dobson et al., 1985; Mironov et al., 2004; Snapir et al., 2018). These properties will vary due to crop type that have been used in agriculture applications like crop identification (Bargiel, 2017; McNairn et al., 2009a; Van Tricht et al., 2018). In literature, there are cases of using both SAR and optical images to increase the accuracy of crop recognition (De Bernardis et al., 2016; Fieuzal, Baup

and Marais-Sicre, 2013; McNairn et al., 2009a; Zhou et al., 2017). In other cases, remote sensing time series were used to derive statistics (e.g. variance, average) to identify irrigation in the landscape (Gao et al., 2018) or models to simulate the backscatter penetration with crop and soil conditions (Attema and Ulaby, 1978; Baghdadi et al., 2017; Bériaux et al., 2015; Li and Wang, 2018). Different microwave bands (L- C- and X-band) have also been studied to assess vegetation dynamics (Brown et al., 2003; Huang et al., 2015; Snapir et al., 2018). The availability of high-frequency SAR data through Sentinel-1 mean that there is the potential to apply it to derive development stages of the crop (Bargiel, 2017; De Bernardis et al., 2016). Together with crop modelling, such information can be used under a data assimilation scheme to improve crop development simulations (Huang et al., 2019; Jin et al., 2018; Lahoz and Schneider, 2014). In conclusion, the question arises whether the form, e.g. the timing of the rise and fall and the height or intensity of the monitoring curves (VH/VV) has a quantitative meaning.

For our indicator crop, winter wheat, AR can recognise the following field characteristics:

- Soil surface roughness and soil water content
- Crop development, type and biomass
- Crop water status, drought stress and onset of maturation

Some questions arise about the importance and relevance of crop density and height, the latter having a generally negative effect on grain yield (as it lowers the harvest index).

The open questions from state of the art at the beginning of this project were firstly whether the dynamics of the SAR backscatter and especially the VH/VV, would result in a consistent and robust parameter set that related to the productivity (yield) of our indicator crop, winter wheat. The second question was whether these mathematical curve parameters could be consistently related to the phenological and biophysical properties of the crop, and the third question was whether this methodology could be transferred to other areas with different pedo-climatic conditions and crop management.

1.3.3 Potential applications of time series monitoring

Time series analysis of satellite-based high-frequency SAR can provide early warnings of low yields as images allows detailed crop life cycle monitoring, contributing to an improved understanding of the impact of agricultural practices, and (Jin et al., 2018; Veloso et al., 2017). There is a wide range of remote sensing data and associate products (Kansakar and Hossain, 2016; De Sy et al., 2012), and the most appropriate form will depend on the challenge such as understanding water dynamics (stress and irrigation) use for crop classification and vegetation modelling (Steele-Dunne et al., 2017).

Data can also support decisions in farms due to gathering and displaying environmental variables across large areas, providing spatial information (Kasampalis et al., 2018; Steele-Dunne et al., 2017). Crop modelling is affected positively with the increase of frequency and availability of satellite data that can support landscape simulations (Betbeder, Fieuzal and Baup, 2016; Huang et al., 2019; Kasampalis et al., 2018) to improve regional production estimates. However, in most cases, the use of satellite-based data is constrained by the correlation of RS and vegetation growth that varies with the environment (Blaes et al., 2006; Casanova, Judge and Miyoung Jang, 2007; van Emmerik et al., 2015), and with time and location (Vicente-Guijalba, Martinez-Marin and Lopez-Sanchez, 2014).

Water status and plant development are two yield formation components, which are be detectable using SAR. Therefore, it is vital to relate the dynamic rise and fall of the backscatter components, particularly the cross-polarisation backscatter ratio (VH/VV), to the phenological development and plant water status measured in ground measurements.

1.4 Hypothesis and objectives

In summary, this short review points to a clear relationship between SAR and crop development, which is used to identify crops, and the water status of the soil and the crop. The recent release of high-frequency SAR data presents the unique opportunity to explore the quantitative potential of those data. The apparent knowledge gap beyond the crop identification based on phenology is the linkage between phenology-related biomass and the water status of the crop with the process of yield formation.

Therefore, this research aims to use these new high-frequency, high-resolution satellites to derive quantitative parameters from SAR cross-polarisation backscatter data to evaluate crop productivity across different agricultural landscapes.

This thesis, therefore, test the overall hypotheses that:

Firstly, vegetation growth and water dynamics can be monitored using SAR backscatter and secondly, the parameters derived from the VH/VV backscatter ratio curve are significantly related to crop performance.

Specifically, the hypothesis is that the combination of crop establishment and early biomass growth and the onset of senescence and maturation can be related to detailed field observations and upscaled to regional yield monitoring.

The above aim and hypotheses are addressed through the following objectives

1. Develop a robust quantitative approach using SAR remote sensing data to define Crop productivity Indicators (CPI) for winter wheat. (**Chapter 2**)
2. Explain how SAR-derived crop productivity indicators relate to field measurements at the field and plot scale across two different seasons. (**Chapter 3)**
3. Explain how SAR-derived crop productivity indicators relate to inter-regional and inter-annual variability in yield. (**Chapter 4)**
4. Synthesise how crop productivity indicators can be used across scales (plot-field-farm) to improve our understanding of regional crop monitoring. (**Chapter 5)**

1.5 How to read the thesis

In my PhD project, based on the above presented current state of the art (Chapter 1), I used the high temporal frequency SAR backscatter data from Sentinel-1, to test the overall hypothesis that the dynamic of the SAR smoothed VH/VV curve is indicative for the state of crop development and water status, and eventually its productivity. Figure 1-4 below displays the thesis structure with all the steps taken to complete the objectives.

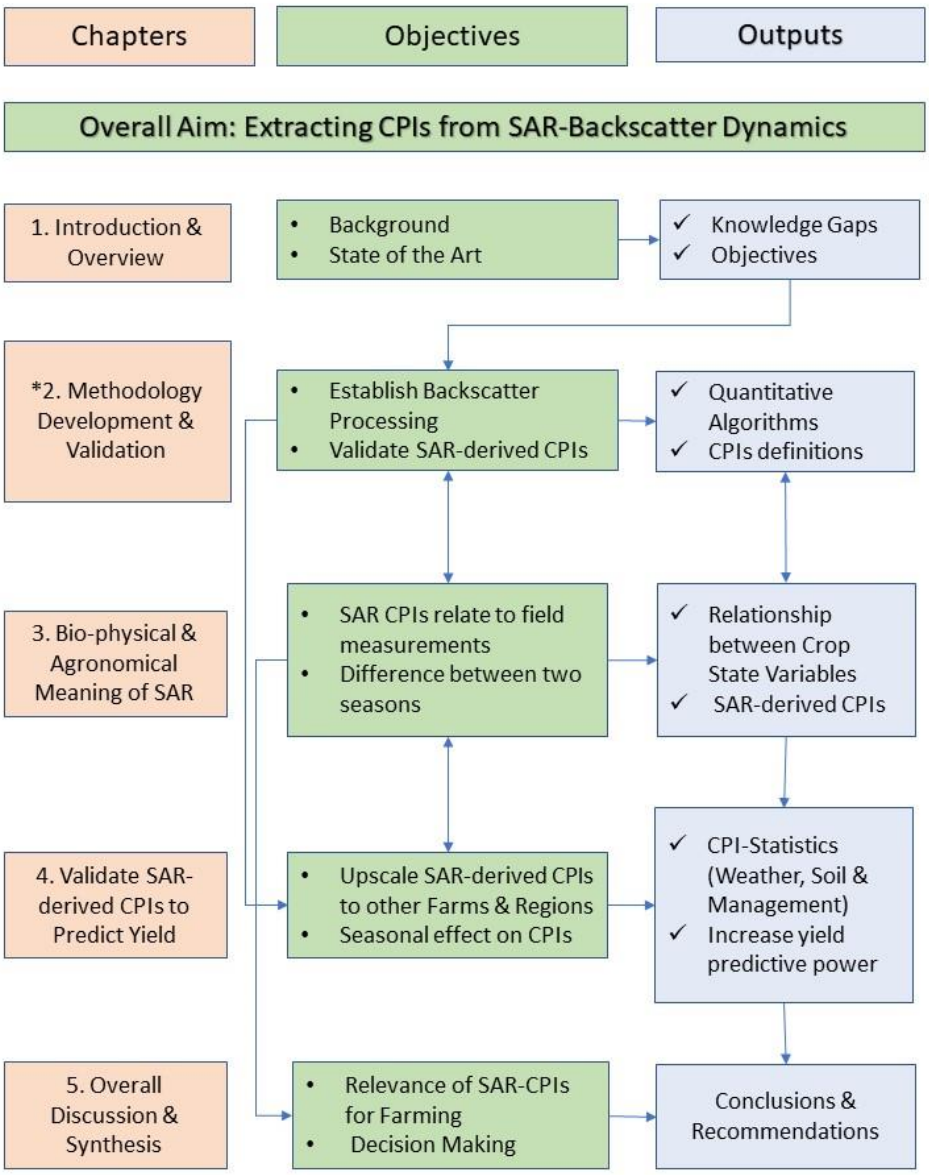


Figure 1-4 Structure of the thesis in terms of objectives and outputs of the different chapters deriving Crop Productivity Indicators (CPI) from high-frequency backscatter curves of Synthetic Aperture Radar (SAR-Cross Pol) and assigning statistics (ANOVA & Multi-Linear Regression) – Place "*" Published in Remote Sensing (Vavlas et al., 2020)

Chapter 2 described the first step, the parameterisation of the dynamic SAR backscatter data using a quantitative curve-fitting approach to characterise unique features of the temporal evolution of the SAR backscatter and how this relates to the wheat crop; the specific crop selected for this research. The temporal analysis of the dynamic curve using logistic curve fitting is at the core of the CPIs extraction procedure to define the timing and magnitude of the signal response for every field included in the analysis. Statistical analysis will be essential to describe the influence of the noise coming from the satellite observation as well as the different conditions of individual fields that may affect the temporal evolution of the SAR curve (drought).

The next step (Chapter 3) is the analysis of the SAR sensitivity to the crop development and biophysical properties, correlating the satellite-derived parameters with *in situ* measurements. Fieldwork was completed for 2018 and 2019 on the Rothamsted Research farm that shows sensitivity with the satellite products. Fieldwork data collected include canopy height, Leaf area index (LAI), soil moisture content (SMC%) (using TDR), and biomass (Fresh and Dry) with plant water content (PWC%) also calculated.

The objective of the fourth part of the thesis (Chapter 4) is the correlation of yield with the CPIs calculated using the methodology on Chapter 2 to observe the SAR parameter sensitivity to field production in other areas. At the same time, soil properties such as soil type are provided by soil maps. A multilinear regression approach will be examined to increase the predictive power of the yield using the important CPIs defined.

In the overall discussion (Chapter 5), I analyse the practical relevance of SAR-derived CPIs for farm-scale decision-making compared to their potential usage in regional yield monitoring. We also discuss whether the proposed method could be turned into an operational tool to give early and near-harvest productivity projections.

2 DERIVING WHEAT CROP PRODUCTIVITY INDICATORS USING SENTINEL-1 TIME SERIES

Published paper:

Vavlas, N., Waine, T.W., Meersmans, J., Burgess, P.J., Fontanelli, G. and Richter, G.M. (2020) 'Deriving Wheat Crop Productivity Indicators Using Sentinel-1 Time Series', Remote Sensing, 12(15), p. 2385

Author Contributions: Conceptualization, N.-C.Vavlas., T. W. Waine., J. Meersmans, and G.M. Richter.; methodology, N.-C. Vavlas; software, N.-C. Vavlas; formal analysis, N.-C. Vavlas; data curation, N.-C. Vavlas and G. Fontanelli.; writing original draft preparation, N.-C. Vavlas; reviewing, T.W. Waine., J. Meersmans, P.J. Burgess., G. Fontanelli, and G.M. Richter; visualization, N.-C. Vavlas. All authors have read and agreed to the published version of the manuscript.

Summary

High-frequency Earth observation (EO) data have been shown to be effective in identifying crops and monitoring their development. The purpose of this paper is to derive quantitative indicators of crop productivity using synthetic aperture radar (SAR). This study shows that the field-specific SAR time series can be used to characterise growth and maturation periods and to estimate the performance of cereals. Winter wheat fields on the Rothamsted Research farm in Harpenden (UK) were selected for the analysis during three cropping seasons (2017 to 2019). Average SAR backscatter from Sentinel-1 satellites was extracted for each field and temporal analysis was applied to the backscatter cross-polarisation ratio (VH/VV). The calculation of the different curve parameters during the growing period involves (i) fitting of two logistic curves to the dynamics of the SAR time series, which describe timing and intensity of growth and maturation, respectively; (ii) plotting the associated first and second derivative in order to assist the determination of key stages in the crop development; and (iii) exploring the correlation matrix for the derived indicators and their predictive power for yield. The results show that the day of the year of the maximum VH/VV value was negatively correlated with yield ($r = -0.56$), and the duration of "full" vegetation was positively correlated with yield ($r = 0.61$). Significant seasonal variation in the timing of peak vegetation ($p = 0.042$), the midpoint of growth ($p = 0.037$), the duration of the growing season ($p = 0.039$) and yield ($p = 0.016$) were observed and were consistent with observations of crop phenology. Further research is required to obtain a more detailed picture of the uncertainty of the presented novel methodology, as well as its validity across a wider range of agroecosystems.

2.1 Introduction

Time series analysis of satellite remote sensing (RS) images allows detailed crop monitoring, an enhanced understanding of the impact of agricultural practices, and can provide early warnings of low yields (Jin et al., 2018; Veloso et al., 2017). There is a wide range of RS data and associate products (Kansakar and Hossain, 2016; De Sy et al., 2012), and the most appropriate form will depend on the challenge such as crop classification, vegetation modelling, and understanding water dynamics such as drought stress and irrigation (Steele-Dunne et al., 2017). Data can also support farming decisions due to the ability to gather and display environmental variables across large areas, providing spatial information (Kasampalis et al., 2018; Steele-Dunne et al., 2017). The increasing availability and frequency of satellite data also support the application of RS in the context of crop modelling (Betbeder, Fieuzal and Baup, 2016; Huang et al., 2019; Kasampalis et al., 2018) to improve regional production estimates. However, in most cases, the use of RS for crop management remains limited as the relationship between the data and crop development and growth varies with the environment (Blaes et al., 2006; Casanova, Judge and Miyoung Jang, 2007; van Emmerik et al., 2015), and with time and location (Vicente-Guijalba, Martinez-Marin and Lopez-Sanchez, 2014).

Synthetic Aperture Radar (SAR) data have become freely available from the European Space Agency (ESA) with the Sentinel-1 (S1) constellation under the Copernicus program (Aschbacher and Milagro-Pérez, 2012). The data comprise high-temporal resolution images, which can be used for RS applications in agriculture (Harfenmeister, Spengler and Weltzien, 2019; Steele-Dunne et al., 2017; Veloso et al., 2017), although its potential has not been fully established (Friesen, Steele-Dunne and van de Giesen, 2012; Paget, Long and Madsen, 2016; Steele-Dunne et al., 2017). The main advantages of SAR over optical sensors are the ability to penetrate the clouds and independence from sun illumination. For these reasons, SAR can provide high-density time series (every 5-6 days). In addition, SAR can be used to monitor the biophysical properties of agricultural fields (Ndikumana et al., 2018), as it is sensitive to changes in the canopy structure and biomass (Harfenmeister, Spengler and Weltzien, 2019; Mattia et al., 2003). As radar is also sensitive to the water content in the observed surface (Alexakis et al., 2017; Steele-Dunne et al., 2017; Vreugdenhil et al., 2018), it can be used to quantify the moisture and structural change in fields (Betbeder, Fieuzal and Baup, 2016; Dobson et al., 1985; Mironov et al., 2004; Snapir et al., 2018).

Another important application of SAR data in agriculture is crop identification (Bargiel, 2017; McNairn et al., 2009a; Van Tricht et al., 2018). In many cases, the SAR time series were combined with optical images to increase the accuracy of crop identification (De Bernardis et

al., 2016; Fieuzal, Baup and Marais-Sicre, 2013; McNairn et al., 2009a; Zhou et al., 2017). In some instances, SAR time series were used to derive metrics (e.g. mean, variance) for identifying irrigated fields (Gao et al., 2018) or models to simulate the backscatter interaction with vegetation and soil conditions (Baghdadi et al., 2017; Bériaux et al., 2015; Li and Wang, 2018) This is well implemented in the water cloud model (Attema and Ulaby, 1978; Baghdadi et al., 2017; Li and Wang, 2018).Some studies used backscatter to describe vegetation dynamics, taking into consideration L- C- and X-microwave bands (Brown et al., 2003; Huang et al., 2015; Snapir et al., 2018) and there is the potential to use SAR to identify crop development stages in the field (Bargiel, 2017; De Bernardis et al., 2016). Such information could be used in combination with crop modelling using different data assimilation techniques (Huang et al., 2019; Jin et al., 2018; Lahoz and Schneider, 2014). The use of time series and logistic based methods to simulate development and growth stages of vegetation can improve the accuracy of the crop development simulations (Che et al., 2014; Klosterman et al., 2014; Son et al., 2016).

In Europe, S1 C-band (5.405 GHz) SAR instruments support operation in dual polarisation (VV+VH), implemented through one vertical transmit chain (V) and two parallel receive chains for H and V polarisation (horizontal and vertical, respectively) over the land (Østergaard et al., 2011). Thanks to the constellation, made by two satellites orbiting in near-polar, sun-synchronous orbits it is possible to benefit of 6-days return time of satellites on the same orbit, leading to a very frequent description of vegetation growth. The VH/VV ratio, which is derived from the two polarisations, shows great potential to describe the dynamics of crop development because it is relatively insensitive to changes in soil moisture (Fieuzal, Baup and Marais-Sicre, 2013; Khabbazan et al., 2019; Veloso et al., 2017; Vreugdenhil et al., 2018). However, so far, there are few quantitative analyses of the dynamic changes of the SAR cross-polarisation ratio (VH/VV) in the literature.

This study is focused on the comparison of VH/VV ratio time series among wheat fields on an experimental farm (Rothamsted Research, UK) across three different years to understand and identify changes in backscatter values that can be related to crop growth and development. The assumption is that the SAR VH/VV ratio can be used to derive indicators that are related to vegetation dynamics, and that the indicators can be used to improve crop management. The objective of the paper is to identify new SAR-derived indicators of wheat crop development and productivity that can provide insights for crop yield formation at field scale.

2.2 Materials and Methods

2.2.1 Study area

In total, 18 winter wheat fields were selected for the analysis of S1 backscatter interaction with vegetation during the 2017 to 2019 agronomic seasons (nine fields were selected in 2017, five fields in 2018, and four fields in 2019 (Figure 2-1). The fields belong to the Rothamsted Research experimental farm in Harpenden, UK (51°48'37.3"N 0°22'36.0"W), located 35 km north of London. The site comprises slightly acid loamy and clayey soils with restricted drainage (Cranfield University, 2021) and the main soil series at Rothamsted is the Batcombe series (Avery and Catt, 1995). Meteorological information about the area on the years of the analysis are displayed on Figure 2-2.

Figure 2-1 Google Earth image of Rothamsted Research experimental farm, with wheat field boundaries from 2017, 2018 and 2019 indicated using red lines

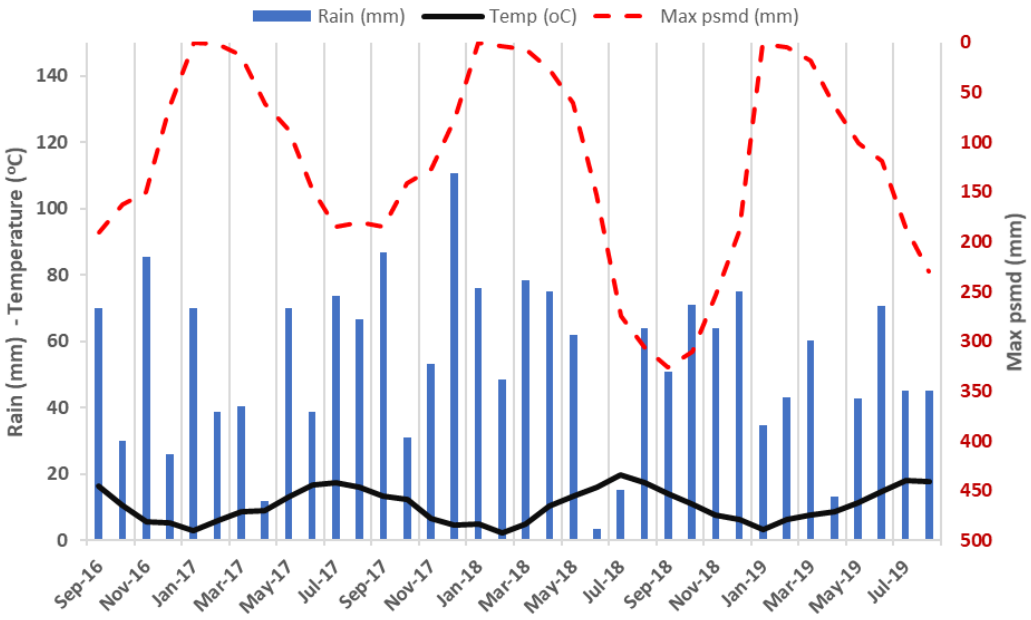


Figure 2-2 Total monthly rainfall (mm) and mean air temperatures (°C) on the primary axis (left) and maximum potential soil moisture deficit (PSMD; mm) on the secondary axis (right) at Rothamsted farm from Sep 2016 to August 2019.

2.2.2 Site ground measurements of wheat development and yield

In situ measurements were made weekly during the growing season to monitor the wheat development and growth stages. In each field (Table 2-1), measurements of grain yield were also taken at the end of each season. These data were used as reference points in the methodology to describe the response of the VH/VV ratio to the dynamics of the winter wheat development and growth and explore the different parameters of the segments with respect to their contribution to yield.

Table 2-1 Wheat field names and number of 10 m x 10 m pixels per field on Rothamsted farm

Field name	No. of S1 Pixels per field	Perimeter (m)	Area (ha)	Ground data collected
Great_Knott_1	333	805	3.37	2017
Great_Knott_2A	146	555	1.46	2017
Great_Knott_3A	112	457	1.12	2017
Little_Knott_1	122	482	1.22	2017
Osier_1_2_3	479	1046	4.75	2017
Sawyers_3	228	645	2.27	2017, 2019

Whitehorse_2B	191	649	1.90	2017, 2019
Bones_Close	437	810	4.39	2018
Sawyers_2	116	469	1.16	2018
Sawyers_4	131	465	1.30	2017, 2018
Stackyard	276	739	2.77	2017, 2018
West_Barnfield_1_2	365	910	3.62	2018, 2019
Drapers	399	839	3.93	2019

2.2.3 SAR data to obtain VH/VV time series in each wheat field

The Sentinel-1 SAR data acquired for the Rothamsted Research farm, over the 3-year study period were derived from orbit 132 (ascending) with a temporal resolution of 6 days. One of the reasons for selecting a single orbit is the orientation of the view of the satellite, which plays a significant role in the direction of the emitted beam from the SAR antenna (Mladenova et al., 2013; Vaudour, Baghdadi and Gilliot, 2014). Using the same orbit also had the advantage of a similar incidence angle (range of 5°) for adjacent fields on even terrain (Mladenova et al., 2013). The ascending orbit 132, instead of descending 81, was also used to avoid early morning measurements when dew may become a confounding factor (Harfenmeister, Spengler and Weltzien, 2019; Khabbazan et al., 2019). Temporal profiles of the S1 VH/VV backscatter ratio were plotted on each field to define the curves' shapes and amplitude, which were related to ground observations of wheat at key growth stages, assisting in the definition of periods defined by the calculated parameters **(**Table 2-2**)**.

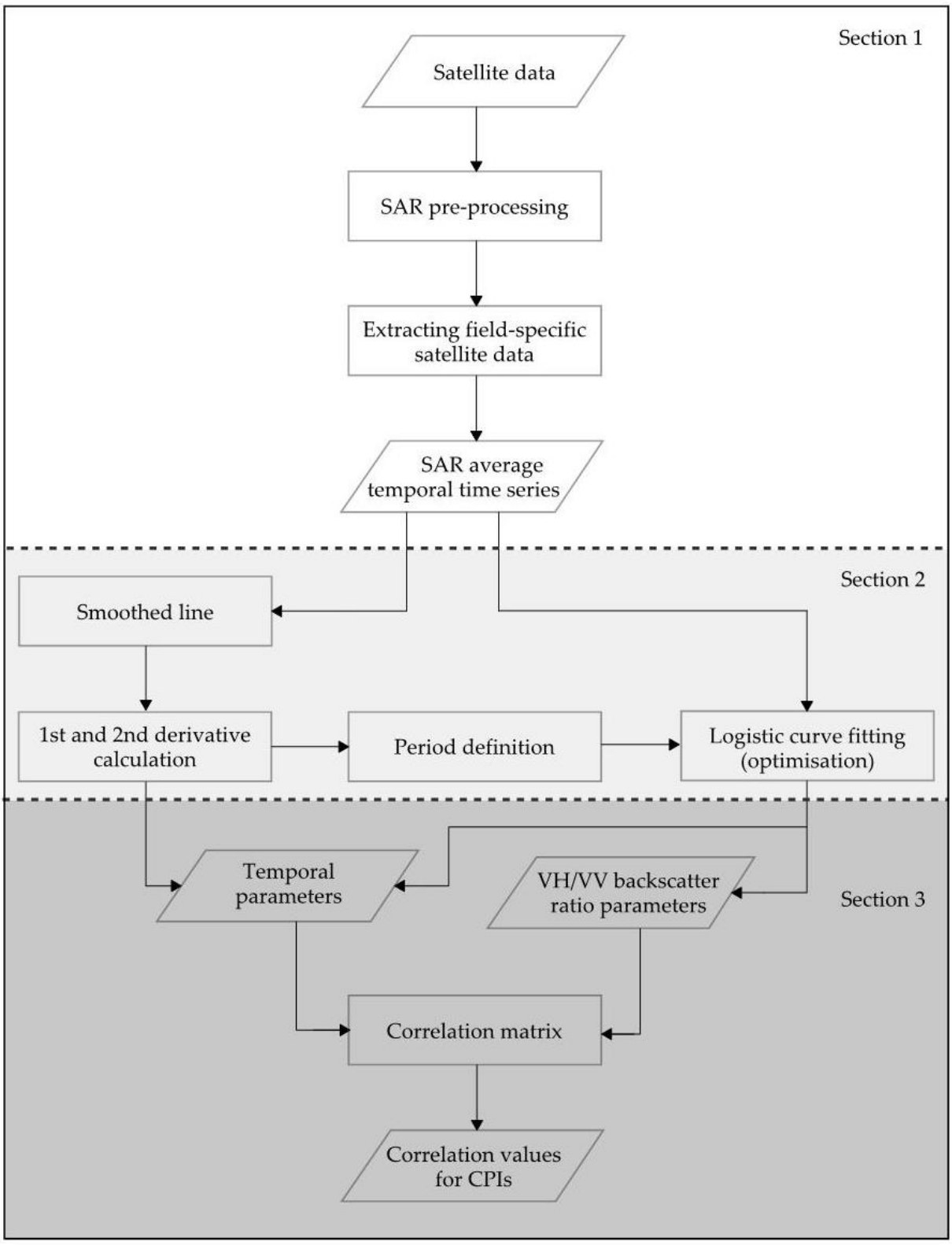


Figure 2-3 Flow diagram explaining the steps for the temporal analysis of SAR time series, including the data and process

2.2.4 SAR pre-processing for field specific VH/VV time series

The SAR data were derived from the Level-1 Ground Range Detected Interferometric Wide Swath (GRD-IW) product from Sentinel-1 satellite. The data have been pre-processed by Environment Systems Ltd. (www.envsys.co.uk) and delivered on the cloud-based Amazon platform. The Sentinel-1 toolbox was used to pre-process the data, including border and thermal noise removal, slice assembly, radiometric calibration, terrain flattening, speckle filtering using refined Lee filter with a 3X3 window and terrain correction to produce gamma naught (γ°) data in VH and VV polarisations. Lastly, each scene was clipped, the γ° VH/VV ratio was calculated. Based on the scale selected for the analysis, a mean value for the whole field (Table 2-1) was used to minimise the speckle effect, and the area of each field was buffered to minimise the influence of surrounding fields. All of the calculations were conducted using untransformed data in a linear form. In the presentation of the results, the initial graphs were developed to allow the reading of both linear and decibel (dB) values, and the final results were presented on decibel axes. In addition, temporal filtering (in the form of smoothing, Savitzky–Golay filter) was used to reduce localised weather patterns, such as heavy rainfall, ice or snow, that can temporarily lower the SAR backscatter. This step avoided the use of additional data (e.g. rainfall used to clean the data) for automation.

2.2.5 Temporal analysis and wheat development definition

The curve of the VH/VV ratio time series was used to produce indicators that could describe wheat crop growth and development across the season. Smoothing approaches for time series of SAR (Canisius et al., 2018; Molijn et al., 2014) and optical data (Klosterman et al., 2014; Zimmermann and Kohler, 2013) have been used in various studies to reduce noise or fill the gaps of the datasets. Here, the Savitzky–Golay filter was applied over the whole period with a moving window of two months and using a second-order polynomial function (Savitzky and Golay, 1964). Then the associated first and second derivatives were calculated to define additional wheat growth stages, linked to the SAR temporal characteristics. Key periods of these three curves were identified, as shown as vertical lines in Figure 2-4**,** and matched with crop development observed in the field.

The next stage was to further parameterise and automate the extraction of these dates and test if they were linked to wheat development. The smoothed VH/VV curve was divided into two periods: i) growth and ii) maturation and senescence to address an observed change in maximum values in the middle of the season. The splitting of the analysis into two periods has also been used by Che et al. (2014), who analysed the temporal evolution of the leaf area index (LAI) of vegetation during development and senescence in Shandong Province in China. The breakpoint of two curves occurs during the flowering period (anthesis) when the wheat

has reached its full height. The calculation of the breakpoint specifies the start for the VH/VV ratio stabilisation period close to the end of spring.

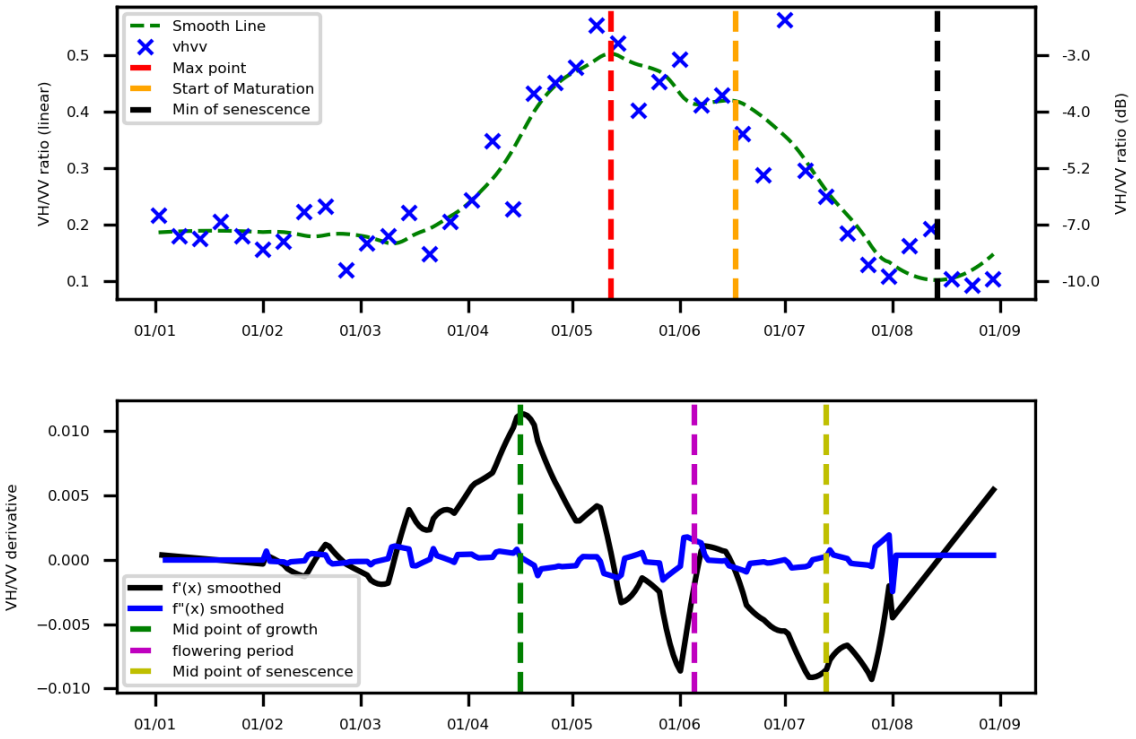


Figure 2-4 a) Mean VH/VV ratio from Sentinel-1 orbit 132 of wheat field (for the Swayers_4 field in 2018) together with a smoothed curve (Savitzky–Golay filter) of the mean (with the date of the maximum point indicated) and b) first (f'x) and second (f''x) derivative (bottom) of the smoothed field-averaged line. The date of the mid-point of the growth period, and the date of the start, mid-point and minimum value of the maturation period are indicated as vertical dotted lines. The secondary y axis in the first graph displays the VH/VV ratio on a decibel scale for reference.

2.2.6 Logistic curve fitting

In order to quantify the characteristics of the temporal curve of VH/VV ratio for winter wheat, we assumed a logistic pattern during both the growth and maturation phases. Two separate curves, rather than a single double logistic curve, was chosen because the starting value of the VH/VV ratio of the maturation phase (which coincided with flowering) tended to be lower than the finishing point of the growth stage (which coincided with booting) in Figure 2-5. In addition, the baseline at the beginning of the season can differ from the end. The two sigmoid logistic curves describing the VH/VV ratio in relation to time t, for each stage was characterised by four parameters (2-1):

$$f(t) = base + \frac{Max}{1 + e^{-b(t-t_0)}} \quad (2\text{-}1)$$

where base is the minimum VH/VV ratio value, Max is the maximum VH/VV ratio value, b is the steepness of the logistic curve and t_0 is the x-value of the sigmoid inflexion point. The calculated values are sensitive to a clear definition of the period represented by each sigmoid curve (Figure 2-5). The growth period curve was defined from the beginning of the season (Parameter 1 in Figure 2-5) up to the time of maximum value of the smoothed curve (Figure 2-5). The maturation period curve was defined as the period after flowering to the minimum value in the maturation period of the temporally smoothed curve.

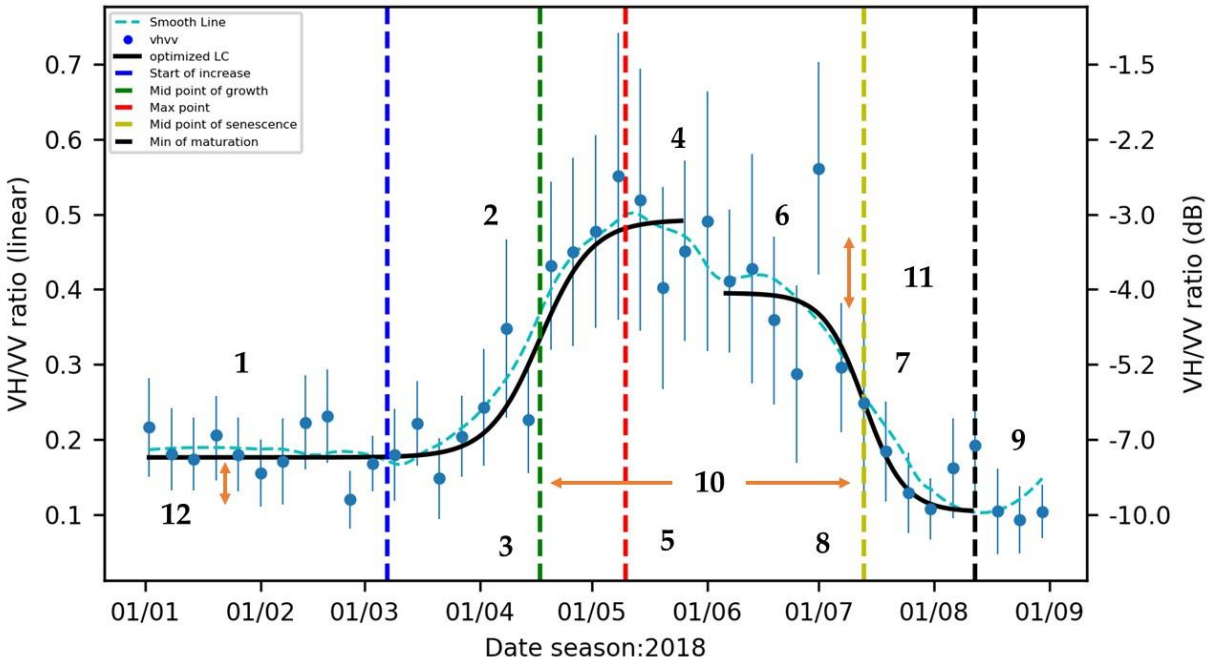


Figure 2-5 Mean SAR VH/VV ratio for a wheat field (Swayers_4 in 2018) with 228 pixels (1.3 ha), between January and September 2018, showing the standard deviation for each date (n = 41), and displaying 12 parameters as calculated using a smoothed line (Savitzky–Golay filter) and two logistic curves (see Table 2-2). All calculations were completed using a linear scale and the secondary y axis displays the decibel scale for reference.

Combining parameters extracted from the two logistic curves allowed us to derive three additional vegetation development-related parameters (10, 11 and 12 in Figure 2-5). The period between the two midpoints of the logistic curves (Parameter 10 in Figure 2-5) was defined as the duration of “full” vegetation which starts with stem elongation and ends with the reduction of the green canopy and translocation of carbohydrates into the grain during maturation. The difference in baselines is related to the backscatter from the biomass of the crop during tillering. The summary of the extracted parameters and their definitions as well as the associated crop development stage is given in Table 2-2.

Table 2-2 Definitions of the VH/VV ratio curve parameters and the anticipated associated crop development stage (BBCH or Zadoks scale)

No.	Symbol	Parameter name	Definition	Derived from	Associated crop development stage
1	G_base	Baseline value for the growth stage	VH/VV ratio at the beginning of the season	Logistic curve	Tillering (GS20)
2	G_steep	Steepness of logistic curve for growth period	Rate of coefficient in eq. 1 (b G)	Logistic curve	Stem elongation
3	G_midP	Time of midpoint of growth period (t_0, G)	DOY when the midpoint of the logistic curve occurs in the growth period	Logistic curve	Stem elongation
4	G_max	Max value for growth stage	Maximum VH/VV ratio value for the full season	Logistic curve	End of stem elongation (GS39) and booting (GS49)
5	TZmax	Time of maximum point	DOY of maximum smoothed value of VH/VV	Smoothed curve	Time of booting, flag leaf unrolled
6	S_max	Value at the start of grain filling	Period of backscatter stabilisation	Logistic curve	Post anthesis: start of grain filling (GS71)
7	S_steep	Steepness of logistic curve for maturation period	Rate of coefficient in eq. 1 (b S)	Logistic curve	Maturation rate
8	S_midP	Time of midpoint of maturation (t_0, S)	DOY when the midpoint of the logistic curve occurs in the maturation period	Logistic curve	Ripening (GS 85-89)
9	S_base	Baseline value at the end of the season	Background value of the VH/VV ratio	Logistic curve	After harvest Period with soil exposed
10	Duration	Duration of “full” vegetation to maturation	Time difference between midpoints (3, 8)	Combination	Period of most of the photosynthate accumulation and translocation
11	D_max	Structure change (Inflorescence)	VH/VV ratio value differences between booting and grain filling periods (4-6)	Combination	Backscatter change during the period when the ear emerges
12	D_base	Tillering backscatter	VH/VV ratio value differences between tillering and bare soil (1-9)	Combination	Tillering with reduced impact of soil

2.2.7 Parameter optimisation

The calculation of the logistic curve fitting parameters in Table 2-2 was based on a weighted least squares (WLS) estimator, which considers the uncertainty of each point by minimising the sum of the squared difference divided by the respected standard deviation (σ) in each point (2-2). The incorporation of the standard deviation in the estimator was used to minimise the outlier effects. The VH/VV ratio points were selected as the observation values in each defined period.

$$min \sum_{i} \left(\frac{obs_i - y_i}{\sigma_i} \right)^2 \quad (2\text{-}2)$$

2.2.8 Automatic curve extraction and correlation analysis

By scripting the SAR processing using Python, it was possible to automatically derive a smoothed VH/VV ratio curve and two logistic curves. Then the 12 parameters were derived for each field site across 3 years. A Pearson correlation coefficient (r) was calculated for each pair in a matrix. This approach allows the illustration of the interactions between the parameters as well as the direct effect on final crop yield. A significant test of the r was performed and highlighted in the plots.

2.3 Results

2.3.1 Annual analysis of VH/VV ratio curve parameters, 2017 to 2019

Two logistic curves for the nine fields selected in 2017 (Figure 2-6), the five fields selected in 2018 (Figure 2-7), and the four fields selected in 2019 (Figure 2-8), were plotted. The smoothed time series were used to identify the start and end of (i) the growth (start of the year to red line) and (ii) the maturation periods (purple to min senescence value). Then the midpoints were calculated by fitting the logistic curves in these periods, creating the time parameters shown in Figure 2-6 to Figure 2-8.

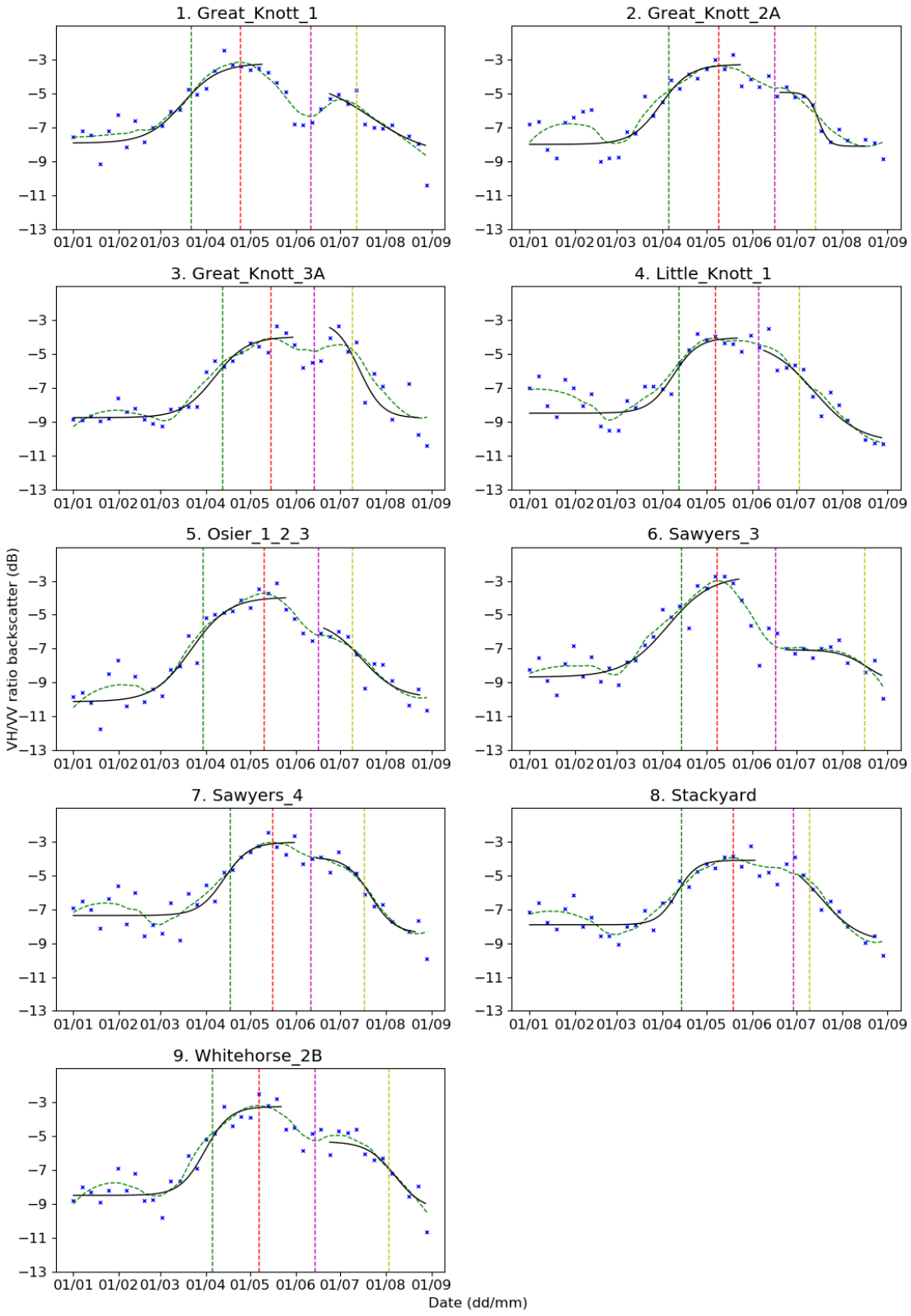


Figure 2-6 VH/VV ratio time series (blue dots) with the smoothed curve (dotted green line) and two logistic curves fitted (black line) for nine fields during 2017, orbit 132. The vertical dotted lines indicate the dates of G_midP (green), TZmax (red), inflorescence (purple), and S_midP (yellow) as determined from the VH/VV backscatter time series analysis

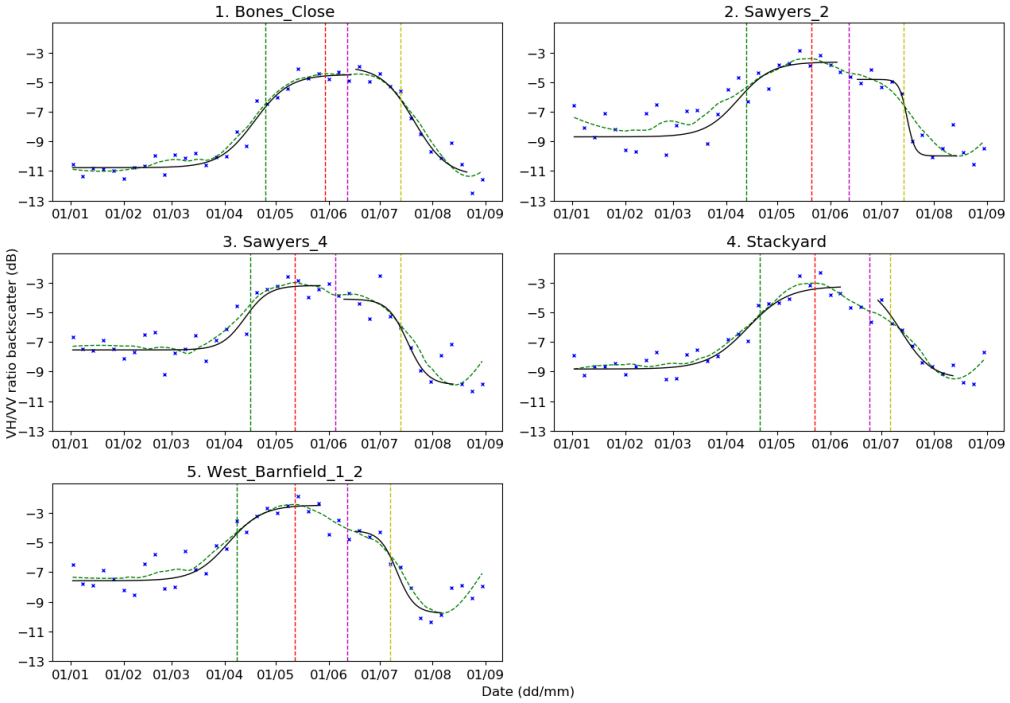


Figure 2-7 VH/VV ratio time series (blue dots) with the smoothed curve (green dashed line) and two logistic curves fitted (black line) for five fields during 2018, orbit 132. The vertical dotted lines indicate the dates of G_midP (green), TZmax (red), inflorescence (purple), and S_midP (yellow) as determined from the VH/VV backscatter time series analysis

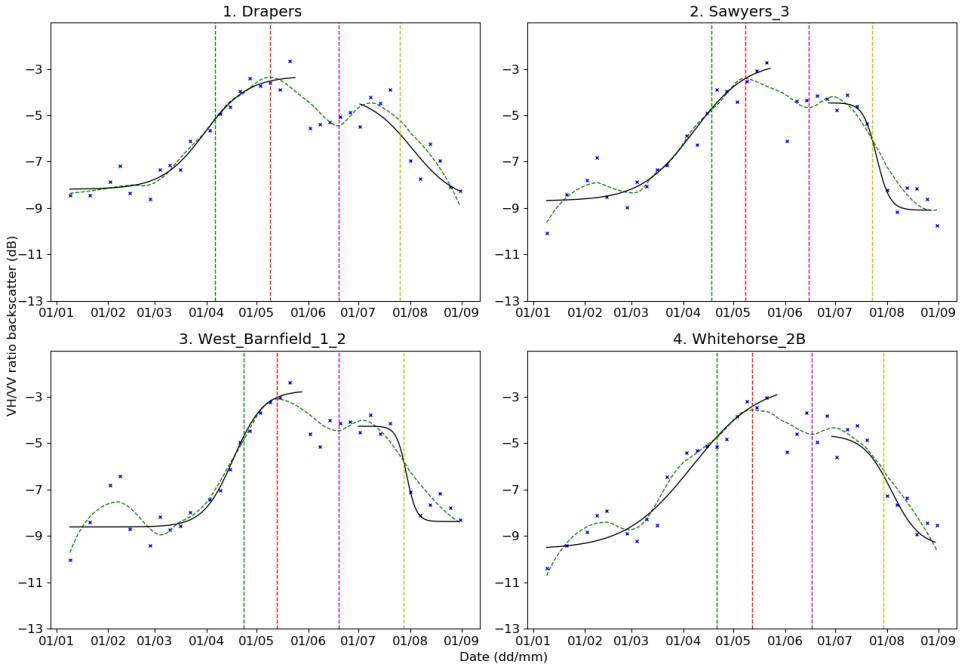


Figure 2-8 VH/VV ratio time series (blue dots) with the smoothed curve (green dashed line) and two logistic curves fitted (black line) for four fields in 2019, orbit 132. The vertical dotted lines indicate the dates of G_midP (green), TZmax (red), inflorescence (purple), and S_midP (yellow) as determined from the VH/VV backscatter time series analysis

The timing of the mean maximum VH/VV ratio (TZmax) was relatively consistent varying from day-of-year (DOY) 129 in 2017, to DOY 140 in 2018, and DOY 131 in 2019 (Table 2-3). The mean values (Table 2-4) of the three parameters of the logistic curve for the growth period were also relatively consistent from year to year: G_base (range 0.13-0.15), G_steep (range: 0.08-0.11), and G_midP (range: DOY 99 to 109). There was a greater variation in G_max which ranged from a mean of 0.31 in 2017 to 0.40 in 2019. The mean values of two of the parameters of the logistic curve for the maturation period were also relatively similar across the three years: S_base (range: 0.10-0.13), and S_midP (range: DOY 199 to DOY 208). By contrast the mean values of S_max (range: 0.23-0.30) and S_steep (range: -0.12 to -0.21) showed greater inter-annual variation (Table 2-4).

The root mean square error (RMSE) of the growth (G) and maturation (S) logistic curves are shown in Table 2-3 for the seasons 2017, 2018 and 2019. Statistical comparison between the parameters of different years showed that TZmax ($p = 0.042$), the G_midP ($p = 0.037$), S_base ($p = 0.030$) are significantly different at 95% confidence using ANOVA or the Kruskal-Wallis H-test if there was a non-normal distribution. There was also a significant inter-annual variation in the Duration ($p = 0.039$) and yield ($p = 0.016$).

Table 2-3 Values of derived parameters (linear scale), yield data, and the root mean standard error of the growth (RMSEG) and maturation (RMSES) stages, for each of the winter wheat fields in 2017, 2018 and 2019

Field	Season	$RMSE_G$	$RMSE_S$	G_base	G_steep	G_max	S_max	S_steep	S_base	G_midP	S_midP	TZmax	Duration	D_max	D_base	Yield
										DoY	DoY	DoY	days			t/ha
Great_Knott_1	2017	0.051	0.004	0.16	0.077	0.36	0.25	-0.05	0.14	85	194	114	109	0.11	0.020	8.67
Great_Knott_2A	2017	0.092	0.022	0.16	0.115	0.28	0.17	-0.29	0.15	93	196	129	103	0.12	0.010	6.43
Great_Knott_3A	2017	0.020	0.047	0.13	0.11	0.24	0.37	-0.13	0.13	99	190	135	91	-0.13	0.000	9.26
Little_Knott_1	2017	0.090	0.033	0.14	0.108	0.31	0.28	-0.07	0.10	106	184	127	78	0.02	0.040	6.17
Osier_1_2_3	2017	0.054	0.004	0.10	0.072	0.34	0.20	-0.08	0.10	93	190	130	97	0.14	0.000	6.63
Sawyers_3	2017	0.070	0.012	0.14	0.066	0.45	0.07	-0.11	0.12	108	228	128	120	0.38	0.020	6.36
Sawyers_4	2017	0.047	0.004	0.18	0.115	0.33	0.26	-0.12	0.14	108	199	136	90	0.07	0.040	6.85
Stackyard	2017	0.037	0.002	0.16	0.159	0.22	0.30	-0.08	0.13	103	191	139	88	-0.08	0.030	5.04
Whitehorse_2B	2017	0.061	0.035	0.14	0.143	0.30	0.18	-0.11	0.11	93	215	127	122	0.12	0.030	9.59
Bones_Close	2018	0.021	0.017	0.08	0.098	0.28	0.33	-0.11	0.07	117	194	150	78	-0.05	0.010	4.09
Sawyers_2	2018	0.118	0.035	0.13	0.102	0.31	0.23	-0.45	0.10	104	196	141	91	0.08	0.030	6.58
Sawyers_4	2018	0.051	0.069	0.18	0.12	0.40	0.29	-0.18	0.10	111	195	132	84	0.11	0.080	5.49
Stackyard	2018	0.044	0.009	0.13	0.071	0.49	0.39	-0.12	0.11	121	187	143	66	0.10	0.020	4.81
W_Barnfield_1_2	2018	0.057	0.012	0.17	0.094	0.43	0.28	-0.20	0.11	101	189	132	88	0.15	0.060	7.90
Drapers	2019	0.018	0.047	0.15	0.079	0.32	0.26	-0.08	0.13	97	207	129	110	0.06	0.018	9.65
Sawyers_3	2019	0.007	0.026	0.14	0.065	0.41	0.24	-0.26	0.12	108	205	128	97	0.17	0.012	7.99
W_Barnfield_1_2	2019	0.030	0.036	0.14	0.114	0.40	0.23	-0.34	0.15	113	209	133	96	0.17	-0.008	8.28
Whitehorse_2B	2019	0.021	0.029	0.11	0.053	0.47	0.23	-0.12	0.11	112	211	132	99	0.23	-0.002	10.17

Table 2-4 Mean and standard deviation values of 12 derived parameters (linear scale) and the wheat yield for each season of the analysis

Season		G_base	G_steep	G_max	S_max	S_steep	S_base	G_midP DoY	S_midP DoY	TZmax DoY	Duration days	D_max	D_base	Yield t/ha
2017	Mean	0.15	0.11	0.31	0.23	-0.12	0.12	99	199	129	100	0.08	0.02	7.22
	Std dev	0.02	0.03	0.07	0.09	0.07	0.02	8	14	7	15	0.14	0.02	1.56
2018	Mean	0.14	0.10	0.38	0.30	-0.21	0.10	111	192	140	81	0.08	0.04	5.77
	Std dev	0.04	0.02	0.09	0.06	0.14	0.02	8	4	8	10	0.08	0.03	1.50
2019	Mean	0.13	0.08	0.40	0.24	-0.20	0.13	107	208	131	101	0.16	0.01	9.02
	Std dev	0.02	0.03	0.06	0.01	0.12	0.01	7	3	2	7	0.07	0.01	1.00

2.3.2 Relationship between SAR-derived parameters and Yield

The interactions between each of the parameters and yield were further displayed using a correlation matrix (Figure 2-9). There were various significant correlations between parameters. As would be expected, the inflorescence indicator (D_max = G_max – S_max) was correlated with the two parameters used in the calculation (G_max and S_max) with r values of 0.76 and -0.75 respectively. At the same time, significant correlations were found between D_max and G_steep (r = -0.54), S_midP (r = 0.68) and Duration (r = 0.49). By contrast, D_base displayed no significant correlations with the other parameters with the exception of G_base (r = 0.65) that was used for its calculation. The definition of D_base is G_base minus S_base, where S_base represents the VH/VV ratio value on bare soil (that should remain relatively consistent) and the value of G_base is determined by the combination of soil and low vegetation during the tillering period.

Starting from the first row in Figure 2-9, the baseline of growth (G_base) shows a significant positive correlation (r = 0.58) with the baseline of maturation (S_base). This is because G_base depends on the shift of the VH/VV ratio during the initial bare/sparse/small vegetation period and the soil background (S_base). In the second row, the slope of the midpoint of growth (G_steep) has a negative correlation with the VH/VV ratio at G_max value (r = -0.68). In the third row, the mid-point of growth (G_midP) is highly correlated with the timing of the max (TZmax) smoothed VH/VV ratio value (r = 0.69) highlighting the interaction of stem elongation and booting stage. G_midP was also correlated with the calculated Duration (r = -0.68) and the maturation mid-point (S_midP; r = 0.78). This is because the Duration value is defined by the difference between these two time-indicators. The value of TZmax, the timing of the maximum value of VH/VV, was positively correlated with S_max (r = 0.50), but negatively correlated with low Duration values (r = -0.64).

During the maturation period, S_max was negatively correlated with the mid-point S_midP (r = -0.72) and Duration (r = -0.80). The only correlation shown by the steepness of the curve in this period was with the yield. The value of S_midP was positively correlated with Duration (r = 0.76) and D_max (r = 0.68). The value of S_base was not significantly correlated with any of the indicators except G-base.

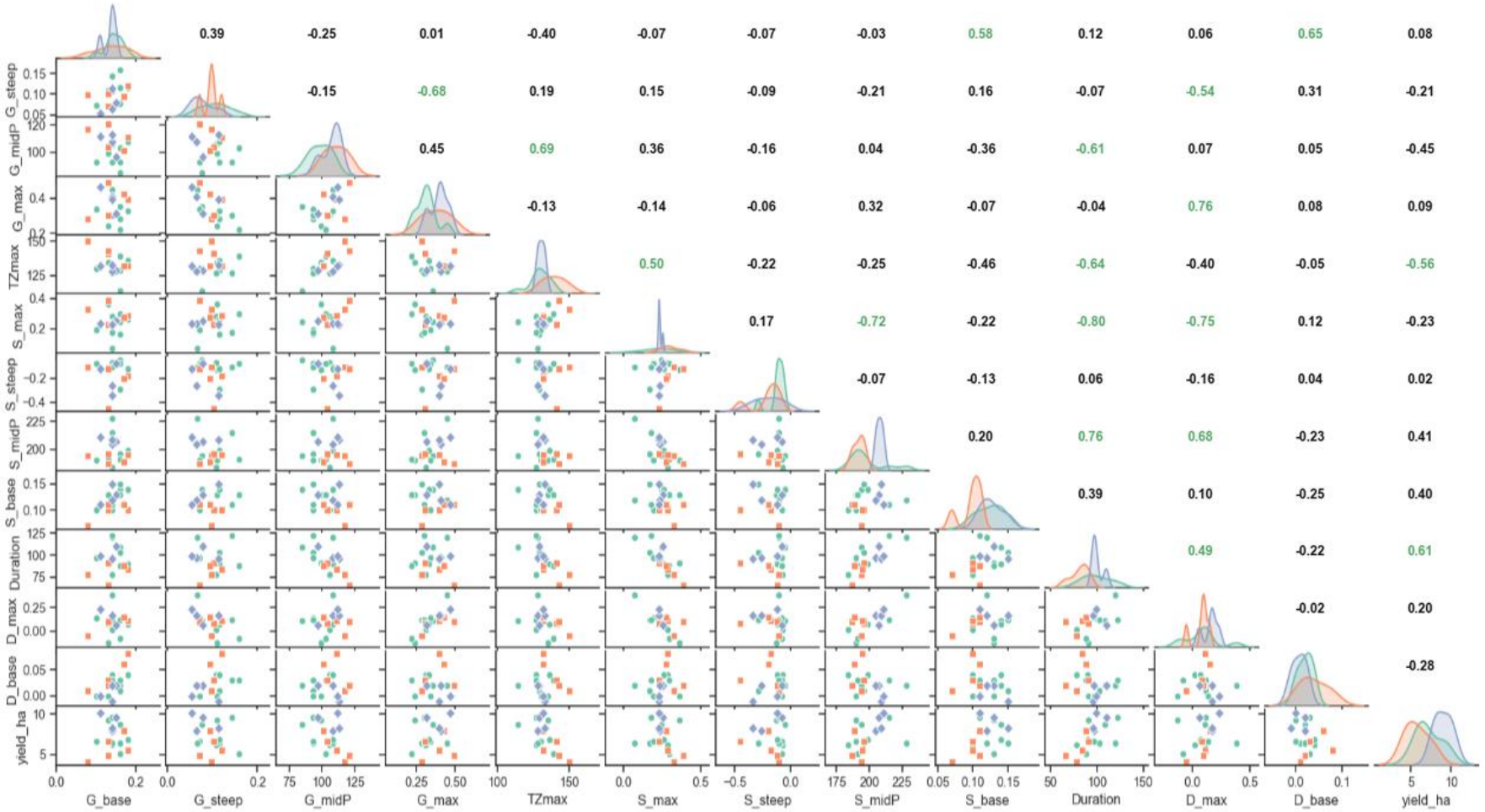


Figure 2-9 Correlation matrix with 2017 (green circles), 2018 (red squares) and 2019 (blue diamond) using the 17 fields. r values are colour-coded green when there was a 95% level of significance for a two-tailed test.

As illustrated by the matrix (Figure 2-9), multiple correlations exist among the individual parameters and between individual parameters and yield (last column). The results suggest that the two parameters, G_midP and TZmax (which are correlated), had the strongest association with grain yield, i.e. r = -0.45 and r = -0.56, respectively, after Duration (r = 0.61). The negative correlation suggests that earlier growth, i.e. lower values of G_midP and TZmax, was associated with higher end of season grain yields. Figure 2-10 displays the two most significant parameters which are sensitive to yield.

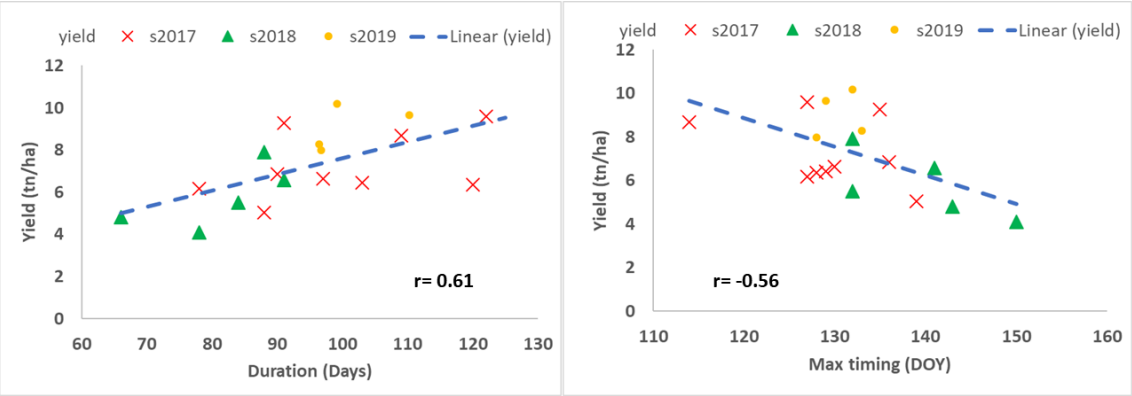


Figure 2-10 Scatter plot showing the relationship between Yield and a) Duration and b) the timing of the maximum value of VH/VV (TZmax). The r and best fit linear line for each relationship, across the three seasons, is shown

2.4 Discussion

Using the example of winter wheat, we showed for data over three growing seasons that it was possible to derive field-specific annual coefficients of a generalised two-component empirical model of crop productivity from the SAR VH/VV ratio. Due to the ubiquitous and high-frequency availability of S1 data, this approach could offer new opportunities for monitoring and predicting productivity in areas affected by frequent cloud cover. The characterisation of VH/VV response over a season used three stages: i) the creation of a smooth VH/VV time series for each field, ii) the definition of the timing of the maximum value using temporal characteristics of the time series and growth and maturation period using first and second derivatives and iii) the creation of two logistic curves using the observation in the predefined periods of growth and maturation (Che et al., 2014). The results are discussed in terms of the robustness of our mathematical method over three seasons to consider multiple factors of variation, second, the gain of biophysical understanding using this method and finally, the agronomic potential of our results.

2.4.1 Technical and statistical process, the goodness of fit, robustness, uncertainty

Since 2016, SAR data from Sentinel-1 have become twice as frequent as data used in previous studies (Veloso et al., 2017; Vreugdenhil et al., 2018), which improves the robustness of the possible analysis. However, to reduce the variability of SAR data we chose to use imagery from only the ascending satellite passing as images from the descending orbit were often affected by morning dew. Similar observations have been made in other studies (Harfenmeister, Spengler and Weltzien, 2019; Khabbazan et al., 2019), indicating greater certainty when using the ascending rather the descending phase of the satellite for predicting crop state variables in this area. Even so, the results demonstrate substantial variation of SAR backscatter, which can be due to differences in topography affecting the incidence angle or changes in canopy architecture (Mladenova et al., 2013; Vaudour, Baghdadi and Gilliot, 2014)

As pointed out in an earlier review (Steele-Dunne et al., 2017) SAR variability is also connected to diurnal variation in plant-soil-water relations, which can affect the erectness of the crop canopy and the dielectric properties of the system. To reduce the effect of such fluctuations in the SAR time series, our approach used a step of smoothing the data, which resulted in a set of relatively consistent dynamic curves for winter cereals (Vreugdenhil et al., 2018), and limited the extreme values of the VH/VV (Figure 2-6). The subsequent logistic fitting approach enabled the calculation of a specific backscatter value and time-specific parameters from the VH/VV ratio data (Figure 2-5). Accounting for the standard deviation during the fitting of the logistic curves, made it possible to give lower weight to data with high uncertainty and improved the precision of the derived curves. The average RMSE values of the logistic curve fitting in growth (G) and senescence/maturation (S) periods were 0.05 and 0.07 respectively.

The creation of the fitted curves allowed the derivation of 12 parameters which can be related to the development, growth and yield of winter wheat. Two temporal parameters from the analysis (TZmax and Duration) were significantly correlated with yield (Figure 2-9). Further analysis could focus on the opportunity of using a combination of parameters or a combination of VV and VH/VV values. The approach was effective in each of three years which were characterised by different environmental conditions and yields. In 2018, there was minimal rainfall in June and July and the grain yields obtained in 2018 (4.09-7.01 t ha^{-1}) were substantially lower than in 2017 (5.04-9.59 t ha^{-1}) and 2019 (7.99-10.17 t ha^{-1}).

2.4.2 Environmental and biophysical understanding

The sensitivity of SAR backscatter to surface changes in the structure and the soil and crop water content means that SAR can be an effective tool to monitor crop development (Bargiel, 2017; Steele-Dunne et al., 2017) starting from early growth, to the increase in the water volume of the crop through stem elongation, and its decrease during maturity. In this paper, we assigned two separate logistic curves during the cropping season to examine whether the resulting parameters were correlated with observations in the field. A specific feature of creating two curves is the derivation of the D_max value, which would not be captured by assuming a single curve. Our interpretation is that the change in the VH/VV value at flowering is a real effect, that equates to a change in the structure of the crop (Figure 2-6) during the inflorescence period as mentioned by other authors (Harfenmeister, Spengler and Weltzien, 2019; Mattia et al., 2003; Veloso et al., 2017). Even so the value of D_max was variable, ranging from -0.13 to 0.38 for an individual field and year (Table 3). Additional research, including perhaps data focused on structure, is required to understand this variability further.

The correlation analysis (Figure 2-9) was able to identify two time-based parameters, one derived from the growth period (G_midP) and one derived from the smoothed VH/VV curve (TZmax), which were negatively associated with the final grain yield (r = -0.45 and -0.56, respectively). This means that a delayed occurrence of the mid-point of the growth curve and the timing of the maximum VH/VV ratio was associated with lower grain yields. The late or slow growth of the crop canopy is likely to result in a reduced capacity of the crop to intercept solar radiation and use photosynthesis to produce the assimilates required for final yield. For example, in 2018, the late occurrence of G_midP and TZmax was associated with an unusually dry period in June and July 2018 (Figure 2-2) which reduced canopy growth and photosynthesis and resulted in substantially lower grain yields than in the other two years.

The analysis also showed that the duration of the period of “full” vegetation (Duration) was positively related to grain yield (r = 0.61). This can be explained by the yield benefitting from a longer period of photosynthesis and a longer grain-filling period when assimilates are transferred from green tissue to the grain (Biscoe and Willington, 1984; Monteith, 2007). The above results suggest that using the VH/VV ratio to identify the timing of specific crop periods, which are related to the structural change of the wheat canopy [2, 14], may be a more useful determinant of yield than the actual values of the VH/VV ratio per se.

2.4.3 Opportunities for agronomic management and modelling

The analysis raises two questions for crop management: the first is what can be learnt about crop dynamics from monitoring satellite-based SAR data, dynamics and amplitude? The second is whether there is potential for a farmer or agronomist to use SAR-derived indicators for spatially varying crop management and inputs between fields? The growth and development of a wheat crop depends on the interaction between the weather (e.g. solar radiation, rainfall, and temperature), crop genetics, field characteristics (e.g. aspect and soil type), and field management (e.g. previous cropping history and the timing and type of cultivation and drilling). The results from analysing the SAR data indicate that it is possible to identify seasonal differences in the timing of crop growth, and this may be useful in identifying the most appropriate schedule of fertilizer and agrochemical application between fields. The analysis also suggests that it may be possible to improve the prediction of the eventual crop yield from an analysis of SAR characteristics.

A potential limitation with the method is that some of the parameters are determined retrospectively from a full temporal dataset, for example the calculation of TZmax, the timing of the maximum value of VV/VH, requires data that occurs after TZmax. Likewise, the value of S_steep, which occurs during the maturation of the crop, requires data that occurs after that point. Developing a really effective predictive tool that can be used to modify field management instantaneously will require parameters that can be estimated in real-time. Further work is needed to improve the estimation of parameters as well as the application of the methodology on other farms to examine the validity and usefulnessof the method. The correlation of the productivity indicators with field measurements also has the potential to assist in better model simulations of crop development and growth. At the same time, temporal characteristics, as well as connection with the biophysical properties of the field, could be examined to improve the calibration and its incorporation in a data assimilation framework to estimate more efficiently the field production (Betbeder, Fieuzal and Baup, 2016; Huang et al., 2019; Jin et al., 2018).

2.5 Conclusions

This paper describes a new methodology to derive wheat productivity indicators from Sentinel-1 VH/VV time series. The temporal curves, their 1st and 2nd derivatives, and logistic curve fitting for growth and maturation periods were used to define 12 pheno-morphological parameters. These provide information about the growth, development and yield of wheat crop at field scale, offering an alternative approach to ground survey or yield estimation.

The analysis of the VH/VV ratio time-series and the correlation matrix indicates that the time-based parameters appear to be related to biophysical changes in the field. In particular the time period of “full” vegetation (Duration) was positively correlated with yield (r=0.61), and a delay in the timing of the maximum VH/VV value (TZmax) was negatively correlated with yield (r=-0.56).

Automation of the SAR image analysis was possible, as the only inputs required were backscatter data from VH and VV polarisations. Future work will explore the use of this method to estimate commercial wheat production in other farming landscapes in the UK. It could also be adapted to monitor growth and development as well as yield prediction for other arable crops, potentially allowing the remote quantitative assessment of environment and management impacts.

3 BIOLOGICAL INTERPRETATION OF SATELLITE RADAR DATA FOR ASSESSING CROP DEVELOPMENT AND YIELD FORMATION

Summary

Plant phenological development and biophysical (water) status are strongly related to backscatter values measured by radar satellite. The overall objective of this chapter is to validate these relationships with field measurements and explore the quantitative relationship between Sentinel-1 SAR and yield formation, by relating sub-field time series of Sentinel-1 backscatter (VV, VH and VH/VV) to field observations. Nine winter wheat fields were sampled in the 2017-2018 and 2018-2019 seasons to determine structural information (canopy height, tiller density, leaf area, biomass) and plant water content (PWC%).

The SAR temporal backscatter curves responded clearly to the volume increase and water content changes during the growth and maturation phases, respectively. The crop development stages were clearly related to the dynamics of the SAR backscatter curves. The temporal evolution of the backscatter for the polarisations and the ratio displayed similar seasonal trend, confirming other studies. The development shift towards earlier maturation during the drought in 2018 compared to 2019 was recognised in an earlier decline of VH/VV ratio.

Observations at plot scale within fields were more variable (CV %) than at field scale but it was still possible to fit the logistic curves and determine the CPIs. Due to some erratic and overall variable VH/VV-curves within most fields, the derivation of the CPIs and their relationship with the observations within fields was more difficult than at field scale. The resulting CPIs from the plots showed no clear relation between yield and crop performance indicators (CPIs). The direct comparison of CPIs with *in situ* data at plot scale showed significant correlations ($p<0.05, n=50$) between G_max and the above ground biomass (AGB) in the growth phase ($r=0.32$), G max and the tiller density ($r=0.39$), and S_max with the plant water content ($r=0.39$). The effect of plant water content (PWC) and early maturation was also displayed as a negative correlation between S_midP and PWC ($r=-0.76$) that varied significantly across the seasons ($p<0.001$).

The results demonstrate that SAR dynamics and CPI values are affected by the structural changes during the growth stage, and low PWC values during the maturation phase. However, the SAR inherent variation (speckle) found at the within field scale impairs robust CPI derivation and reduces the significance of correlations found at the field scale.

3.1 Introduction

In the microwave region of the electromagnetic spectrum (1 mm < λ < 1 m), satellite-based sensors operate from the L Band (1 to 2 GHz; 15 to 30 cm) to X Band (8 to 12 GHz, 2.5 to 3.75 cm). Synthetic Aperture Radar (SAR) of Sentinel 1 operates C-band (~5.5 cm wavelength), emitting energy in the vertical polarisation and receiving its backscatter in vertical (VV) and horizontal (VH) polarisation (Torres et al., 2012). Several studies have displayed the suitability of C-band SAR to monitor the vegetation dynamics. The C-band is ideal for the identification of low biomass crops such as wheat (Liu et al., 2013; McNairn et al., 2009b). A strong correlation between C-band backscatter and leaf area index (LAI) has been found (McNairn et al., 2012). Other studies observed correlations between biomass and C-Band SAR (Harfenmeister, Spengler and Weltzien, 2019; Ndikumana et al., 2018; Paloscia et al., 1999; Veloso et al., 2017; Vreugdenhil et al., 2018; Wiseman et al., 2014). Polarimetric SAR has been also used to monitor phenological stages for different crops (Bargiel, 2017; Harfenmeister, Spengler and Weltzien, 2019; Mascolo et al., 2015; Nunziata et al., 2015; Shang et al., 2013; Veloso et al., 2017). At the same time, using the ratio of different polarisations, studies were able to correlate this with wheat biomass (Mattia et al., 2003; Veloso et al., 2017), and LAI (Satalino, Dente and Mattia, 2006).

The intensity of the energy scattered by vegetation is a function of the dielectric properties of the crop canopy (Steele-Dunne et al., 2017), mainly determined by the canopy structure and the water content, as well as the soil contribution of water and surface roughness/texture (Baghdadi et al., 2018; Bousbih et al., 2017; Wang et al., 1987). Different levels of attenuation and scattering can occur in response to changes in the size, shape and orientation of canopy components such as leaves, stem, fruit and plant density (Karam et al., 1992; Karam and Fung, 1989; Senior, Sarabandi and Ulaby, 1987; Yueh et al., 1992). The level of attenuation and scattering is dependent on the size of the object compared to the wavelength. Larger and more complex canopy components scatter more in contrast to smaller components that tend to reduce the backscatter signal. Biological parameters that reflect a change in the canopy structure and biomass

can be monitored using SAR (Harfenmeister, Spengler and Weltzien, 2019; Mattia et al., 2003). As radar is also sensitive to the vegetation and water content in the surface (Alexakis et al., 2017; Steele-Dunne et al., 2017; Vreugdenhil et al., 2018), it can be used to quantify the moisture and structural change of the field surface (Betbeder, Fieuzal and Baup, 2016; Dobson et al., 1985; Mironov et al., 2004; Snapir et al., 2018). *In situ* data are being used to improve the interpretation of the satellite signal and the field conditions (Alexakis et al., 2017; Baghdadi et al., 2017; Bousbih et al., 2017; Ndikumana et al., 2018; Snapir et al., 2018; Stevens et al., 2010)

The objective of this chapter is to relate properties of the SAR backscatter curves, in particular the VH/VV-ratio, to the plant growth dynamics and other observations within wheat fields, using field observations from the 2018 and 2019 growing seasons. The correlation of crop performance indicators (CPIs) with crop properties such as biomass and plant water content can help in the interpretation of the characteristics of the SAR temporal profiles and provide information about the variability of crop performance/status within fields. The first step is to define the sample approach and type of measurements appropriate for the analysis based on SAR sensitivity biophysical properties of the crop. To do this, it will be helpful to detangle the contributions from the structure, soil and water in the polarisation temporal signatures (VH/VV ratio) across the different crop development stages. The second step is the characterisation of the SAR temporal curve based on *in situ* measurements and crop growth trends during the season. At the same time, a third step is to examine the sensitivity of VH, VV polarisations and the ratio to changes in the crop characteristics. The last step is to determine the sensitivity of the CPIs to biological parameters and help to identify CPIs that can be used to approximate within field observations. This would indicate possible CPI candidates that could potentially be combined to increase the correlation with crop productivity. Part of this objective will also examine the different scale effects on the CPIs calculation (field average and sub-field average).

3.2 Materials and Methods

3.2.1 Study site

In total, nine winter wheat fields (Figure 3-1) were selected from the Rothamsted Research experimental farm (51°48'37.3"N 0°22'36.0"W) for the analysis of SAR backscatter interaction with vegetation during the 2018 and 2019 agronomic seasons (four fields were selected in 2018, and six fields in 2019, with one field used in both years; Table 3-1). The soil texture in the area comprises slightly acid loamy and clayey soils with restricted drainage (Cranfield University, 2021). The main soil series at Rothamsted is the Batcombe series (Avery and Catt, 1995). The sow and harvest date don't display huge variation as the common timings are September (>DOY 270) and August (DOY> 215), respectively.

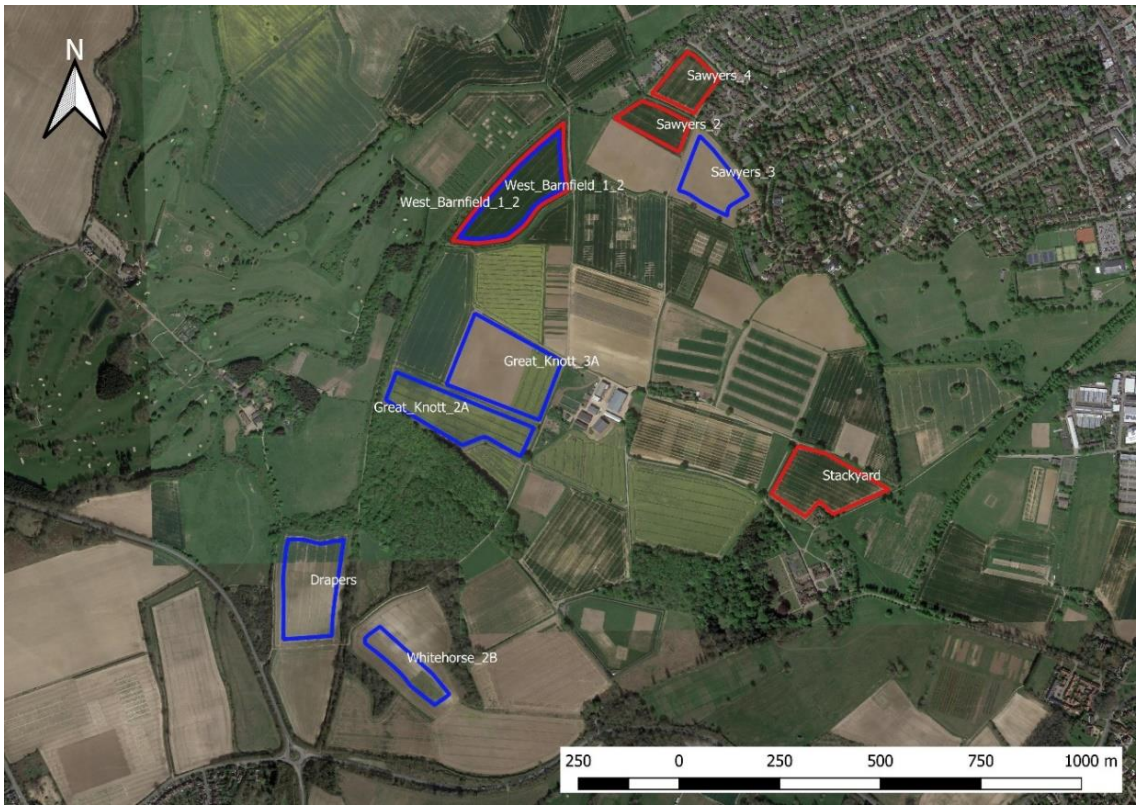


Figure 3-1 Google Earth image of Rothamsted Research experimental farm, with wheat field boundaries from 2018 (red) and 2019 (Blue)

Table 3-1 Field names and management information from the Rothamsted experimental farm; *Nitram fertiliser with 34.5% N;

Field name	Harvest season	Wheat variety	Density seeds/m^2	Applied N* kg/ha	Previous crop	Yield (t/ha)
Sawyers_2	2018	Siskin	300	110	W. Wheat	6.6
Sawyers_4	2018	Siskin	300	110	W. Wheat	5.5
Stackyard	2018	Siskin	350	100	W. Wheat	4.8
West_Barnfield1_2	2018	Siskin	300	120	W Oilseed	7.9
West_Barnfield1_2	2019	Siskin	350	180	W. Wheat	8.3
Sawyers_3	2019	Graham	325	180	S. Barley	8.0
Whitehorse_2	2019	Graham	325	150	W. Beans	10.2
Great_knott_3	2019	Graham	325	150	W Oilseed	11.0
Great_knott_2	2019	Graham	325	150	W Oilseed	11.0
Drapers	2019	Graham	325	150	W Oilseed	9.7

3.2.2 Field data and observations

The selection of the monitoring variables in the field was based on those variables identified in the literature (Section 1.3) as able to capture the dynamics of wheat development (Bargiel, 2017; Harfenmeister, Spengler and Weltzien, 2019; Mandal et al., 2020; Veloso et al., 2017; Vreugdenhil et al., 2018). The six selected variables were canopy height, crop density (plants and tillers per square metre), growth stage, leaf area index, biomass and plant water content. Field measurements were taken to characterise the field status and derive correlations with the characteristics of the crop development. The temporal sampling was designed where possible to enable non-destructive procedures and to consider the time requirements. GPS points were taken for each sample point. Ground truth data were collected during the vegetation periods of 2017/2018 and 2018/2019. Information about the measurements collected from the fields are presented in Table 3-2 and Figure 3-2. The field monitoring was based on weekly observations of the crop development at a defined “phenology point” in each field. The weekly assessment of the wheat involved crop development stage on a BBCH scale (Lancashire et al., 1991) and height measurement. Every 2-3 weeks, based on the change in the main growth stages of the crop, biomass measurements were taken in five positions across the field (Patel, Srivastava and Navalgund, 2006), resembling a “W” pattern, for eight dates between tillering and harvest. Random sampling was carried out

within an approximately 15 m radius around a GPS-marked position, with five sample replicates. The radius varies in the literature from 10 m (Harfenmeister, Spengler and Weltzien, 2019) to 20 m (Vreugdenhil et al., 2018). After the sample collection, three variables are measured in the lab, including wet, dry biomass, to calculate the plant water content. A defined set of measurement points on winter wheat fields were monitored from the end of March until the harvest. Leaf area index (LAI) was measured in May at the phenological stage of “booting” when all the leaves are fully grown, the flag leaf is fully unfolded, and the ear is about to emerge. Five plants were sampled randomly at each point to reduce the disruption in the sampling position across the season and taking between 20 to 25 tillers per sample. The tiller density (No. tillers/m * rows/m) and the dry biomass per tiller (*TWT*) measured in the laboratory (g/tiller) were used to calculate the above ground biomass (AGB) per m^2 (Equation 3-1)

$$\mathrm{AGB}\left(\frac{\mathrm{g\,DM}}{\mathrm{m2}}\right) = tiller\ density\left(\frac{No.tiller}{\mathrm{m2}}\right) * TWT\left(\frac{\mathrm{g\,DM}}{\mathrm{tiller}}\right) \qquad (3\text{-}1)$$

The plant water content was measured from the measurements of the fresh above ground biomass (Fresh AGB) and the dry above ground biomass (Dry AGB) (Equation 3-2).

$$Plant\ Water\ Content\ = \frac{Fresh\ AGB\left(\frac{\mathrm{g}}{\mathrm{m2}}\right) - Dry\ AGB\left(\frac{\mathrm{g}}{\mathrm{m2}}\right)}{Fresh\ AGB\left(\frac{\mathrm{g}}{\mathrm{m2}}\right)} \qquad (3\text{-}2)$$

During the field campaigns, various measurements were taken on five subplots in each field between late March until shortly before harvest. Plant height and crop density and the phenological stages of plant development (Lancashire et al., 1991) were determined weekly. The plant material, sampled eight times, was dried in an oven at 80°C for at least 36 hours and the AGB dry matter of each plot was determined (Equation 3-1). The dry seed matter determines the final yield at each point on the samples before the harvest. The farm management also supplies the average yield value in t/ha after the machine harvest.

Table 3-2 Description of the field observation types the frequency and replicates as well as the instruments used during the fieldwork

Variable	Units	Instrument	Frequency	Replicates	Description / Reference
Canopy height	cm	Tape	Weekly	5	Vertical assessment of canopy
Crop density	Tillers m^{-2}	Manual	End of season (3 times)	3	No. of tillers in 1 m * No. of rows in 1 m width
Phenological stage	BBCH scale	Optical	Weekly	5	The development stage of the crop
Leaf Area Index (LAI)	m^2 m^{-2}	Sunscan	At booting stage	5	
Aboveground biomass (AGB)	g m^{-2}	Scale, Oven	2-3 weeks	5 plants per sample	Fresh & Dry matter (FM, DM) from 5 plants
Yield	t ha^{-1}	Scale and harvest machinery	End of Season	5 plants per sample	Dry seeds biomass/ combine harvest machinery
Plant Water Content (PWC)	%	Oven	2-3 weeks	5 plants per sample	Derived from biomass at 80°C for at least 36 hr

Figure 3-2 Sample collection from Sawyers_3 field creating a W shape across the field with the total samples (blue) from different dates. The centroid (red) is calculated by the repeated blue dots assigned in each point (red). This centroid provides the area with a radius of 15 m creating plots (orange) to extract the SAR polarisations for the individual points in the field. The field is buffed 10 m to reduce the effect of the field boundaries in SAR.

A summary of the description of the development stages can be found in (Earth Observation and Research Branch Team, 2011):

Foundation phase

1. **Leaf development (GS 10-19):** the first leaf has broken through the coleoptile
2. **Tillering (GS 20-29):** When the fourth leaf emerges from the stem that starts the development of primary tillers

Construction phase

3. **Stem elongation (GS 30-39):** This stage is the period by which internodes lengthen the tillers. The monitoring procedure is based on counting the nodes on the main stem felt by hand. Close to the end, the flag leaf is emerged at the top as the last leaf to develop (least three nodes detected).
4. **Booting (GS 41-49):** At the beginning the flag leaf extends and then starts to swell until the sheath opens, exposing the seed head. This stages ends when the awns are visible.

5. **Inflorescence emergence (GS 50-59):** This is the period when the head emerges at the top of the stems exposing the seeds.

Production phase

6. **Flowering/Anthesis (GS 61-69):** The presence of anthers signals the flowering period, starting days after the full emerge of the seed head.
7. **Development of fruit (GS 71-79):** This period is characterised as early milk stage grains green colour, and lot of water accumulated.
8. **Ripening (GS 83-89):** Soft dough is created in the kernels and impressions done with a fingernail will be held less as the time progress
9. **Senescence/Maturity (GS 92-99):** Very hard seeds that signal the harvest period.

Figure 3-3 Key growth stages (GS) observed in the fields considered in this study across the growing season and the three phases, foundation (red), construction (green) and production (blue) phase.

3.2.3 Satellite data and time series

The same procedure was followed for the pre-processing data as in section 2.2.4 to produce the VH/VV time series. Both Sentinel 1A and 1B satellites were used and the temporal evolution of the signal was created keeping the same orbit (132, ascending) to eliminate inconsistency with the incident angle and the time of passing. That means the revisit time of each satellite is around 12 days, and the temporal repeat frequency of the two satellites will be 6 days.

Temporal filtering using smoothing (Savitzky and Golay, 1964) was used to reduce the effect of local weather events, such as heavy rainfall, ice or snow, that can temporarily change the SAR backscatter intensity. SAR backscatter data is sensitive to the presence of water in the surface of soil and vegetation that results in changes in the amplitude of the signal. It was also vital to recognise the time-series trends and define the development stages of the crop.

3.3 Results

3.3.1 Variability of field measurements

Fieldwork was crucial to understand development stages and how the temporal evolution of the SAR backscatter responds to the changes in the wheat structure. The results from the fields showed the variability between the fields and the 2017-2018 and 2018-2019 growing seasons. In the rest of the chapter, the seasons are referred to by the harvest year.

In Figure 3-4, the mean LAI recorded across the ten sample sites in either 2018 or 2019 are shown in Figure 3-4a. The LAI at the booting stage (Figure 3-4a) across the ten fields follows a pattern similar to the tiller density (Figure 3-4b) and the mass of the leaf in April (Figure 3-5a). For example, in 2018, the lowest LAI and tiller density values were recorded in Sawyers 4, and the highest values were in West Barnfield. It is also worth mentioning that the wheat at West Barnfield was cultivated after oilseed rape (OSR), and hence, the high biomass in April (Figure 3-5) may be related to a residual nitrogen effect. Further examination of this effect needs more fields across farms and years to properly display the management influence on the crop.

There was a contrast between the LAI and tiller density values for West Barnfield between 2018 and 2019, with lower values in 2019 than 2018. By contrast, the general trend was for higher LAI and tiller density values in 2019 than 2018. At the construction phase, LAI (Figure 3-4a) and biomass (Figure 3-5a) display lower values in 2018. The slow development is reflected in the grain measurements in Figure 3-5b, with the fields in 2019 having higher values in comparison. The fields of 2019 displayed higher water content values (g m^{-2}) in the crop (with the exception of Sawyer_3) that can be explained by the dry conditions affecting the 2018 observations.

The comparison between fields within the same season versus the same field across the two years can be helpful to understand the sensitivity of SAR backscatter to changes in the monitored fields.

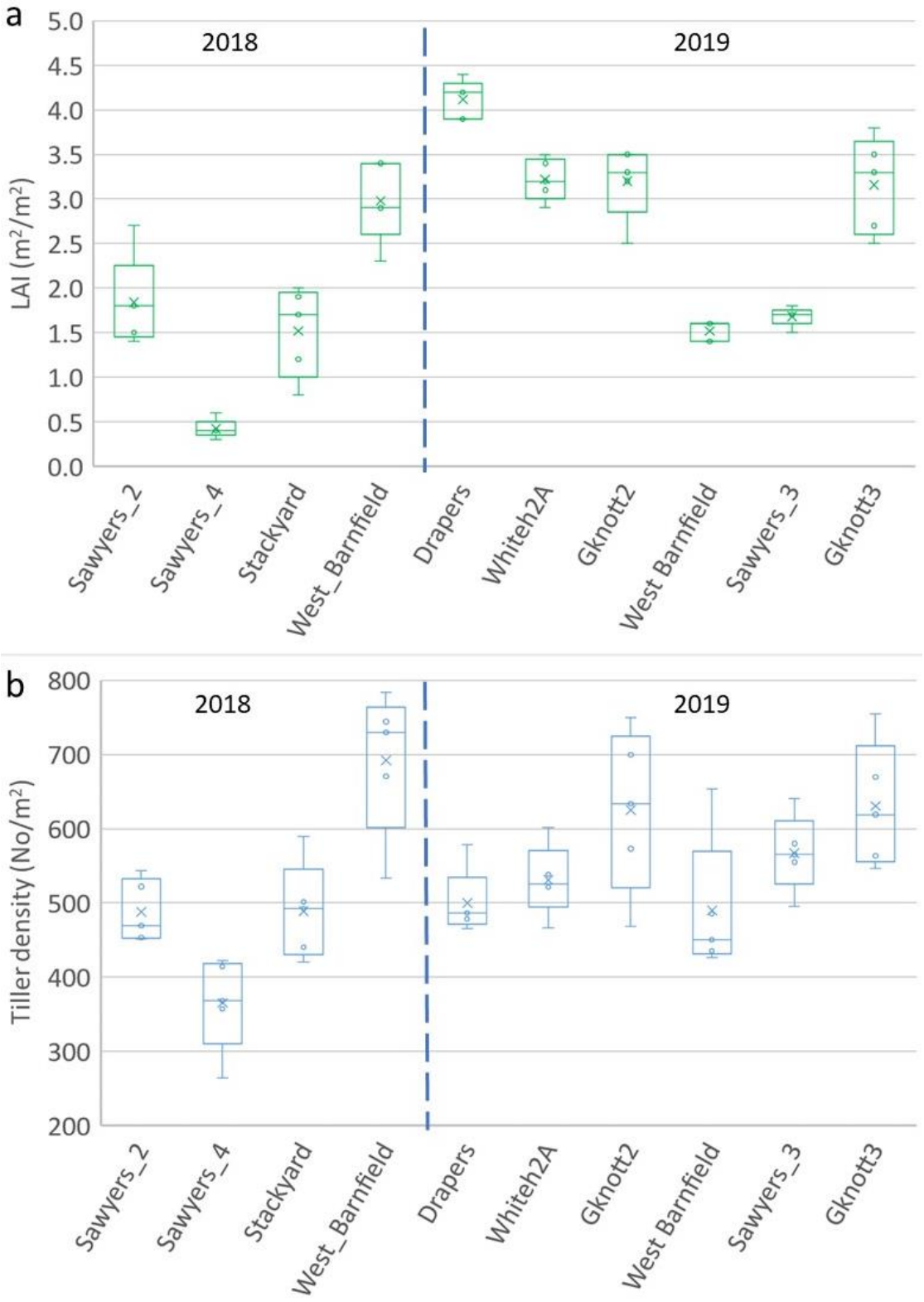


Figure 3-4 Mean values and associated variability of a) the LAI at the booting stage and b) the tiller density at the maturation phase in selected fields across two seasons: 2018 (Sawyers 2, Sawyers_4, Stackyard and West Barnfield) and 2019 (Drapers, Whitehall, Gknott 2 and 3, Sawyers_3, and West Barnfield) using the five sample points per field.

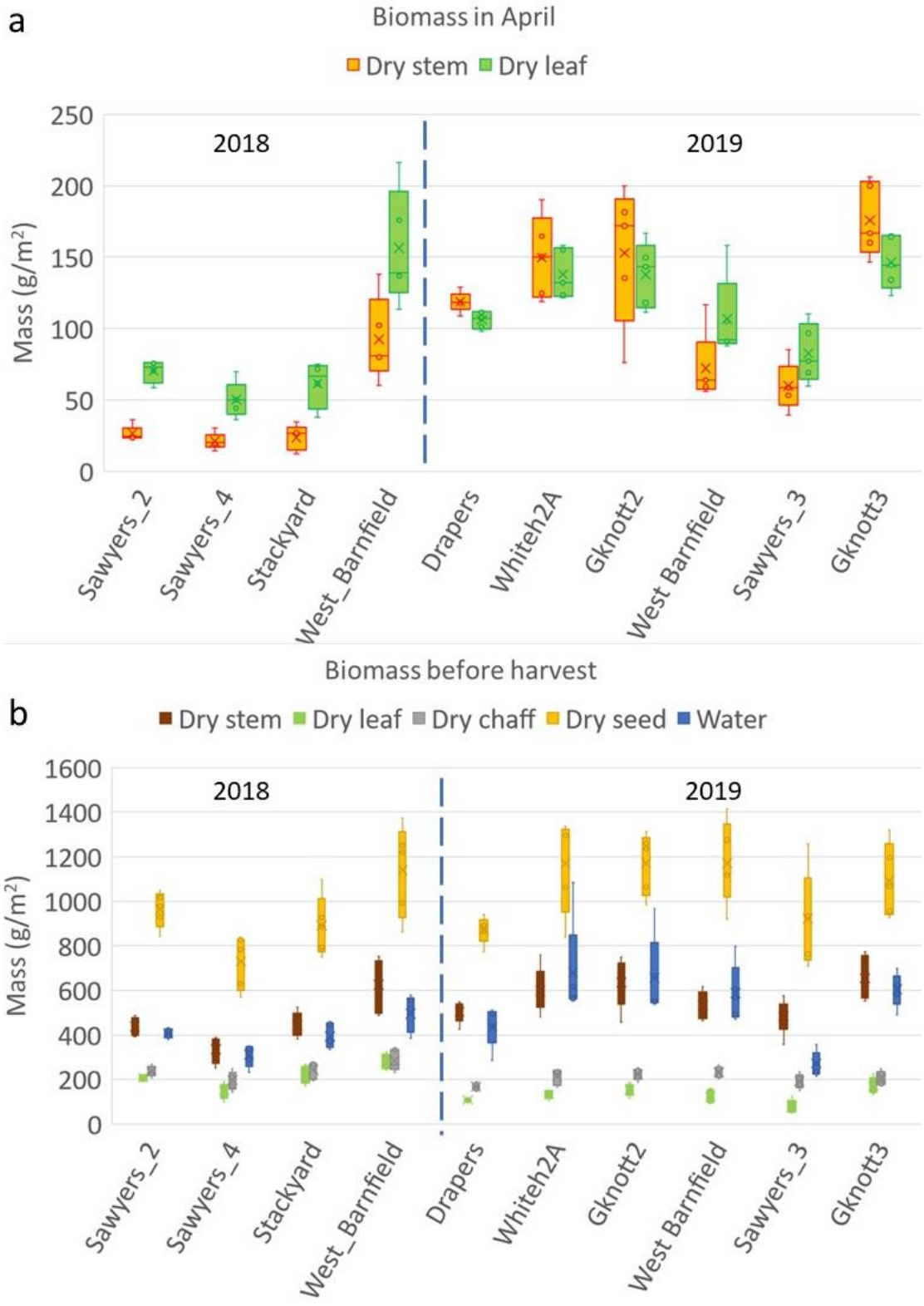


Figure 3-5 Mean values and associated variability of dry biomass distribution between different components at a) stem elongation in April and b) before harvest across two seasons (2018, 2019) using the five sample points per field.

3.3.2 SAR temporal development for the different polarisations and their response based on key crop development stages

The description of the main stages of development for cereals, based on the BBCH scale, can be used to identify the main stages of development for wheat. Figure 3-3 displays pictures from the fieldwork across the season to better illustrate the changes of the structure in the same common wheat field in the two seasons that was considered in this case. The timing of the development stages was broadly the same across the farm in each year as the same management was applied in the commercial fields each growing season.

The temporal evolution of the backscatter from the three different polarisations (i.e. VH, VV, VH/VV) of Sentinel-1 SAR (Figure 3-6) can be related to the development stage (Harfenmeister, Spengler and Weltzien, 2019; Veloso et al., 2017). Observations that are part of the plant structure as well as describe the water availability in the system are also displayed at the bottom of the figure. Figure 3-6 shows that both VH and VV polarisations are reduced in value during the tillering period up to BBCH 31 (start of stem elongation). On the other hand, the VH/VV ratio increases during the same period of time, indicating that VV is decreasing faster than the VH. This can be due to the tillers changing the surface of the field. The VH/VV ratio and the VV polarisation reach their highest and lowest values, respectively, close to the booting stage (BBCH 41). Plant height and biomass follows a similar pattern.

The same trend continues for the field observations up to the beginning of the flowering stage (BBCH 61), but there is a stabilisation of the VH/VV ratio that is mainly related to a change in the top layer of the canopy. The ear emerges and the change in the direction of the flag leaf from vertical to be more inclined to the ground as well as the increased moisture accumulated in the grain, affect the VV sensitivity to water balance. Eventually, the wheat structure reaches its final form at the end of the flowering stage (BBCH=70). After this point, the principal canopy change is a reduction of the plant water content, which is associated with a reduction in the VH/VV ratio.

A difference in the VH polarisation was observed between the two seasons. In 2018, the maturation phase displayed lower values compared to the following year. The reduced water content in the dry summer of 2018 affected the structure of the top layer of the canopy. Harfenmeister et al. (2019) have displayed a positive correlation of VH with the PWC% (R^2 = 0.51). The change can be seen in Figure 3-7 and a similar pattern was noticed in other fields in the same years.

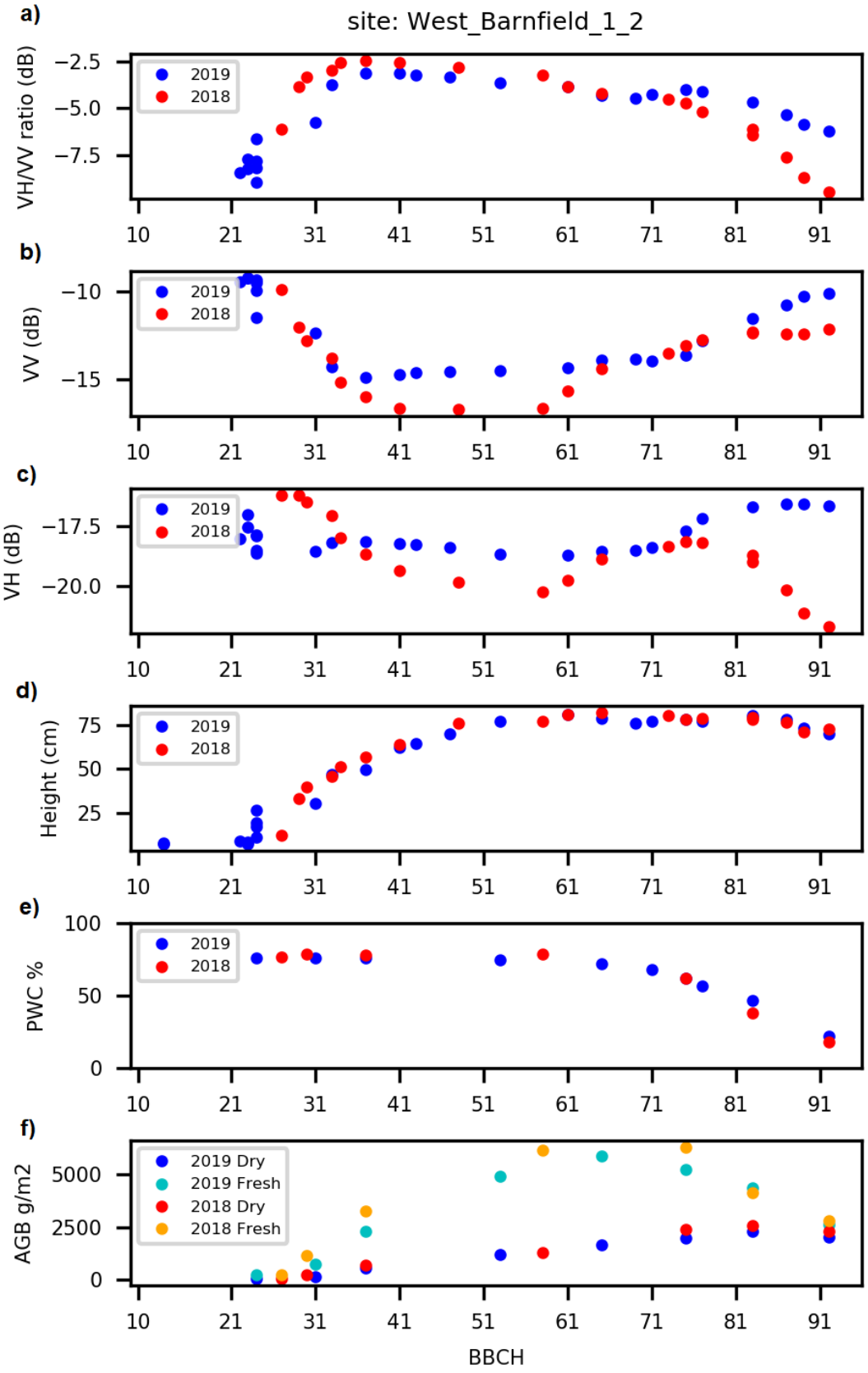


Figure 3-6 SAR backscatter polarisations of a) VH/VV, b) VV, c) VH) and field measurements of d) height, e) plant water content (PWC), and f) above ground biomass (AGB) against growth stages on the BBCH scale for the West Barnfield field for two periods: 2017-18 and 2018-19.

Figure 3-7 images at the ripening stage of wheat between the two seasons of the same field (West Barnfield) displaying the difference in colour, development stage and ear maturation level as 2018 matured a week earlier, is more yellow and ear is more inclined to the ground than 2019.

3.3.3 Detection of annual differences in vegetation dynamics and yield formation based on the VH/VV ratio

The differences in the dynamics of VH/VV ratio curves can be noticed in the West Barnfield wheat field between 2018 & 2019 (Figure 3-8). West Barnfield was sown with wheat in both years near the 1st of October and harvested at the beginning of August. Table 3-3 describes the key development stages of these crops on the same field and the dates in order to compare the development. Figure 3-8 shows the time shift in crop development due to different seasonal weather patterns and differences in agricultural management. Drought conditions at the end of the 2018 (Figure 2-2) growing season affected the crop development and yield. The field observations of the development stage for the two seasons presented in Table 3-3 shows that the difference in growing degree days (GGD) happens after the flowering period (DOY>150). At the same time, Figure 3-8 displays the earlier maturation in the warmer season (2018) using the VH/VV ratio as an independent observation.

The two smoothed curves display similar development taking into consideration the midpoint of the growth curve, the booting stage (Max) and the inflorescence (DOY 150). The canopy structure is also the same during this period as canopy height, biomass and PWC% in Figure 3-8 are at similar values. The yield difference between the two consecutive years in this field was 0.4 t/ha.

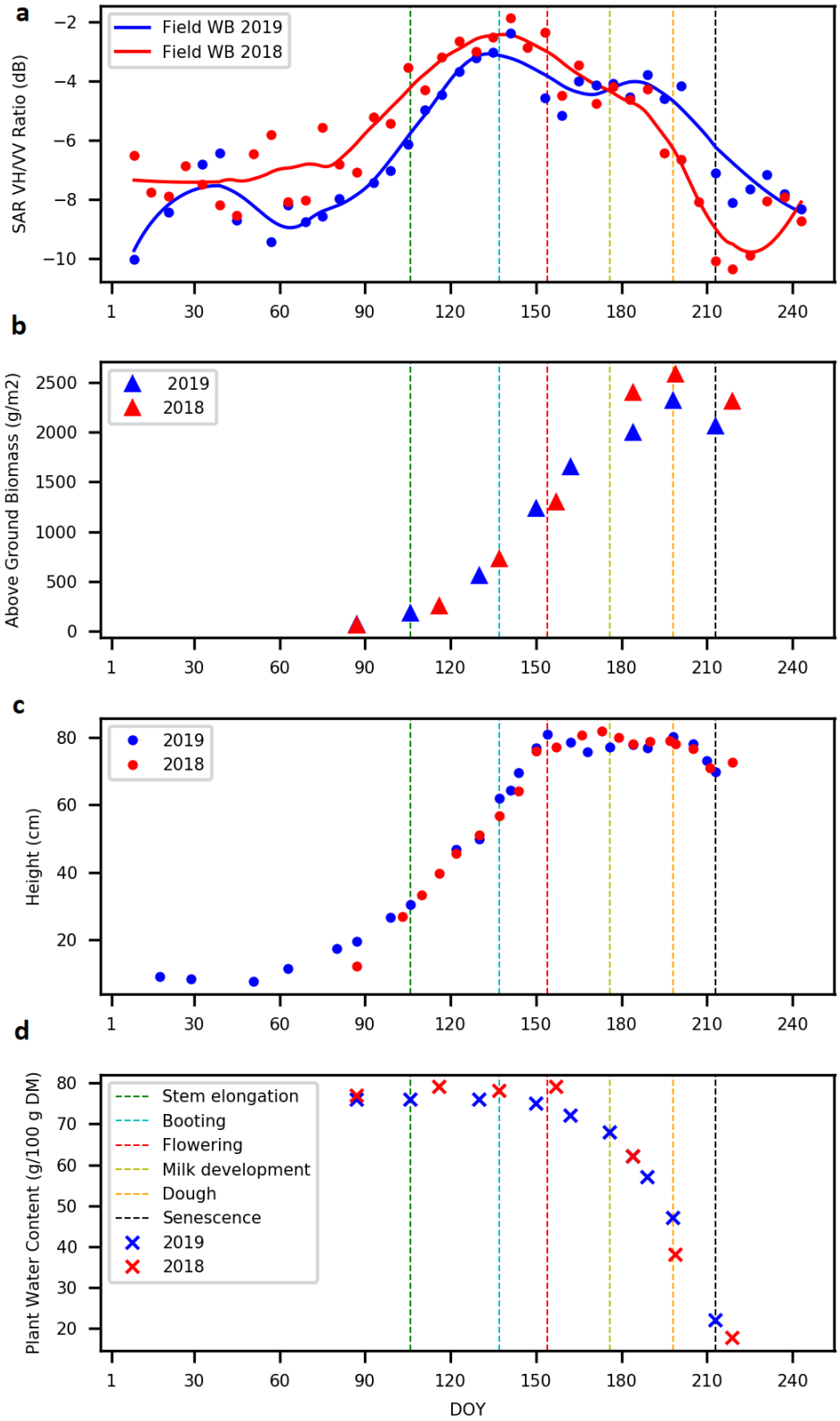


Figure 3-8 a) Means of SAR backscatter polarisation ratio from Sentinel-1 for 2018 (**red**) and 2019 (**blue**) together with the smoothed lines in West Barnfield and b) field measurements for AGB (▲), c) Canopy Height (o) and d) PWC (×) for the experimental in relation to the wheat phenological development characterised below Table 3-3 -

Table 3-3 Key crop development stages of winter wheat of “West Barnfield” site in the two growing seasons and the final yield provided at harvest

Info	BBCH scale	2018 (DOY)	2019 (DOY)
Drilling (autumn)	01	27/09/2017	09/10/2018
Start of Stem Elongation	31	19/04 (109)	16/04 (106)
Booting	41	17/05 (137)	17/05 (137)
Flowering	61	08/06 (159)	03/06 (154)
Fruit development	71	21/06 (172)	25/06 (176)
Ripening (Dough)	81	09/07 (190)	17/07 (198)
Senescence	91	31/07 (212)	01/08 (213)
Yield (t/ha)	-	7.9	8.3

3.3.4 Dynamics of VH/VV-ratio time series on fields from the same year with different productivity

As the backscatter is mainly affected by the stage of the crop structure as well as the water content of the object, the time series of observations and backscatter can give an insight into the crop development. Considering 2018 two fields with contrasting biomass measurements and final yields (West Barnfield and Sawyers 4) were compared (Figure 3-9b). West Barnfield had a yield of 7.9 t ha^{-1}, whereas Sawyers 4 had a yield of 5.5 t ha^{-1}. The ratio at the beginning of the season for both fields is mainly affected by the weather conditions in the area that set an average value of -8 dB. The value of VH/VV remains at a similar value until the tillering period in winter and more significantly, the signal changes when the crop enters the stem elongation stage in April (DOY≈107). The change in volume of the target increases the ratio, as it is noticed in Figure 3-8 until it arrives at the booting stage in May (DOY≈137). The difference in biomass between the average values (Δ) (Figure 3-9b) does not translate to the difference in the backscatter throughout the season (Figure 3-9a). This could be a result of the height being similar for both fields, which define the volume (Figure 3-9c) as well as the plant water content was at a high level (78%) during the same period (Figure 3-9d). However, the highest values of the VH/VW ratio was observed in West Barnfield (blue) during the growth period, and this could be a response of a higher density compared to Sawyers 4 (Figure 3-4) as the development stage is the same across all fields and positions in the same year (Figure 3-10). The CPI describing the max value of the VH/VV ratio (G_max) is significantly correlated with the tiller density (Table 3-6). After anthesis, the AGB difference is stabilised as well as the canopy height arrives at its maximum during the summer. The evolution of the ratio curve follows the reduction of the plant water content during grain development until harvest.

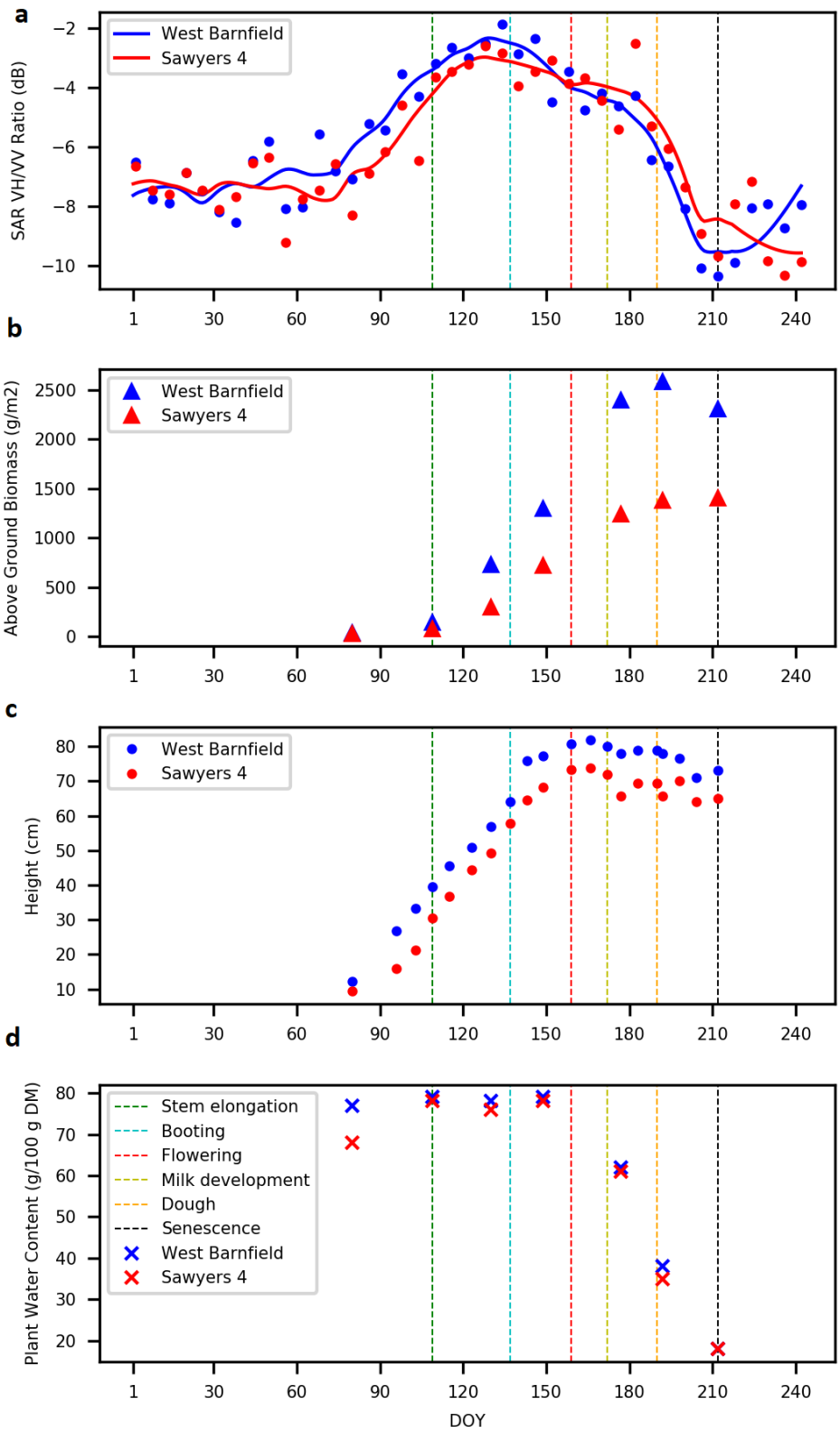


Figure 3-9 Field data and temporal evolution of SAR using the average of two wheat fields in 2018, West Barnfield (blue) and Sawyers 4 (red), with a significant difference in biomass. Field measurements for b) above ground biomass (AGB), c) canopy height, and d) plant water content together with the development stages (vertical lines), Temporal VH/VV ratio SAR smoothed with the development stages (down)

The relation between smoothed VH/VV ratio curve and the canopy height of the two fields is displayed in the Figure 3-11. A positive correlation was observed during stem elongation. After booting, although the height continues to increase, this is not associated with an increase in the VH/VV ratio. As the crop loses water during grain filling, the penetration increases and the VH/VV ratio declines.

3.3.5 Relationship between CPIs and biophysical variables

3.3.5.1 Scale effect on CPIs

The different spatial level at which SAR data is averaged (i.e. field or plot scale) plays a crucial role in the variability in the VH/VV temporal curves, because as more pixels are averaged there will be less noise due to speckle. Table 3-4 compares the field level averaged values of the data from Section 2.3.1. These CPI calculations are based on two different scales, a) the field average from five fields in 2018 (mean area = 2.2 ha) and six fields in 2019 (mean area = 2.4 ha) and b) corresponding sample point average (derived from the surrounding area of the 15 m, area = 0.07 ha).

On average the mean coefficient of variation for the CPIs in 2018 was 32% when derived at a field level and 40% when derived at a within field level. In 2019, the corresponding mean coefficients of variation were 25% at a field level and 36% for within field samples. (Table 3-4). The results show that working at the plot scale (average of 6-9 pixels) the variability of the CPIs is greater than the field scale (average of all pixels in the field). However, for some CPIs, the opposite was true such as the steepness of the senescence period (S_steep) and the D_max value which characterise the structural change of the wheat between booting and flowering.

For most of the CPIs, the mean values were not that different when comparing the two different scales for the same growing season (Table 3-4). The important differences ($p<0.001$, Table 3-5) can be observed in time and development stages related indicators that field observations showed the same development across the farm, but the standard deviation is around nine days for field and 15 days for the plot size derived values. These significant differences were noticeable in both scales when the two different years were compared. The duration, as well as the midpoint of maturation (S_midP), had significant differences ($p<0.001$) that were not displayed when the scales were compared each year. The timing of the booting stage displayed no significant difference in all comparisons on Table 3-5 except the field level comparison between the two years ($p<0.05$). Similarly, G_max shown no significant difference across all comparisons.

Table 3-4 Comparisons of the two scales (field and plot) between the CPIs derived from a) 2018 using the mean values of VH/VV for five fields and of VH/VV values for five plots within five fields, and b) 2019 using the mean values of VH/VV for six fields and of VH/VV for five plots (15 m radius circle, approximate 7 pixels) within the six fields , by displaying the mean, standard deviation (SD) and coefficient of variation (CV). The yield measurement for the field was derived from the combine harvester and for the within field samples, the yields was derived from the hand harvesting of individual plants before harvest. N is number of observations and A is the mean area (ha) of the sampling unit.

Year	SAR Sampling Unit	Statistics	G_base	G_steep	G_max	S_max	S_steep	S_base	G_midP DoY	S_midP DoY	TZmax DoY	Duration days	D_max	D_base	Yield t/ha
2018	Field	Mean	0.14	0.1	0.38	0.3	-0.21	0.1	111	192	140	81	0.08	0.04	5.77
	N=5	SD	0.04	0.02	0.09	0.06	0.14	0.02	8	4	8	10	0.08	0.03	1.5
	A =2.2	CV	29%	20%	24%	20%	67%	20%	7%	2%	6%	12%	100%	75%	26%
	Plot (within	Mean	0.16	0.12	0.36	0.30	-0.12	0.10	103	192	137	89	0.06	0.07	9.3
	field)	SD	0.04	0.06	0.11	0.11	0.03	0.02	9	8	9	11	0.12	0.04	2.0
	N=25 A=0.07	CV	22%	49%	31%	36%	26%	24%	8%	4%	7%	12%	204%	54%	22%
2019	Field	Mean	0.13	0.08	0.4	0.24	-0.2	0.13	107	208	131	101	0.16	0.01	9.02
	N=6	SD	0.02	0.03	0.06	0.01	0.12	0.01	7	3	2	7	0.07	0.01	1
	A =2.4	CV	15%	38%	15%	4%	60%	8%	7%	1%	2%	7%	44%	100%	11%
	Plot (within	Mean	0.15	0.12	0.37	0.25	-0.13	0.11	105	214	134	109	0.12	0.07	10.7
	field)	SD	0.04	0.06	0.11	0.06	0.03	0.04	15	11	15	21	0.13	0.05	2.0
	N=30 A=0.07	CV	26%	50%	30%	26%	25%	36%	14%	5%	11%	19%	109%	78%	19%

Table 3-5 Levels of significant difference: 0.05<p<0.1(o), 0.01<p<0.05(*), 0.001<p<0.01 (**) and p<0.001 (***) between field and plot scale in each year and between the two years on the same scale

	G_base	G_steep	G_max	S_max	S_steep	S_base	G_midP	S_midP	TZmax	Duration	D_max	D_base	Yield
Field vs plot in 2018							o					o	***
Field vs plot in 2019	o	*				*		*				***	**
18 vs 19 Field level				o		*		***	*	**		o	**
18 vs 19 plot level				*				***		***	o		*

3.3.5.2 CPIs correlation with biophysical properties of wheat

The results in Chapter 2 show that specific indicators, such as the duration of the high vegetation (stem elongation to dough development) and the timing of the booting stage significantly correlate with the final yield at a field scale (Chapter 2.5). This chapter focuses on the use of the 50 sample points (within and across fields and seasons) to correlate specific field variables with the 12 CPIs. These CPIs were calculated for each point extracted using the average values of all pixels in a radius of 15 m (as pixel size is 10 m) around the centroid of the sample (Figure 3-2).

Despite the higher variation when the CPIs were calculated based on the smaller averaging area, in this analysis, there was a significant level of correlations in Table 3-6. The correlations (r) were displayed and the significant values are coloured red or green. The information that can be extracted by these correlations can be described using the two periods defined to calculate the CPIs. The first five CPIs describe the growing period (i.e. foundation and construction phase) and the 6 to 9 production phase. The last three (10 to 12) are a combination of parameters from the two periods.

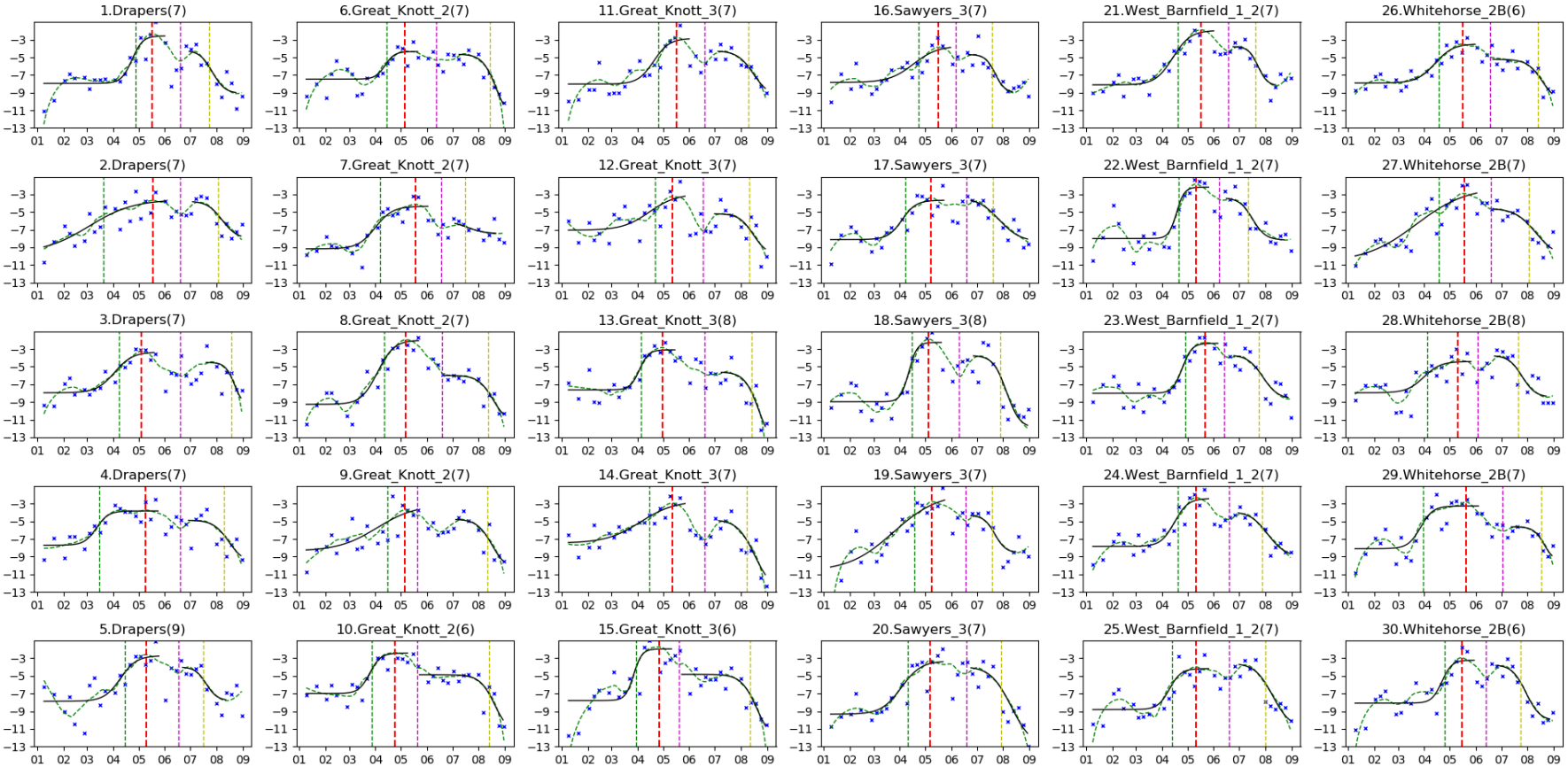


Figure 3-10 CPI calculation using the two logistic curve fitting for 2019 harvest season. Every column represents the same field with its five sample positions and the parenthesis displays the number of pixel used to average the VH/VV ratio.

As shown in Table 3-6, tiller density is significantly correlated ($p<0.05$) with the G_max (maximum. value in growing period) with $r = 0.39$, as well as the D_max (the difference between the two maxima of the growth and maturation period) with $r = 0.32$. This latter response is driven by G_max as the maximum value in the senescence stage, S_max is not correlated ($r = 0.01$) with tiller density. Similarly, the above ground biomass (AGB) is also positively correlated with G_max ($r = 0.32$, $p<0.05$) in May but not the final values before harvest with any CPI. These results can justify the difference noticed in Figure 3-9 where the field with higher AGB and tiller density has a higher intensity of VH/VV ratio but no difference in the maturation period as S_max is not sensitive to the AGB difference. Leaf mass in May, (which is related to the LAI and of AGB) is also positively correlated with G_max ($r = 0.34$, $p<0.05$) (Table 3-6).

The negative correlation between G_base and end of growth period leaf mass ($r = -0.34$, $p<0.05$) as well as the final canopy height at the end of the structural development in May shows the effect of tillering period on the canopy structure during the development phase. There is a correlation of the canopy height (which got the highest value after the end of May) with CPIs (No 8,10 and 11) derived from the maturation/senescence period (from $r= 0.29$ up to $r = 0.5$) of which the AGB seems to be not related with them (Table 3-6).

The senescence parameters (S_max, S_midP) are expected to be the CPIs most affected by the reduction of water content in the fields. Starting with the total vegetation water amount in the field (g/m^2 in July), there is a correlation first with the CPI G_steep referring to the steepness of the growth curve. G_steep is also associated with the PWC% in July ($r = -0.29$, $p<0.05$) as the only source of water because the dry conditions of the summer do not allow the soil to remain wet in the top layer. The lower values of PWC in the maturation period affect the timing of the different stages. For example, low rainfall periods (reduce water availability) will result in earlier senescence of the plant, as shown in Figure 3-8 with the drier condition in 2018 compared to 2019. The S_midP and duration display a positive correlation with water availability at the end of the season ($r = 0.46$ and $r= 0.29$ respectively with $p<0.05$). This is in contrast to the water content at the end of the growing season, which shows a significant negative correlation with these two parameters $r= -0.76$ and $r= -0.54$ with $p<0.05$. This effect might result from increased development up to May (as PWC is almost 80% of the total AGB), which restricts the water later in the season, especially during an arid summer, when water availability is limited for the crop. This can also be observed through the positive correlation of S_midP

and duration with PWC in July, indicating that the maturation is observed later with higher values of S_midP and increased duration.

Despite the significant correlation of CPIs with biophysical properties of the field, none of them displayed correlation with the yield, but three indicators showed sensitivity with the harvest index (harvested seeds/AGB ratio). The negative response of G_max to harvest index was expected as this indicator displayed sensitivity to AGB at the beginning of the season. The positive correlation with the S_midP and also with duration (Duration =S_miP - G_miP), with $r= 0.39$, $p<0.05$ and $r= 0.32$, $p<0.05$ respectively, showed that the later maturation/senescence starts due to water availability and growth development, the longer the growth and grain fill duration will be (Biscoe and Willington, 1984; Monteith, 2007).

Table 3-6 Correlation coefficient (r) between CPIs and field observations (50 points in total) using both growing seasons (2018 & 2019), significant r (P level>0.05) coloured green and red. details of each indicator definition can be found in Table 2-2.

CPI No.	CPI	tillers m-2	AGB May g/m2	AGB Harvest g/m2	Leaf May g/m2	Height May	Max Height	Water July g/m2	PWC May	PWC June	PWC July	yield t/ha	Harvest Index
1	G_base	-0.16	-0.17	-0.19	-0.34	-0.06	-0.33	-0.23	0.21	0.03	-0.17	-0.18	0.00
2	G_steep	-0.19	-0.17	-0.27	-0.21	-0.38	-0.18	-0.34	-0.06	-0.02	-0.29	-0.23	0.10
3	G_midP	0.06	0.04	0.01	0.15	0.07	0.13	0.08	-0.03	0.11	0.04	-0.01	-0.06
4	G_max	0.39	0.32	0.20	0.34	0.19	0.17	0.10	0.03	-0.14	-0.03	0.10	-0.29
5	TZmax	0.05	-0.04	-0.03	-0.03	0.11	0.12	-0.15	0.06	0.05	-0.23	-0.05	-0.05
6	S_max	0.01	-0.07	-0.01	-0.02	-0.02	-0.24	-0.17	0.39	-0.05	-0.25	-0.06	-0.15
7	S_steep	0.00	-0.01	0.15	0.01	-0.15	-0.17	0.08	0.22	0.14	0.00	0.18	0.09
8	S_midP	0.09	0.38	0.13	0.19	-0.10	0.50	0.46	-0.76	0.00	0.57	0.24	0.39
9	S_base	0.00	0.04	-0.12	-0.03	0.10	0.06	-0.19	-0.02	0.12	-0.22	-0.13	-0.02
10	Duration	0.03	0.25	0.09	0.05	-0.11	0.29	0.29	-0.54	-0.07	0.39	0.18	0.32
11	D_max	0.32	0.32	0.18	0.30	0.18	0.30	0.19	-0.24	-0.08	0.14	0.12	-0.14
12	D_base	-0.11	0.01	-0.05	-0.16	-0.20	-0.17	0.04	-0.09	-0.05	0.14	-0.02	0.07

3.4 Discussion

3.4.1 Temporal variation of SAR profiles and field observations

The values of SAR backscatter can be reliably related to different development stages of wheat between sowing in October and harvesting in August (Bargiel, 2017; Harfenmeister, Spengler and Weltzien, 2019; Veloso et al., 2017). Thus, the temporal evolution of the SAR curve can be explained using distinct periods of different signal interaction, as it is affected by the plant growth development and moisture changes (Harfenmeister, Spengler and Weltzien, 2019). At the beginning of the season in the autumn, the main target in the field is bare soil and the crop height is less than the C-band wavelength of ~5.5 cm. For this reason, the values are primarily determined by the soil moisture and any surface water that remained in the field (Harfenmeister, Spengler and Weltzien, 2019; Khabbazan et al., 2019) . From March until May, the crop moves from tillering (15 cm height) to the next development stage of stem elongation (up to 65 cm). In Figure 3-6, the increase of the canopy volume and the biomass in the fields causes the decrease of the two polarisations. The VV is dominated by the wheat's vertical structure that contributes to the signal attenuation (Veloso et al., 2017; Vreugdenhil et al., 2018). The VH reduction is less than the VV in this period, as the primary influence is the volumetric scattering and double bounce produced by the crop growth (Brown et al., 2003; Nasrallah et al., 2019). During the stem elongation, the increase of the volumetric backscatter is counterbalanced by the attenuation of the double bounce. This difference between the two polarisations results in the increase of the VH/VV ratio (Khabbazan et al., 2019; Vreugdenhil et al., 2018).

The following distinct stage is the booting, which coincides with the maximum VH/VV ratio value timing (TZmax). The combination of the flag leaf and the start of ear emergence affects the direct backscatter (VV) due to the increase of the water in the top layer of the canopy as well as the structural change of the flag leaf opening (Harfenmeister, Spengler and Weltzien, 2019). The timing of max ratio value has been reached typically between the full ear emerge in mid of May and the flowering period in the first weeks of June, and subsequently, continue to reduce because of the steady increase in VV values as compared to stable VH values. Note that in chapter 2.2.5 we demonstrated that this feature in the temporal curve of the VH/VV ratio is particularly useful in order to define the point at which the max value will be reached, or in other words, the booting stage of the wheat.

The VH/VV backscatter ratio stabilizes before the crop reaches the maturation phase, which is mainly a consequence of its final structure. This stability is determined by its ear above the canopy and high values of water content in the fruit during grain-filling before it starts to reduce the VH/VV-ratio during maturation. Figure 3-11 displays how the relationship between SAR smoothed curve values and canopy height (structural characteristic) is affected by the development stage of wheat. High correlation is displayed during the stem elongation, but the sensitivity is reduced as the crop moves to maturation phase.

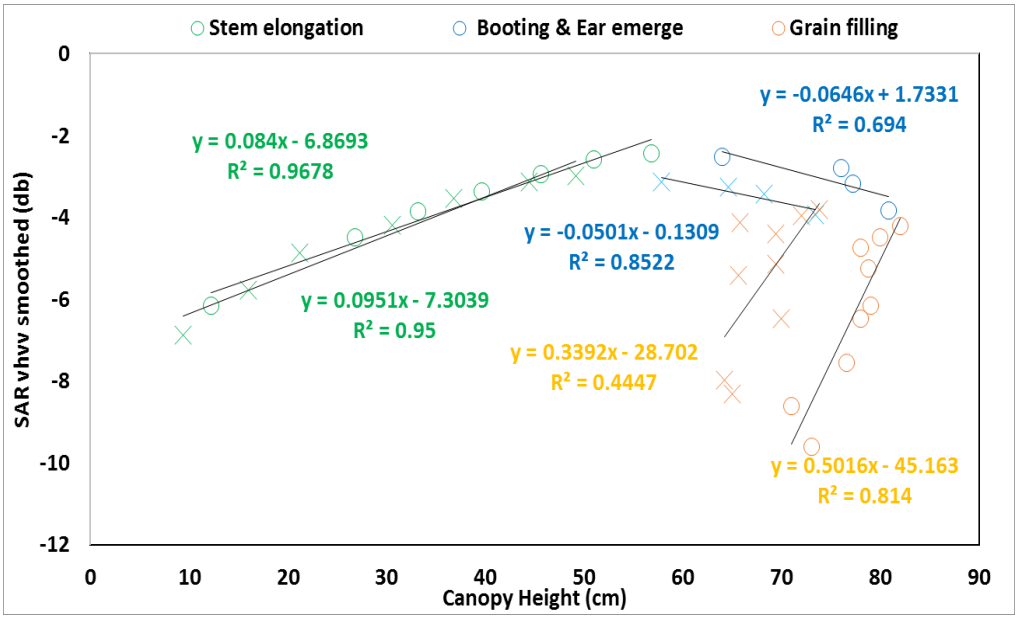


Figure 3-11 The relationship between the SAR VH/VV ratio and the canopy height for three wheat development periods: stem elongation, booting and ear emergence, and grain filling. The x and o symbols display two different fields in 2018.

The last period is during the reduction of the water content in the crop during July and August. The plant water content is significantly reduced throughout this period and the VH/VV ratio follows the same trend as the soil contributions start to affect the backscatter. The ratio falls from -5 dB to -8 dB (or lower), arriving at the winter levels. The decrease of VH values during this period in 2018, in contrast to 2019, can be justified by ear inclination to the ground because of the difference in the development stage (Harfenmeister, Spengler and Weltzien, 2019). The harvest took place at the beginning of August when the increase of the noise was noticed in the SAR data, indicating that the satellites acquire images on base soil/stubble remaining in the field.

The temporal analysis of the SAR was successful to describe changes in the structure and development stage of wheat across the farm. The polarisation trends follow the

changes of the biophysical properties of the fields, displaying potential of monitoring crops in the agricultural landscape.

3.4.2 CPIs connection to biophysical properties of wheat

The correlation of the CPIs with the field parameters is useful to identify sensitive indicators that can give information about the status of the crop using only SAR data. The analysis displayed what these indicators can tell us about the crop. The most sensitive CPI from the growing period is G_max that has a significant correlation with the tiller density and the AGB at the booting stage of the plant development. This sensitivity also gives a negative correlation with the harvest index as the denominator is the biomass regardless of the fact that the G_max is not sensitive to the AGB before harvest. The reason could be that strong early development during the construction phase will result in increased biomass at the end of the season. However, this could also be explained by the loss of leaves of the crop as being affected by the water availability and hence, reducing the connection of biomass at the end of the season.

The baseline in the foundation phase (G_base), which has a significant negative correlation with structural characteristics of leaf and height of the plant, needs to be examined further as no direct relationship could be found. The steepness of the logistic growth curve (G_steep) also interacts negatively with the height during the stem elongation and the water balance at the end of the season (July). This can be justified by the vigorous growth at the beginning of the season that can reduce the water availability closer to harvest due to increased water demand. Both midpoints of growth (G_midP) and max timing (TZmax) were connected with the different development stages as the first displays the time of stem elongation and the second the time of booting. The primary influence in crop development stages is the temperature of the area (De Bernardis et al., 2016).

The CPIs derived by the second half of the crop development (production phase) are sensitive to the canopy's water characteristics as the wheat's structural development is completed (Harfenmeister, Spengler and Weltzien, 2019). The S_max is positively correlated with the plant water content in the late stages of the construction phase ($r= 0.39$, $p<0.05$). This is important as the max in the senescence period is noticed during the flowering stage. Thus, this indicator can provide information about the water content of the plant before the start of the senescence. The next important indicator is the midpoint of the senescence (S_midP) that also refers to the timing of the dough

development in the plant and is highly correlated with the water content on that stage (r= 0.57, p<0.05), which is, in contrast, significantly negatively correlated with the water content at the end of construction phase. Combined with the correlation with the AGB in May (r= 0.38, p<0.05), these connections show an interaction of the conditions in the construction phase with the water availability at the end. High values of S_midP indicate elongation of the fruit development period, allowing more translocation of carbon in the grain (Biscoe and Willington, 1984; Monteith, 2007), which is also illustrated by the positive correlation with the harvest index. The baseline at the end of the season characterises the conditions of the field after the harvest period and there is no sensitivity with any crop parameter as expected.

The last three indicators, derived from other indicators, are highly influenced by their sensitivity to field observations. Duration is an example of the difference between the midpoints that show sensitivity to water content more than the early development of the crop. In contrast, the D max is affected by the G_max more than the S_max as the field data show more variability in AGB than the water content in the period of the construction phase.

This methodology has the potential to assist on crop model applications (Huang et al., 2019; Kasampalis et al., 2018), but it is crucial to develop further the ability to invert the SAR time series to parameters associated with the soil and crop. The most promising connection is with the development stages that can help the model better simulate development differences without the need for temperature and sowing date as being used for the growing degree days calculation (De Bernardis et al., 2016). The plant water content change can help to improve the spatial water variability and be connected with the soil properties across the farm at the field scale. Moreover, the use of CPIs as biomass proxy for each field can help to improve model simulations.

3.4.3 Sampling approach

An objective in this chapter was to understand the underlying biology of high temporal frequency SAR and its relation to yield formation, by sampling the biophysical properties of wheat fields and correlate them with CPIs. Different sample approaches have been used in literature to collect vegetation parameters. Random sampling in each field for each date was considered for height and LAI (Bousbih et al., 2017) and Patel et al. used also five positions in the field to cover as much possible area (Patel, Srivastava and Navalgund, 2006). In Bhyiyan et al. the biomass was collected using 10 plats and scale

the value using the plant density (Bhuiyan et al., 2018).Harfenmeister et al. monitor the phenology on BBCH scale, height with random plants and biomass by destructive sampling of 1m rows (Harfenmeister, Spengler and Weltzien, 2019). A combination of the above sampling approaches was used in this work. The sampling approach had two essential criteria, first was the applicability and time management that one person or the farmers can also apply for the necessary monitoring procedure. Second, the minimal disturbance of the sampling area was vital as multiple samples were taken throughout the year, which could affect the surrounding surface structure. This disturbance can potentially influence the SAR backscatter and the commercial value of the area should not be reduced by removing production from farmers' fields.

The SAR temporal curves derived from the plot scale have similar signature as the average of the field, allowing the use of the methodology from Chapter 2 to be applied across all sample points in the farm. Figure 3-12 shows the lines from 5 plots and the average of that field to compare the signatures. This result is encouraging when the description of the biophysical changes needed to be described by different observations across the field by the farmers. Figure 3-10 displayed how variable the different plot curves can be that increase the uncertainty of the CPIs (Table 3-4) but with no significant difference when most of the CPIs means between field and plot scale are compared in Table 3-5. One exception is the D_base that has significant difference ($p<0.001$) in 2019 that is clear how variable the baselines can be when less pixels were averaged, as the soil (moisture) will affect the radar response (Baghdadi et al., 2018; Bousbih et al., 2017; Gherboudj et al., 2011) . The baseline variation at plot scale is visible in Figure 3-10.

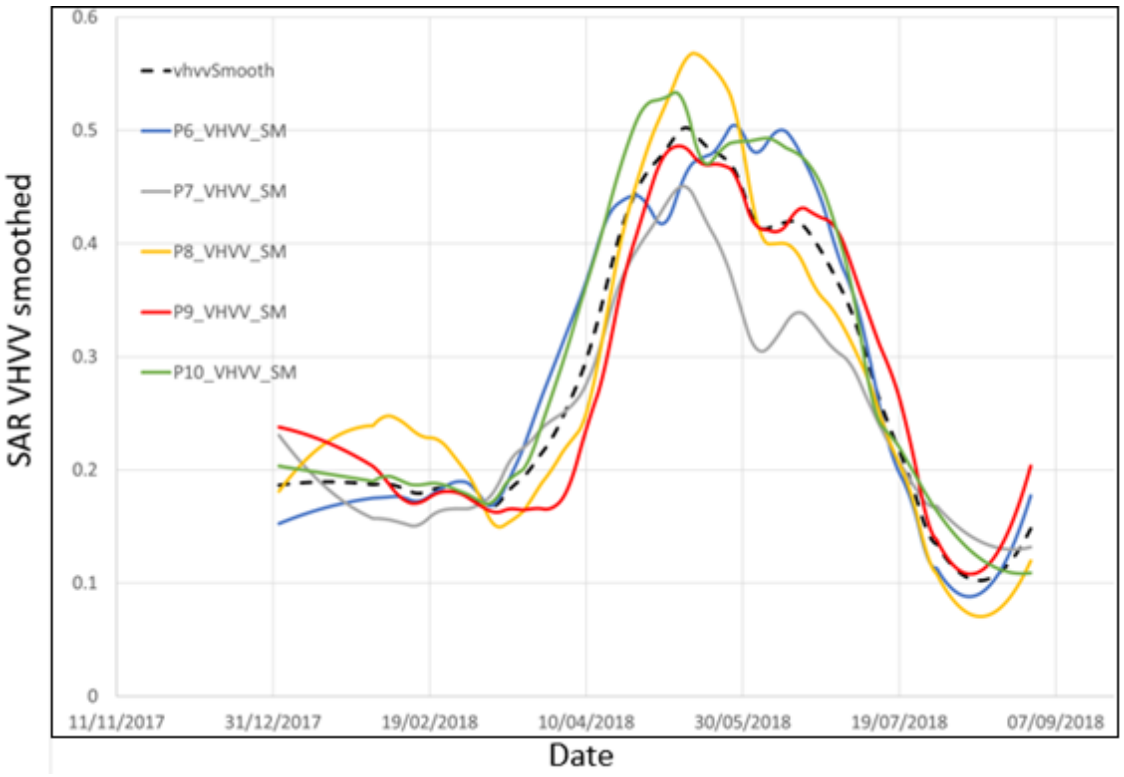


Figure 3-12 Display of the curve signature similarity of field average value of VH/VV (hatched line) smoothed curve with five individual plots smoothed lines (P6 to P9) in the same wheat field.

The correlation between the CPIs and the localised within-field yield shows no sensitivity with the field production but with the harvest index instead, in contrast to the section 2.3.2 where two CPIs (TZmax/Booting timing and duration) displayed a significant correlation (r= -0.56 and r= 0.61) with the yield. The first issue is the difference in the scale, because in this study we have been using field averaged values for the calculation of the temporal curve, which are characterized by less variability and noise as compared to point measurements where the SAR curve is the average of the pixels around the centroid of the accumulative samples in the five positions in the field (Table 3-4).

The second issue is the different sampling processes between the field average from harvest and hand sampling that resolve to different losses of biomass material. Combine harvester losses are less than 1% (Setiyono et al., 2019) and potential bias between hand and combine harvester method is shown as less than 4% of the total height resulting to 1% variation on harvest index (Dai et al., 2016). Figure 3-13 displays a comparison between the two different yield calculation processes. The low yield fields are overestimated in the field sampling procedure (Figure 3-13), affecting the correlations in Table 3-6. This can be why CPIs are sensitive to yield at field scale (duration and booting time) but are not significantly correlated in sub-field (plot) scale.

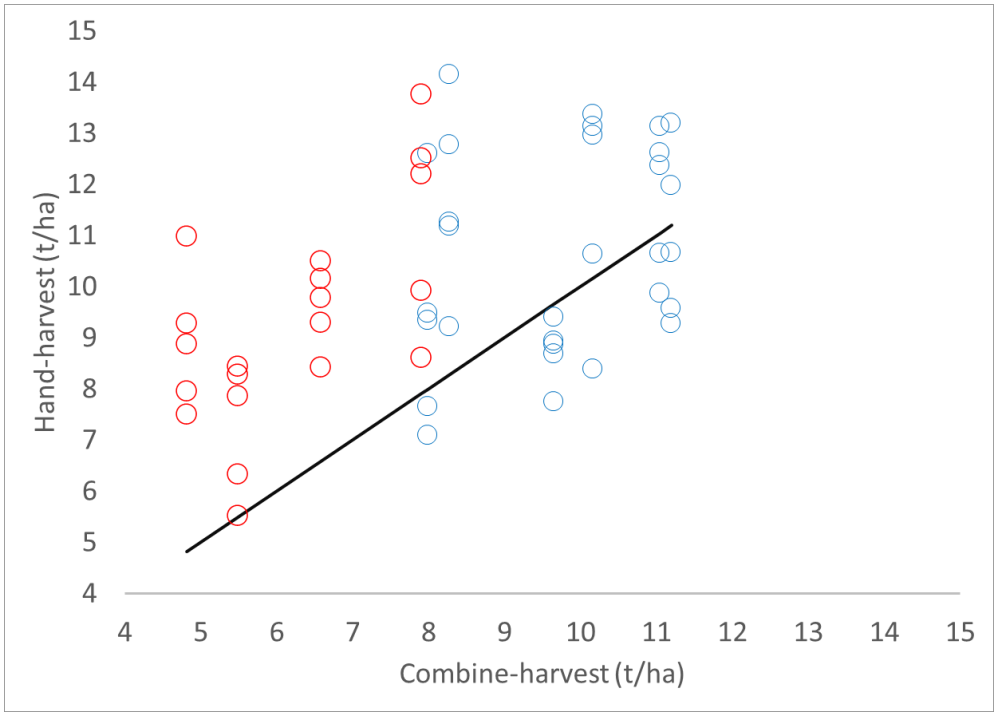


Figure 3-13 Yield comparison between the combine harvest from the machinery as field average in t/ha (x-axis) and the hand harvest in each sample point by five points per field (y-axis) as well as the 1:1 line (black) for the two different years (2018-red, 2019-blue)

The sensitivity to harvest index helps to normalise this effect as it is defined as the ratio of yield/AGB that makes it unrelated to the total amount and displays the plant efficiency simultaneously. This index displays a significant correlation with the duration of the vegetation and G_max (Table 3-6). That is expected, as in the previous chapter duration was significantly correlated (r= 0.61) to yield component of the index and G_max is sensitive to biomass (r= -0.29).

3.5 Conclusions

The formulation of distinct correlations between the SAR-derived CPIs and individual in-field observations remains challenging, as there are confounding factors that affect the correlations (scale and noise). However, the sensitivity of the temporal curve analysis across the growing season can provide helpful information about the crop's development trend (Figure 3-9). Key crop development stages are recognised using the dynamic SAR backscatter curves to monitor crops, allowing the interpretation of the growth.

The SAR backscatter ratio (VH/VV) is sensitive to changes in the crop architecture during the construction phase (stem elongation) and inflorescence (ear emergence). At the same time, VH/VV ratio is sensitive to relative plant water content (PWC %). The analysis between the two seasons displays the temporal shift between backscatter curves that could be related to crop density at the beginning of the season and the water scarcity during the summer of 2018 which made the growing season shorter than 2019 (Figure 3-8a).

Most important CPIs displaying useful interaction with the field parameters is the G_max that correlates with the structural development at the beginning of the season and S_max ($r= 0.39$, $p<0.05$) and S_midP ($r= 0.57$, $p<0.05$) as water indicators. All three parameters indicate sensitivity with the harvest index that can describe the efficiency of the plant with respect to the yield. The within-field sampling provided valuable insight into correlation to productivity, but the variation of the CPIs is much higher than the field average (Table 3-4). The results show that despite the similar average values of the CPIs the CV is higher at the plot scale as compared to the field scale (25% at a field level and 36% for plot level).

For future work, it will be useful to apply the CPIs across multiple fields and farms to increase the sample size and be able to approximate the biophysical properties in a way that this can be quantified. The latter will be helpful in order to use the CPIs in a modelling scheme.

4 EFFECT OF MANAGEMENT AND PEDOCLIMATIC CONDITIONS ON SAR-DERIVED CROP PRODUCTIVITY INDICATORS

Summary

The advantage of using spaceborne remote sensing technologies is the ability to monitor large areas in order to examine the heterogeneity due to different weather, soil and management. The objective of this chapter is to apply the CPI methodology to other areas and understand how SAR-derived CPIs relate to inter-regional and inter-seasonal yields. Two case studies help to understand the regional and seasonal effects. The first (CS1) focused on validating the multi-seasonal application in a different area (Norfolk), the second (CS2) applied the methodology across four farms with different weather and soil characteristics.

The correlation matrices displayed similar results to those at the field scale on Rothamsted Farm (Chapter 2) where CPIs were correlated with wheat yield. The two CPIs identified initially (Duration, Booting) also displayed significant correlations across four seasons in Norfolk farms (CS1) with yield (8-11 t/ha; $p < 0.05$, $n = 153$), however, weak with $r = 0.17$ and $r = -0.21$, respectively. The matrix for CS2 also showed significant correlations between yield and booting timing (TZmax $r = -0.59$) and duration ($r = 0.53$) across yields ranging from 5 to 11 t/ha ($p < 0.05$, $n = 76$). Several other CPIs had significant correlations in CS2, with $r = -0.23$ (S_max) and $r = -0.51$ (G_midP). The variation of the key CPIs was examined, and the significant difference between the four farms has been tested using the analysis of variance (ANOVA). The analysis displayed a significant effect of soil, year (weather pattern) and management (preceding crop) on SAR-derive indicators as well as the combination of year (weather) and soil type factors that translates into drought effect on CPIs.

The most promising CPIs derived from the ANOVA analysis were used to combined in a multilinear regression approach to increase their predictive power of yield. The two case studies displayed different responses to the regression performance. The model including four important CPIs (G_max, S_max, G_midP, S_midP) performed best ($R^2_{adj} = 0.49$, AIC = 204.4). Different years for model calibration were also tested in Norfolk using the rest of individual years as validation, and the results displayed RMSE between 1 t/ha and 2.2 t/ha in all seasons and Pbias between -19% and 23% displaying considerable overestimation and underestimation of this approach. The difficulty to

predict the yields in one year based on the crop response of another year is due to the fact that it will be greatly affected by management and weather differences.

4.1 Introduction

The core of remote sensing applications in agriculture is the ability to monitor large areas over time (Earl et al., 2003). The radar advantage of no cloud or daylight limitations increases the ability to use this technology in the highly dynamic agricultural landscape (Khabbazan et al., 2019). This is the reason that SAR is commonly used for crop classification with dense temporal profiles (Bargiel, 2017; Gomez-Chova et al., 2015; Skriver et al., 2011). The crop classification is closely connected to the identification of key crop development stages that can give important information on the vegetation dynamics of the different crops (Khabbazan et al., 2019). The recognition of the growth stages can be particularly useful to assimilate the information to models (Che et al., 2014; Klosterman et al., 2014; Son et al., 2016) that can be used across large areas. The ability of SAR to define the development stages (Bargiel, 2017; Betbeder et al., 2016; Harfenmeister, Spengler and Weltzien, 2019; Mercier et al., 2020; Schlund and Erasmi, 2020; Veloso et al., 2017) can contribute to phenology mapping that can be used in areas where on-field measurements are not existing or have a high cost (Nasrallah et al., 2019).

SAR can be used to assess the changes in the biophysical properties of the fields due to the sensitivity of radar to structure and water (Ndikumana et al., 2018). Attempts to estimate the field productivity mainly use of optical data and indices (NDVI) (Campos et al., 2019) in wheat (Mashaba et al., 2017) or combination with optical and SAR (Fieuzal and Baup, 2017) or SAR and models (Betbeder, Fieuzal and Baup, 2016; Setiyono et al., 2019). SAR only analysis is less common by using the polarisations regression with yield (Molijn et al., 2014). There is an increase of machine learning and remote sensing applications aiming to increase the yield predictive power (Hunt et al., 2019) as well as the extraction of biophysical information (Attarzadeh et al., 2018; Ndikumana et al., 2018; Van Tricht et al., 2018; Vreugdenhil et al., 2018). The yield estimation across the agricultural landscape is challenging as data availability is low.

There could be real advantages of providing information on crop development and growth at a large-scale using SAR data, but the approach needs to be tested across contrasting farms and regions. Hence, in this study, we applied the approach as presented in Chapter 2 across four farms based in two regions within the UK

(Hertfordshire and Norfolk). The difference between the two areas in terms of soil type and climate was used to understand seasonality and topography effects on SAR derived crop productivity indicators (CPIs). The temperature and the water balance of the areas is an important driver of development in the wheat fields. It is essential to monitor the differences in development stages as well as in biomass and water availability to be able to proxy the actual field production at the end of the growing season.

This chapter seeks to apply the methodology that defines the CPIs, described in Chapter 2 and the understanding of important CPIs of Chapter 3 to other areas, allowing comparison of farms with different soil and weather patterns. Analysis of variance (ANOVA) has been used to examine the significant effect of farm management, soil characteristics and climate variables. Significant effects on the CPIs are important for extrapolating the indicators across the agricultural landscape that involve different soils, climate, and management. Finally, the CPIs have been combined for a more robust correlation with the field production. The use of multi-linear regression analysis increases the yield prediction power by using the SAR derived indicators that significantly correlate with the yield (Chapter 2) and crop characteristics (Chapter 3). The combination of different indicators across the growing season helps to cover more time-related effects on yield than using individual CPIs that describe part of the crop life circle. At the end of this chapter, the results will display whether the approach is quantitative, transferable and predictive.

4.2 Materials and methods

4.2.1 Study sites

In this chapter, crop management and yield data were collected from four farms: Rothamsted and East Hall in Hertfordshire and Salle and Hydehall in Norfolk (Figure 4-1). SAR data were compiled for these four farms, representing two different soil types, landscapes, and weather patterns. Four seasons (from 2017 to 2020) of data were available from Norfolk in East Anglia, UK. The two farms in Norfolk are next to each other and follow similar crop management routines. Information about crop management, such as crops sowing and harvest dates, were collected for each growing season from the four farms. Data collected from the Hertfordshire farms were complete only in 2017 as technical problems prevented yield data collection at one of the farms during the following years.

Figure 4-1 Google Earth images from the four farms that display fields available for wheat cultivation in each area. Rothamsted Farm, Hertfordshire (a), East Hall, Hertfordshire (b), Hydehall, Norfolk (c) and Salle Farm, Norfolk (d)

4.2.1.1 Farm structure crop diversity and management

Overall, the potential number of wheat fields across the four farms (Table 4-1) varies considerably based on available fields, rotations and market, and the final number of fields used for the analysis have been reduced after quality assurance of the SAR-derived CPIs (see circles in Figure 4-5). The fields with values flagged as outliers were examined for the goodness of fit representing the parameterisation level.

Table 4-1 Total number of wheat fields provided for each season and farm before quality assurance. (* no yield data available); proposed boxes mark the subsample in the analysis (two case studies)

Season	Rothamsted	East Hall	Salle	Hydehall	Total
2017	11	25	24	21	81
2018	9	24*	34	19	86
2019	13	25*	19	18	75
2020	No data	No data	27	15	42
Total	33	74	104	73	

Part of the management of the farm is the preceding crop that determines the residual mineral nitrogen (N_m) in the soil that is likely to affect the development of the following

wheat crop. The residual N_m in the field can be categorised based on three crop groups. Group 1 (low residual N_m effect) covers cereals like winter wheat and low input crops, Group 2 (medium) includes oilseed rape, and Group 3 (high) includes legumes like beans. In addition, the harvest date of preceding crops can affect the sowing date of the wheat (e.g. late harvest of sugar beet).

Table 4-2 Different groups of residual soil N effect (https://www.fas.scot/downloads/tn731-nitrogen-recommendations-for-cereals-oilseed-rape-and-potatoes/)

Level of Nitrogen residual	Crop
Group 1	Wheat, barley, oats, linseed, sugar beet
Group 2	Oilseed rape
Group 3	Winter or spring beans, grass

Crop rotation affects several aspects of farm management. The soil residual N_m and sowing/harvest dates may vary within and among the farms, as they are based on the type of the previous crops in the field. Together with other factors such as soil hydrological and meteorological variables interact with crop management (N fertiliser application and leaching losses). For example, the delayed harvest from sugar beet will affect the sowing date of winter wheat in the latter part of the autumn, even winter, which together with a low residual Nm could cause weaker and later development.

Table 4-3 Number of fields for each farm based on the residual soil N group

Nitrogen status of previous crop	Rothamsted	East Hall	Salle	Hydehall
Group 1	12	34	14	14
Group 2	6	0	31	21
Group 3	2	39	20	17

Being part of the crop rotation, the crops found at the farms are sugar beet, spring beans, oilseed rape, barley, oats, and wheat. In Hertfordshire, the rotation is based on cereals with break crops of oilseed rape and beans, but in Norfolk follows a more spread rotation of seven years between the crops.

4.2.1.2 Soil heterogeneity

The main soil series at the Rothamsted research farm is Batcombe, a deep loam to clay soil, with silty plateau drifts and flints (Avery and Catt, 1995). The soil is characterised in the Soils Guide (Cranfield University, 2021) as “Fine silty over clayey and fine loamy over clayey soils with slowly permeable subsoils and slight seasonal waterlogging”. The unsaturated chalk on the subsoil affects the vertical movement of the rainwater as excess

water drains vertically down, entering the chalk. Cereals are the preferable crops and liming is needed as the soil is characterised as acid.

The main soil series at East Hall is Hornbeam 2, which is characterised as “Slightly acid loamy and clayey soils with impeded drainage” (Cranfield University, 2021). The soil is very flinty and is cultivated mainly with cereals and oilseed rape. Most of the area is slightly droughty depending on the crop, as grass can be most affected by this type of soil. The two farms in Norfolk (Salle and Hydehall) are dominated by the Beccles 1 soil series, which is described as “Slowly permeable seasonally wet slightly acid but base-rich loamy and clayey soils” (Cranfield University, 2021). Crops are affected by relatively impermeable subsoil that casing seasonal waterlogged soils and the drought effect are slight in summer. A good percentage of the farms is characterised by Wick 2 soil type described as well-drained deep loam. Table 4-4 and Table 4-5 display different soil types in the winter wheat fields of the four farms selected for the upscaling analysis. The variability of the farms can be examined with the main soil types described below:

Table 4-4 Number of fields per farm and soil characteristics (Cranfield University, 2021) across the four selected farms with wheat as a crop.

Soil texture	Rothamsted	East Hall	Salle	Hydehall
Deep loam	0	10	21	17
Deep loam to clay	19	49	2	7
Deep silty to clay	1	2	0	1
Loam over chalk	0	6	0	0
Loam over gravel	0	1	1	0
Seasonally wet deep loam to clay	0	5	41	25
Seasonally wet deep sand	0	0	0	2
Total	20	73	65	52

Table 4-5 Soil properties (average) describing the dominant soil series under arable management in each of the four farms (Cranfield University, 2021)

Soil properties	Rothamsted Batcombe	East Hall Hornbeam 2	Salle / Hydehall Beccles 1
Clay (%)	24	21	25
Sand (%)	18	32	46
Available water content (mm)	135	135	120
Saturated hydraulic conductivity Ksat (cm/day)	58.4	102.5	100.4
Depth (cm) to gley	60	45	25
Depth (cm) to slowly permeable layer	45	46	38
pH	6.5	7.8	7.4

4.2.1.3 Seasons and weather patterns

Between September 2016 and August 2019, the mean air temperature at Rothamsted and Salle followed the same climatic pattern, with the mean temperature in June and July to be greater at Rothamsted than Salle (Figure 4-2a). Data from the two areas were compared to understand the differences between the two regions of interest and the growing seasons in each area. At Coltishall (nearest station to the Norfolk farms, 10 miles), the annual 30-year mean (1981 to 2010) is 674 mm and at the Rothamsted site, the value is 712 mm. The annual average maximum temperature is 13.8°C and minimum of 6.4°C in Coltishall compared to Rothamsted of 13.7°C and 6°C. Sunshine duration in Rothamsted is 1585 hours per year and 1610 hours per year at Coltishall (Met Office, 2021).

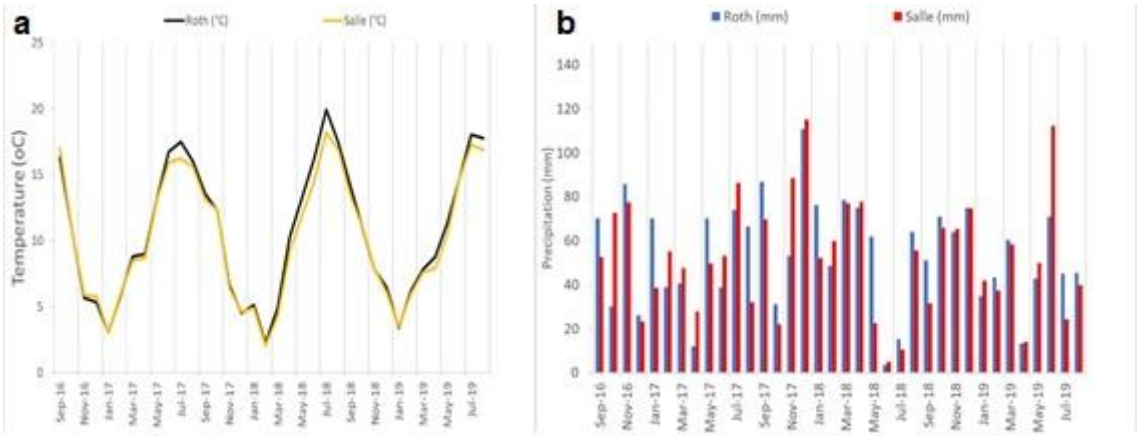


Figure 4-2 Monthly mean temperature (a) and monthly total precipitation (b) for Rothamsted (Roth) in Hertfordshire and Salle in Norfolk

Differences in soil type, soil water retention and precipitation between Hertfordshire and Norfolk will affect the soil water availability on the respective farms. The trend of the accumulated seasonal precipitation is similar between the two regions (Figure 4-2b); therefore, the porosity and infiltration rate of the soil is likely to affect the moisture content creating differences between the farms. Comparing the three years and the four calendar seasons, one can notice that the 2018 summer was drier compared to the other two years, which created significant water scarcity in both regions (Figure 4-3). This would affect the crops on shallow and light-textured soils more strongly. At the same time, the winter and spring of that year were considerably higher than the rest, affecting the crop establishment in areas with high clay content.

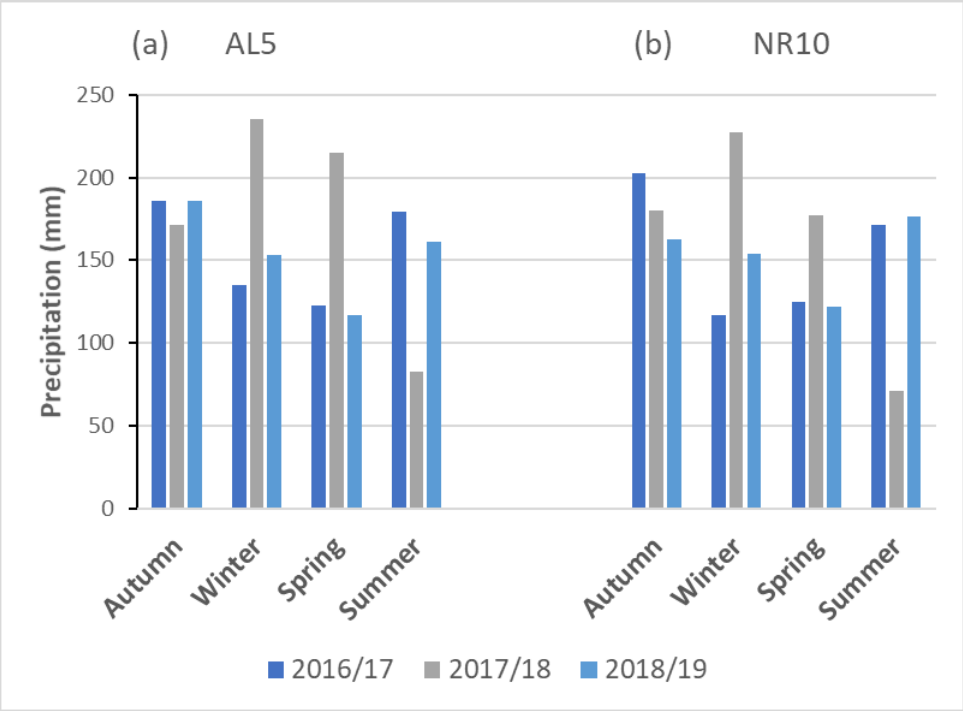


Figure 4-3 Seasonal variations of the total precipitation across the three growing seasons in the two regions (a) Rothamsted – AL5 and (b) Reepham/Salle – NR10 using the local weather stations in the farms.

4.2.2 Selected CPIs for the upscale

The calculation of the CPIs across the four farms and the available growing season was carried out using the SAR time series from the Sentinel-1 satellite. The CPIs were defined at the field scale, averaging pre-processed SAR backscatter data in terms of cross-polarisation ratio (VHVV), smoothing before the fitting procedure (Chapter 2). It was important for the analysis to examine the ranges of the selected CPIs in each farm as a first step. These calculations assisted in identifying possible outliers in the analysis based on each specific farm and growing season. The most critical CPIs considered are given in the previous chapters and illustrated in Figure 4-4 and Table 4-6.

Chapter 2 demonstrated that it was possible to use SAR data to calculate 12 crop productivity indicators. One CPI (5) was the day of the maximum VH/VV value. There were then four CPIs (1-4) describing a logistic curve from a low VH/VV value at the start of the season to a high value in mid-season, and four CPIs (6-9) describing a logistic curve from a high value to a low value at the end of the cropping season. Lastly, there were three CPIs derived from the two logistic curves: the duration between the mid-points (10) and the difference between the maximum (11) and minimum VH/VV values (12) for the two logistic curves.

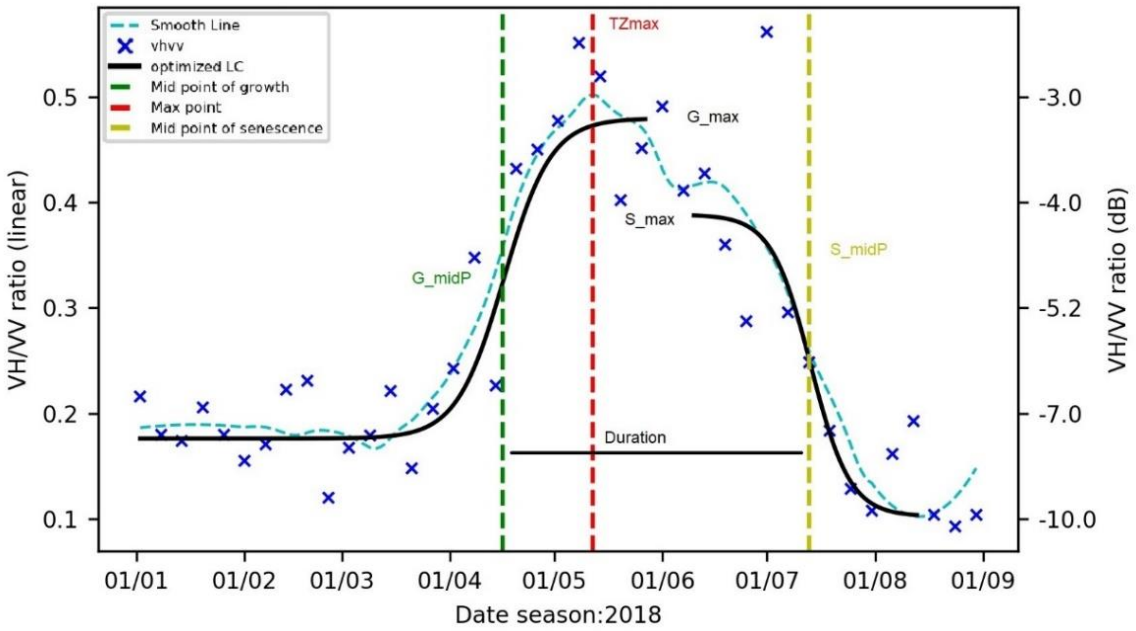


Figure 4-4 Presentation of important CPIs derived from the previous chapters that were used for the statistical analysis in the different farms.

Table 4-6 Selected CPIs definitions that displayed contribution towards biophysical properties of wheat taken from previous chapters

No.	Symbol	Parameter name	Definition
3	G_midP	Time of midpoint of the growth period	DOY when the midpoint of the logistic curve occurs in the growth period
4	G_max	Max value for growth stage	Maximum VH/VV ratio value for the full season
5	TZmax	Time of maximum point	DOY of maximum smoothed value of VH/VV
6	S_max	Value at the start of grain filling	Period of backscatter stabilization, correlated with plant water content
8	S_midP	Time of midpoint of maturation	DOY when the midpoint of the logistic curve occurs in the maturation period
9	S_base	Baseline value at the end of the season	Background value of the VH/VV ratio
10	Duration	Duration of “full” vegetation to maturation	Time difference between midpoints (3, 8)

4.2.3 Quality assurance for the CPI outlier definition and elimination

The data used for analysis was checked for potential outliers using the interquartile range (IQR), whitch measures variability based on quartiles derived from the data. In descriptive statistics, the IQR value describes the spread of the middle half of the distribution. The IQR is useful for data sets with outliers because it is based on the middle half of the distribution, extreme values less influence it. The fields flagged up by the IQR analysis were then assessed visually by multi-plots of all fields in each year and farm. The need for visual assessment was necessary as, in some instances, the time series of SAR justified the extreme values of the CPIs. In this case, the field that was reacting to the radar signal different from the rest remained in the analysis if the shape of the logistic curve remained close to the rest of the fields.

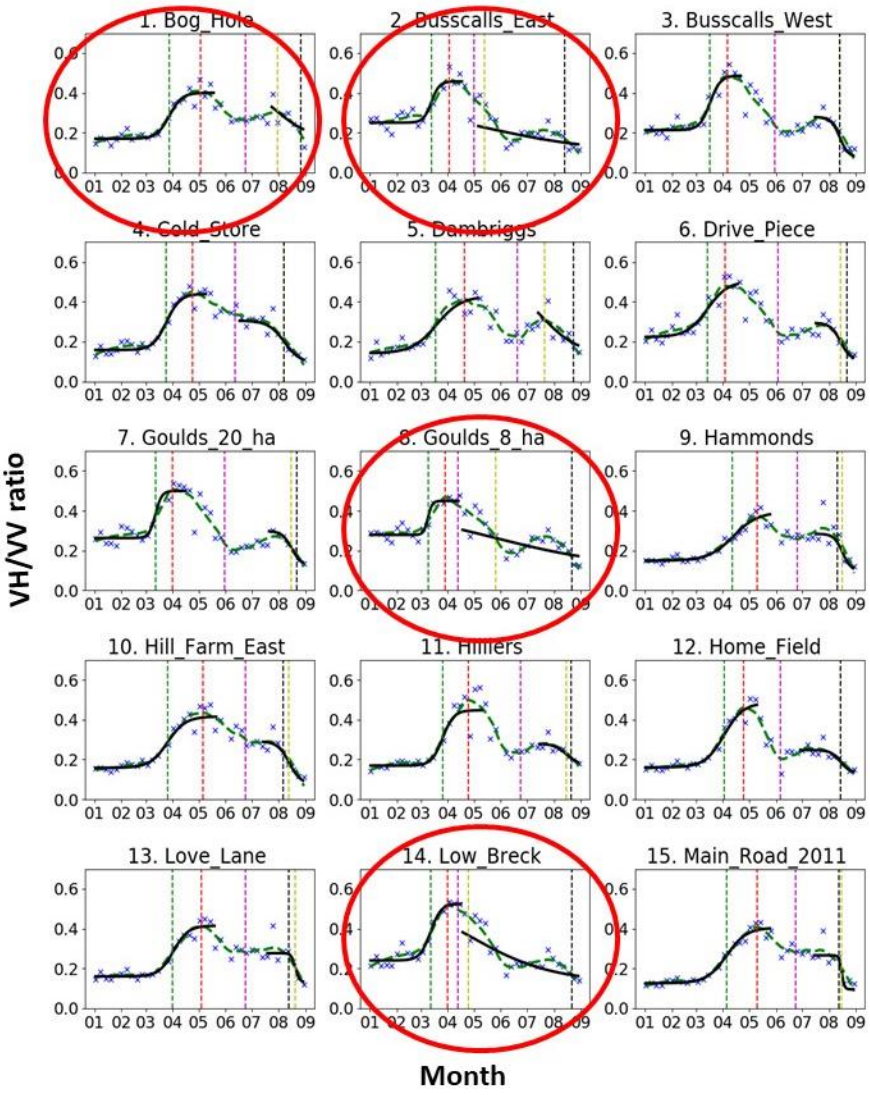


Figure 4-5 Examples of poor curve fitting that the IQR application on the senescence CPIs flagged up the fields (red circles) that were then excluded from the analysis in Hydehall farm.

The first layer of this process was the stepwise verification of the CPIs starting from the temporal indicators that should not vary among winter wheat fields in the same years, farm and region as the number of growing degree days (GDD) in the same area should be similar. It assumes that the sowing dates in the same farm should not be far apart and varieties have similar development. For example, spring wheat was detected as its CPI related to stem elongation (G_midP) had a much later date than the rest of the farm, whilst winter barley was considerably earlier. Thus, such fields were removed from the analysis.

The next step was the curve shape that sometimes diverted from the expected form. The fields with values flagged as outliers were examined for the goodness of fit that represents the level of parameterisation. A few fields had to be eliminated due to the noise in the SAR-CP data points not allowing clear identification of time-related indicators (S_midP, the timing of Senescence/Maturation) and the visual assessment was also able to identify the fields for which the shape of the logistic curve did not follow the expected smoothed curve (Figure 4-5).

4.2.4 Sensitivity of CPIs

A correlation matrix was used to validate the sensitivity of the CPIs. Two sets of data were analysed to explore the ability of the CPIs defined in Chapter 2 to describe and predict yield. First, the approach was applied across four farms by adding a second farm in Hertfordshire relatively close to Rothamsted farm and two farms in Norfolk region. The 2017 season was selected due to missing yield information for the rest of the seasons in Hertfordshire (technical difficulties related to the harvest management). Second, four growing seasons (harvest in 2017 to 2020) were selected from sites in Norfolk in Eastern England, representing two farms next to each other following similar management.

4.2.5 Analysis of variance (ANOVA) for CPIs related to the biophysical environment and crop management

To answer the question of how seasonal and regional pedo-climatic and management factors affected the values of the CPIs, we conducted an ANOVA. The different examined categories were defined by the data availability across the farms (number of fields, yield). The unbalanced design was a consequence of the different number of fields in each farm and year and the heterogeneity of the farm in respect to soil management and weather conditions. The first two are growing season (covers 2017, 2018 and 2019 season) and farm (four farms in the two regions). Then, the different soil types are

extracted as the dominant soil type per field and the different management was defined by the group of the preceding crop. The equation used for the analysis incorporated the following categories (Table 4-7):

Table 4-7 Information about the different factors considered in the ANOVA based on the data available from the selected farms.

Factor Categories	Description
Farm	representing the overall management of each farm
Year	Harvest year for different growing seasons/ weather conditions
Pr. Crop N	Different groups that give information about the management and residual soil N_m are carried over by the preceding crop type in the wheat field.
Soil type	soil texture affects the available water capacity (AWC) and infiltration rate

In addition, interaction factors were defined to display the relationships between the factors. An interaction effect represents when the impact of one factor depends on the other factor's level. For example, the selected interactions for the analysis are the following:

- soil type * Pr. Crop N = effect of soil type depending on residual N availability
- year * soil type = effect of soil depending on the weather patterns during the growing season.

This analysis was focused on understanding the effects of each site and management category on the specified CPIs (Table 4-6). This helped to narrow down the number of CPIs used in the multiple linear regression analysis. The statistical software used for this analysis was GenStat (VSN International, 2020).

4.2.6 Multiple linear regression analysis (MLR)

The regression analysis was used to improve the yield prediction using the important CPIs identified from the previous chapters. The most useful CPIs identified in Chapter 2 and 3 related to yield were the duration and timing of the booting stage (TZmax). The yield data covering all four farms at the same period was only available in the 2017 season. For that reason, the initial MLR analysis was focused on the 2017 data as the

first option to cover all four farms and then the seasonality was tested in Norfolk farms for the four available growing seasons. The ANOVA results also helped towards these formats on the year and region options for the MLR training approach. The regression estimator selected was least squares and the dependable variable was the yield (t/ha).

The first test was the use of the CPIs identified in Chapter 2 as the most sensitive to yield. The first formula is defined by Equation 4-1:

$$yield = \alpha * Duration + \beta * TZmax + \gamma \quad (4\text{-}1)$$

where α, β and γ are the coefficients of the regression, Duration is the duration between the mid-points of the growth and senescence stages, and TZ max is the booting timing.

The next regression incorporated sensitive CPIs from Chapter 3 related to biomass and pant water content. Also, the duration is represented by the two midpoints in the growth and senescence period as well as α, β, γ, δ, and ε are the coefficients of the regression.

$$yield = \alpha * G\_midP + \beta * G\_max + \gamma * S\_max + \delta * S\_midP + \varepsilon \quad (4\text{-}2)$$

The addition of the interactions of the two midpoints and the interaction of the biomass and water indicators were also examined to see whether they could improve the considered statistical measures.

$$yield = \alpha * G\_midP + \beta * G\_max + \gamma * S\_max + \delta * S\_midP + \varepsilon * G\_midP * S\_midP * + f * G\_max * S\_max + z \quad (4\text{-}3)$$

Where α, β, γ, δ, ε, f, and z are the coefficients of the regression.

Two criteria were selected for the comparison of the different models in the multiple linear regression.

The Akaike information criterion (AIC) was used to evaluate how well a model fits the generated data (Bozdogan, 1987). In statistics, AIC compares different possible models and determines which one is the best fit for the data, considering the degree of model complexity. This way, it penalizes the degree of model complexity. AIC is calculated from:

$$AIC = 2M - 2\ln(L) \quad (4\text{-}4)$$

where M is the number of independent variables used to build the model, and ln(L) is the model likelihood (log).

R^2 displays the goodness of fit between points and lines. Adjusted R^2 does the same thing but takes into consideration the number of terms. By adding more and more variables to a model, adjusted R^2 value will change based on how useful the variables are. For that reason, R^2_{adj} will always be less than or equal to R^2.

$$R^2_{adj} = 1 - \left[\frac{(1 - R^2)(n-1)}{n-k-1} \right] \tag{4-5}$$

where: n is the number of points in the data sample, and k is the number the number of variables in your model, without taking into consideration the constant.

The adjusted R^2 will penalize the increase of independent variables (K) that do not fit the model(Ratner, 2009). This allowed the opportunity to compare the goodness of fit across different models and options of the training data. The AIC was selected as model selection criteria for the best model, but R^2 adjusted values were also displayed for easy comparison among the different models.

The descriptive statistics used to evaluate the best selected model are the Root Mean Square Error (RMSE) and the bias, which are commonly used statistical measures to evaluate model performance (Moriasi et al., 2015)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{4-6}$$

where P is the predicted yield from the regression, O is the observed yield provided by the farms and n is the number of fields/observations.

$$\mathrm{PBIAS} = \frac{\sum_{i=1}^{n}\left(Y_i^{\mathrm{obs}} - Y_i^{\mathrm{sim}}\right)}{\sum_{i=1}^{n}\left(Y_i^{\mathrm{obs}}\right)} \tag{4-7}$$

Where Y is the value of observed (obs) and simulated (sim) yield

4.3 Results

4.3.1 Relationship between SAR-derived CPIs and yield

The correlation matrix displays the Pearson correlation value (r) for each of the CPIs colour-coded green (two-tailed test with 95% significance level). Two approaches were used to test the methodology in different areas, with different soils and more diverse management:

- **Case Study 1** (CS1)**:** Two farms of high productivity in Norfolk, considering four growing seasons (2017-2020). Scaling up over time within a region to examine the effect of annual differences in weather
- **Case Study 2** (CS2)**:** Four farms across the two regions of interest (Hertfordshire and Norfolk) in growing season 2016-2017, when yield data from all farms are available. Scaling up across space in a single year (effect of soil and management)

The CS1 is part of the validation of the methodology used in Chapter 2 at Rothamsted farm. The same 12 indicators described were used across multiple years to examine the correlation with the yield provided by the two farms in Norfolk (Salle, Hydehall). In total, 153 out of the 177 fields of winter wheat between the two farms and the four years were used to calculate the CPIs and 148 were used for the analysis as yield data were available from the harvest. The correlation matrix also displays the correlation between the CPIs that can eliminate highly correlated indicators (e.g. G_midP and TZmax) to be used together in the later stage of multiple linear regression analysis. The second approach (CS2) is the calculation of the CPIs across the four farms. Case study 2 was used to help understand the regional effect as well as the different soil type effects on the CPI values. In total, 76 out of the 81 fields CPIs were calculated for winter wheat between the four farms in 2017, of which 69 fields had available yield data.

The correlation matrixes displayed a significant relationship between the CPIs and yield in both options explained above. IQR was used to identify the outliers that were examined visually, and any potential outliers had been eliminated. In the multi-seasonal case, the dominant CPIs are the ones related to time. This signified that the changes in weather patterns that affect the growth stages and the duration of high-value vegetation that can use the available radiation for photosynthesis. CPIs representing the maturation period did not show sensitivity with the yield except the S_steep (rate of maturation, r =0.25) and S_base (field background value at the end of the growing season, r = 0.31).

Compare with Chapter 2.3.2, the time-related indicators such as TZmax (booting) and duration were showing similar trends, i.e. r = - 0.21 and r = 0.17, respectively. In contrast to Rothamsted farm, the G_midP (stem elongation) also displayed a significant correlation, i.e. r = 0.21, and strongly correlated with the booting stage r = 0.81.

Figure 4-6 Correlation matrix for the 2017-2020 season using the two Norfolk farms displaying the correlation among the CPIs. The correlation values (r) are colour-coded green when there was a 95% level of significance for a two-tailed test

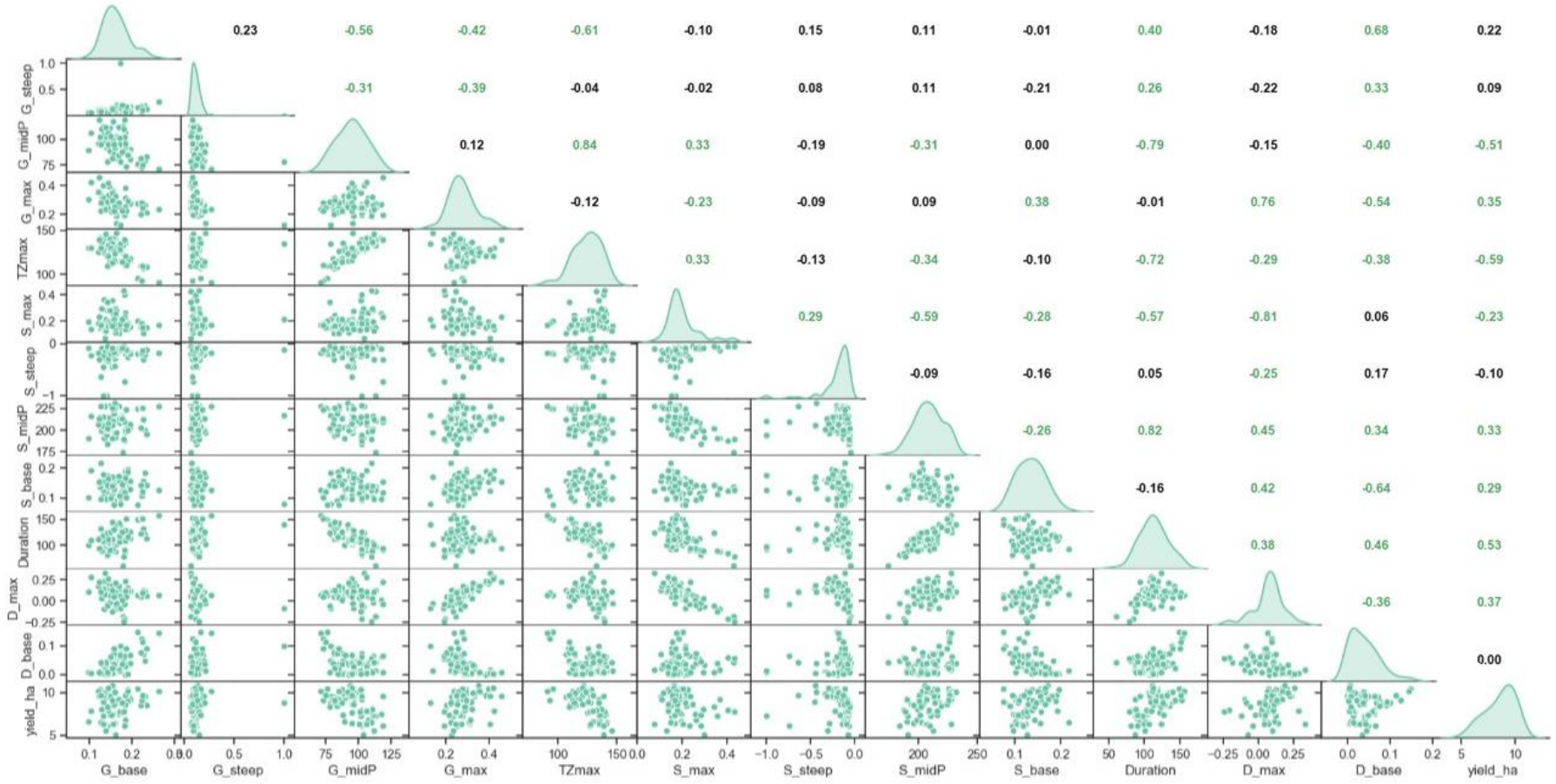


Figure 4-7 Correlation matrix for the 2017 season across the four farms displaying the correlation among the CPIs. The correlation values (r) are colour-coded green when there was a 95% level of significance for a two-tailed test

4.3.2 Variation of key CPIs across locations

The averages of the CPIs across the four farms in 2017 (CS2) displayed potential differences between the two regions under the same climate conditions (Table 4-8). The coefficient of variation (CV) varied between 3% and 40%, with the most variable indicator in the same year being the Smax (CV 17-51%), which is related to the plant water content (See Chapter 3), potentially affected by the soil heterogeneity in each farm. The temporal CPIs, which are related to the crop development stages (G_midP, S_midP, TZmax), had less than 10% variation in the same year, except at the Hyde Hall farm where the senescence period was more variable. This was also visible in the duration, with Salle having less variability than Hyde Hall (7% compared to 24%). The in-farm variability of yield was lower in Norfolk farms, and the average yields were also higher than in Hertfordshire farms. There were striking temporal differences between the two regions. On average, crops developed almost two weeks later on the Hertfordshire than the Norfolk farms (G_midP of 101 DoY compared to 88 DoY). This effect is also visible in the value of TZmax, which is related to the timing of booting and the Duration, which was two to four weeks longer on the Norfolk than the Hertfordshire farms. In the 2017 data, there is already a difference between Salle and Hyde Hall Farm, especially in the variation of the time related CPIs.

Table 4-8 Mean and standard deviation (SD) of the important CPIs from the four different farms during 2017 to compare the regional effects on CPIs* and yield data.) TZmax = booting stage, G_midP = midpoint of growth, S_midP = midpoint of senescence, G_max = biomass proxy at the end of the construction phase, Smax = water proxy at the start of maturation (see Table 4-6)

Farm	Stats	Gmax	Smax	Sbase	GmidP DoY	SmidP DoY	TZmax DoY	Duration days	Yield t/ha
Roth (Herts)	Mean	0.29	0.20	0.13	98	199	131	101	7.1
	SD	0.06	0.10	0.02	11	13	9	14	1.6
	CV	20%	51%	17%	12%	7%	7%	14%	22%
East Hall (Herts)	Mean	0.23	0.23	0.12	104	205	136	101	7.5
	SD	0.04	0.09	0.03	10	11	6	17	0.9
	CV	19%	39%	25%	10%	5%	4%	17%	12%
Salle	Mean	0.33	0.17	0.16	92	206	120	114	9.7
(Norfolk)	SD	0.07	0.03	0.02	9	6	8	7	0.7
	CV	21%	17%	11%	10%	3%	6%	7%	7%
Hyde.	Mean	0.26	0.21	0.12	83	210	111	127	9.4
hall	SD	0.04	0.08	0.03	10	34	14	30	0.6
(Norfolk)	CV	14%	40%	23%	12%	16%	12%	24%	7%

4.3.3 Analysis of variance for the selected CPIs

The Analysis of variance (ANOVA) displayed the sensitivity of each of the CPIs to the environmental and management categories and their interactions. Different runs of ANOVA were applied for each of the CPIs using the selected categories. The selected ANOVA applied an unbalanced design analysis using GenStat regression for the Duration indicator across the four farms in the 2017-19 period (i.e. three seasons). Table 4-9 summarises the results of the individual CPIs.

The results of ANOVA across the selected CPIs can be used to identify if there are significant effects of area and management. The effect of the farm management had significant effects on the CPIs, especially in the construction phase of the growing season. The parameter Year had a significant effect on all the indicators ($p<0.001$). The weather patterns are crucial for the crop's development stage as well as for the water input (precipitation) and evapotranspiration rate. The importance of the residual nitrogen from the previous crop indicates a significant effect on the midpoint of growth that is strongly correlated with the stem elongation stage of the wither wheat. The soil type had a similar effect on the stem elongation phase ($p<0.001$) as well as $p<0.1$ in relation to the plant water content at the beginning of the maturation phase of the crop. This was expected as water availability in the plant is affected by the water content of the soil.

The soil type had a significant effect on growth midpoint with a lower level of significance on duration and max value of senescence that represents water capacity of the plant, but also this effect is based on the year the CPI is calculated as the primary input in water balance in the soil in non-irrigated fields is the amount and frequency of the precipitation in the area. The S_base indicator is also connected with the soil status after the harvest was significantly affected by the seasonality of the region.

The results showed that the different management (farm) characteristics affects the CPI as well as the effects described by the period (Year) and soil interaction. The nitrogen levels from the previous crop type did not display a significant effect on duration even when taking into consideration the soil type.

Table 4-9 ANOVA results using three harvest years (2017-2019) for the four selected farms, testing the influence of annual meteorology (Year), management (Farm), residual nitrogen (Pr_crop_N), Soil type and the interactions of soil season and nitrogen residual and soil on the important CPIs at four different levels of significance: 0.05<p<0.1(º), 0.01<p<0.05(*), 0.001<p<0.01 (**) and p<0.001 (***) (number of samples =210)

Factors	G_midP	G_max	S_midP	S_max	S_base	Duration
Farm	***	***	**	**		***
Year	***	***	***	***	***	***
Pr_crop_N	***	*	º			**
Soil_type	***			º		*
Pr_crop_N*Soil_type				*		
Year*Soil_type	***		*	***	*	***

4.3.4 Multiple linear regression (MLR) to estimate wheat yield

The results displayed in this section are based on the two different subsets of CPIs representing two case studies (CS) covering temporally (CS1) and spatially (CS2) variable yield data. The different models selected to predict the yield are presented for the different options. All combinations of the important indicators were used in all potential combinations and the results of the level of fit are defined by the selected criteria (AIC).

The analysis was divided into different sections, starting with the two case studies: multi-season temporal analysis for the Norfolk farms (CS1) and spatial variability analysis for the four farms in 2017 (CS2). Based on these results, the next MLR application used the best model across the Norfolk region in multiple years to identify potential correlations with yield. Finally, the analysis concluded with the recalibration of the model in rotation for each year and applied that across the other growing seasons in the same region.

4.3.4.1 Multi-season temporal analysis for the Norfolk farms (CS1)

The multi-season temporal analysis was based on the approach presented in Chapter 2, of which the application is focused on one farm in Hertfordshire across multiple years of winter wheat fields. The results of the Norfolk analysis covered four years, based on the defined MLR, did not display satisfactory results to estimate yields from the best model. Also, the analysis of the individual farms did not improve the statistical criteria selected. The best model based on the AIC had an $R^2_{ajd} \approx 0.1$ even using the individual farms to reduce potential management influence. Table 4-10 displays the results from the main formulas for the same region for the four available seasons with yield data.

Table 4-10 Multiple linear regression results for the three main models at four different levels of significance: 0.05<p<0.1(°), 0.01<p<0.05(*), 0.001<p<0.01 (**) and p<0.001 (***) for the Norfolk farms in years 2017-2020

	Factors	Coefficient		P>|t|
Model 1	Intercept	11.36	±2.085	***
	Duration	0.002	±0.007	
	TZmax	-0.017	±0.011	
Model 2	Intercept	12.46	±2.527	***
	G_midP	-0.036	±0.009	***
	G_max	3.36	±1.089	**
	S_max	0.49	±1.414	
	S_midP	-0.003	±0.01	
Model 3	Intercept	8.90	±14.80	
	G_midP	-0.016	±0.15	
	G_max	6.56	±3.66	°
	S_max	5.63	±5.92	
	S_midP	0.009	±0.07	
	G_midP:S_midP	-0.0001	±0.001	
	G_max:S_max	-13.75	±15.06	

The simple model (Model 1) was based on the two most significant CPIs from the Chapter 2 methodology. The level of significance of each independent variable was low as only the intersect coefficient had $p < 0.05$. In Model 2, the use of the two midpoints instead of the duration and the addition of the two maximums indicators (G_max and S_max) aimed to represent the development as well as the physical properties of the crop (biomass and water content, respectively). At the same time, the high correlation of G_midP and TZmax removed the second as it is derived from the smooth line instead of the logistic curve. The two factors representing the construction phase of the crop had a significant effect on the regression in contrast to the maturation phase of the plant. In addition, the G_midP had a negative effect between the independent variable and the dependent variable ($p<0.05$), indicating that the late development has a negative impact on the yield in model 2. In Model 3, the additional factors added the interaction of the two midpoints (G_midP*S_midP) and the biomass and water content (G_max*S_max) did not increase the predictive power and all factors in Model 3 displayed no significant correlation except the G_max ($p<0.5$).

4.3.4.2 Spatial variability analysis for the four farms in 2017 (CS2)

The spatial analysis results for the 2017 growing season was more promising than the temporal analysis across two farms within one region. The R^2_{adj} values were much higher than the previous case study (CS1). Table 4-11 displays the factors and the coefficients as well as the p-values for the three main models discussed previously (Section 4.2.6).

Table 4-11 Results of multiple linear regression for the three main models at four different levels of significance: 0.05<p<0.1(o), 0.01<p<0.05(*), 0.001<p<0.01 (**) and p<0.001 (***) for the four farms during the 2017 growing season (CS2).

	Factors	Coefficient		p>|t|
Model 1	Intercept	12.96	±3.11	***
	Duration	0.016	±0.011	
	TZmax	-0.050	±0.017	**
Model 2	Intercept	4.85	±3.26	
	G_midP	-0.067	±0.011	***
	G_max	10.15	±2.10	***
	S_max	3.71	±2.37	
	S_midP	0.03	±0.013	*
Model 3	Intercept	20.23	±22.78	
	G_midP	-0.22	±0.252	
	G_max	4.52	±6.275	
	S_max	-2.66	±8.549	
	S_midP	-0.04	±0.108	
	G_midP:S_midP	0.0007	±0.001	
	G_max:S_max	28.44	±31.38	

The result of Model 1, yield = f(Duration, TZmax), showed a correlation with R^2_{adj} = 0.35 and an AIC value of 215.5. The early booting stage (TZmax) is a significant CPI ($P < 0.01$) as well as the intercept with $p < 0.001$. Duration did not have a significant effect in this model. The next stage was therefore to examine if the use of developmental stages could be more valuable than using just the duration derived from the two mid-points.

Model 2, which used two midpoints instead of the duration, displayed a very high correlation between the midpoint of growth and the booting stage, and hence TZmax was removed. In addition, the biomass and water information from CPIs is added to increase the predictive power related to yield as important properties of the crop canopy. Model 2, yield = f (Gmid, Smid, Gmax, S_max) gave values of R^2_{adj} = 0.46, and AIC = 204.4. Adding the interaction factors (Gmid*Smid and Gmax*Smax) in Model 3 did not increase the statistical measures to justify using this formula. More specifically, R^2_{adj} = 0.45 and the AIC value is increased to 206.9. None of the coefficients displayed a significant effect in this regression.

4.3.4.3 MLR summary of the different models and options

Table 4-12 below summarises the results of the two options in the MLR analysis. The analysis followed a stepwise increase of the variables in the linear regression to examine the statistical criteria improvement. AIC has been used as model selection criteria, i.e. the model with the lowest AIC value has been considered as the best model. However, we also present R^2 adjusted to have an extra measure to compare the models. The results are displayed in Table 4-12 together with the calculations of individual farms in Norfolk, where yield data across the four years were available.

Using AIC as model selection criteria, the best model is different for each of the options. For example, the 3rd formula is the best model for Salle farm in contrast to the 5th formula for the neighbouring Hydehall farm. However, some other more complex models seem to have slightly higher R^2_{adj} values. The two options selected for the analysis displayed different formulas as the best fit based on AIC.

It is difficult to get a good fit across years as every season will change the CPI's response based on the weather conditions. At the same time, different fields every growing season affect the yield as we analysed in the ANOVA, soil type and N residual contribute to changes in crop performance, especially with the weather patterns difference among years. In contrast, using a single year across the farms offered a much better fit on the statistical criteria as the main driver of weather patterns will not differ in the same season as across years.

Table 4-12 Adjusted coefficient of determination R^2_{adj} and AIC of the multiple linear regressions of selected variables relative to wheat yield across four farms in one season and across four seasons in two farms in Norfolk using the important CPIs (Duration, TZmax=booting stage, Gmid=midpoint of growth, Smid=midpoint of senescence, Gmax=biomass proxy at the end of the construction phase, Smax=water proxy at the start of maturation). In bold the best model in each case

Formula variables	Four farms in growing season 2016-17 (N=69)		Four growing seasons in Norfolk (N=148)		Four growing seasons Salle (N=83)		Four growing seasons Hydehall (N=65)	
	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC	R^2_{adj}	AIC
Duration, TZmax	0.346	215.5	0.030	424.4	0.065	255.7	-0.019	164.8
Gmid,Smid, TZmax	0.339	217.1	0.030	425.3	0.055	257.5	-0.035	166.8
Gmid,Smid,Gmax	0.446	205.0	**0.091**	**415.7**	**0.107**	**252.9**	0.034	162.3
Gmid,Smid,Smax	0.272	223.8	0.031	425.2	0.084	255.0	0.000	164.5
Gmid,Smid,Gmax,Smax	**0.458**	**204.4**	0.085	417.6	0.115	253.1	**0.064**	**161.1**
Gmid,Smid,Gmax,Smax, Gmid*Smid	0.454	205.8	0.079	419.6	0.116	253.9	0.058	162.4
Gmid,Smid,Gmax,Smax, Gmax*Smax	0.458	205.3	0.084	418.7	0.116	253.9	0.048	163.1
Gmid,Smid,Gmax,Smax, Gmid*Smid, Gmax*Smax	0.452	206.9	0.078	420.7	0.128	253.6	0.043	164.4

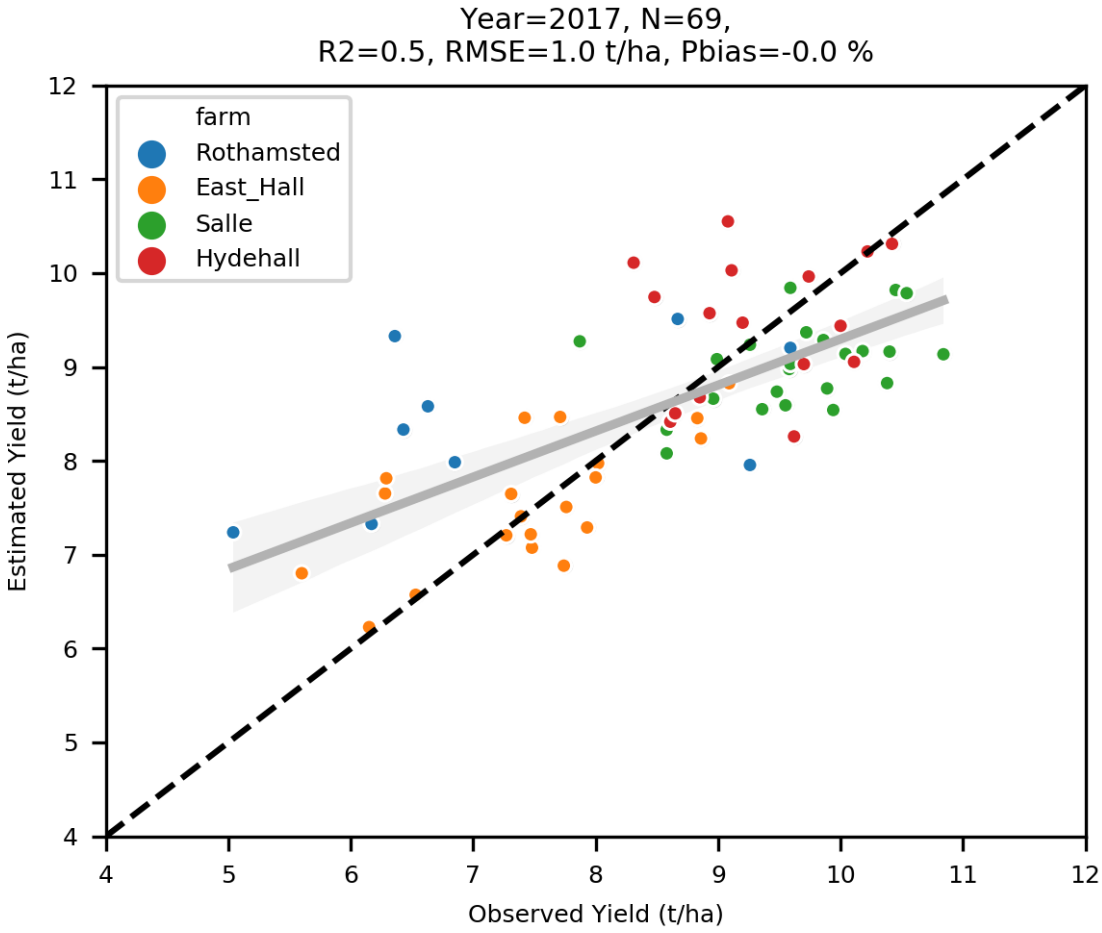


Figure 4-8 Actual (calibration) vs predicted values of yield (t/ha) for the four different farms in the 2016-17 growing season using the best model yield =f (Gmid, Smid, Gmax, Smax) based on the AIC criterion. The black line represents the 1:1 line and the grey is the regression with R^2=0.5. The 95% confidence limits are highlighted as faded regions.

The regression analysis results displayed in Figure 4-8 presents a model selected for the upscaling purpose of the analysis that can be applied across farms in the same growing season. The regression line showed that with a Pbias of 0%, the low yield areas are overestimated by the model and the high yield areas are underestimated. The RMSE was estimated at 1 t/ha based on all farms in 2017 corresponding to a yield range of 5-11 t/ha. It should also be noticed that the two farms in Norfolk have higher yields than those in Hertfordshire. For that season, there was a cluster of points on the high end of the actual yield compared to the fewer values on the low end. The regression also showed that the Rothamsted values are overestimated by the model (blue dots) compared to the East Hall farm located in the same region.

4.3.4.4 Use the model from 2017 (CS2) to predict the 2018-2020 yield (CS1)

Figure 4-9 below displayed the attempt to use the best-fitted model from the MLR analysis in the Norfolk region to examine the potential of the calibrated model of 2017 across the four farms. The hypothesis tested was that using the best-fitted model in the 2017 year would be able to simulate the rest of the years as different ranges of yield, soil, and management can simulate better productivity in other years than the CPI differences can be connected to these changes. The results showed negative Pbias of the model across the four seasons in the two farms in this area, indicating an underestimation. The 2017 underestimation was expected as the high yield areas were underestimated in the model. The RMSE value varied from 0.9 t/ha in 2017 (that were part of the calibrated dataset) to 1.4 t/ha in 2019. This showed the difficulty of working across multiple years, combining different indicators affected in different ways based on weather patterns. The ANOVA also displayed this conclusion on the year as factor effect on CPIs.

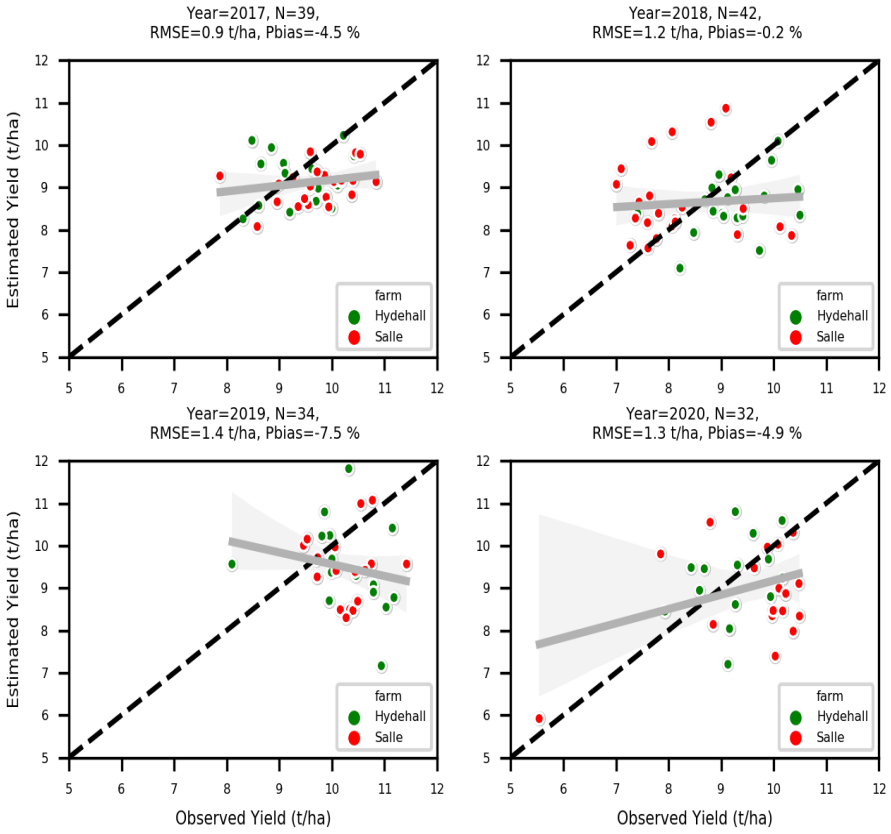


Figure 4-9 Yield prediction for the four different years in Norfolk based on the 2017 four farm regression model (CS2) and comparison with the 1:1 line (black). The regression line is displayed as blue and the 95% confidence limits are highlighted as faded regions.

4.3.4.5 Prediction of Norfolk farm yield based on individual years

The last application to attempt the prediction of yield across seasons has the format of one season calibration with the model selected in this option and validate on the rest of the years available in the database. Figure 4-10 displays an example of this approach. The results of all runs across seasons in Norfolk and the rotation of the calibration year are shown in Table 4-13.

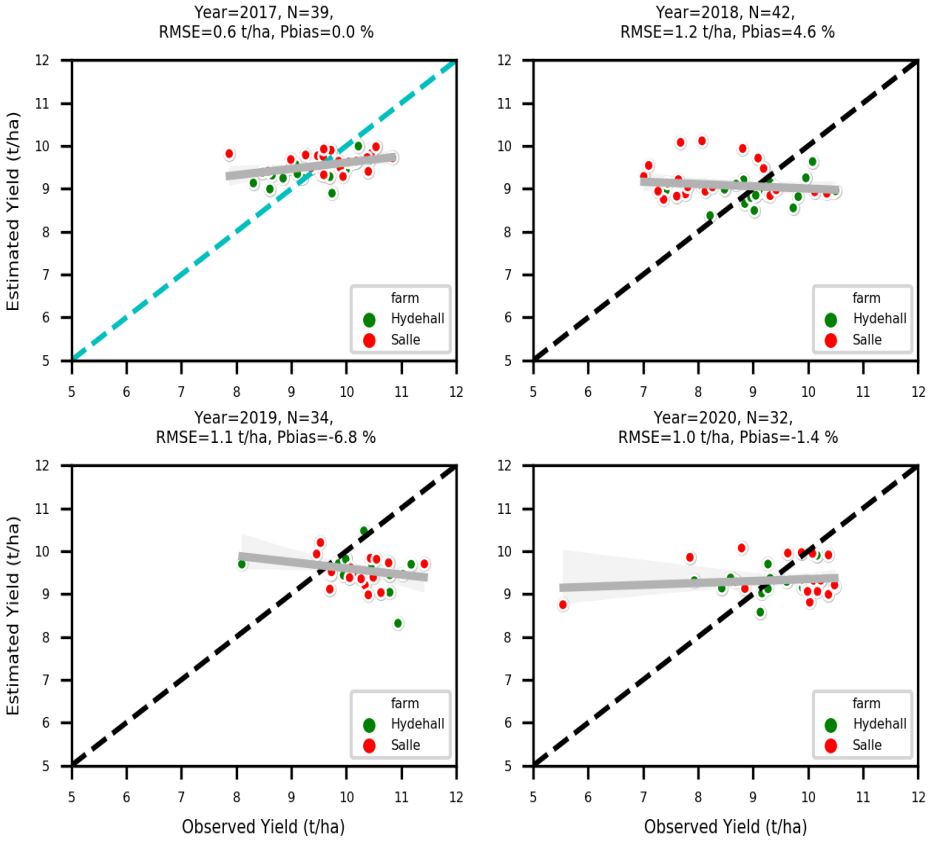


Figure 4-10 Comparison of predicted yields against actual yields for the Norfolk region using the 2017 growing season as calibration (cyan 1:1 line) for the multiple linear regression model and the rest of the years for comparison (black lines 1:1) as well as the regression lines for each year (grey lines). The 95% confidence limits are highlighted as faded regions. model yield=f(Gmid,Smid,Gmax,Smax)

The rotation of the calibration year displayed the effect of the different weather patterns across the four years in this region. 2018 has a different water balance from the rest of the growing season that was also observed on the Pbias values across the validation row in Table 4-13. The 2018 calibration values were the highest across the validation years. The negative sign displayed the underestimation due to lower yields in that period.

Table 4-13 RMSE and Pbias for the different years in Norfolk farm based on re-calibration of the individual year (different rows) of the best-selected formula. The calibration data is shown in italic.

Validation year		2017 (N=39)		2018 (N=42)		2019 (N=34)		2020 (N=32)	
		RMSE (t/ha)	Pbias %	RMSE (t/ha)	Pbias %	RMSE (t/ha)	Pbias %	RMSE (t/ha)	Pbias %
Calibration year	2017	*0.6*	*0.0*	1.2	4.6	1.1	-6.8	1.0	-1.4
	2018	1.3	-9.2	*0.9*	*0.0*	2.0	-18.7	1.4	-11.6
	2019	1.0	6.0	2.2	23	*0.6*	*0.0*	1.3	7.8
	2020	1.0	0.2	1.5	14	1.2	-8.9	*0.8*	*0.0*

4.4 Discussion

The use of the methodology of the SAR-derived crop productivity indicators in other regions requires an understanding of the effect of the environment on the quantification procedure. The validation using different approaches based on data availability showed that individual CPIs were correlated with the yield and the responses were similar to those observed in Chapter 2. Quality assurance in combination with the level of fitness of the curve fitting is essential to eliminate potential outliers and be able to use correlation matrices to display relationship of CPIs with the yield in other areas. The ANOVA results displayed how significant the effect of individual categories and the interactions on the CPIs are. This was focused on the influence of the soil, season, previous crop nitrogen residue and farm (different management). The multiple linear regression analysis (MLR) aimed to increase the yield prediction level by combining the significant CPIs selected based on previous results.

4.4.1 Landscape and management factors effects on CPIs

The different pedoclimatic conditions in the two regions of interest, Hertfordshire and Norfolk, and the different factors affected by management were examined using the ANOVA. The analysis gave insight on how the important CPIs defined by the previous chapters and the validation in section 4.3.1, was affected by the management and environmental factors in Table 4-7. The ANOVA displayed the sensitivity of the diverse conditions on weather and soil type as well as the management of each farm as significant contributors in the CPI (Duration) variability. At the start of each growing season, the soil moisture and temperature play a crucial role in germination and crop emergence (Rinaldi et al., 2005). The significant interaction of soil and year ($p<0.001$)

indicates that the impact of soil type depends on the year (or, in other words, the weather conditions). Same conditions were noticed with the year alone but in contrast to the individual soil category that showed a less significant effect, for example, on duration ($p<0.05$). In contrast, the soil with the preceding crop interaction did not display significance related to the different N applications adjusted based on previous crop information that allows more efficient N application (Liu et al., 2003). This way, the field with crops in the previous season had less residual nitrogen in the soil usually received higher amounts of fertiliser to compensate for the difference. The farm management can affect the duration due to different harvest and sowing periods, but these dates are usually determined by the weather, soil workability, and grain moisture. The interesting point is that the G_midP is significantly affected by all individual categories as well as the year*soil interaction, concluding that the management, weather and soil can affect the establishment (Rinaldi et al., 2005) and the timing of the construction phase.

4.4.2 Combination of the CPIs to increase yield prediction

The combination of selected important indicators can provide a more robust analysis of the field production. These results display the potential of the CPIs to quantify yield without using any data from the farm. Multiple linear regression has been used to estimate height and biomass in rice fields, for example (Ndikumana et al., 2018) and even compared with ANN (artificial neural networks) (Kumar et al., 2018; Palakuru, Adamala and Bachina, 2020). Applications of Sentinel-1 and MLR are described to simulate soil moisture (Chatterjee, Huang and Hartemink, 2020). There is a noticeable increase in literature to use machine learning approaches to improve the performance of the models to simulate the biophysical properties of the crops.

Multiple linear regression models were used based on the important CPIs derived from the previous chapters. The use of the important CPIs in the regression can connect the different effects of the biophysical parameters to the production. That can make this easier to be upscaled in large areas, as the driving force of the yield description using MRL is based on the crop biological changes. At the same time, the increase in variables at the multiple linear regression is not efficient at the level of importing the interaction of some of the indicators in the formula. Using only the two most important CPIs from the previous analysis in Chapter 2 (Duration, booting timing) had a low potential to predict the yield. The increase in efficiency in the two statistical criteria (R^2_{adj} and AIC) shows that the best way to describe the yield is the formula that involves the crucial development stages and the indicators that describe important crop characteristics such as the

biomass and the plant water content. The introduction of the two interaction factors did not improve the results like the increase of factors penalised the R^2_{adj} values in contrast with the adding effect of those that were not significant.

The analysis of the two case studies based on yield data availability displayed adequate predictive power using regression in the same growing season. The results show the regression analysis is a helpful tool for extrapolating regions in the agricultural landscape in the same growing season. However, individual CPIs work better across multiple years in the same area based on the significant correlations displayed in the correlation matrix (Figure 4-6). The attempt to predict the yield calibrating the best-selected model with one year and apply that in the rest of the growing seasons did not display high prediction power with the Pbias varying from -18.7 % to 23% in the different combinations summarised in Table 4-13 and the RMSE was constantly above 1 t/ha and up to 2.2 t/ha in the validation years. The combination of the indicators in the regression models (Table 4-12, Table 4-13), displays that the seasonality effect on individual CPI varies, making the correlation coefficients getting low values. This reduces the applicability of the methodology using regression moving from one year to the next in the same area that is not optimal for individual farms with low variation in yield (Norfolk) but still can inform better the region with higher yield variation across seasons or across multiple areas in the same year (Chapter 2).

4.4.3 Applications of SAR-derived CPIs

The idea of mapping regions using indicators can assist in monitoring the agricultural fields in a specific area and even be able to locate fields with a significant difference from the surrounding landscape (Chatterjee, Huang and Hartemink, 2020). The development stages CPIs are good indicators that can be easily displayed across farms in map format. At the same time, the G_max and S_max can also potentially show the biomass and plant water capacity respectively of the winter wheat across the fields, but the lack of available in-situ measurements makes it difficult to get a good corelation with the CPIs. This mapping can assist irrigation strategies as Sentinel-1 data are sensitive to the crop's water status and soil moisture (Bastiaanssen, Molden and Makin, 2000; Gao et al., 2018; El Hajj et al., 2016). A biomass proxy can be used on fertiliser applications as the estimates of the biomass are already used for that. The results of correlation with the biophysical properties need to be improved to be able apply in the farm landscape and the current noise due to speckle effect of the radar signal and the resolution, don’t allow the methodology to be applicable. Most common is NDVI (Vizzari, Santaga and

Benincasa, 2019) or other bands from optical sensors like Sentinel-2 (Söderström et al., 2017). Even the combination using the regression of the CPIs can assist in yield prediction mapping in the same year (R^2=0.5, which is higher than any individual CPI from the correlation matrix). These results can assist in a yield estimation framework that can include other sources of information, tools (models) as well as other sensors as the current methodology can't stand by itself on yield prediction. This proposed CPI method could be an additional tool that will contribute to the improvement of monitoring crop performance with limited predictive power but is easily used across large areas (satellite data).

Applying the methodology across other farms and the increase of data points and diversification of the different categories used in the ANOVA may help increase the correlation value and improve yield prediction. Further research is needed to invert the CPI values to biomass and water content and other biophysical characteristics at field and subfield scale across farms to assist precision farming. The increase of the spatial resolution could potentially improve the predictions, improving the representation of the variability, as lack of resolution and speckle makes this satellite-based SAR method unreliable for small or oddly shaped fields. Larger fields will have less noise due to more pixels averaged across the area. Currently, the methodology lacks the predictive power at the subfield scale and does better work at the field scale, which makes it possible to draw comparisons at the farm level. More research involving different sensors will be needed to increase the accuracy of the current methodology.

4.5 Conclusions

The results of this chapter's improved the understanding of how SAR-derived crop productivity indicators relate to inter-regional and inter-seasonal yield variation and their potential to improve regional monitoring. The analysis focused on the validation of the previously established methodology (Chapter 2) in different farms, areas and growing seasons. The correlation matrices of two case studies were able to display similar results with the Rothamsted multi-annual analysis (Chapter 2). Both duration and TZmax displayed low but significant correlation ($p<0.05$) with the inter-annual case ($r= 0.17$ and $r= -0.21$ respectively). In the case of inter-regional the correlations were also significant with $r= 0.53$ for duration and $r= -0.59$ for TZmax.

The effects of the year (weather) displayed in the ANOVA was persistent, having high significance across all key CPIs from the correlation matrices ($p<0.001$). Similar results had been shown by the farm factor concluding that CPIs affected by different management (Table 4-9). The soil type also showed an effect ($p < 0.001$) with the stem elongation indicator (G_midP) as well as less effect on plant water capacity indicator S_max ($p < 0.1$) that can be related to the establishment and water availability at the end of the season, respectively. That is also connected with effect from the interaction factor between year and soil type ($p<0.001$), indicating that the soil effect, on these two CPIs as well as on duration, is based on the weather of each year.

Multiple linear regression was developed to determine the optimum combination of SAR-derived indicators to estimate yield. The combination with the best selection criteria (AIC=204.4 in the inter-annual case) was the use of the two time related CPIs (G_midP, S_midP) and the biomass (G_max) as well as water related CPI (S_max).The analysis focused on one year had better results (R^2_{adj} =0.46 in CS2 compare to R^2_{adj}=0.01 in CS1 in Table 4-12) as expected considering the ANOVA indicated year significantly affect all the important CPIs ($p<0.001$). The results of the multi-season analysis using the best-selected model in one year displayed high bias (Pbias from -18.7 % to 23%) on the rest of the years in the same region (Table 4-13). Future work should include multiple farms across different soil types and climate conditions that can understand the prons of the CPIs and limitations and potential ranges per factor.

5 SYNTHESIS

The launch of the Sentinel-1a satellite in 2014 gave new opportunities for radar applications in landscape monitoring, which increased with its twin (1b), launched in 2016. High temporal frequency SAR backscatter data with both satellites in orbit from April 2016 became available across the globe, boosting opportunities for research into the use of radar applications in agriculture. The research work for this thesis started in 2017, as dense time series of SAR backscatter from the Sentinel-1 satellites became available which could be used to examine the growing season for winter wheat (Sep-Aug). The temporal approach of the analysis created the need to quantify the in-field changes to interpret specific characteristics of the time series of SAR-backscatter. The mathematical parameters of these dynamic characteristics aimed to describe the construction and maturation phases of crop development, establishing a framework of CPIs. These indicators of the VH/VV backscatter ratio were explored in terms of their relation to the development, growth and yield of winter wheat crops.

The research gap was form around the explanation and use of time dense SAR dynamics to correlate with wheat growth cycle and derive quantitative indicators for crop monitoring. Thus, the aim of the thesis is to describe the vegetation dynamics using SAR backscatter across fields and correlate the time-series characteristics derived from the VH/VV backscatter ratio (CPIs) curve with crop production and enable monitoring of crop changes across regions.

The objectives described in the introduction are displayed below:

1. Define CPIs from the temporal characteristics of the VH/VV ratio by developing a quantitative approach. (**Chapter 2**)
2. Connect the CPIs with *in situ* information to explain the sensitivity to biophysical properties of wheat. (**Chapter 3)**
3. Explore the inter-regional and inter-annual variability of CPIs with the yield. (**Chapter 4)**

The objective of this Chapter is to synthesise how CPIs can be used across scales (plot-field-farm) to improve our understanding of regional crop monitoring. To achieve this, I integrate the results of these three objectives and discuss the key findings in the light of other studies, and finally, draw conclusions how this could be implemented in practice and further improved in future research.

5.1 Deriving quantitative CPIs from SAR-backscatter dynamics

The framework for parameterising the temporal curve of SAR-backscatter was the initial step to relate SAR-derived indicators with winter wheat crop growth characteristics (Chapter 2). The thesis has demonstrated that CPIs can be derived from the curve characteristics for the periods of growth and senescence of VH/VV time series. The defined CPIs include the timings of these main growth phases, slopes of development and the values, like maximum and minimum of the amplitude. The identification of the respective development stages was crucial to identify the construction and maturation periods using field observations at first, to enable the split of the temporal curves using only the curve characteristics. The parameter values from each logistic curve were compared with the yield in each field across the three seasons of available data (2017-2019). The two CPIs with the significant correlation to the field level yield were the Duration (derived from the distance between the two midpoints of the two logistics curves) with r = 0.61 and the TZmax (timing of booting derived from the max of the smoothed curve) with r = -0.56 based on the correlation matrix. The positive relationship of yield with duration is supported by previous work that has related high yields with a longer periods of grain filling (Biscoe and Willington, 1984; Monteith, 2007). The booting timing works in a similar way as the earlier the value, the earlier the crop develops, which is closely related to the temperature in the area (De Bernardis et al., 2016; Rinaldi et al., 2005).

The shape of the time series was instrumental to relate field observations to the inflexion points or maxima, for example, when the maximum occurred across the year, which was observed to coincide with the booting period for winter wheat (BBCH = 41-49), also confirmed by Harfenmeister, Spengler and Weltzien, 2019; Mercier et al., 2020. Other researchers also found this temporal pattern that can assist in crop classification (Bargiel, 2017; Veloso et al., 2017). This observation helps to split the year into two periods, each of which can be described by a logistic curve. Crop growth has been simulated with logistic curves before (Villegas, 2001) in work reported for optical sensors using different indicators (Che et al., 2014; Son et al., 2016; Zhang et al., 2003), but there were no previous study assessing Sentinel-1 time-series using this approach. The usage of two curves instead of one double logistic curve was selected due to the structural change in May. Ear emergence after the booting period changes the VH-VV ratio to a lower value. The structural change in the wheat canopy is described by D_max (G_max-S_max) and significant correlation with height is not present anymore on this period, as the change

in shape at the top layer of wheat (ear) influence the scattering mechanism. Based on the literature in Chapter 1, during low vegetation (G_base) or bare soil periods (S_base), the effect of soil roughness and moisture change is visible to the radar backscatter (Baghdadi et al., 2018; Bousbih et al., 2017). The logistic curve's use helps define the baseline (in each field) in Chapter 2, considering that one value describes the tillering period. This means that sudden changes, e.g. rainfall, will not affect the value, as such events increase the variability of the signal (higher standard deviation of the mean values of that day) that gives less weight in the curve fitting method (Section 2.2.7).

To understand the response of the SAR-backscatter signal to field biophysical properties (biomass, LAI, height, PWC, phenology) in Chapter 3, especially at plot scale, the uncertainty sources of SAR should be taken into consideration. Variation composed by orbits, pre-processing and spatial averaging of backscatter values should be minimised, as it will influence the extraction of the CPIs from the satellite data. The different orbits with different incidence angles, and morning (descending) and evening (ascending) became apparent (Figure 5-1) as the trends of the VH/VV ratio are affected during the period of high vegetation before maturation. Using one orbit to produce the VH/VV ratio temporal curve is required to reduce the variation in the backscatter values across the season (Weiß et al., 2021). The selection of one orbit reduces the revisit time of the satellites above the specific area of interest to 6 days, but the incidence angle remains in the range of 5° (depends on the local slopes), and the orientation of the satellite view remains constant (Mladenova et al., 2013; Vaudour, Baghdadi and Gilliot, 2014). In the UK, the ascending orbit (evening) was selected to avoid the interference of early morning dew visible in the morning acquisition that will affect the backscatter due to sensitivity to water in the surface (Harfenmeister, Spengler and Weltzien, 2019; Khabbazan et al., 2019).

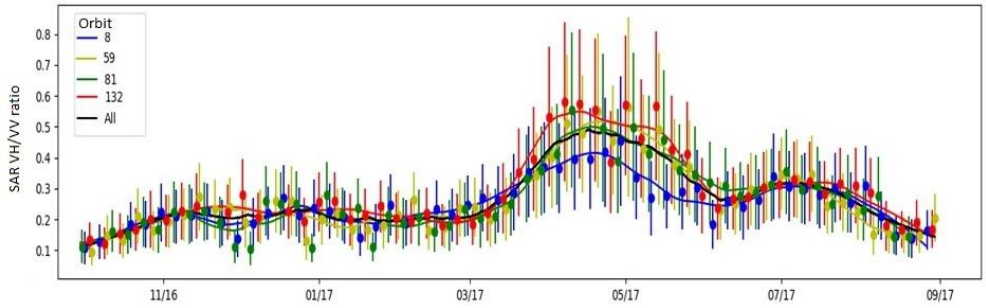


Figure 5-1 Example of the VH/VV temporal curves of the same field with the VH/VV ratio field average points and standard deviation, based on four different orbits (8, 81 for descending and 59, 132 for ascending) passing from the same farm in Norfolk displaying differences during the vegetation period.

Even with the use of one orbit, there is still variation in the SAR temporal curve as the acquisition on different days means that the state of the crop, soil conditions, and weather will affect the attenuation and reflectance of the radar signal. For example, defining the maximum VH/VV ratio will be very difficult as fluctuating values close to the high vegetation period will shift the crucial timing used to define the two crop growth periods. The use of a smoothing curve (Savitzky and Golay, 1964) helps to identify and separate the trends of the backscatter (Torbick et al., 2017) from weather variation. This will improve to identify key development stages in the field also using the derivatives of this curve (Figure 2-4). High standard deviation is noticed in high values of the VH/VV ratio time series (Figure 2-5) compared to the low values in the tillering period (BBCH=20-30).The variability in each field average daily value of the polarisations is incorporated during the logistic curve fitting. This method minimises the square of the points' distance with the line that is divided by the standard deviation (Section 2.2.7). This way, the outlier points due to weather conditions or other random outliers have a small effect on the curve fitting.

In Chapter 3, the sub-field plot scale affected the calculation of specific CPIs (Figure 3-10) as the difference in the pixel number averaging played crucial role on the noise and variation of the indicators. Part of the pre-possessing of SAR images in the lee filter averages based on a window of pixels. A common use is 7 x 7 pixels (Barbouchi et al., 2016; Denize et al., 2018; Liu et al., 2013; Mercier et al., 2020; Van Tricht et al., 2018) or even 5 x 5 (Filgueiras et al., 2019; Laurin et al., 2018; Li and Wang, 2018). In this thesis, the use of a “3 x 3” Lee filter (Weiß et al., 2021; Zhou et al., 2017) was selected as the GPS points of *in situ* data (chapter 3) in each position are covered mostly by a plot (15 m radius), meaning that 3 x 3 will be more representative to the sample area. in addition, the relatively small field size in Rothamsted experimental farm plays a role in the effect of the surroundings. Small fields translated to small number of pixels and (depends on the shape) higher percentage of border pixels per field. In the case of 7 x 7, the pixels close to the borders would have been affected from more pixels outside the field. There is more significant effect on the CPIs scale comparison in 2019 compared to 2018 displayed in Table 3-5. The most affected CPIs are related to the baseline that less averaged pixels will influenced the mean value as soil contributes more in low vegetation (Harfenmeister, Spengler and Weltzien, 2019). It is important to decide the scale of the analysis (calculate CPIs) from the pre-processing step, during the speckle filtering, as the area that is needed to be averaged for the CPI calculations will be affected by the filtering.

5.2 CPIs and biophysical properties of field and plot scale

In Chapter 3, key crop development stages were identified based on weekly field observations that helped to define the temporal curve periods. The ability to estimate the timing of the different development stages can help in different ways, depending on the application. The comparison of key stages within samples of the same farm can identify differences in management that affect crop development rates as well as response to mitigating water availability of soil types as it would be independent from weather. That can also be applied when comparing different regions where soil and weather will have different effects on development. Another way to apply the phenology sensitivity of the methodology, is to examine changes between years in development in the same area (soil properties will remain the same). Comparing the two summers of 2018 and 2019 in Rothamsted farm (Figure 3-8) with *in situ* data displayed the desiccation effect of a drier season (2018), resulting in a shorter grain-fill period and lower plant water content in July (r = -0.76 between S_midP and PWC%). The use of the SAR-derived development stages on calibrating or validating models can assist in landscape monitoring as more information on a large scale is available using satellites and reduce the amount of *in situ* information needed for monitoring crops.

The results of the fieldwork displayed a sensitivity of the signal with canopy height and biomass in the growth period until the booting stage of the cereal (,Figure 3-9). It also showed that plant water content (%) drives the reduction of the signal response during the senescence period (Figure 3-8). The individual polarisations also arrived at their minimum at this period creating a phenological stage saturation at BBCH 51 to describe height (60 cm) and AGB (below 600 g m^{-2}) that agrees with other literature results that there is no further increase in the VH/VV signal in spite of the fact that ear emergence increases height by another 20 cm (Betbeder et al., 2016; Fieuzal, Baup and Marais-Sicre, 2013; Harfenmeister, Spengler and Weltzien, 2019). This effect is described by the D_max indicator. The limitation of the biomass mapping due to water reduction at the maturation period affects the correlation between CPIs and yield. Most of the applications at the end of the harvest season focus on the water balance and management monitoring of the harvest area such as soil moisture, harvest, ploughing and sowing dates (Bastiaanssen, Molden and Makin, 2000; Forkuor et al., 2017; Snapir et al., 2018; Son et al., 2016; Vaudour, Baghdadi and Gilliot, 2014). The proposed methodology can assist using the water related CPIs (Table 2-2) together with the trend of the SAR VH/VV ratio to estimate moisture at harvest, but the problem raised is that

weather events close to harvest period will affect the estimation. The actual harvest date needed to be determined from weather and management information (when the machinery is available).

An objective of the plot scale analysis (Chapter 3) was to relate the CPIs with the biophysical properties represented by the five sample points in each field. The definition of the plot inside the field was based on a centroid of randomly sampled points (across the season in five positions) and a radius of 15 m to cover most of the adjacent pixels. The low but still statistically significant values of correlations are highly influenced by the SAR noise (speckle) of using fewer pixels in averaging at the plot compared to the field scale. CPIs that displayed biological meaning in their correlations with the crop characteristics can be differentiated based on the period they represent. The CPIs from the logistic growth curve (construction) were significantly correlated with the *in situ* observations in May involving dry AGB, leaf mass and tiller density. The maximum amplitude of the SAR Backscatter Ratio (G_max) displayed a significant correlation with dry AGB in May (r = 0.32), leaf mass (r = 0.34) and tiller density (r = 0.39). In contrast, no correlation was established with the height measured in May but in contrast, height change during stem elongation (Figure 3-11) is visible during the temporal evolution of SAR polarisations that period (Nasirzadehdizaji et al., 2019; Veloso et al., 2017; Vreugdenhil et al., 2018).

Maturation-related CPIs correlated with water-related properties as the dry conditions in 2018 greatly affected the development stages and the region's and season's water limitation (S_midP and PWC% r = -0.76). The harvest index (HI) had a better relationship with the CPIs than the yield data: G_max, S_midP and Duration were significantly correlated with the Harvest index (r = -0.29, 0.39, 0.32, respectively). That is an interesting result that can help identify the crop's efficiency to produce the grain and is strongly related to the different varieties and growing conditions (Dai et al., 2016). More data from the different growing seasons will contribute to the increase of the correlations.

Direct correlation with the differences in phenological stages was not applicable in this case as *in situ* data were taken from a single farm with homogenous management and temperature ranges, forcing all wheat fields to develop at a similar rate. The timing related CPIs had a standard deviation around 6-14 days at field scale in most cases across the same farm in chapter 3 (Table 3-4) and across farms in Chapter 4 (Table 4-8) that is very close to the revisit time of the satellites (6 days). Further improvement should

involve the incorporation of temperature through quantitative, process-based modelling (De Bernardis et al., 2016).

The CPIs displayed significant correlations with the crop characteristics, which is in line with other literature (Harfenmeister, Spengler and Weltzien, 2019; Veloso et al., 2017; Vreugdenhil et al., 2018). However, the estimation of crop biophysical values (e.g. biomass) is challenging and needs more *in situ* data to calibrate the relationship between CPI and the desired parameter. Different approaches have been used in the literature to estimate biophysical properties from RS and EO data, such as yield (Fieuzal and Baup, 2017), development stages (De Bernardis et al., 2016; Nasrallah et al., 2019; Schlund and Erasmi, 2020), biomass, height, LAI (Betbeder, Fieuzal and Baup, 2016; Chang, Shoshany and Oh, 2018; Fassnacht et al., 2021) and water content (Han et al., 2019) that involved regressions, machine learning approaches, data fusion with optical (De Bernardis et al., 2016) model assimilation (Huang et al., 2019; Jin et al., 2018; Kasampalis et al., 2018) and *in situ* data to emulate the properties from a variety of crops. Different approaches can be used with CPIs to improve the predictive power in regards to crop properties, for example regression, crop modelling (data assimilation) or even machine learning, but these approaches need a large volume of in *situ* data to cover large ranges of properties.

5.3 Potential benefits of regional monitoring and agricultural applications

SAR has the advantage of cloud penetration help on the high temporal resolution of information, especially in regions where optical data cannot be retrieved and assist this away on better monitoring of changes in fields. High temporal frequency data have the potential to detect dynamic patterns across large areas (Bargiel, 2017). Another advantage of the proposed methodology is that eventually, the calculation of indicators can be based solely on SAR images with no information from the fields (i.e. fully automated). In Chapter 4, field-specific temporal analysis of different farms has been conducted to relate the radar-derived CPIs to crop productivity and how different local conditions may influence the sensitivity of these indicators.

The analysis of four farms showed that yield was correlated with the individual CPIs (Figure 4-6, Figure 4-7), confirming the results, thus validating the approach established on the Hertfordshire farm of Rothamsted Research (Figure 2-9). The same indicators (duration and TZmax) have shown sensitivity to yield using Norfolk area across multiple

seasons (CS1) and for the season of 2017 across four farms in Hertfordshire and Norfolk (CS2). The different soil, year and management practice (tested using ANOVA) significantly affected the selected CPIs, that displayed relevance for yield formation (duration, TZmax) and biophysical properties (G_midP, G_max, S_max, S_midP). Soil type influence in crops is based on physical, chemical, and biological conditions that characterise most soil quality indicators (Bünemann et al., 2018). It is essential to point out that the soil interaction with the yield sensitive CPIs was significant ($p<0.001$) when the year was considered in the ANOVA. Those results have shown that the dynamics of the soil characteristics and its interaction with annual weather patterns play a crucial role on the crop monitoring (Table 4-9). Soil processes will be affected by biophysical changes caused by weather conditions (water balance, climate) and farm management (Vereecken et al., 2016).

The indicators selected for the multilinear regression analysis were based on criteria using the results from all chapters. Selected indicators based on:

- The CPI definitions and results with yield correlation on Chapter 2,
- Their sensitivity for biophysical soil and crop properties (Chapter 3),
- Validation on other farms from the correlation matrices with yield (Chapter 4), and
- The environmental and management effects (ANOVA in Chapter 4)

The results of Chapter 2 displayed the significant correlation of TZmax and Duration (S_midP - G_midP) with the yield (at field-scale). Chapter 3 also added the structural sensitivity of G_max and the plant water relationship with S_max and displayed the importance of time-related indicators again. The validation using correlation matrices in Chapter 4 reinforced the duration and TZmax with lower but significant correlations. In the same chapter, the ANOVA showed the sensitivity of the selected indicators (based on the previous results) to changes in management, weather (year) and soil (based on the year) that can assist the regression as these factors affect the crop production across the agricultural landscape.

There are limitations to the methodology that includes the resolution (due to speckle effect), need for in-situ data (regression) and the pedoclimatic information (different soil types and weather patterns).The yield in Norfolk covers a smaller range and high productivity (average >9 t/ha; CV 7%) in 2017 compared to Hertfordshire farms with underperforming fields (average >7 t/ha; CV 12 - 22%) during the same year (Table 4-8). The results of the multilinear regression shown the potential of the methodology to

recognise high and low yield areas whilst the differentiation within a high performing farm can be difficult as it will be affected by low variability of yield.

The potential benefit for regional monitoring using the proposed methodology is focused on the sensitivity of the indicators for yield and other biophysical properties. Phenology is recognised as one of the crop characteristics that can be monitored using the time series of radar (Bargiel, 2017; Nasrallah et al., 2019; Schlund and Erasmi, 2020) and optical satellites (De Bernardis et al., 2016) and is applicable to the methodology displayed in the thesis. The proposed methodology defined CPIs sensitive to the plant water content in maturation as demonstrated by others (Han et al., 2019; Paloscia et al., 1999). This thesis showed that this property can assist in monitoring the maturation rate of the crop. Biomass is another biophysical parameter in the crop growth that displayed correlation with SAR (Chang, Shoshany and Oh, 2018; Ndikumana et al., 2018; Paloscia et al., 1999) and especially with the CPIs (G_max) defined in the construction phase of wheat. The temporal signature of VH/VV ratio can also assist on land classification frameworks (Bargiel, 2017; Veloso et al., 2017) and the management approaches in crop rotation across the agricultural landscape.

5.4 Research Impact

The initial approach was to focus on radar signal to understand how it related to the field status of the crop, winter wheat. It is essential to display the correlation between the field observations and SAR backscatter to connect the wheat biological properties. I have developed a novel approach for curve fitting and crop parameter determination from the dynamic SAR backscatter polarisation ratio (VH/VV) in Chapter 2. This provides the ability to identify yield components, define crop productivity indicators (CPI's) related to the *in situ* data, and become the base for developing a management tool to provide farmers with field information. This can contribute to better understanding the use of this kind of indicator, as well as it will display the limitations and the sensitivity in crop monitoring.

The correlation of these indicators with the field observations displays valuable information of the SAR interaction with the wheat canopy such as development stage, LAI, biomass, PWC (Canisius et al., 2018; Han et al., 2019; Harfenmeister, Spengler and Weltzien, 2019; Wiseman et al., 2014). Monitoring the soil-crop interaction using soil maps to group response of the signal will also provide a more generalised approach of signal sensitivity in crop and soil variability (Forkuor et al., 2017). This will contribute

towards the increase of current knowledge of SAR interaction with crop characteristics in different conditions across time and space. High spatial and temporal resolution data will enable better comparison between areas. The ability to upscale gives the advantage to understand the influence of management in the distinct stages of crop development. Combining that with the correlation of the crop physical characteristics with the radar backscatter could extract more information about the individual fields across the landscape (Fassnacht et al., 2021; Nasrallah et al., 2019).

The combination of potential CPIs and the spatial variability across different scales will enable us to estimate the condition of crops under different management and conditions (soil types, weather). These indicators can help map different areas displaying locations that may need different management to improve crop development and yield at the end of the season. In parallel, correlation of *in situ* data that are being collected for signal understanding could potentially translate the satellite data into physical properties of the crop (LAI, moisture, biomass) (Canisius et al., 2018; Han et al., 2019; Nasrallah et al., 2019; Ndikumana et al., 2018; Paloscia et al., 1999) that can assist process-based models (Palakuru, Adamala and Bachina, 2020) to simulate the system (under a data assimilation scheme) and extract useful information about the crop growth (Betbeder, Fieuzal and Baup, 2016; Jin et al., 2018; Kasampalis et al., 2018).

The remote sensing data can offer information on a large scale that could assist in improving the crop simulation in the agricultural landscape (Fassnacht et al., 2021; Liu et al., 2013; Nasrallah et al., 2019). This process will contribute toward the application of SAR data in agriculture and achieve a higher level of farm monitoring by using the indicators as biophysical proxies to drive the crop models. For example, the time related CPIs can assist on the phenology simulation of the crop. The ability to extract useful information at a large scale with good spatial resolution could have an enormous impact on farming: Taking advantage of the satellite data's sensitivity to crop changes could help to improve monitoring of crop development and eventually improve yield forecasts. The advantage of the methodology is the use of SAR images with no field data needed for the CPI extraction, making the application in other areas with the same revisit times from Sentinel-1 satellites possible.

The crop classification was not part of the objectives in this project, but as different crops will respond differently based on the physical characteristics of the canopy, there is potential to use the described methodology to assist in classification frameworks. The temporal SAR change can easily recognise the increase in volume when the crop enters

the construction phase and the water reduction due to the maturation, providing times of crop development that can distinguish crop types. Sentinel-1 time series have already been used for crop classification, especially in areas where the cloud cover dominates the growing season (Bargiel, 2017; McNairn et al., 2009a).

In conclusion, the potential application of the research work can be summarised below:

- Estimation of yield production for wheat fields across farms in different regions
- Calculation of key development stages across fields from farm management
- Indication about Biomass and Plant water content at field scale
- Understanding of the whole growing season using temporal changes
- Crop model improvement using data assimilation

5.5 Recommendations for future work

The use of the two logistic approach displayed the potential to describe the temporal dynamics, though with limitations. This is especially important at the area where the two curves intersect. The only indicator that describes the ear emerge in this case is the difference in the two maximums that didn't display significant correlation with the field production. The disconnection is a bigger problem as at smaller scales the SAR speckle affect the curve fitting approach, reducing the sensitivity of the CPIs. Further research can be conducted to improve the connection with in-situ data and test different curves that may improve the fitting.

The next steps for the methodology to be improved should involve more *in situ* data to test across sites and additional observed years and fields are required to better understand the variability of CPIs across the landscape. The transfer of the methodology to other crops needs to simulate the SAR temporal signatures as different crop types have different structures and development stages (timings). Crop classification is based on that to map the crops across large areas. The methodology can be transferred easily to other cereals, like barley, for example, as its signature is very close to that of wheat and can also be represented by two logistic curves. In contrast, oilseed rape has different development that other curves may be considered to successfully replicate the development of the crop (Veloso et al., 2017; Vreugdenhil et al., 2018).

The combination of the SAR with optical data can unlock more spaceborne potential, depending on the limitations due to cloud cover of the location. Field-specific zonation of crop productivity using yield or NDVI maps (Sentinel-2) will give the spatial variability that

can be linked to changes in soil texture to provide indicators using the SAR-derived productivity indicators. Segmentation of fields based on dependent datasets to SAR temporal curve will be the basis for distinguishing the fields' heterogeneity. NDVI or/and soil maps could be used as the base of this. The CPIs calculated based on NDVI derived segments in specific fields could be tested on sensitivity with the yield similar to the plot scale in this analysis. The difference will be that the homogeneity based on optical indices can potentially reduce the variability in the plot, not be limited by the area surrounding the sample. Further examination of the CPIs on soil heterogeneity by calculating the indicator's response across a large range of soil types could potentially help to derive soil properties that affect the crop dynamics. Figure 5-2 displays an example of NDVI segmentation and VH/VV backscatter ratio smoothed curve on different segments.

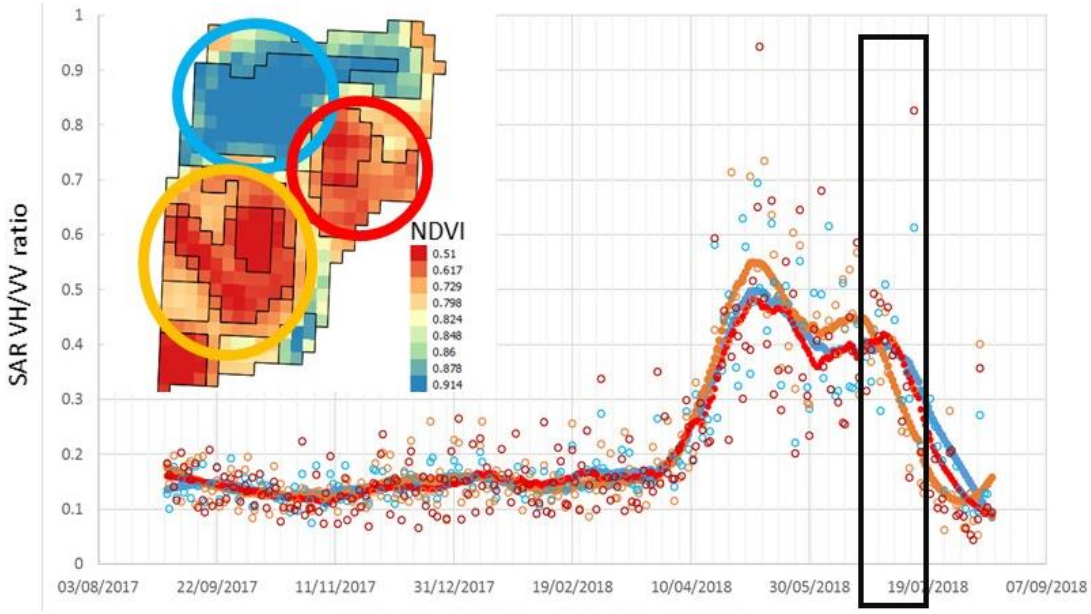


Figure 5-2 Example of using NDVI as the base to define areas with similar characteristics that can be examined using the SAR temporal analysis. The black box shows the time of the optical image acquisition and the smoothed lines display the SAR VH/VV ratio temporal evolution on the three zones in 2018.

Analysis should be conducted about the sensitivity of this methodology to the different incidence angles (Figure 5-1) of the observations (topography), weather conditions in the fields related to very low temperatures in winter and the radar sensitivity to water in the system that affect the transfer and validity of the approach to other sites. The timings of the crop will be affected less than the amplitude as they follow the temporal changes in the growing season connected to development changes, but CPIs connected with biomass and plant water content, the incidence angle will affect the backscattering (Weiß

et al., 2021). Another challenge to keep in mind is the use of the methodology in areas with a 12-day revisit time (outside of Europe) can be challenging to get a clear curve and the climate difference may affect the curve's parameters and how the growing season can be shorter or longer.

5.6 Conclusions

This research was able to use SAR data (Sentinel-1) to quantify the temporal characteristics of the SAR VH/VV-ratio successfully and derive 12 CPIs (from the two logistic curves), aiming to monitor this way the whole growing season of wheat. The objective was achieved by using SAR-only data and assuming a logistic crop development function to extract significant ($p<0.05$) CPIs for yield prediction (Duration, $r=0.61$ and the booting indicator TZmax, $r=-0.56$).

These CPIs were also related to crop developmental and biophysical properties using *in situ* data collected in 2018 and 2019 at Rothamsted experimental farm. The canopy structure and water content related to VH/VV ratio development over time and the derived CPIs displayed correlation with field observations. G_max was correlated with the structural changes in the construction phase ($r = 0.32$, $p<0.05$ with Biomass in May). S_max was positively correlated with the plant water content in the late stages of the construction phase ($r= 0.39$, $p<0.05$) as well as with S_midP ($r= 0.57$, $p<0.05$) that was affected by the arid conditions in the summer of 2018.

The methodology was successfully applied across other farms located in different regions and the CPIs duration and TZmax displayed significant corelation ($p<0.05$) with the yield in inter-annual (CS1) and inter-regional (CS2) analysis. Each of the different categories of information regarding management, weather, and soil with year/weather interaction significantly affected the selected CPIs (Table 4-9). The multilinear regression used to combine temporal (development stage) and state (biomass, water) CPIs resulted in a stronger correlation with yield. The method increased the yield prediction power in the same season across farms ($R^2= 0.5$) but was unable to make prediction across multiple years for the higher yielding farms of Norfolk ($R^2_{ajdj} <0.1$).

The use of CPIs, derived from SAR backscatter ratio, will create opportunities to improve farm and landscape management through Earth observation. It allows to improve monitoring crop growth and development and in particular maturation, even if in-field variability is challenging due to high fluctuation in SAR backscatter and random temporal variation by weather. This enables the user of the method to gain insight into yield formation. Temporal and spatial information from SAR backscatter ratio, therefore, paves the way to improve field specific crop model applications at the farm and regional scale.

REFERENCES

Alexakis, D.D., Mexis, F.D.K., Vozinaki, A.E.K., Daliakopoulos, I.N. and Tsanis, I.K. (2017) 'Soil Moisture Content Estimation Based on Sentinel-1 and Auxiliary Earth Observation Products. A Hydrological Approach', *Sensors*, 17(6), p. 1455.

Alvarez-Mozos, J., Casali, J., Gonzalez-Audicana, M. and Verhoest, N.E.C. (2006) 'Assessment of the operational applicability of RADARSAT-1 data for surface soil moisture estimation', *IEEE Transactions on Geoscience and Remote Sensing*, 44(4), pp. 913–924. Available at: 10.1109/TGRS.2005.862248 (Accessed: 30 September 2021).

Amoah, J.K.O., Amatya, D.M. and Nnaji, S. (2013) 'Quantifying watershed surface depression storage: determination and application in a hydrologic model', *Hydrological Processes*, 27(17), pp. 2401–2413.

Aschbacher, J. and Milagro-Pérez, M.P. (2012) 'The European Earth monitoring (GMES) programme: Status and perspectives', *Remote Sensing of Environment*, 120(2012) Elsevier Inc., pp. 3–8.

Attarzadeh, R., Amini, J., Notarnicola, C. and Greifeneder, F. (2018) 'Synergetic use of Sentinel-1 and Sentinel-2 data for soil moisture mapping at plot scale', *Remote Sensing*, 10(8), pp. 1–18.

Attema, E.P.W. and Ulaby, F.T. (1978) 'Vegetation modeled as a water cloud', *Radio Science*, 13(2), pp. 357–364.

Avery, B.W. and Catt, J.A. (1995) *The Soil at Rothamsted The Soil-map Units*.

Baghdadi, N., El Hajj, M., Choker, M., Zribi, M., Bazzi, H., Vaudour, E., Gilliot, J.-M. and Ebengo, D. (2018) 'Potential of Sentinel-1 Images for Estimating the Soil Roughness over Bare Agricultural Soils', *Water*, 10(2), p. 131.

Baghdadi, N., El Hajj, M., Zribi, M. and Bousbih, S. (2017) 'Calibration of the Water Cloud Model at C-Band for Winter Crop Fields and Grasslands', *Remote Sensing*, 9(9), p. 969.

Barbouchi, M., Chokmani, K., Abdelfattah, R. and Ben Aissa, N. (2016) 'Yield estimation of the winter wheat using Radarsat 2 polarimetric SAR reponse', *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. IEEE, pp. 555–560.

Bargiel, D. (2017) 'A new method for crop classification combining time series of radar

images and crop phenology information', *Remote Sensing of Environment*, 198 Elsevier Inc., pp. 369–383.

Bastiaanssen, W.G.., Molden, D.J. and Makin, I.W. (2000) 'Remote sensing for irrigated agriculture: examples from research and possible applications', *Agricultural Water Management*, 46(2), pp. 137–155.

Behrenfeld, M.J. and Falkowski, P.G. (1997) 'Photosynthetic rates derived from satellite-based chlorophyll concentration', *Limnology and Oceanography*, 42(1), pp. 1–20.

Bériaux, E., Waldner, F., Collienne, F., Bogaert, P. and Defourny, P. (2015) 'Maize Leaf Area Index Retrieval from Synthetic Quad Pol SAR Time Series Using the Water Cloud Model', *Remote Sensing*, 7(12), pp. 16204–16225.

De Bernardis, C., Vicente-Guijalba, F., Martinez-Marin, T. and Lopez-Sanchez, J.M. (2016) 'Contribution to Real-Time Estimation of Crop Phenological States in a Dynamical Framework Based on NDVI Time Series: Data Fusion With SAR and Temperature', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8), pp. 3512–3523.

Betbeder, J., Fieuzal, R. and Baup, F. (2016) 'Assimilation of LAI and Dry Biomass Data From Optical and SAR Images Into an Agro-Meteorological Model to Estimate Soybean Yield', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6), pp. 2540–2553.

Betbeder, J., Fieuzal, R., Philippets, Y., Ferro-Famil, L. and Baup, F. (2016) 'Contribution of multitemporal polarimetric synthetic aperture radar data for monitoring winter wheat and rapeseed crops', *Journal of Applied Remote Sensing*, 10(2), p. 026020.

Bhuiyan, H.A.K.M., McNairn, H., Powers, J., Friesen, M., Pacheco, A., Jackson, T.J., Cosh, M.H., Colliander, A., Berg, A., Rowlandson, T., Bullock, P. and Magagi, R. (2018) 'Assessing SMAP Soil Moisture Scaling and Retrieval in the Carman (Canada) Study Site', *Vadose Zone Journal*, 17(1), p. 180132.

Biscoe, P. V and Willington, V.B.A. (1984) Environmental effects on dry matter production *Nitrogen requirement of cereals : proceedings of a conference organised by the Agricultural Development and Advisory Service, September 1982*. London : H.M.S.O., 1984.,

Blaes, X., Defourny, P., Wegmuller, U., Della Vecchia, A., Guerriero, L. and Ferrazzoli,

P. (2006) ‘C-band polarimetric indexes for maize monitoring based on a validated radiative transfer model’, *IEEE Transactions on Geoscience and Remote Sensing*, 44(4), pp. 791–800.

Bousbih, S., Zribi, M., Lili-Chabaane, Z., Baghdadi, N., El Hajj, M., Gao, Q. and Mougenot, B. (2017) ‘Potential of Sentinel-1 Radar Data for the Assessment of Soil and Cereal Cover Parameters’, *Sensors*, 17(12), p. 2617.

Bozdogan, H. (1987) ‘Model selection and Akaike’s Information Criterion (AIC): The general theory and its analytical extensions’, *Psychometrika*, 52(3), pp. 345–370.

Brown, S.C.M., Quegan, S., Morrison, K., Bennett, J.C. and Cookmartin, G. (2003) ‘High-resolution measurements of scattering in wheat canopies - Implications for crop parameter retrieval’, *IEEE Transactions on Geoscience and Remote Sensing*, 41(7 PART I), pp. 1602–1610.

Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., de Goede, R., Fleskens, L., Geissen, V., Kuyper, T.W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J.W. and Brussaard, L. (2018) ‘Soil quality – A critical review’, *Soil Biology and Biochemistry*, 120(February) Elsevier, pp. 105–125.

Bush, T.F. and Ulaby, F.T. (1978) ‘An evaluation of radar as a crop classifier’, *Remote Sensing of Environment*, 7(1), pp. 15–36.

Campos, I., González-Gómez, L., Villodre, J., Calera, M., Campoy, J., Jiménez, N., Plaza, C., Sánchez-Prieto, S. and Calera, A. (2019) ‘Mapping within-field variability in wheat yield and biomass using remote sensing vegetation indices’, *Precision Agriculture*, 20(2) Springer US, pp. 214–236.

Canisius, F., Shang, J., Liu, J., Huang, X., Ma, B., Jiao, X., Geng, X., Kovacs, J.M. and Walters, D. (2018) ‘Tracking crop phenological development using multi-temporal polarimetric Radarsat-2 data’, *Remote Sensing of Environment*, 210(February) Elsevier, pp. 508–518.

Casanova, J.J., Judge, J. and Miyoung Jang (2007) ‘Modeling Transmission of Microwaves Through Dynamic Vegetation’, *IEEE Transactions on Geoscience and Remote Sensing*, 45(10), pp. 3145–3149.

Chang, J.G., Shoshany, M. and Oh, Y. (2018) ‘Polarimetric Radar Vegetation Index for Biomass Estimation in Desert Fringe Ecosystems’, *IEEE Transactions on Geoscience*

and Remote Sensing, 56(12) IEEE, pp. 7102–7108.

Chatterjee, S., Huang, J. and Hartemink, A.E. (2020) 'Establishing an Empirical Model for Surface Soil Moisture Retrieval at the U.S. Climate Reference Network Using Sentinel-1 Backscatter and Ancillary Data', *Remote Sensing*, 12(8) Multidisciplinary Digital Publishing Institute, p. 1242. Available at: 10.3390/rs12081242 (Accessed: 1 September 2021).

Che, M., Chen, B., Zhang, H., Fang, S., Xu, G., Lin, X. and Wang, Y. (2014) 'A New Equation for Deriving Vegetation Phenophase from Time Series of Leaf Area Index (LAI) Data', *Remote Sensing*, 6(6), pp. 5650–5670.

Choker, M., Baghdadi, N., Zribi, M., El Hajj, M., Paloscia, S., Verhoest, N., Lievens, H. and Mattia, F. (2017) 'Evaluation of the Oh, Dubois and IEM Backscatter Models Using a Large Dataset of SAR Data and Experimental Soil Measurements', *Water*, 9(1), p. 38.

Cranfield University (2021) *The Soils Guide*., *Cranfield University* Available at: www.landis.org.uk (Accessed: 13 May 2021).

Dai, J., Bean, B., Brown, B., Bruening, W., Edwards, J., Flowers, M., Karow, R., Lee, C., Morgan, G., Ottman, M., Ransom, J. and Wiersma, J. (2016) 'Harvest index and straw yield of five classes of wheat', *Biomass and Bioenergy*, 85 Elsevier Ltd, pp. 223–227.

Denize, J., Hubert-Moy, L., Betbeder, J., Corgne, S., Baudry, J. and Pottier, E. (2018) 'Evaluation of Using Sentinel-1 and -2 Time-Series to Identify Winter Land Use in Agricultural Landscapes', *Remote Sensing 2019, Vol. 11, Page 37*, 11(1) Multidisciplinary Digital Publishing Institute, p. 37. Available at: 10.3390/RS11010037 (Accessed: 1 September 2021).

Dobson, M., Ulaby, F., Hallikainen, M. and El-rayes, M. (1985) 'Microwave Dielectric Behavior of Wet Soil-Part II: Dielectric Mixing Models', *IEEE Transactions on Geoscience and Remote Sensing*, GE-23(1), pp. 35–46.

Earl, R., Taylor, J.., Wood, G.., Bradley, I., James, I.., Waine, T., Welsh, J.., Godwin, R.. and Knight, S.. (2003) 'Soil Factors and their Influence on Within-field Crop Variability, Part I: Field Observation of Soil Variation', *Biosystems Engineering*, 84(4), pp. 425–440.

Earth Observation and Research Branch Team (2011) *Crop Identification and BBCH Staging Manual: SMAP-12 Field Campaign*.

van Emmerik, T., Steele-Dunne, S.C., Judge, J. and van de Giesen, N. (2015) ‘Impact of Diurnal Variation in Vegetation Water Content on Radar Backscatter From Maize During Water Stress’, *IEEE Transactions on Geoscience and Remote Sensing*, 53(7), pp. 3855–3869.

European Space Agency (2007) *ASAR Product Handbook*.

FAO (2011) ‘Food , Agriculture and Cities. Challenges of food and nutrition security, agriculture and ecosystem management in an urbanizing world’, *FAO, Rome, Italy*, , p. 48.

Fassnacht, F.E., Poblete-Olivares, J., Rivero, L., Lopatin, J., Ceballos-Comisso, A. and Galleguillos, M. (2021) ‘Using Sentinel-2 and canopy height models to derive a landscape-level biomass map covering multiple vegetation types’, *International Journal of Applied Earth Observation and Geoinformation*, 94(April 2020), p. 102236.

Fieuzal, R. and Baup, F. (2017) ‘Forecast of wheat yield throughout the agricultural season using optical and radar satellite images’, *International Journal of Applied Earth Observation and Geoinformation*, 59 Elsevier B.V., pp. 147–156.

Fieuzal, R., Baup, F. and Marais-Sicre, C. (2013) ‘Monitoring Wheat and Rapeseed by Using Synchronous Optical and Radar Satellite Data—From Temporal Signatures to Crop Parameters Estimation’, *Advances in Remote Sensing*, 02(02) Scientific Research Publishing, pp. 162–180. Available at: 10.4236/ars.2013.22020 (Accessed: 7 September 2021).

Filgueiras, R., Mantovani, E.C., Althoff, D., Fernandes Filho, E.I. and Cunha, F.F. da (2019) ‘Crop NDVI Monitoring Based on Sentinel 1’, *Remote Sensing*, 11(12) Multidisciplinary Digital Publishing Institute, p. 1441. Available at: 10.3390/rs11121441 (Accessed: 1 September 2021).

Fischer, W.A., Hemphill, W.R. and Kover, A. (1976) ‘Progress in remote sensing (1972–1976)’, *Photogrammetria*, 32(2) Elsevier, pp. 33–72.

Forkuor, G., Hounkpatin, O.K.L., Welp, G. and Thiel, M. (2017) ‘High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models’, *PLOS ONE*, 12(1), p. e0170478.

Franceschetti, G. and Lanari, R. (1999) ‘Synthetic aperture radar processing’, *Electronic*

engineering systems series, CRC Press

Friesen, J., Steele-Dunne, S.C. and van de Giesen, N. (2012) 'Diurnal Differences in Global ERS Scatterometer Backscatter Observations of the Land Surface', *IEEE Transactions on Geoscience and Remote Sensing*, 50(7), pp. 2595–2602.

Gao, Q., Zribi, M., Escorihuela, M., Baghdadi, N. and Segui, P. (2018) 'Irrigation Mapping Using Sentinel-1 Time Series at Field Scale', *Remote Sensing*, 10(9), p. 1495.

Gherboudj, I., Magagi, R., Berg, A.A. and Toth, B. (2011) 'Soil moisture retrieval over agricultural fields from multi-polarized and multi-angular RADARSAT-2 SAR data', *Remote Sensing of Environment*, 115(1), pp. 33–43.

Giannakis, G.V., Nikolaidis, N.P., Valstar, J., Rowe, E.C., Moirogiorgou, K., Kotronakis, M., Paranychianakis, N.V., Rousseva, S., Stamati, F.E. and Banwart, S.A. (2017) 'Integrated Critical Zone Model (1D-ICZ)', in *Advances in Agronomy*. 1st edn. Elsevier Inc., pp. 277–314.

Gomez-Chova, L., Tuia, D., Moser, G. and Camps-Valls, G. (2015) 'Multimodal Classification of Remote Sensing Images: A Review and Future Directions', *Proceedings of the IEEE*, 103(9), pp. 1560–1584.

El Hajj, M., Baghdadi, N., Zribi, M., Belaud, G., Cheviron, B., Courault, D. and Charron, F. (2016) 'Soil moisture retrieval over irrigated grassland using X-band SAR data', *Remote Sensing of Environment*, 176, pp. 202–218.

Han, D., Liu, S., Du, Y., Xie, X., Fan, L., Lei, L., Li, Z., Yang, H. and Yang, G. (2019) 'Crop Water Content of Winter Wheat Revealed with Sentinel-1 and Sentinel-2 Imagery', *Sensors 2019, Vol. 19, Page 4013*, 19(18) Multidisciplinary Digital Publishing Institute, p. 4013. Available at: 10.3390/s19184013 (Accessed: 1 September 2021).

Harfenmeister, K., Spengler, D. and Weltzien, C. (2019) 'Analyzing Temporal and Spatial Characteristics of Crop Parameters Using Sentinel-1 Backscatter Data', *Remote Sensing*, 11(13), p. 1569.

Huang, J., Gómez-Dans, J.L., Huang, H., Ma, H., Wu, Q., Lewis, P.E., Liang, S., Chen, Z., Xue, J.H., Wu, Y., Zhao, F., Wang, J. and Xie, X. (2019) 'Assimilation of remote sensing into crop growth models: Current status and perspectives', *Agricultural and Forest Meteorology*, 276–277(17) Elsevier, p. 107609.

Huang, W., Sun, G., Ni, W., Zhang, Z. and Dubayah, R. (2015) 'Sensitivity of multi-source SAR backscatter to changes in forest aboveground biomass', *Remote Sensing*, 7(8), pp. 9587–9609.

Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W. and Rowland, C.S. (2019) 'High resolution wheat yield mapping using Sentinel-2', *Remote Sensing of Environment*, 233, pp. 1–37.

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G. and Wang, J. (2018) 'A review of data assimilation of remote sensing and crop models', *European Journal of Agronomy*, 92(November 2017) Elsevier, pp. 141–152.

Kanevsky, M.B. (2008) *Radar imaging of the ocean waves*. Elsevier.

Kansakar, P. and Hossain, F. (2016) 'A review of applications of satellite earth observation data for global societal benefit and stewardship of planet earth', *Space Policy*, 36 Elsevier Ltd, pp. 46–54.

Karam, M.A. and Fung, A.K. (1989) 'Leaf-shape effects in electromagnetic wave scattering from vegetation', *IEEE Transactions on Geoscience and Remote Sensing*, 27(6), pp. 687–697.

Karam, M.A., Fung, A.K., Lang, R.H. and Chauhan, N.S. (1992) 'A microwave scattering model for layered vegetation', *IEEE Transactions on Geoscience and Remote Sensing*, 30(4), pp. 767–784.

Kasampalis, D., Alexandridis, T., Deva, C., Challinor, A., Moshou, D. and Zalidis, G. (2018) 'Contribution of Remote Sensing on Crop Models: A Review', *Journal of Imaging*, 4(4), p. 52.

Khabbazan, S., Vermunt, P., Steele-Dunne, S., Arntz, L.R., Marinetti, C., van der Valk, D., Iannini, L., Molijn, R., Westerdijk, K. and van der Sande, C. (2019) 'Crop monitoring using Sentinel-1 data: A case study from The Netherlands', *Remote Sensing*, 11(16), pp. 1–24.

Klosterman, S.T., Hufkens, K., Gray, J.M., Melaas, E., Sonnentag, O., Lavine, I., Mitchell, L., Norman, R., Friedl, M.A. and Richardson, A.D. (2014) 'Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery', *Biogeosciences*, 11(16), pp. 4305–4320.

Kumar, P., Prasad, R., Gupta, D.K., Mishra, V.N., Vishwakarma, A.K., Yadav, V.P., Bala, R., Choudhary, A. and Avtar, R. (2018) 'Estimation of winter wheat crop growth parameters using time series Sentinel-1A SAR data', *Geocarto International*, 33(9) Taylor & Francis, pp. 942–956. Available at: 10.1080/10106049.2017.1316781 (Accessed: 1 September 2021).

Lahoz, W.A. and Schneider, P. (2014) 'Data assimilation: making sense of Earth Observation', *Frontiers in Environmental Science*, 2(May), pp. 1–28.

Lancashire, P.D., Bleiholder, H., Boom, T. Van Den, Langelüddeke, P., Stauss, R., Weber, E. and Witzenberger, A. (1991) 'A uniform decimal code for growth stages of crops and weeds', *Annals of Applied Biology*, 119(3), pp. 561–601.

Laurin, G.V., Balling, J., Corona, P., Mattioli, W., Papale, D., Puletti, N., Rizzo, M., Truckenbrodt, J. and Urban, M. (2018) 'Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data', *Journal of Applied Remote Sensing*, 12(01), p. 1.

Li, J. and Wang, S. (2018) 'Using SAR-derived vegetation descriptors in a water cloud model to improve soil moisture retrieval', *Remote Sensing*, 10(9), pp. 11–14.

Liu, C., Shang, J., Vachon, P.W. and McNairn, H. (2013) 'Multiyear crop monitoring using polarimetric RADARSAT-2 data', *IEEE Transactions on Geoscience and Remote Sensing*, 51(4) IEEE, pp. 2227–2240.

Liu, X., Ju, X., Zhang, F., Pan, J. and Christie, P. (2003) 'Nitrogen dynamics and budgets in a winter wheat-maize cropping system in the North China Plain', *Field Crops Research*, 83(2), pp. 111–124.

Mandal, D., Kumar, V., Ratha, D., Dey, S., Bhattacharya, A., Lopez-Sanchez, J.M., McNairn, H. and Rao, Y.S. (2020) 'Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data', *Remote Sensing of Environment*, 247(June) Elsevier, p. 111954.

Mascolo, L., Lopez-Sanchez, J.M., Vicente-Guijalba, F., Mazzarella, G., Nunziata, F. and Migliaccio, M. (2015) 'Retrieval of phenological stages of onion fields during the first year of growth by means of C-band polarimetric SAR measurements', *International Journal of Remote Sensing*, 36(12), pp. 3077–3096.

Mashaba, Z., Chirima, G., Botai, J.O., Combrinck, L., Munghemezulu, C. and Dube, E.

(2017) ‘Forecasting winter wheat yields using MODIS NDVI data for the Central Free State region’, *South African Journal of Science*, 113(11–12), pp. 1–6.

Mattia, F., Thuy Le Toan, Picard, G., Posa, F.I., D’Alessio, A., Notarnicola, C., Gatti, A.M., Rinaldi, M., Satalino, G. and Pasquariello, G. (2003) ‘Multitemporal c-band radar measurements on wheat fields’, *IEEE Transactions on Geoscience and Remote Sensing*, 41(7), pp. 1551–1560.

McNairn, H., Champagne, C., Shang, J., Holmstrom, D. and Reichert, G. (2009a) ‘Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories’, *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(5) Elsevier B.V., pp. 434–449.

McNairn, H., Shang, J., Jiao, X. and Champagne, C. (2009b) ‘The contribution of ALOS PALSAR multipolarization and polarimetric data to crop classification’, *IEEE Transactions on Geoscience and Remote Sensing*, 47(12), pp. 3981–3992.

McNairn, H., Shang, J., Jiao, X. and Deschamps, B. (2012) ‘Establishing Crop Productivity Using Radarsat-2’, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B8(September), pp. 283–287.

Mercier, A., Betbeder, J., Baudry, J., Le Roux, V., Spicher, F., Lacoux, J., Roger, D. and Hubert-Moy, L. (2020) ‘*Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages*’

Met Office (2021) *UK climate averages*. Available at: https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-climate-averages (Accessed: 1 June 2021).

Mironov, V.L., Dobson, M.C., Kaupp, V.H., Komarov, S.A. and Kleshchenko, V.N. (2004) ‘Generalized refractive mixing dielectric model for moist soils’, *IEEE Transactions on Geoscience and Remote Sensing*, 42(4), pp. 773–785.

Mladenova, I.E., Jackson, T.J., Bindlish, R. and Hensley, S. (2013) ‘Incidence Angle Normalization of Radar Backscatter Data’, *IEEE Transactions on Geoscience and Remote Sensing*, 51(3), pp. 1791–1804.

Molijn, R.A., Iannini, L., Mousivand, A. and Hanssen, R.F. (2014) ‘*Analyzing C-band SAR polarimetric information for LAI and crop yield estimations*’, Neale, C. M. U. and Maltese,

A. (eds.), Vol.2, p. 92390V.

Monteith, J.L. (2007) ‘Climatic variation and the growth of crops’, *Quarterly Journal of the Royal Meteorological Society*, 107(454), pp. 749–774.

Moriasi, D.N., Gitau, M.W., Pai, N. and Daggupati, P. (2015) ‘Hydrologic and water quality models: Performance measures and evaluation criteria’, *Transactions of the ASABE*, 58(6), pp. 1763–1785.

Nasirzadehdizaji, R., Sanli, F.B., Abdikan, S., Cakir, Z., Sekertekin, A. and Ustuner, M. (2019) ‘Sensitivity analysis of multi-temporal Sentinel-1 SAR parameters to crop height and canopy coverage’, *Applied Sciences (Switzerland)*, 9(4)

Nasrallah, A., Baghdadi, N., El Hajj, M., Darwish, T., Belhouchette, H., Faour, G., Darwich, S. and Mhawej, M. (2019) ‘Sentinel-1 data for winter wheat phenology monitoring and mapping’, *Remote Sensing*, 11(19)

Ndikumana, E., Ho Tong Minh, D., Dang Nguyen, H., Baghdadi, N., Courault, D., Hossard, L. and El Moussawi, I. (2018) ‘Estimation of Rice Height and Biomass Using Multitemporal SAR Sentinel-1 for Camargue, Southern France’, *Remote Sensing*, 10(9), p. 1394.

Nunziata, F., Migliaccio, M., Sanchez, J.M.L., Mascolo, L., Mazzarella, G. and D’Urso, G. (2015) ‘C-band polarimetric SAR measurements for the monitoring of growth stages of corn fields in the piana DEL Sele zone’, *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, pp. 3377–3380.

Østergaard, A., Snoeij, P., Navas Traver, I., Ludwig, M., Rostan, F. and Croci, R. (2011) ‘C-band SAR for the GMES Sentinel-1 mission’, *European Microwave Week 2011: ‘Wave to the Future’, EuMW 2011, Conference Proceedings - 8th European Radar Conference, EuRAD 2011*, (January), pp. 234–240.

Paget, A.C., Long, D.G. and Madsen, N.M. (2016) ‘RapidScat Diurnal Cycles Over Land’, *IEEE Transactions on Geoscience and Remote Sensing*, 54(6), pp. 3336–3344.

Palakuru, M., Adamala, S. and Bachina, H.B. (2020) ‘Modeling yield and backscatter using satellite derived biophysical variables of rice crop based on Artificial Neural Networks.’, *Journal of Agrometeorology*, 22(1)

Paloscia, S., Macelloni, G., Pampaloni, P. and Sigismondi, S. (1999) ‘The potential of C-

and L-band SAR in estimating vegetation biomass: the ERS-1 and JERS-1 experiments', *IEEE Transactions on Geoscience and Remote Sensing*, 37(4), pp. 2107–2110.

Patel, P., Srivastava, H.S. and Navalgund, R.R. (2006) '*Estimating wheat yield: an approach for estimating number of grains using cross-polarised ENVISAT-1 ASAR data*', Valinia, A., Uratsuka, S. and Misra, T. (eds.), Vol.6410, p. 641009.

Petropoulos, G., Srivastava, P., Piles, M. and Pearson, S. (2018) 'Earth Observation-Based Operational Estimation of Soil Moisture and Evapotranspiration for Agricultural Crops in Support of Sustainable Water Management', *Sustainability*, 10(2), p. 181.

Pinker, R.T. and Laszlo, I. (1992) 'Global Distribution of Photosynthetically Active Radiation as Observed from Satellites', *Journal of Climate*, 5(1), pp. 56–65.

Rahman, M., Moran, M., Thoma, D., Bryant, R., Holifield Collins, C.D., Jackson, T., Orr, B.J. and Tischler, M. (2008) 'Mapping surface roughness and soil moisture using multi-angle radar imagery without ancillary data', *Remote Sensing of Environment*, 112(2), pp. 391–402. Available at: 10.1016/j.rse.2006.10.026 (Accessed: 30 September 2021).

Ratner, B. (2009) 'The correlation coefficient: Its values range between 1/1, or do they', *Journal of Targeting, Measurement and Analysis for Marketing*, 17(2), pp. 139–142.

Rayfuse, R. and Weisfelt, N. (2012) *The Challenge of Food Security*. Edward Elgar Publishing.

Rinaldi, M., Di Paolo, E., Richter, G.M. and Payne, R.W. (2005) 'Modelling the effect of soil moisture on germination and emergence of wheat and sugar beet with the minimum number of parameters', *Annals of Applied Biology*, 147(1), pp. 69–80.

Sahebi, M.R., Bonn, F. and Gwyn, Q.H.J. (2003) 'Estimation of the moisture content of bare soil from RADARSAT-1 SAR using simple empirical models', *International Journal of Remote Sensing*, 24(12) Taylor and Francis Ltd., pp. 2575–2582. Available at: 10.1080/0143116031000072948 (Accessed: 30 September 2021).

Satalino, G., Dente, L. and Mattia, F. (2006) 'Integration of MERIS and ASAR Data for LAI Estimation of Wheat Fields', *2006 IEEE International Symposium on Geoscience and Remote Sensing*. IEEE, pp. 2255–2258.

Savitzky, A. and Golay, M.J.E. (1964) 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures', *Analytical Chemistry*, 36(8), pp. 1627–1639.

Schlund, M. and Erasmi, S. (2020) ‘Sentinel-1 time series data for monitoring the phenology of winter wheat’, *Remote Sensing of Environment*, 246, p. 111814.

Senior, T.B.A., Sarabandi, K. and Ulaby, F.T. (1987) ‘Measuring and modeling the backscattering cross section of a leaf’, *Radio Science*, 22(6), pp. 1109–1116.

Setiyono, T.D., Quicho, E.D., Holecz, F.H., Khan, N.I., Romuga, G., Maunahan, A., Garcia, C., Rala, A., Raviz, J., Collivignarelli, F., Gatti, L., Barbieri, M., Phuong, D.M., Minh, V.Q., Vo, Q.T., Intrman, A., Rakwatin, P., Sothy, M., Veasna, T., Pazhanivelan, S. and Mabalay, M.R.O. (2019) ‘Rice yield estimation using synthetic aperture radar (SAR) and the ORYZA crop growth model: development and application of the system in South and South-east Asian countries’, *International Journal of Remote Sensing*, 40(21), pp. 8093–8124.

Shang, J., Jiao, X., McNairn, H., Kovacs, J., Walters, D., Ma, B. and Geng, X. (2013) ‘Tracking crop phenological development of spring wheat using synthetic aperture radar (SAR) in northern Ontario, Canada’, *2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*. IEEE, pp. 517–521.

Skriver, H., Mattia, F., Satalino, G., Balenzano, A., Pauwels, V.R.N., Verhoest, N.E.C. and Davidson, M. (2011) ‘Crop Classification Using Short-Revisit Multitemporal SAR Data’, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(2), pp. 423–431.

Snapir, B., Waine, T.W., Corstanje, R., Redfern, S., De Silva, J. and Kirui, C. (2018) ‘Harvest Monitoring of Kenyan Tea Plantations With X-Band SAR’, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3), pp. 930–938.

Söderström, M., Piikki, K., Stenberg, M., Stadig, H. and Martinsson, J. (2017) ‘Producing nitrogen (N) uptake maps in winter wheat by combining proximal crop measurements with Sentinel-2 and DMC satellite images in a decision support system for farmers’, *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, 67(7), pp. 637–650. Available at: 10.1080/09064710.2017.1324044 (Accessed: 1 September 2021).

Son, N.-T., Chen, C.-F., Chang, L.-Y., Chen, C.-R., Sobue, S.-I., Minh, V.-Q., Chiang, S.-H., Nguyen, L.-D. and Lin, Y.-W. (2016) ‘A logistic-based method for rice monitoring from multitemporal MODIS-Landsat fusion data’, *European Journal of Remote Sensing*, 49(1), pp. 39–56.

Srivastava, H., Patel, P., Manchanda, M.L. and Adiga, S. (2003) ‘Use of multiincidence angle RADARSAT-1 SAR data to incorporate the effect of surface roughness in soil moisture estimation’, *IEEE Transactions on Geoscience and Remote Sensing*, 41(7), pp. 1638–1640. Available at: 10.1109/TGRS.2003.813356 (Accessed: 30 September 2021).

Steele-Dunne, S.C., McNairn, H., Monsivais-Huertero, A., Judge, J., Liu, P.-W. and Papathanassiou, K. (2017) ‘Radar Remote Sensing of Agricultural Canopies: A Review’, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(5), pp. 2249–2273.

Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L. and van Wesemael, B. (2010) ‘Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy’, *Geoderma*, 158(1–2) Elsevier B.V., pp. 32–45.

De Sy, V., Herold, M., Achard, F., Asner, G.P., Held, A., Kellndorfer, J. and Verbesselt, J. (2012) ‘Synergies of multiple remote sensing data sources for REDD+ monitoring’, *Current Opinion in Environmental Sustainability*, 4(6), pp. 696–706.

Torbick, N., Chowdhury, D., Salas, W. and Qi, J. (2017) ‘Monitoring rice agriculture across myanmar using time series Sentinel-1 assisted by Landsat-8 and PALSAR-2’, *Remote Sensing*, 9(2)

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. and Rostan, F. (2012) ‘GMES Sentinel-1 mission’, *Remote Sensing of Environment*, 120, pp. 9–24.

Van Tricht, K., Gobin, A., Gilliams, S. and Piccard, I. (2018) ‘Synergistic use of radar Sentinel-1 and optical Sentinel-2 imagery for crop mapping: a case study for Belgium’, *Remote Sensing*, , pp. 1–22.

Vaudour, E., Baghdadi, N. and Gilliot, J.M. (2014) ‘Mapping tillage operations over a peri-urban region using combined SPOT4 and ASAR/ENVISAT images’, *International Journal of Applied Earth Observation and Geoinformation*, 28(1) Elsevier B.V., pp. 43–59.

Vavlas, N., Waine, T.W., Meersmans, J., Burgess, P.J., Fontanelli, G. and Richter, G.M. (2020) ‘Deriving Wheat Crop Productivity Indicators Using Sentinel-1 Time Series’,

Remote Sensing, 12(15), p. 2385.

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J.-F. and Ceschia, E. (2017) 'Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications', *Remote Sensing of Environment*, 199, pp. 415–426.

Vereecken, H., Schnepf, A., Hopmans, J.W., Javaux, M., Or, D., Roose, T. and Vanderborght, J. (2016) 'Modelling Soil Processes : Key challenges and new perspectives', *Vadose Zone Journal*

Vicente-Guijalba, F., Martinez-Marin, T. and Lopez-Sanchez, J.M. (2014) 'Crop Phenology Estimation Using a Multitemporal Model and a Kalman Filtering Strategy', *IEEE Geoscience and Remote Sensing Letters*, 11(6), pp. 1081–1085.

Villegas, D. (2001) 'Biomass Accumulation and Main Stem Elongation of Durum Wheat Grown under Mediterranean Conditions', *Annals of Botany*, 88(4), pp. 617–627.

Vizzari, M., Santaga, F. and Benincasa, P. (2019) 'Sentinel 2-Based Nitrogen VRT Fertilization in Wheat: Comparison between Traditional and Simple Precision Practices', *Agronomy*, 9(6) Multidisciplinary Digital Publishing Institute, p. 278. Available at: 10.3390/agronomy9060278 (Accessed: 1 September 2021).

Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C. and Strauss, P. (2018) 'Sensitivity of Sentinel-1 Backscatter to Vegetation Dynamics: An Austrian Case Study', *Remote Sensing*, 10(9), p. 1396.

VSN International (2020) *Genstat for Windows 21st Edition* 21. VSN International, Hemel Hempstead

Wang, J., Engman, E., Mo, T., Schmugge, T. and Shiue, J. (1987) 'The Effects of Soil Moisture, Surface Roughness, and Vegetation on L-Band Emission and Backscatter', *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(6), pp. 825–833.

Weiß, T., Ramsauer, T., Jagdhuber, T., Löw, A. and Marzahn, P. (2021) 'Sentinel-1 backscatter analysis and radiative transfer modeling of dense winter wheat time series', *Remote Sensing*, 13(12)

Wiseman, G., McNairn, H., Homayouni, S. and Shang, J. (2014) 'RADARSAT-2 Polarimetric SAR Response to Crop Biomass for Agricultural Production Monitoring',

IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(11), pp. 4461–4471.

Yueh, S.H., Kong, J.A., Jao, J.K., Shin, R.T. and Le Toan, T. (1992) 'Branching model for vegetation', *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), pp. 390–402.

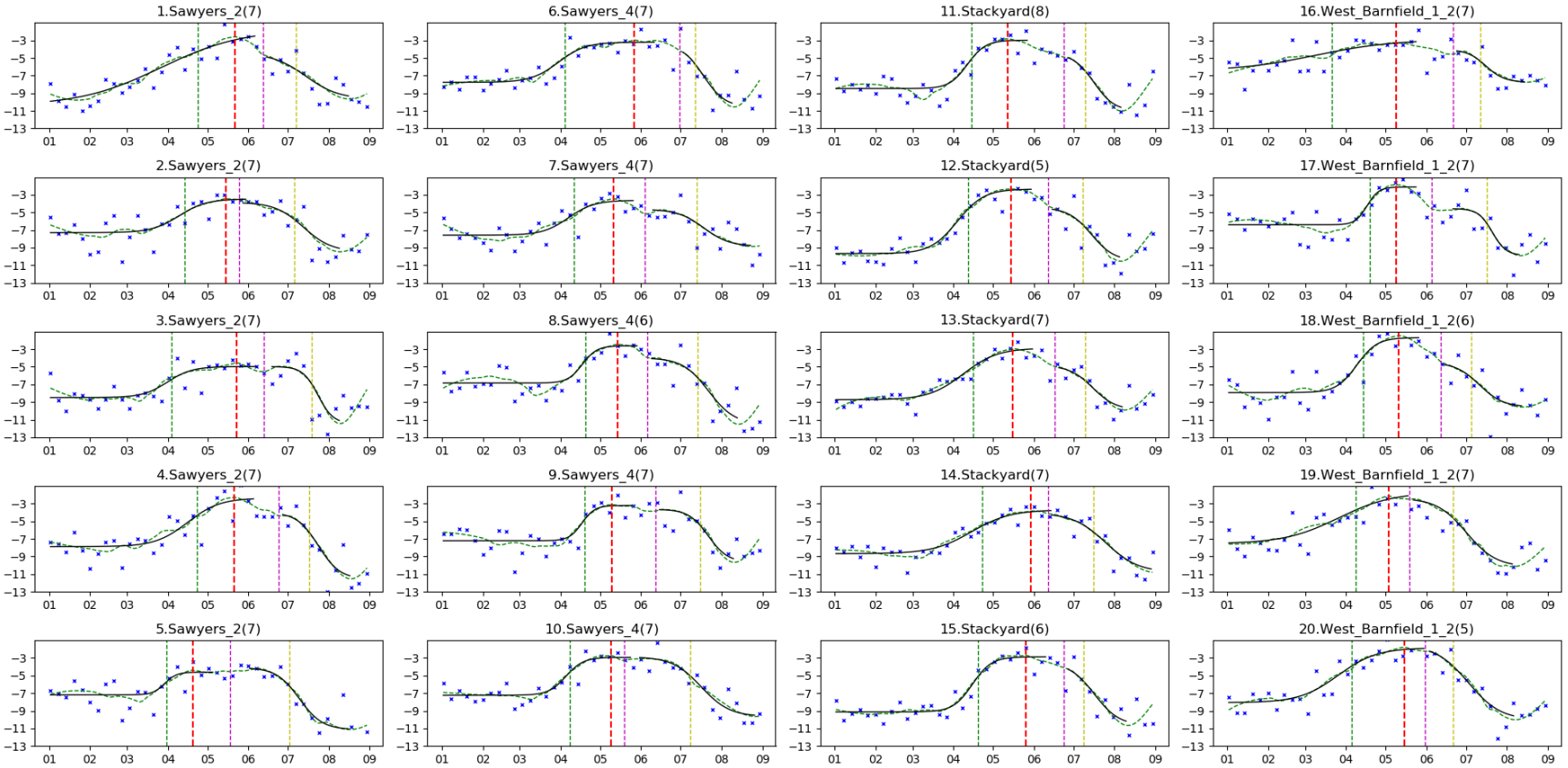
Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C. and Huete, A. (2003) 'Monitoring vegetation phenology using MODIS', *Remote Sensing of Environment*, 84(3), pp. 471–475.

Zhou, T., Pan, J., Zhang, P., Wei, S. and Han, T. (2017) 'Mapping Winter Wheat with Multi-Temporal SAR and Optical Images in an Urban Agricultural Region', *Sensors*, 17(6), p. 1210.

Zimmermann, B. and Kohler, A. (2013) 'Optimizing Savitzky–Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy', *Applied Spectroscopy*, 67(8), pp. 892–902.

APPENDICES

A.1 Plot scale Curve fitting



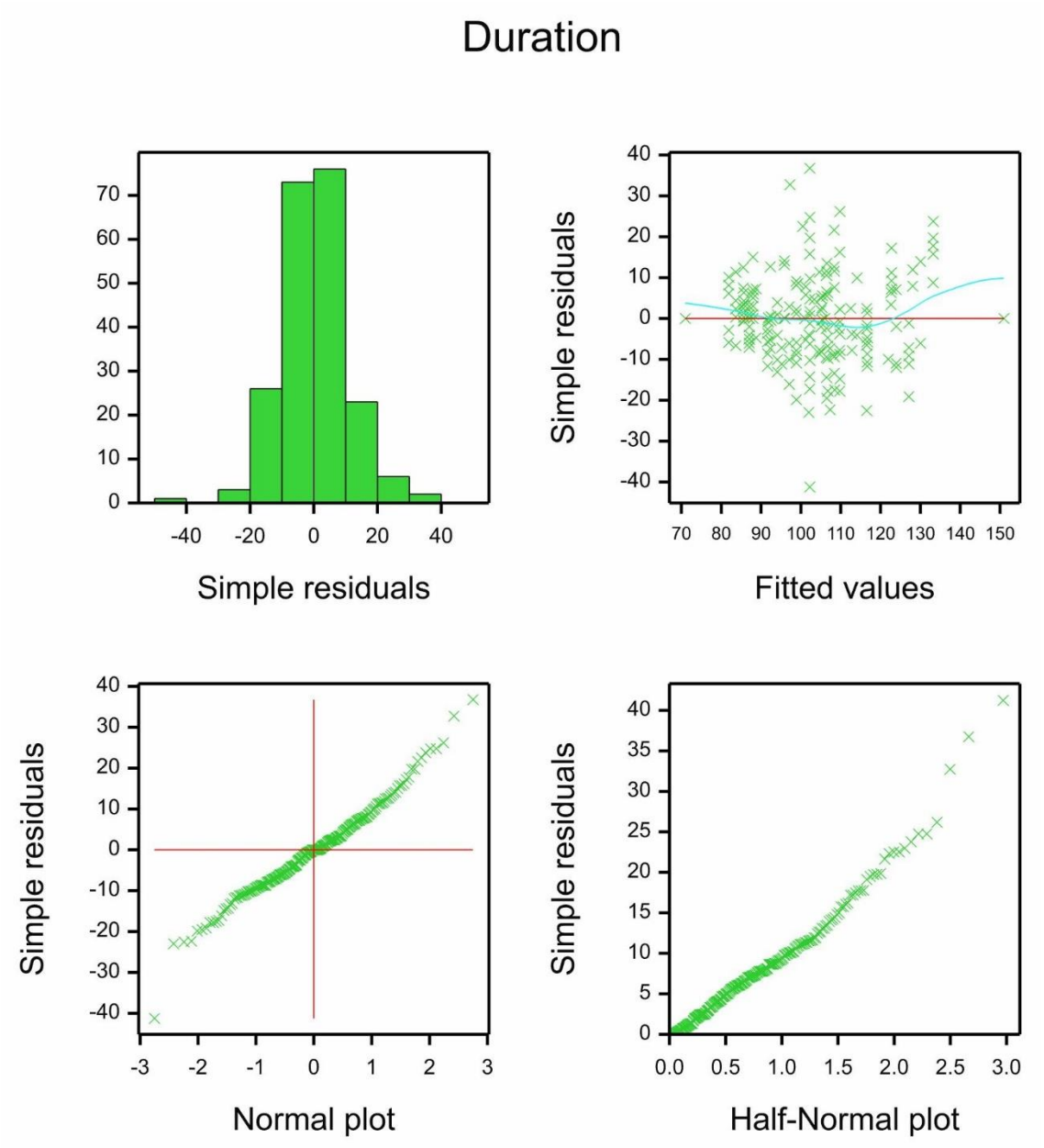
CPI calculation using the two logistic curve fitting for 2018 harvest season. Every column represents the same field with its five sample positions and the parenthesis displays the number of pixel used to average the VH/VV ratio.

A.2 Summary statistics for each season and location

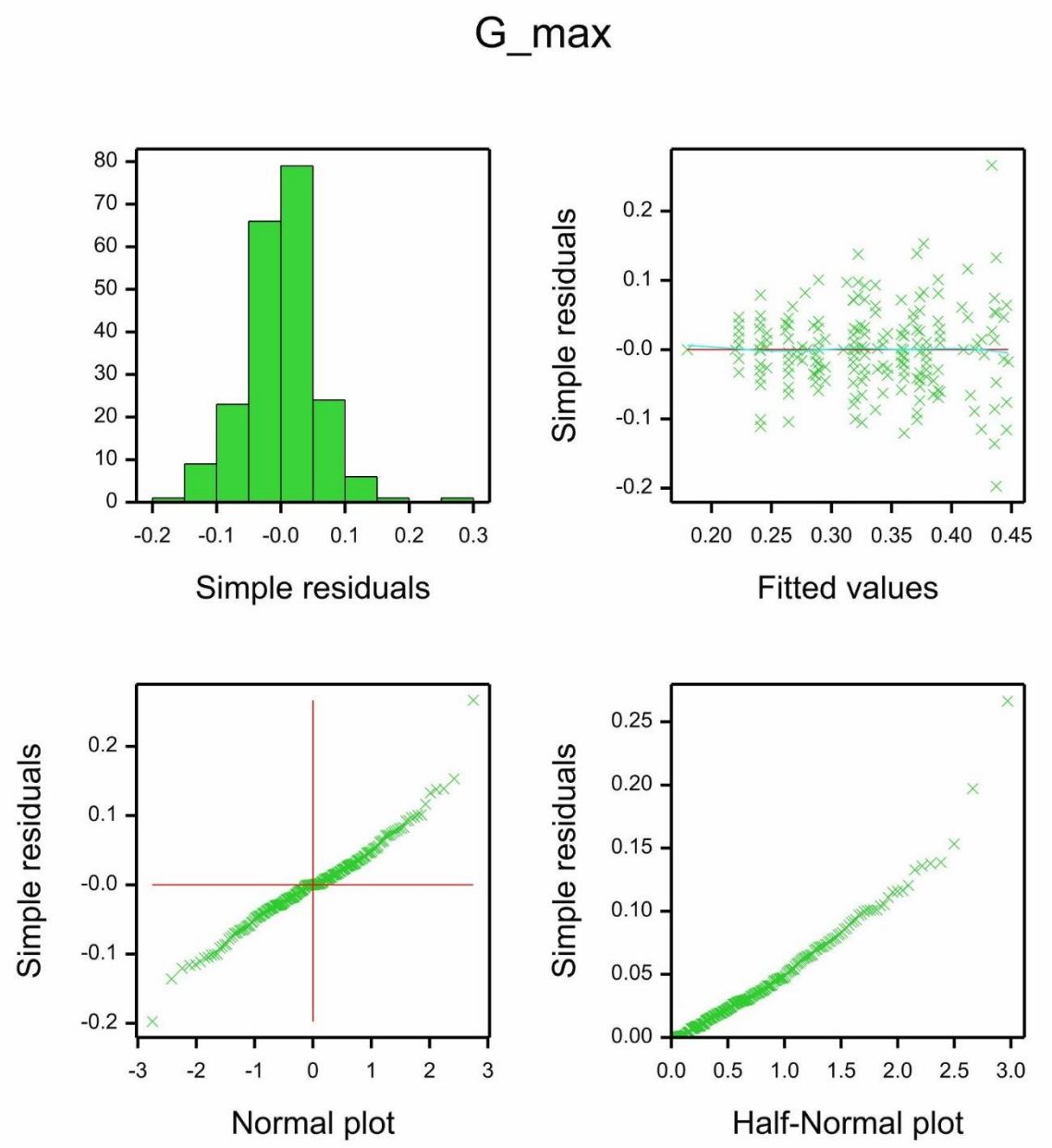
Average values of the defined 12 CPIs calculated form the four different farms for the 2017 season to enable the comparison of the CPIs across the four farms during the common season with available yield data.

Farm	Statistics	G_base	G_steep	G_max	S_max	S_steep	S_base	G_midP	S_midP	TZmax	Duration	D_max	D_base	Yield
								DoY	DoY	DoY	days			t/ha
Roth	Mean	0.15	0.12	0.29	0.20	-0.28	0.13	98	199	131	101	0.0887	0.02	7.1
	Std dev	0.02	0.04	0.06	0.10	0.36	0.02	11	13	9	14	0.1241	0.01	1.6
	CV	15%	35%	20%	51%	-132%	17%	12%	7%	7%	14%	140%	60%	22%
East Hall	Mean	0.16	0.15	0.23	0.23	-0.20	0.12	104	205	136	101	-0.002	0.04	7.5
	Std dev	0.02	0.18	0.04	0.09	0.28	0.03	10	11	6	17	0.1028	0.02	0.9
	CV	11%	121%	19%	39%	-136%	25%	10%	5%	4%	17%	-5736%	56%	12%
Salle	Mean	0.17	0.12	0.33	0.17	-0.23	0.16	92	206	120	114	0.1513	0.02	9.7
	Std dev	0.04	0.03	0.07	0.03	0.09	0.02	9	6	8	7	0.0639	0.03	0.7
	CV	24%	27%	21%	17%	-40%	11%	10%	3%	6%	7%	42%	135%	7%
Hydehall	Mean	0.19	0.15	0.26	0.21	-0.15	0.12	83	210	111	127	0.0478	0.08	9.4
	Std dev	0.04	0.08	0.04	0.08	0.14	0.03	10	34	14	30	0.0944	0.05	0.6
	CV	23%	53%	14%	40%	-92%	23%	12%	16%	12%	24%	198%	70%	7%

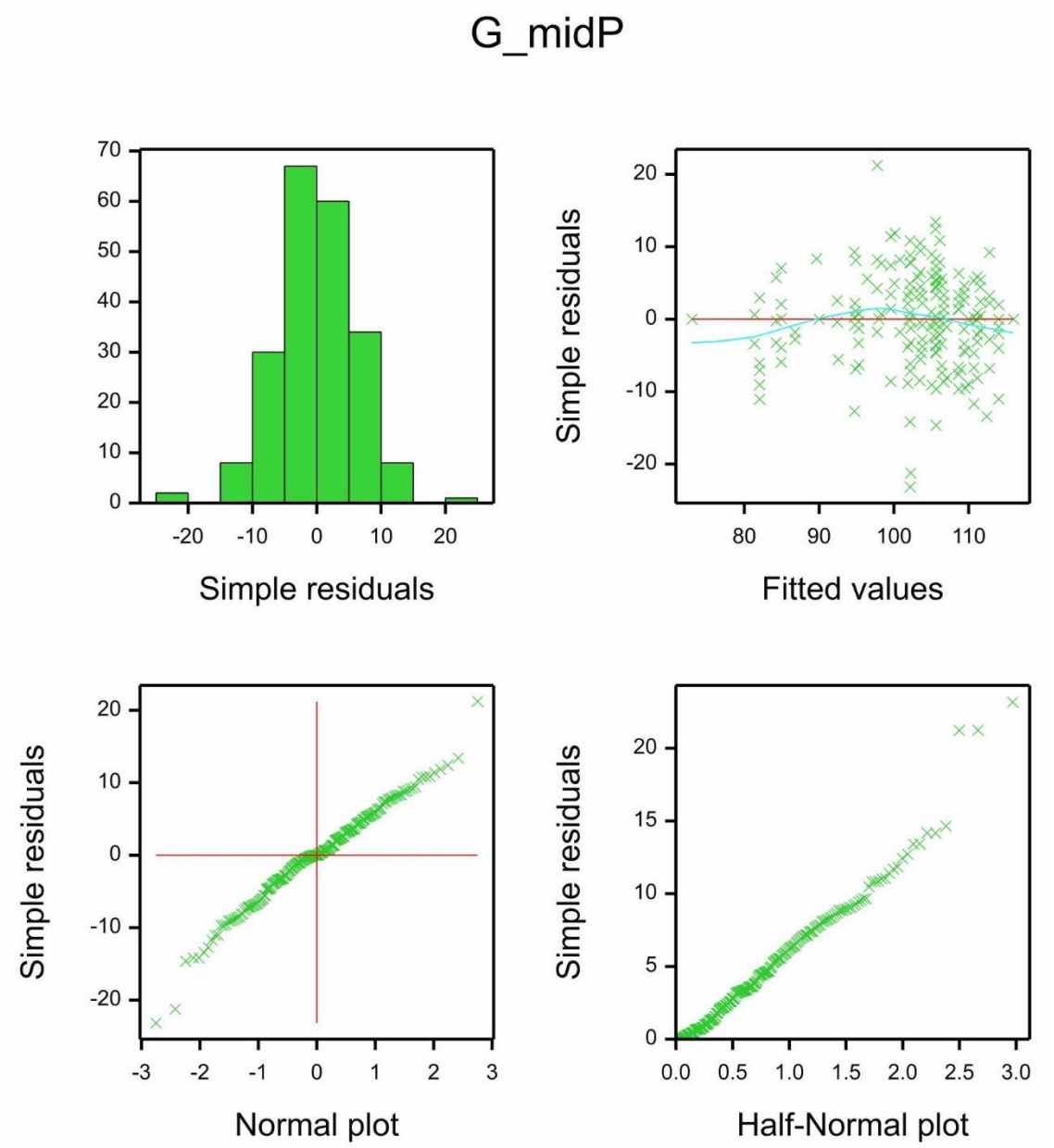
A.3 ANOVA- residual plots



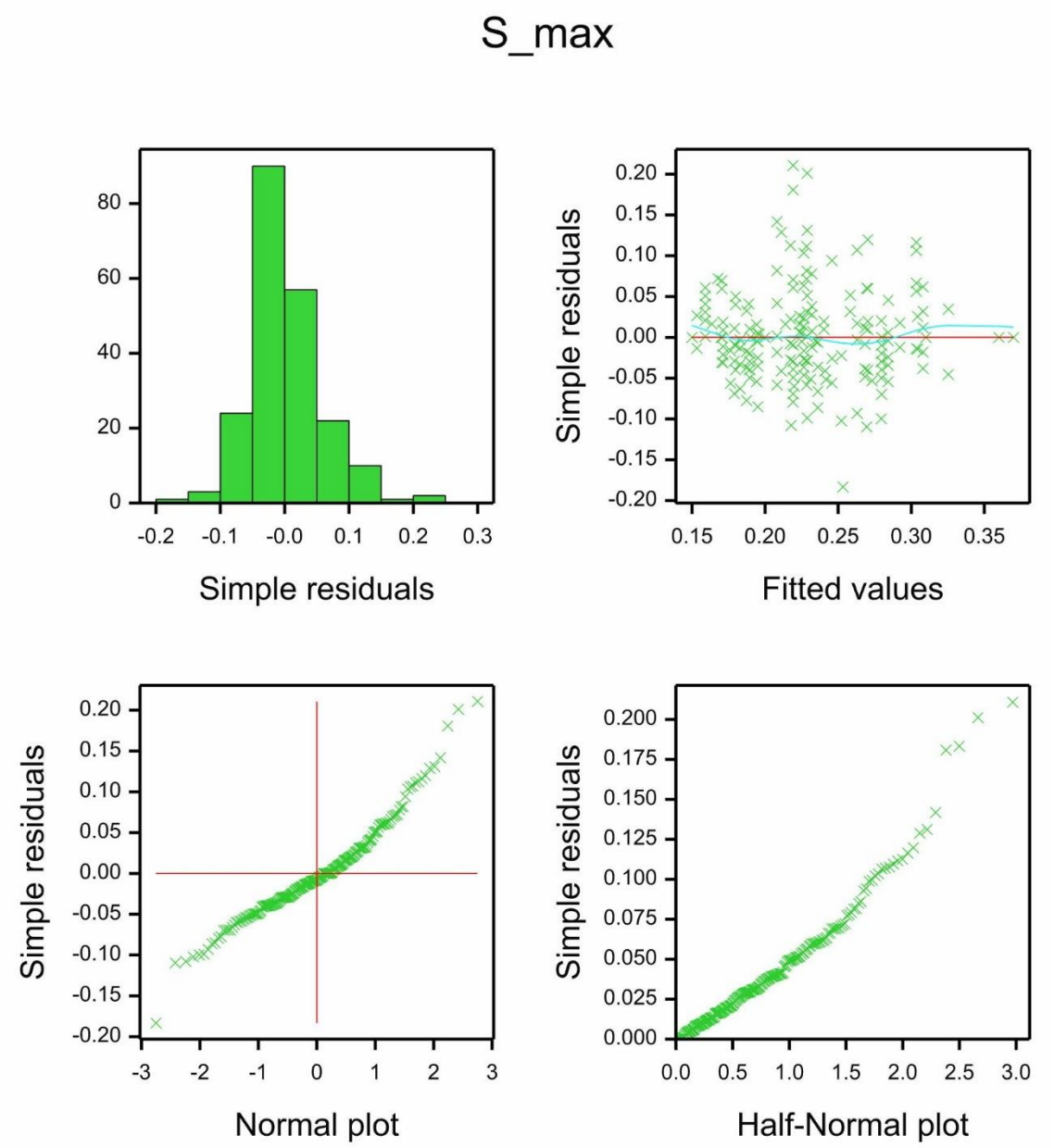
Resedual plots created for the ANOVA based on the CPI duration using Genstat software (VSN International, 2020)



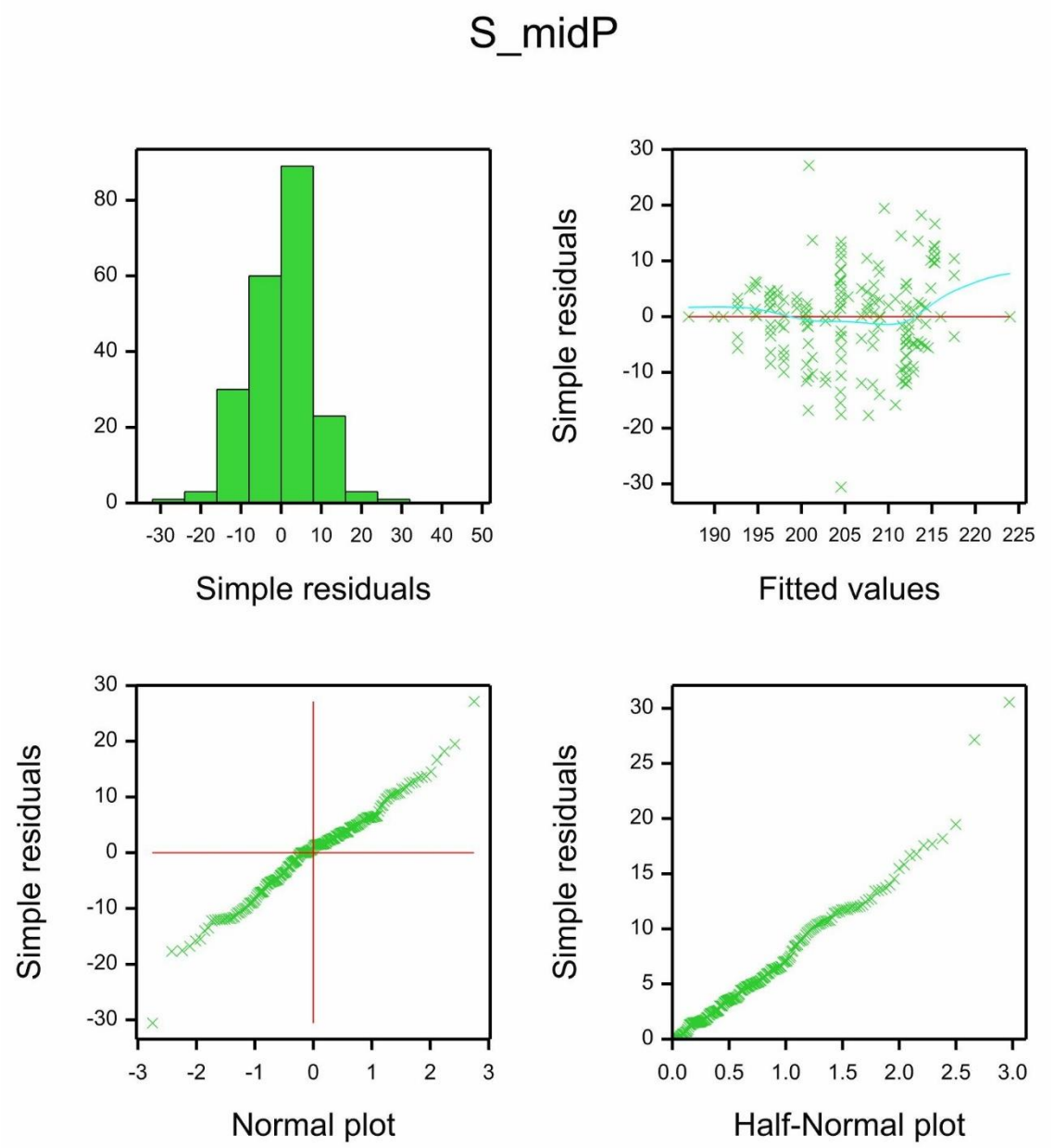
Resedual plots created for the ANOVA based on the CPI G_max using Genstat software (VSN International, 2020)



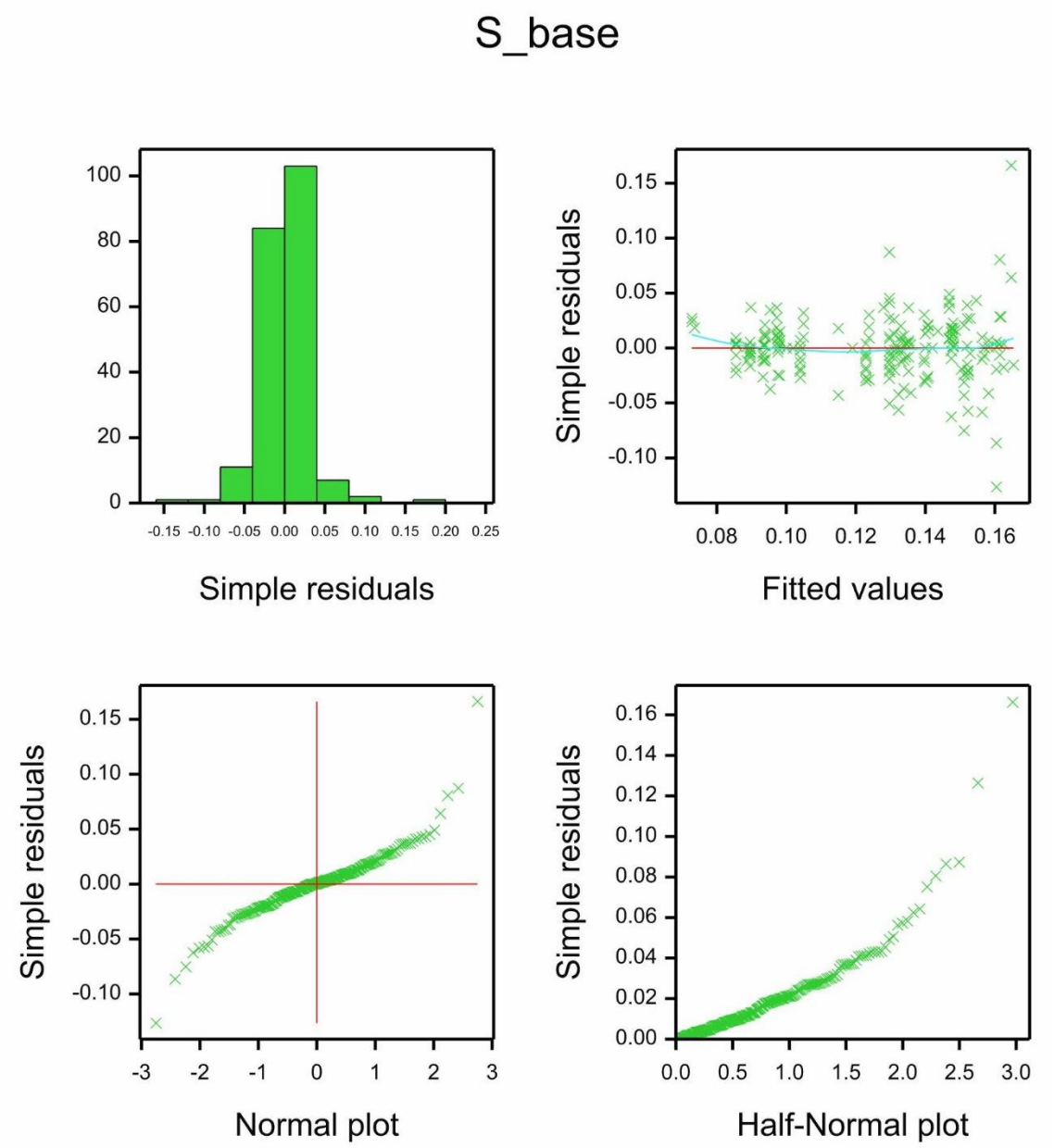
Resedual plots created for the ANOVA based on the CPI G_midP using Genstat software (VSN International, 2020)



Resedual plots created for the ANOVA based on the CPI S_max using Genstat software (VSN International, 2020)



Resedual plots created for the ANOVA based on the CPI S_midP using Genstat software (VSN International, 2020)



Resedual plots created for the ANOVA based on the CPI S_base using Genstat software (VSN International, 2020)

A.4 ANOVA Full table

ANOVA results, testing the influence of season, management (farm), residual nitrogen (Pr_crop_N) , soil type and interactions of soil season and nitrogen residual and soil on the important CPIs

Factors	G_midP	G_max	S_midP	S_max	S_base	Duration
Farm	<.001	<.001	0.002	0.002	0.21	<.001
Year	<.001	<.001	<.001	<.001	<.001	<.001
Pr_crop_N	<.001	0.016	0.09	0.59	0.605	0.007
Soil_type	0.001	0.506	0.173	0.093	0.382	0.011
Pr_crop_N*Soil_type	0.109	0.835	0.608	0.017	0.339	0.285
Year*soil_type	<.001	0.198	0.029	<.001	0.047	<.001

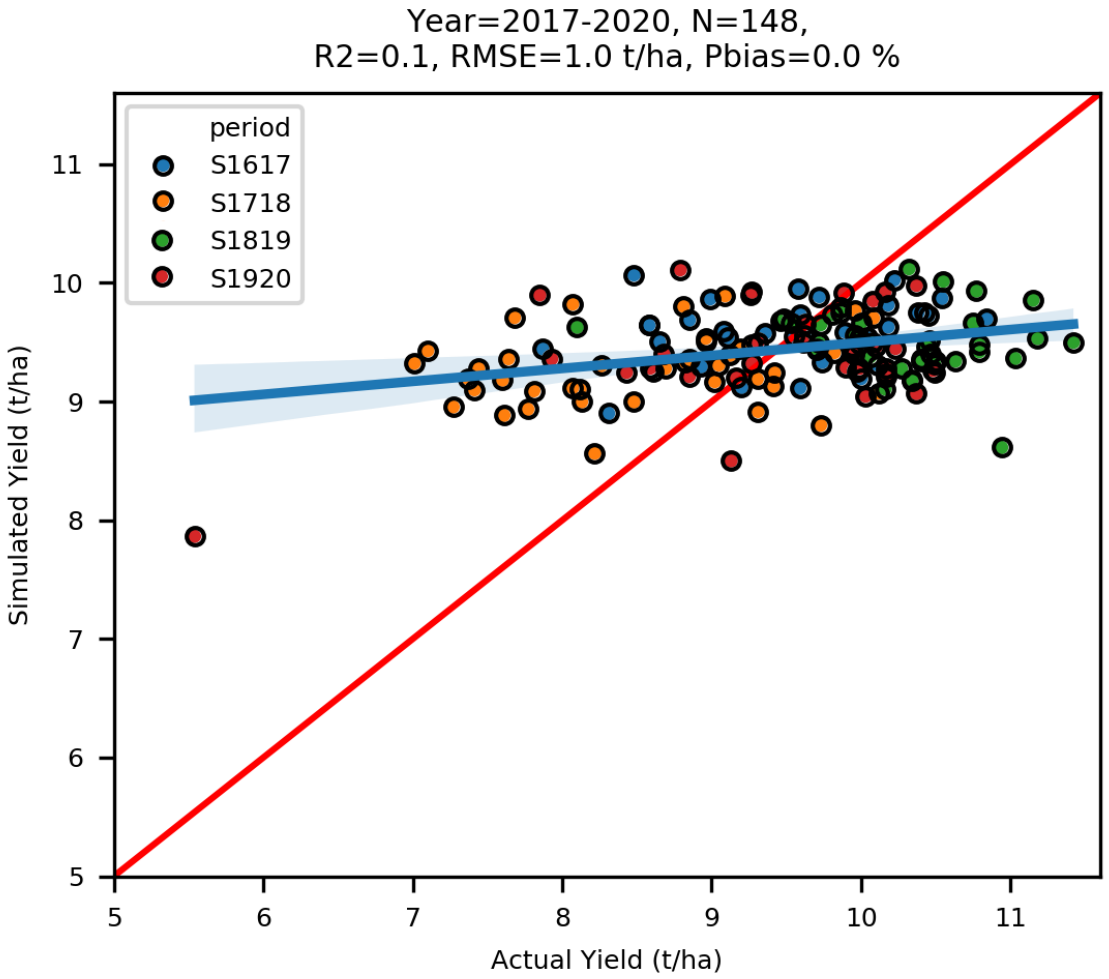
A.5 MLR values

multi-linear regression results for the three main models at four different levels of significance: 0.05<p<0.1(º), 0.01<p<0.05(*), 0.001<p<0.01 (**) and p<0.001 (***) for the Norfolk farms in years 2017-2020

	Factors	Coefficient		P>|t|
Model 1	Intercept	11.3651	±2.085	0.000
	Duration	0.002	±0.007	0.786
	TZmax	-0.0166	±0.011	0.137
Model 2	Intercept	12.4593	±2.527	0.000
	G_midP	-0.0365	±0.009	0.000
	G_max	3.3636	±1.089	0.002
	S_max	0.4919	±1.414	0.728
	S_midP	-0.0031	±0.01	0.76
Model 3	Intercept	8.8986	±14.859	0.55
	G_midP	-0.0158	±0.151	0.917
	G_max	6.5595	±3.663	0.075
	S_max	5.6322	±5.917	0.343
	S_midP	0.009	±0.07	0.898
	G_midP:S_midP	-0.0001	±0.001	0.882
	G_max:S_max	-13.7533	±15.065	0.363

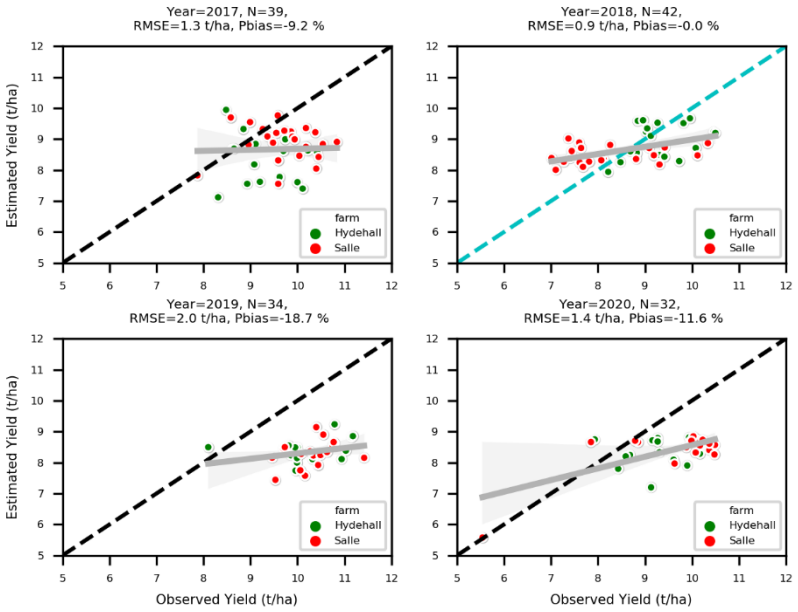
multi-linear regression reuslts for the three main models at four different levels of significance: 0.05<p<0.1(o), 0.01<p<0.05(∗), 0.001<p<0.01 (∗∗) and p<0.001 (∗∗∗) for the four farms during the 2017 growning season.

	Factors	Coefficient		P>|t|
Model 1	Intercept	12.9588	±3.111	0.000
	Duration	0.0163	±0.011	0.137
	TZmax	-0.0498	±0.017	0.004
Model 2	Intercept	4.8521	±3.262	0.142
	G_midP	-0.0671	±0.011	0.000
	G_max	10.1504	±2.103	0.000
	S_max	3.7086	±2.369	0.122
	S_midP	0.0312	±0.013	0.018
Model 3	Intercept	20.2283	±22.775	0.378
	G_midP	-0.2234	±0.252	0.379
	G_max	4.5247	±6.275	0.474
	S_max	-2.6608	±8.549	0.757
	S_midP	-0.0356	±0.108	0.744
	G_midP:S_midP	0.0007	±0.001	0.535
	G_max:S_max	28.4371	±31.381	0.368

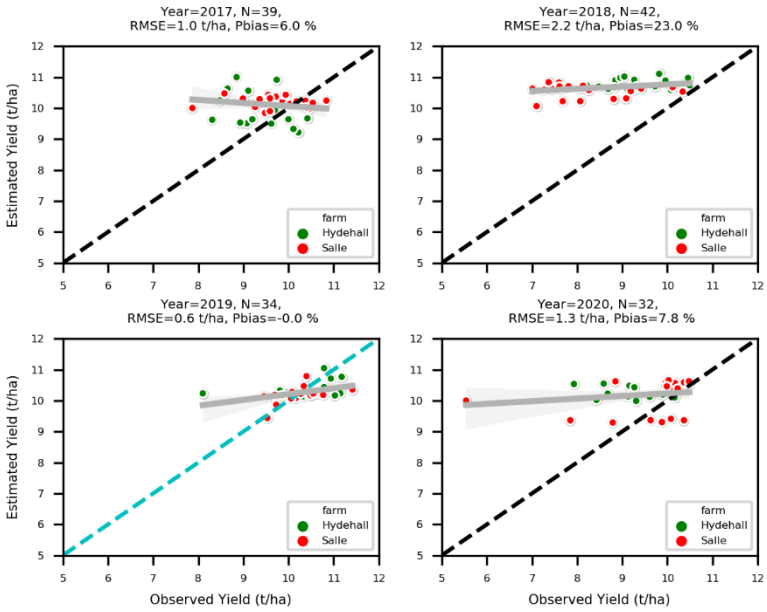


Actual and predicted yield comparison using the linear regression model best fitted for the four seasons in the two Norfolk farms, yield =f (Gmid,Smid,Gmax).

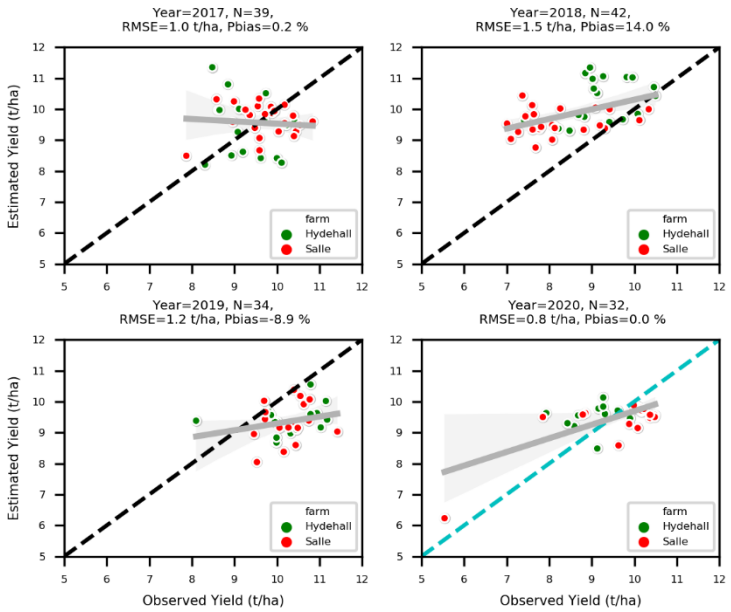
A.6 MLR one-year calibration



Comparison of predicted yields against actual yields for the Norfolk region using the 2018 growing season as calibration (cyan 1:1 line) for the multiple linear regression model and the rest of the years for comparison (black lines 1:1) as well as the regression lines for each year (grey lines). The 95% confidence limits are highlighted as faded regions. model yield=f(Gmid,Smid,Gmax,Smax)



Comparison of predicted yields against actual yields for the Norfolk region using the 2019 growing season as calibration (cyan 1:1 line) for the multiple linear regression model and the rest of the years for comparison (black lines 1:1) as well as the regression lines for each year (grey lines). The 95% confidence limits are highlighted as faded regions. model yield=f(Gmid,Smid,Gmax,Smax)



Comparison of predicted yields against actual yields for the Norfolk region using the 2020 growing season as calibration (cyan 1:1 line) for the multiple linear regression model and the rest of the years for comparison (black lines 1:1) as well as the regression lines for each year (grey lines). The 95% confidence limits are highlighted as faded regions. model yield=f(Gmid,Smid,Gmax,Smax)