Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Prediction of crop biophysical variables with panel data techniques and radar remote sensing imagery

Clara Simón de Blas [a,b,*], Rubén Valcarce-Diñeiro [c,**], Ana E. Sipols [d], Nilda Sánchez Martín [c,e], Benjamín Arias-Pérez [c], M. Teresa Santos-Martín [f]

[a] Department of Computer Sciences and Statistics, Rey Juan Carlos University, Madrid, Spain
[b] IUES, UCM, Madrid, Spain
[c] Department of Cartographic and Land Engineering, University of Salamanca, Ávila, Spain
[d] Department of Applied Mathematics, Materials Science and Engineering and Electronic Technology, Rey Juan Carlos University, Madrid, Spain
[e] CIALE, University of Salamanca, Salamanca, Spain
[f] Institute of Fundamental Physics and Mathematics, Department of Statistics, University of Salamanca, Salamanca, Spain

ARTICLE INFO

Article history:
Received 10 June 2020
Received in revised form
2 December 2020
Accepted 19 February 2021
Published online 9 March 2021

Keywords:
Biophysical variables
Panel data
PCA
Polarimetric SAR
RADARSAT-2

Since the late 1970s, remote sensing techniques have been proven to be suitable for characterizing and monitoring plants and crops. In particular, synthetic aperture radar (SAR) missions contribute considerably to this prediction effort. However, the main issue when using SAR image series together with field observations is the scarcity of data due to the difficulty of acquiring field measurements. This research aimed to contribute to solving this problem with an alternative statistical model that can overcome the lack of a long, robust series of field-based ground truth observations. The main novelty of this research is the evaluation of the potential of a panel data approach to radar remote sensing imagery for predicting crop biophysical variables. For this purpose, RADARSAT-2 imagery was acquired over the study area in central Spain. Simultaneously, a field campaign was deployed to estimate crop parameters in the same area and to validate the results of the modelling. The analysis of the influence of the crop type on the incidence angle and the polarimetric parameters showed a strong influence of the co-polar channels (HH, VV), the entropy (H) and the coherence between the co-polar channels (γ_{HHVV}), with the differences being higher at 25°. The panel data analysis method demonstrated that good predictions, with R^2 greater than 0.78, were achieved for all biophysical variables analysed in this study. Overall, this novel statistical approach with remote sensing data showed great applicability for the prediction of crop variables, even with a short series of observations.

* Corresponding author.
** Corresponding author.
E-mail addresses: clara.simon@urjc.es (C.S. de Blas), ruben.v.d@usal.es (R. Valcarce-Diñeiro), anaelizabeth.garcia@urjc.es (A.E. Sipols), nilda@usal.es (N. Sánchez Martín), benja@usal.es (B. Arias-Pérez), maysam@usal.es (M.T. Santos-Martín).
https://doi.org/10.1016/j.biosystemseng.2021.02.014

1. Introduction

The support that remote sensing techniques provide to environmental and agricultural monitoring is undeniable. In the projected climate-change scenarios, the remote, long-term observations available from satellites seem crucial for the correct diagnosis of a varied set of problems. Focusing on crop monitoring and yield prediction, the temporal variation in C-band synthetic aperture radar (SAR) measurements from remote observations has proven to provide biophysical information about growth condition indicators (biomass, height, etc.) and the underlying soil moisture (Mattia et al., 2003). The ultimate objective of those studies is either to estimate or to monitor the biophysical parameters of crops (biomass, height, phenological stage, yield, etc.). This information is particularly useful for farmers, who are interested in timely information about crop conditions (e.g., particular phenological stages for irrigation and fertilisation purposes and disease detection), water requirements (e.g., for irrigation only when and where necessary) and final crop productivity (Lopez-Sanchez & Ballester-Berman, 2009).

Among the available SAR on-board satellites, the RADARSAT-2 mission by the Canadian Space Agency (CSA) offers the advantage of multiple polarimetric modes and three beam incidence angles (Morena, James, & Beck, 2004), which are key to monitoring changes in vegetation structure (Jiao et al., 2014). Therefore, many experiments with RADARSAT-2 over different crops, such as wheat (Xu et al., 2014), maize (Xie et al., 2017), rice (Choudhury & Chakraborty, 2006; Wu, Wang, Zhang, Zhang, & Tang, 2011) or multiple crops (Wiseman, McNairn, Homayouni, & Shang, 2014), have collected RADARSAT (1 and 2) SAR data at different frequencies, polarisations and incidence angles to link these polarimetric parameters to agronomic field parameters (Lopez-Sanchez & Ballester-Berman, 2009). Many methods have been applied to monitor crops, from the simplest, such as linear regression with field measurements (Canisius et al., 2018; Liu, Shang, Vachon, & McNairn, 2012; McNairn, Ellis, Van Der Sanden, Hirose, & Brown, 2002; Moran et al., 2012; Wiseman et al., 2014) and trend or temporal analysis (Choudhury & Chakraborty, 2006; Jiao et al., 2014; Yang et al., 2015), to the more sophisticated physical models (Gherboudj, Magagi, Berg, & Toth, 2011; Jiao et al., 2011; Yonezawa et al., 2012). In order to estimate vegetation parameters for various crops with full, compact and dual-pol SAR data, researchers have shown it is possible using semi-empirical and physically based model estimations (Beriaux, Lucau-Danila, Auquiere, & Defourny, 2013; Bériaux, Waldner, Collienne, Bogaert, & Defourny, 2015; Chauhan, Srivastava, & Patel, 2018; Mandal, Hosseini et al., 2019; Mandal, Kumar, McNairn, Bhattacharya, & Rao, 2019). However, statistical modelling still remains barely explored, probably due to the difficulty of gathering enough measurements to build a robust model, as recognised in many examples of radar applications (Valcarce-Diñeiro, Lopez-Sanchez, Sánchez, Arias-Pérez, & Martínez-Fernández, 2018). In addition, from a statistical viewpoint, the remote sensing community may lack in-depth knowledge of the statistical tools, leading to an incorrect integration of the radar parameters or a misinterpretation of the results.

In a previous study (Valcarce-Diñeiro et al., 2018), the sensitivity of the radar parameters from RADARSAT-2 (i.e., the backscattering coefficient or any other variable derived from the radar measurements) for monitoring biophysical parameters that describe crop phenology was appraised by means of simple statistical correlations. As stated in Cable, Kovacs, Jiao, and Shang (2014), it is important that analysts understand how polarimetric SAR interacts with targets before attempting more complex analyses. Overall, it was highlighted that the biomass, leaf area and height of rainfed crops may be surveyed using different radar parameters provided by RADARSAT-2. This insight suggests the need for a deeper statistical analysis to further explore these relationships.

Different statistical techniques have been used to study the relationships between field and satellite data. First, a factor analysis was carried out to reduce the dimensionality of the study variables. Then, a factorial experiment analysed the possible interactions between the crop type and the incidence angle in the variables measured by direct observation in the field. Finally, the panel data methodology for time series was applied to study the possible relationships between field and satellite data, allowing the modelling of unobservable heterogeneity. Panel data techniques for short time series are usually applied in economics ((Arellano, 2003; Baltagi, 2008; Engel & Reinecke, 2012; Gourieroux & Monfort, 1993; Hsiao, 2007; Mátyás & Sevestre, 2013) and others) but have not previously been used in remote sensing applications. Typically, remote sensing observations should be validated through *in situ* observations (the so-called "ground truth") that are both time consuming and spatially scarce. Therefore, data availability and high-dimensionality relationships derived from cross-sectional units over time periods may be criticised because of the limitations of the datasets for comparison. In an attempt to overcome this problem, we suggest the use of panel data or longitudinal data that can provide robust comparisons even when applied to short datasets with more complicated clustering or hierarchical structures.

An econometric panel data model includes a sample of economic agents or agents of interest (individuals, companies, banks, cities, countries, etc.) for a given period of time, that is, it combines both types of data (dimension temporal and structural). The main objective of applying and studying panel data is to capture unobservable heterogeneity, either between economic agents or study agents as well as over time, since this heterogeneity cannot be detected either with time series studies or with those of cross section. The use of panel data in the agriculture area can be found in many contributions in the literature. Hu and McAleer (2005) estimated production efficiency in the agricultural sector in China with a panel dataset comprising 30 provinces for the 7-year period, 1991–1997. Mundlak, Butzer, and Larson (2008) presented an empirical analysis of a panel of countries to estimate agricultural production function using a measure of capital in agriculture absent in most studies. Kouser and Qaim (2011) studied the health impacts of farmers resulting from Bt-related changes in the use of chemical pesticides by means of panel data from India to estimate non-toxic effects. Finally, Adah, Kayode, and Victor (2017) used the panel data analysis to examine cereal productivity in West Africa.

The main objective of this paper is to obtain a good characterisation of agronomic parameters using SAR images through a panel-based method never applied before in this field. We hypothesise that the use of panel data techniques can improve the ability of radar remote sensing to accurately predict crop biophysical variables using limited ground-truth data.

The paper has been organised as follows: Section 2 introduces the study area and field measurements. Section 3 presents the satellite datasets and polarimetric parameters derived. Section 4 shows the estimation and forecast methodologies. The results are shown in Section 5, followed by the discussion in Section 6. Section 7 describes the main conclusions of this work and further future research.

2. Area and field measurements

A field campaign in which the relationships between active radar measurements (RADARSAT-2) and agricultural characteristics were explored was developed as part of the activities of a project devoted to the application of active satellite sensors to agriculture (Sánchez et al., 2015; Valcarce-Diñeiro et al., 2018). The study area (Fig. 1) was a subarea included in the soil moisture stations network of the University of Salamanca, a network intended to monitor soil moisture and to validate remote sensing products (Colliander et al., 2017; González-Zamora, Sánchez, Pablos, & Martínez-Fernández, 2019; Sánchez, Martínez-Fernández, Scaini, & Perez-Gutierrez, 2012). REMEDHUS is located in the central sector of the Duero basin in Spain (41.1°–41.5°N; 5.1° to 5.7°W) and is made up of 23 soil moisture stations and four weather stations providing continuous data since 1999. A more detailed description of the sensors and the set up may be found in Sánchez, Martínez-Fernández, Scaini, and Perez-Gutierrez (2012).

Seven plots of different size were selected (Fig. 1) in this agricultural area, namely, F11 (barley), H9 (natural pasture), J12, K10 and L7 (wheat), M9 (barley) and N9 (rye). These crops (although H9 is not actually considered a crop; the plot is used for cattle livestock) represent the most common land use in the Castilla y León region, i.e., rainfed cereals. This region is characterised by a continental Mediterranean climate, with scarce rainfall in the central, semiarid area where REMEDHUS is located (less than 400 mm on average). The main traditional crops are adapted to the climatic conditions. The growing cycle of the rainfed cereals and the pasture spans from seeding in October to harvesting at the end of June; due to the low temperatures during winter, the crops do not begin their development phase until the end of February. Therefore, nine field measurements were acquired every two weeks from February to July 2015. Out of the seven plots, five (F11, H9, J12, L7 and N9) were used for training the model and two (K10 and M9) for validating the results of it.

The selected crop parameters attempted to define the crop growth status over time. Direct ground measurements, such as plant height, sample collection and photographs, were randomly acquired in each plot directly in a frame of one m^2 over the crop. In the laboratory, the leaf area index (LAI), the fraction of vegetation cover (FVC), the fresh and dry weights, the vegetation water content (VWC) and the percentage of water content (PWC) were measured. The green LAI was estimated scanning and scaling the green samples, FVC was retrieved through a supervised classification of zenithal photographs, and VWC was estimated as the difference (or the relative difference for PWC) between wet and dry weights of the sample. Soil moisture (SM) was collected at each REMEDHUS station installed at a depth of 5 cm in each plot, at the same date and time of the rest of measurements. The stations are equipped with Hydra Probes Soil Sensors (StevensVR Water Monitoring System Inc., Portland, OR, USA). These parameters are commonly monitored in the literature aimed at

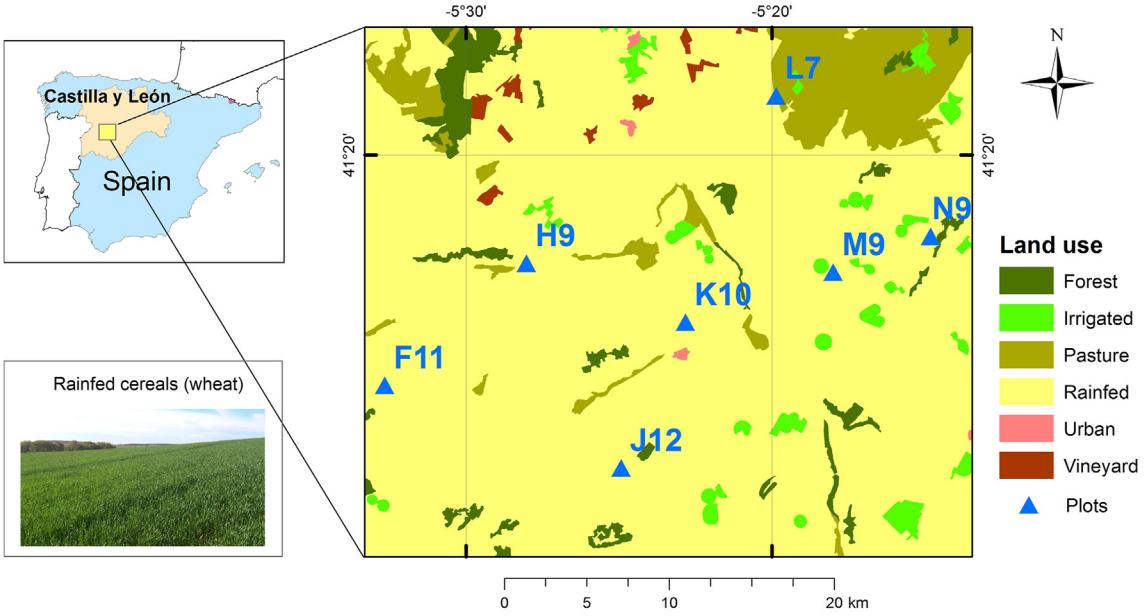

Fig. 1 – Location and selected plots on a land use map, including a photograph of the most common crop in the area, rainfed cereals.

crop monitoring (Jackson et al., 2004) and may be indicators of crop development, yield, water status and evapotranspiration (Lopez-Sanchez & Ballester-Berman, 2009). For a more detailed description of their retrieval, please see Sánchez et al. (2012) and Valcarce-Diñeiro et al. (2018), among others.

3. Imagery: RADARSAT-2 data and pre-processing

The RADARSAT-2 data were acquired over the study area between February and July 2015 and are provided in RADARSAT-2 Fine Quad-Pol Single–Look Complex (SLC) format. Three sets of 7, 7 and 6 RADARSAT-2 images were acquired in three different beam modes: FQ16W, FQ11W and FQ6W, respectively (Fig. 2). These data are delivered with nominal slant range and azimuth resolutions of 5.2 m and 7.6 m. To harmonise the dates of the field measurements with the imagery acquisitions, a spline interpolation of the ground measurements dates was performed, as shown in Valcarce-Diñeiro et al. (2018).

To preprocess RADARSAT-2 data, the Sentinel-1 Toolbox of the Sentinel Application Platform (SNAP) and PolSARpro, both of which were provided by the European Space Agency (ESA), were used. Figure 3 summarises the data-treatment workflow, from which ten polarimetric parameters were derived (Table 1). A more detailed description of the procedure and the parameters may be found in Valcarce-Diñeiro et al. (2018). The symbols used hereafter to denote these observables are shown in Table 1.

4. Methodology applied to the forecast and estimation model

The goal of the present study is to find potential relationships between satellite and field measurements in order to establish a dependence model between both datasets for forecasting purposes. From this model, field parameters of interest in agriculture that are difficult and expensive to monitor by direct observation could be estimated from image observations with a structural equation that expresses the dependence between them.

4.1. *Factor analysis: data reduction*

We use a factor analysis to reduce the dimensionality of the database from the ten radar parameters described in Table 1. These variables were taken from RADARSAT-2 between February 2015 and July 2015 through weekly observations.

This study aims to construct k factors (where k is a number to be determined between 1 and the number of original variables) using the correlation matrix method, which shows that there are significant correlations among all study variables, as presented in Table 2. The new factors will be expressed as linear combinations of the original variables, representing the common shared information between them. The new k-dimensional space will allow the detection of linear relationships between the original variables and, in some cases, will detect irrelevant or redundant variables. To improve the interpretability of the results, we rotated the axis of the space to make the location of the axes fit the actual data points better.

Bartlett's test (Bartlett, 1937) of sphericity, presented in Table 3, is used to determine whether the correlation matrix of the variables in study is an identity matrix, which would indicate that the variables are unrelated and therefore unsuitable for structure detection. Ratios between backscattering coefficients (HH/VV, HV/VV) were excluded from this test, as were combinations of existing variables in the database. A high value of this statistic equals lower statistical significance and indicates that it is unlikely that the correlation matrix is an identity matrix. Thus, the data are adequate for a factor analysis, as there is evidence of possible linear relationships among the variables being studied. The Kaiser-Meyer-Olkin (KMO) test is a statistic that indicates the proportion of variance for the variables in the study that might be caused by underlying factors. Values close to one indicate that the component analysis is adequate.

As such, we conclude that the principal components analysis has an adequate sample (Kaiser, 1974) and that the identity proximity of the correlation matrix approximation hypothesis is rejected. Table 4 presents the initial and final communalities (the proportion of each variable's variance that can be explained by the factors) and shows that the majority of the information is explained by the proposed factors, as the associated coefficients for all variables are close to one.

Table 5 presents the total variance explained by each factor. We observe that constructing four factors preserves 98% of the information in terms of the cumulative variance provided by the first 4 components.

Therefore, this analysis constructs four factors with the following parameters: HH, HV, VV, HH/VV, HV/VV, γ_{HHVV}, PPD, H, α_1 and γ_{P1P2}. To improve the scores of the contributions of the original variables in the factors, we selected an Equmax

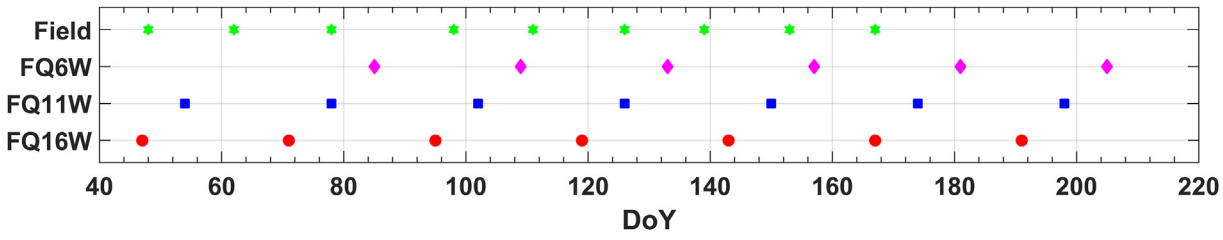

Fig. 2 – Acquisition dates and beam modes for the RADARSAT-2 dataset.

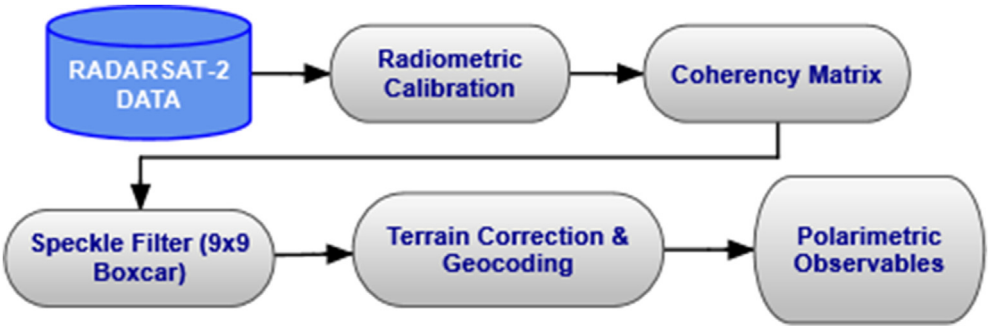

Fig. 3 – RADARSAT-2 processing workflow.

Table 1 – Polarimetric parameters derived from RADARSAT-2.

Polarimetric parameters	Symbol
Backscattering coefficient at HH, HV and VV channels	σ^0_{HH}, σ^0_{HV}, σ^0_{VV}, or simply HH, HV and VV
Ratio of backscattering coefficients at HH, HV and VV channels	HH/VV, HV/VV
Normalised correlation (coherence) between the copolar channels (HH and VV)	γ_{HHVV}
Polarisation phase difference between the copolar channels (HH and VV)	PPD
Entropy and dominant alpha angle (from the eigen decomposition of the coherency matrix)	H, α_1
Normalised correlation (coherence) between the 1st two channels in the Pauli basis (HH + VV and HH-VV)	γ_{P1P2}

Table 3 – Bartlett's sphericity test and KMO.

Statistic		Value
Sample adaptation measure of Kaiser-Meyer-Olkin.		0.391
Bartlett's sphericity test	Approximate Chi-squared	1996.740
	Degrees of freedom	28
	Significance	0.000

Table 4 – Initial and final communalities.

Variables	Initial	Extraction
HH (dB)	1.000	0.994
VV (dB)	1.000	0.992
HV (dB)	1.000	0.990
γ_{HHVV} (unitless)	1.000	0.961
PPD (degrees)	1.000	0.998
H (unitless)	1.000	0.990
α_1 (degrees)	1.000	0.960
γ_{P1P2} (unitless)	1.000	0.970
HH/VV (dB)	1.000	0.970
HV/VV (dB)	1.000	0.981

rotation. In Eq. (4.1), we present the constructed factors with the parameters associated with each variable according to the values presented in Table 6.

$F_1 = -0.06HH + 0.029VV + 0.354HV - 0.11HH/VV + 0.28HV/VV - 0.22\gamma_{HHVV} - 0.167PPD + 0.369H + 0.043\alpha_1 - 0.228\gamma_{P1P2}$

$F_2 = 0.18HH - 0.127VV - 0.144HV + 0.388HH/VV - 0.027HV/VV - 0.009\gamma_{HHVV} - 0.067PPD - 0.189H + 0.208\alpha_1 + 0.446\gamma_{P1P2}$ (4.1)

$F_3 = 0.489HH + 0.38VV + 0.3054HV + 0.046HH/VV - 0.029HV/VV + 0.061\gamma_{HHVV} - 0.059PPD - 0.073H - 0.015\ \alpha_1 + 0.072\gamma_{P1P2}$

$F_4 = -0.01HH + 0.021VV - 0.074HV - 0.089HH/VV - 0.081HV/VV - 0.016\gamma_{HHVV} + 1.006PPD + 0.038H + 0.032\alpha_1 + 0.063\gamma_{P1P2}$

Table 2 – Correlation matrix of variables.

		HH	VV	HV	HH/VV	HV/VV	γ_{HHVV}	PPD	H	α_1	γ_{P1P2}
Correlation	HH	1	0.708[a]	0.566[a]	0.190[b]	−0.056	0.079	0.132	−0.148	0.054	0.268[a]
	VV	0.708[a]	1	0.267[a]	−0.558[a]	−0.543[a]	0.570[a]	0.035	−0.454[a]	−0.582[a]	−0.437[a]
	HV	0.566[a]	0.267[a]	1	0.293[a]	0.664[a]	−0.567[a]	0.254[a]	0.676[a]	0.461[a]	0.227[a]
	HH/VV	0.190[b]	−0.558[a]	0.293[a]	1	0.689[a]	−0.700[a]	0.107	0.457[a]	0.873[a]	0.923[a]
	HV/VV	−0.056	-.0543[a]	0.664[a]	0.689[a]	1	−0.936[a]	0.194[b]	0.941[a]	0.854[a]	0.537[a]
	γ_{HHVV}	0.079	0.570[a]	−0.567[a]	−0.700[a]	−0.936[a]	1	−0.256[a]	−0.911[a]	−0.927[a]	−0.578[a]
	PPD	0.132	0.035	0.254[a]	0.107	0.194[b]	−0.256[a]	1	0.284[a]	0.253[a]	0.210[b]
	H	−0.148	−0.454[a]	0.676[a]	0.457[a]	0.941[a]	−0.911[a]	0.284[a]	1	0.723[a]	0.304[a]
	α_1	0.054	−0.582[a]	0.461[a]	0.873[a]	0.854[a]	−0.927[a]	0.253[a]	0.723[a]	1	0.817[a]
	γ_{P1P2}	0.268[a]	−0.437[a]	0.227[a]	0.923[a]	0.537[a]	−0.578[a]	0.210[b]	0.304[a]	0.817[a]	1

[a] Significant correlation for level 0.01 (bilateral).
[b] Significant correlation for level 0.05 (bilateral).

Table 5 – Total variance explained by each factor.

Component	Initial eigenvalues			Squared saturation aggregation of the extraction			Squared saturation aggregation of the rotation		
	Total	% Variance	% Cumulative	Total	% Variance	% Cumulative	Total	% Variance	% Cumulative
1	5.483	54.827	54.827	5.483	54.827	54.827	3.410	34.100	34.100
2	2.103	21.028	75.854	2.103	21.028	75.854	3.141	31.409	65.509
3	1.331	13.312	89.166	1.331	13.312	89.166	2.099	20.991	86.500
4	0.890	8.902	98.068	0.890	8.902	98.068	1.157	11.567	98.068
5	0.111	1.107	99.175						
6	0.058	0.582	99.757						
7	0.022	0.225	99.982						
8	0.002	0.018	100						
9	3.6E-14	3.6E-13	100						
10	2.8E-14	2.8E-13	100						

Table 6 – Factor parameters. Extraction method: Principal component analysis.

Variables	Component 1	Component 2	Component 3	Component 4
HH (dB)	−0.029	0.233	0.967	0.055
VV (dB)	−0.238	−0.477	0.841	0.022
HV (dB)	0.787	0.078	0.588	0.142
γ_{HHVV} (unitless)	0.297	0.938	−0.032	0.034
PPD (degrees)	0.871	0.438	−0.140	0.106
H (unitless)	−0.804	−0.493	0.192	−0.185
α_1 (degrees)	0.084	0.047	0.070	0.992
γ_{P1P2} (unitless)	0.940	0.172	−0.179	0.211
HH/VV (dB)	0.579	0.760	−0.119	0.184
HV/VV (dB)	0.117	0.964	0.052	0.156

4.2. *Factorial experiment: effects of factors*

To study the possible interactions between the type of crop and the incidence angle in the variables measured by direct observation in the field, a factorial experiment with one factor was considered in this research. This methodology allows us to confirm or reject the hypothesis that the average value, in all the levels determined by the given factor, is the same for the selected variable under study. Eq. (4.2) presents the mathematical model assumed for the studied variable (Y) as an explanatory response from the given factor:

$$Y = \mu + \alpha_i + \varepsilon_i \tag{4.2}$$

where μ represents the global average value for the model, α_i represents the average value of the factor at level i, with $i = 1,\ldots,a$ and a is the number of different levels considered for the factor variable, and ε_i is the error, assuming that $\varepsilon_i \sim N(0, \sigma^2)$, i.e., ε_i is independently and identically distributed with the normal distribution.

The average value, μ, is a score for all the given observations. When no differences are found from a given factor, we assume that all the observations will have the same value, which corresponds to the average value with a small random fluctuation. To study the effects of a given factor, it is better to consider a case in which all the considered groups composed of the different levels of the factor are homogeneous in variance.

The Levene test contrasts the null hypothesis of variances homogeneity against the alternative that at least one of them is significantly different from the others. If interactions were found between a factor and a selected variable in the study, the Duncan and Scheffe test were used to determine the different homogeneous groups. Both tests are multiple comparison procedures. Duncan's method uses a studentised range statistic to compare sets of means. Duncan's MRT is especially protective against false negative (Type II) error at the expense of having a greater risk of making false positive (Type I) errors, whereas Scheffé's method adjust significance levels in a linear regression analysis to account for multiple comparisons. The thresholds considered in Duncan's test are lower than in Scheffé's test and, therefore, it is easier to find differences between the compared means.

The result of the test is a set of subsets of means, where in each subset means have been found not to be significantly different from one another.

4.3. *Panel data: model estimation and forecast*

A useful method for studying the possible relationships between field and satellite measures over a short period of time is the panel data approach. This econometric model allows the combination of structural and time dimensions in an estimation model for forecast purposes. The main objective of panel data is to capture unobservable heterogeneity, either between the agents under study or in time, due to the impossibility of applying time series analysis or cross-sectional techniques. Eqs. (4.3) and (4.4) present the general mathematical model used for the studied variable as a dynamic explanatory response in time when the lagged dependent variable violates strict exogeneity, that is, when

endogeneity may occur and we consider a random effects model:

$$Y_{it} = \gamma Y_{i,t-1} + \beta X_{it} + u_{it} \quad (4.3)$$

$$u_{it} = \theta_i + \upsilon_{it} \quad (4.4)$$

where γ is the parameter associated with the lag of the dependent variable, β is the parameter associated to the explicative variables (X), θ_i are time-invariant effects that are fixed over time, t represents the time period (t = 1, 0.9) and υ_{it} is a time-varying random component. Endogeneity is said to exist when there is a correlation between the parameter or variable and the error term. Endogeneity can arise as a result of measurement error, autoregression with autocorrelation of errors, simultaneity and omitted variables. In general terms, a causal link between the independent and the dependent variables of a model leads to endogeneity.

In case u_t is believed to be not correlated with any of the independent variables, we consider a fixed-effects model as given by Eq. (4.5).

$$Y_{it} = \theta_{it} + \beta X_{it} + u_t \quad (4.5)$$

The selection of the model was derived from the Hausman test, and when possible, we considered a unique general model for the crops evaluated in the present study. The Hausman test can be used to differentiate between fixed effects model (null hypothesis) and random model in the panel analysis (alternative) testing whether the unique errors (u_t) are correlated with the regressors.

4.4. *Validation*

Wheat (K10) and barley (M9) were used to validate the model fit accuracy, and compute general measures such as the root-mean-square error (RMSE) and the mean absolute error (MAE) for validation. When the estimation model follows a random model, the predicted values are calculated by substituting the values of the explanatory variables multiplied by the estimated parameters. For example, the SM for wheat (K10) would be estimated through H and HV parameters as detailed in Eq. (4.6):

$$SM_{K10,\,t} = \beta_0 + \beta_1{*}HH + \beta_2{*}VV + \beta 3{*}HV + \beta_4{*}\gamma_{HHVV} + \theta_i + \gamma_t + \varepsilon_{it} \quad (4.6)$$

When the estimation model follows a field fixed effect, the predicted values are calculated by substituting the values of the explanatory variables multiplied by the estimated parameters adding the average field fixed effect, for instance, the FVC for wheat (K10) would be estimated as expressed in Eq. (4.7):

$$FVC_{K10,t} = \beta_1{*}F_1 + \beta_2{*}F_3 + \bar{x}_{crops} + \varepsilon_t \quad (4.7)$$

Finally, if the estimation model follows a timed and field fixed effect, the predicted values are calculated by substituting the values of the explanatory variables multiplied by the estimated parameters adding the average field fixed effect and the corresponding time effect, for instance, the height for wheat (K10) would be estimated as expressed in Eqs. (4.8) and (4.9):

$$Height_{K10,2} = \beta_1{*}\gamma_{P1P2(K10,\,2)} + \bar{x}_{crops} + \gamma_{K10,1} + \varepsilon_{K10,\,2} \quad (4.8)$$

$$Height_{K10,3} = \beta_1{*}\gamma_{P1P2(K10,\,3)} + \bar{x}_{crops} + \gamma_{K10,2} + \varepsilon_{K10,\,3} \quad (4.9)$$

Parameters and estimation models were estimated by means of SAS Enterprise Guide 9.4.

5. Results

5.1. *Factorial experiment*

The possible effects of the crop type and of the field measures were analysed in this study (see Fig. 4). Due to the large number of results of possible combinations of effects in the variables observed, only those results in which the associated Anova table reached significant effects were presented. The associated Anova table showed no significant differences in LAI (p-value 0.562), biomass (p-value 0.332) or PWC (p-value 0.819) among the different crops. No significant differences in height were found (p-value 0.234) among the crops, although rye presented higher values than barley. Furthermore, the Levene test rejected the null hypothesis of variance homogeneity (p-value 0.005), and the height value was larger for rye. In the case of FVC (p-value 0.04), the Scheffe test rejected the existence of differences among the crops, although the Duncan test found larger average values for pasture. Significant differences were found for soil moisture (p-value $\approx$0), with a higher average for pasture and a lower average for barley (F11) and wheat (K10) crops (see Table A.1 for the soil moisture statistics by crop).

When considering the possible effects of the incidence angle, the crop type and the polarimetric parameters, the results show the following (Tables A.2 and A.3):

1. No significant differences were found between the incidence angle and HV (p-value$_{angle}$ 0.656), HH/VV (p-value$_{angle}$ 0.390), HV/VV (p-value = 0.057), PPD (p-value = 0.692), α_1 (p-value = 0.077) or γ_{P1P2} (p-value = 0.833).
2. Significant differences were found in the HV (p-value$_{crop}$ $\approx$0) and HH/VV (p-value$_{crop}$ = 0.003) for the different crops, and HV and HH/VV were larger for wheat (L7) than for barley (F11). Significant differences were found in the HV/VV (p-value$_{crop}$ $\approx$0) for the different crops, and HV/VV was larger for natural pasture (H9) than for barley (F11). Significant differences were found in the PPD (p-value$_{crop}$ $\approx$0) values for the different crops, and PPD was larger for natural pasture (H9) than for wheat (K10). Significant differences were found in incidence angles (p-value$_{crop}$ = 0.017) for the different crops, though the Scheffe test rejected the existence of differences between crops. The Duncan test found larger average values of the incidence angle for the natural pasture (H9). Significant differences were found in γ_{P1P2} (p-value$_{crop}$ $\approx$0) for the different crops, and γ_{P1P2} was larger for wheat (J12) than for barley (F11).

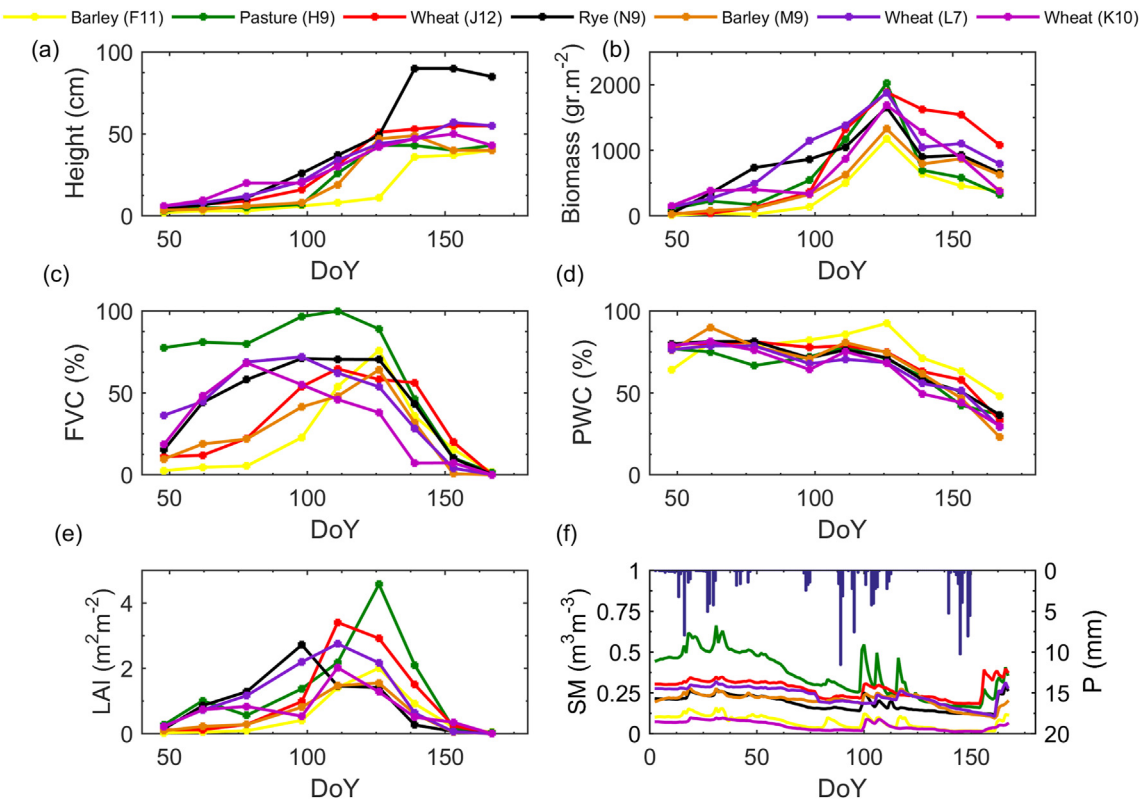

Fig. 4 – Field measurements used in the panel data for the sampled crops. (a) height; (b) biomass; (c) fraction of vegetation cover (FVC); (d) percentage of water content (PWC); (e) leaf area index (LAI); (f) daily precipitation (P) and soil moisture (SM) for the 7 agricultural plots. DoY (Day of Year).

3. Significant differences were found in HH (p-value$_{angle}$ = 0.001; p-value$_{crop}$ ≈0), which was higher at the 25° incidence angle than at the 36° incidence angle and was larger for wheat (L7) than for pasture (H9). For VV, significant differences were found (p-value$_{angle}$ = 0.001; p-value$_{crop}$ ≈0); VV was higher at the 25° incidence angle than at the 36° incidence angle and was larger for wheat (L7) than for wheat (J12). For γ_{HHVV}, significant differences were found (p-value$_{angle}$ = 0.001; p-value$_{crop}$ ≈0); γ_{HHVV} was larger at the 25° incidence angle than at the 36° incidence angle and was larger for barley (F11) than for the natural pasture (H9). For the entropy (H), significant differences were found (p-value$_{angle}$ = 0.003; p-value$_{crop}$ ≈0); H was larger at the 36° incidence angle than at the 25° incidence angle and was larger for natural pasture (H9) than for barley (F11).

When considering the possible effect of the incidence angle, the type of crop and the polarimetric parameters grouped by factors, the results show the following (Tables A.4 and A.5):

1. No significant differences were found between the incidence angle, the second factor (p-value = 0.761) and the fourth factor (p-value = 0.738) previously constructed.
2. Significant differences were found for the second factor (p-value$_{crop}$ ≈0) and the type of crop was larger for wheat (J12) than for natural pasture (H9). Significant differences were found for the fourth factor (p-value$_{crop}$ = 0.003), and the type of crop was larger for natural pasture (H9) than for wheat (K10).
3. Significant differences were found for the first factor constructed (p-value$_{angle}$ = 0.025; p-value$_{crop}$ ≈0), being larger for 36° incidence angle with respect to 25° incidence angle and larger for natural pasture (H9) with respect to barley (F11). The results are coherent, as the first factor is mainly due to the contribution of HV, HV/VV, γ_{HHVV} and H. For the third factor constructed, significant differences were found (p-value$_{angle}$ ≈0; p-value$_{crop}$ ≈0) being larger for the 25° incidence angle with respect to the 36° incidence angle and larger for wheat (L7) with respect to natural pasture (H9). The results are coherent, as the third factor is mainly due to the contribution of HH and VV backscattering coefficients.

5.2. *Panel data*

Several forecast and estimation models were developed using panel data to estimate field measurements using the satellite measurements and the four constructed factors. For each crop parameter, the estimation model for the original panel data is presented. Figure 5 shows the model results as a function of satellite parameters, and Fig. 6 shows the model results as a function of the factors derived from satellite parameters. A

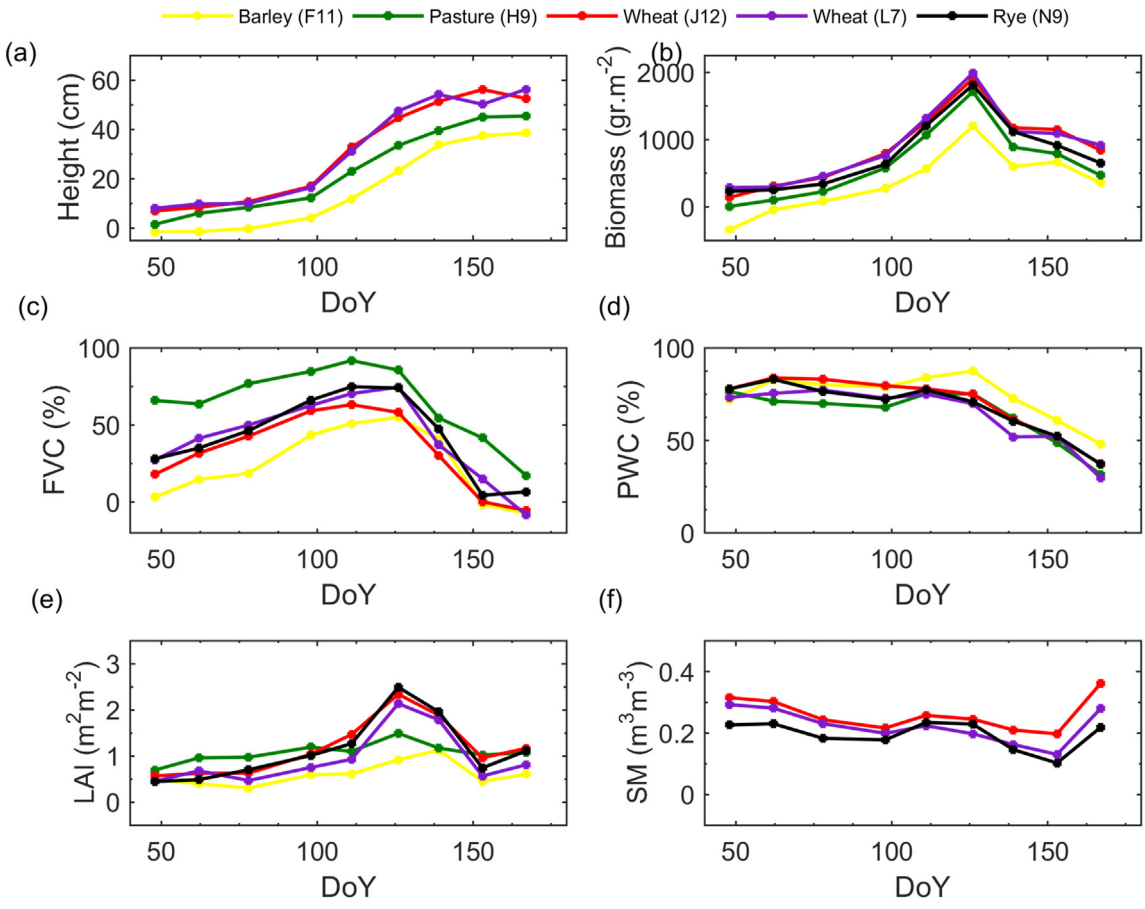

Fig. 5 – Estimation models for the original panel data for biophysical variables as a function of satellite parameters. (a) height; (b) biomass; (c) fraction of vegetation cover (FVC); (d) percentage of water content (PWC); (e) leaf area index (LAI); (f) daily precipitation (P) and soil moisture (SM) for the 7 agricultural plots. DoY (Day of Year).

general model was considered when possible, although in some cases significant differences between the crops in the field measurements lead to specific models. Later, the final model equations without nonsignificant parameters are described.

5.2.1. *Height*

Rye (N9) showed different behaviour from the other crops in the panel data, as also shown in Fig. 4. Therefore, this crop was left out of the general estimation model. To avoid collinearity, the following variables were eliminated from the analysis: HH, VV, HV, HH/VV, HV/VV, PPD, γ_{HHVV} and α_1. The Hausman test assumed the null hypothesis (p-value = 0.128). Therefore, a model with two fixed effects was used. Eq. (5.1) represents the estimation model for the height of all crops except rye. The estimated parameters θ_i and γ_t for the individual-specific, time-invariant effects are given in Table A.6. The associated R^2 value for the model in Eq. (5.1) was 0.95.

$$Height_{it} = 33.77*\gamma_{P1P2(it)} + \theta_i + \gamma_t + \varepsilon_{it} \quad (5.1)$$

The residuals of the model proposed in Eq. (5.1) followed a normal distribution (p-value = 0.837), validating the adequacy of the proposed structural modelling equation. When considering the four constructed factors, the Hausman test assumed the null hypothesis (p-value = 0.999); therefore, a model with fixed effects was used. Eq. (5.2) represents the estimation model for height. The estimated parameters θ_i and γ_t for the individual-specific, time-invariant effects are given in Table A.7. The associated R^2 value for the model in Eq. (5.2) was 0.95.

$$Height_{it} = 4.05*F_2 + \theta_i + \gamma_t + \varepsilon_{it} \quad (5.2)$$

The residuals of the model proposed in Eq. (5.2) followed a normal distribution (p-value = 0.96), validating the adequacy of the proposed structural modelling equation.

5.2.2. *FVC*

No significant differences were found in FVC among the crops (see Fig. 4); therefore, a general estimation model was created. To avoid collinearity, the following variables were eliminated from the analysis: HH/VV, HV/VV, γ_{HHVV} and α_1. The Hausman test supported the null hypothesis (p-value = 0.829); therefore, a model with fixed effects was generated. Eq. (5.3) represents the estimation model for FVC considering barley (F11), natural pasture (H9), wheat (J12 and L7), and rye (N9) crops. The estimated parameters θ_i and γ_t for the individual-specific, time-invariant effects are given in Table A.6. The associated R^2 value for the model in Eq. (5.3) was 0.82.

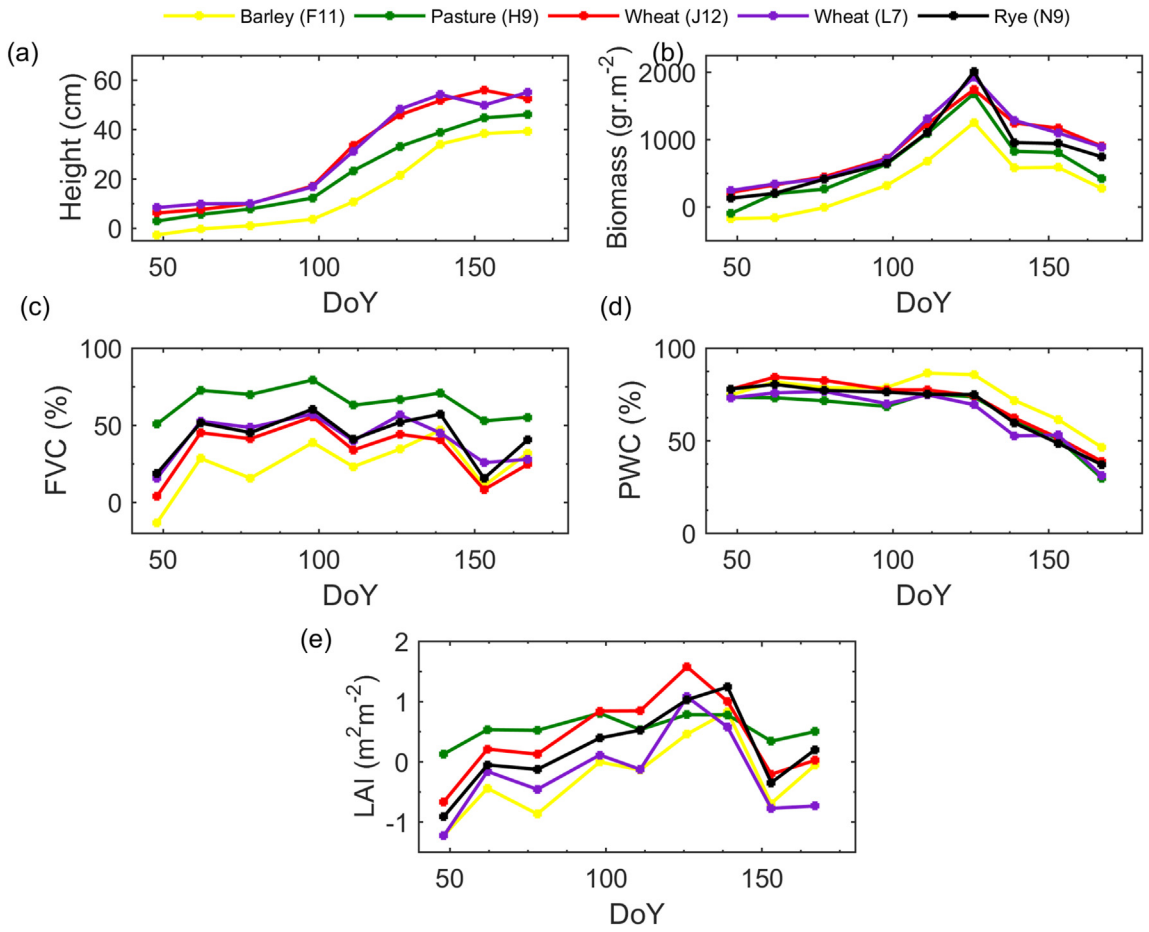

Fig. 6 – Estimation models for the original panel data for biophysical variables as a function of factors derived from a lineal combination of satellite parameters. (a) height; (b) biomass; (c) fraction of vegetation cover (FVC); (d) percentage of water content (PWC); (e) leaf area index (LAI). DoY (Day of Year).

$$FVC_{it} = -6.74*HV + 177.02*H + \theta_i + \gamma_t + \varepsilon_{it} \quad (5.3)$$

The residuals of the model proposed in Eq. (5.3) followed a normal distribution (p-value = 0.95), validating the adequacy of the proposed structural modelling equation. When considering the four constructed factors, the Hausman test supported the null hypothesis (p-value = 0.46); therefore, a model with fixed effects was used. Eq. (5.4) represents the estimation model for FVC. The estimated parameters θ_i for the individual-specific effects are given in Table A.7. The associated R^2 value for the model in Eq. (5.4) was 0.429.

$$FVC_{it} = 7.28*F_1 - 14.9*F_3 + \theta_i + \varepsilon_{it} \quad (5.4)$$

The residuals of the model proposed in Eq. (5.4) followed a normal distribution (p-value 0.637), validating the adequacy of the proposed structural modelling equation.

5.2.3. *LAI*

No significant differences in LAI were found among crops (see Fig. 4); therefore, a general estimation model was used. The Hausman test assumed the null hypothesis (p-value = 0.94); therefore, a model with fixed effects was used. Eq. (5.5) represents the estimation model for LAI considering barley (F11), natural pasture (H9), wheat (J12 and L7), and rye (N9) crops. The estimated parameters θ_i and γ_t for the individual-specific, time-invariant effects are given in Table A.6. The associated R^2 value for the model in Eq. (5.5) was 0.78.

$$LAI_{it} = -0.06*\alpha_1 + \theta_i + \gamma_t + \varepsilon_{it} \quad (5.5)$$

The residuals of the model proposed in Eq. (5.5) followed a normal distribution (p-value = 0.214), validating the adequacy of the proposed structural modelling equation. When considering the four constructed factors, the Hausman test rejected the null hypothesis (p-value <0.0001); therefore, a model with random effects was created. Eq. (5.6) represents the estimation model for LAI. The associated R^2 value for the model presented in Eq. (5.6) was 0.29, which was derived from a low-quality data suitability for this model.

$$LAI_{it} = 0.47*F_1 + 0.37*F_2 - 0.31*F_3 + \varepsilon_{it} \quad (5.6)$$

The residuals of the model proposed in Eq. (5.6) followed a normal distribution (p-value = 0.757), validating the adequacy of the proposed structural modelling equation.

5.2.4. Biomass

No significant differences in biomass were found among the crops (see Fig. 4); therefore, a general estimation model was used. The Hausman test assumed the null hypothesis (p-value = 0.879); therefore, a model with fixed effects was created. Eq. (5.7) shows the estimation model for biomass considering barley (F11), natural pasture (H9), wheat (J12 and L7), and rye (N9) crops. The estimated parameters θ_i and γ_t for the individual-specific, time-invariant effects are given in Table A.6. The associated R^2 value for the model in Eq. (5.7) is 0.87.

$$Biomass_{it} = -65.35*HV + \theta_i + \gamma_t + \varepsilon_{it} \quad (5.7)$$

The residuals of the model proposed in Eq. (5.7) followed a normal distribution (p-value = 0.691), validating the adequacy of the proposed structural modelling equation. When considering the four constructed factors, the Hausman test assumed the null hypothesis (p-value = 0.96); therefore, a model with fixed effects was used. Eq. (5.8) shows the estimation model for biomass. The estimated parameters θ_i and γ_t for the individual-specific, time-invariant effects are given in Table A.7. The associated R^2 value for the model presented in Eq. (5.8) was 0.88.

$$Biomass_{it} = 103.34*F_4 + \theta_i + \gamma_t + \varepsilon_{it} \quad (5.8)$$

The residuals of the model proposed in Eq. (5.8) followed a normal distribution (p-value = 0.657), validating the adequacy of the proposed structural modelling equation.

5.2.5. PWC

No significant differences were found between crops for PWC (see Fig. 4); therefore, a general estimation model was considered. The Hausman test supported the null hypothesis (p-value = 0.88); therefore, a model with fixed effects was used. Eq. (5.9) shows the estimation model for PWC considering barley (F11), natural pasture (H9), wheat (J12 and L7), and rye (N9) crops. The estimated parameters θ_i and γ_t for the individual-specific, time-invariant effects are given in Table A.6. The associated R^2 value for the model in Eq. (5.9) was 0.95.

$$PWC_{it} = -2.82*VV - 0.62*\alpha_1 - 37.93*\gamma_{P1P2} + \theta_i + \gamma_t + \varepsilon_{it} \quad (5.9)$$

The residuals of the model proposed in Eq. (5.9) followed a normal distribution (p-value = 0.985), validating the adequacy of the proposed structural modelling equation. When considering the four constructed factors, the Hausman test supported the null hypothesis (p-value = 0.99); therefore, a model with fixed effects was used. Eq. (5.10) represents the estimation model for PWC. The estimated parameters θ_i and γ_t for the individual-specific, time-invariant effects are given in Table A.7. The associated R^2 value for the model in Eq. (5.10) was 0.94.

$$PWC_{it} = -4.52*F_2 - 5.28*F_3 + \theta_i + \gamma_t + \varepsilon_{it} \quad (5.10)$$

The residuals of the model proposed in Eq. (5.10) followed a normal distribution (p-value = 0.918), validating the adequacy of the proposed structural modelling equation.

5.2.6. SM

Barley (F11) and pasture (H9) showed different behaviour from the other crops in the panel data, as also shown in Fig. 4. Therefore, these crops were left out of the general estimation model. The Hausman test rejected the null hypothesis (p-value = 0.999); therefore, a model with fixed effects was used. Eq. (5.11) shows the estimation model for SM considering wheat (J12 and L7) and rye (N9) crops. The estimated parameters θ_i and γ_t for the individual-specific, time-invariant effects are given in Table A.6. The associated R^2 value for the model in Eq. (5.11) is 0.96.

$$SM_{it} = 0.046*HH + 0.01*VV - 0.035HV - 0.34*\gamma_{HHVV} + \theta_i + \gamma_t + \varepsilon_{it} \quad (5.11)$$

The residuals of the model proposed in Eq. (5.11) followed a normal distribution (p-value = 0.935), validating the adequacy of the proposed structural modelling equation. When considering the four constructed factors, the Hausman test rejected the null hypothesis (p-value = 0.898); therefore, a model with fixed effects was used. Eq. (5.12) shows the estimation model for SM. The estimated parameters θ_i and γ_t for the individual-specific, time-invariant effects are given in Table A.7. The associated R^2 value for the model in Eq. (5.12) was 0.96.

$$SM_{it} = 0.019*F_1 + 0.018*F_2 + \theta_i + \gamma_t + \varepsilon_{it} \quad (5.12)$$

The residuals of the model proposed in Eq. (5.12) followed a normal distribution (p-value = 0.874), validating the adequacy of the proposed structural modelling equation.

5.3. Validation

We used Eqs. (5.1) and (5.12) to predict direct ground measurements for each crop. Table A.8 shows the RMSE and MAE between the predicted and the observed values, whereas Figs. 7 and 8 present the measured and predicted curves. Comparing both procedures, i.e., models considering the constructed factors and models using direct satellite measures, the results were balanced in terms of errors, since the first performed better for wheat (K10) in case of height, biomass and PWC, whereas the latter performs better for barley (M9) for height, FVC and LAI. Note that the model fails when the constructed factors are considered as explanatory variables for SM since they do not contribute significantly. Table A.6 presents the estimated parameter associated to the contribution in the model for each variable and for two fixed models the time period (denoted by TSi with i∈{1,…,8}) and the average crop type contribution in the panel sample. Therefore, there were no remarkable differences in both methodologies, but the single use of direct measurements has the advantage, apart from its greater simplicity, that it allowed the SM calculation.

Figures 7 and 8 presents the prediction models for the original panel data for field parameters as a function of satellite parameters and as a function of factors derived from a lineal combination of satellite parameters. Graphically the prediction seemed similarly good for both strategies, as well as for both crops, excepting for FVC and LAI. It seems

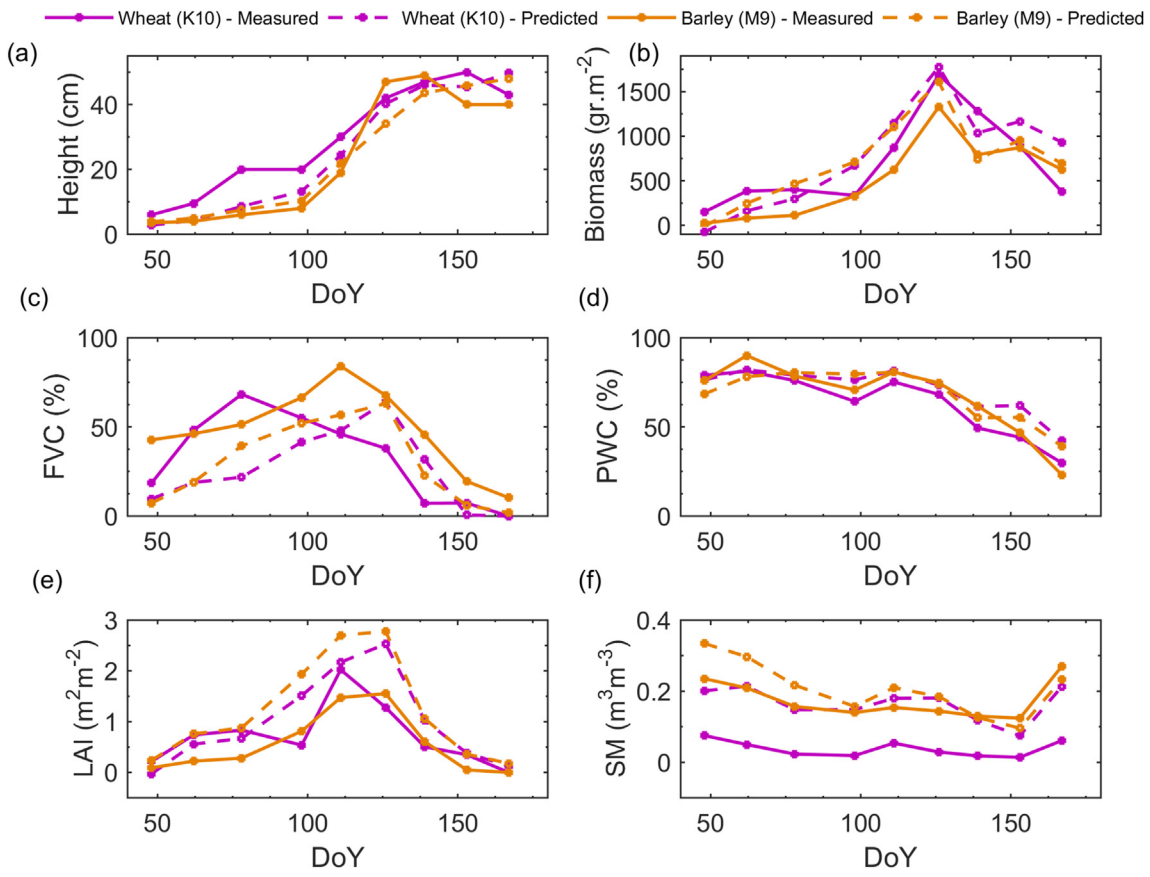

Fig. 7 – Prediction models for the original panel data for field parameters as a function of satellite parameters. (a) height; (b) biomass; (c) fraction of vegetation cover (FVC); (d) percentage of water content (PWC); (e) leaf area index (LAI); (f) daily precipitation (P) and soil moisture (SM) for the 7 agricultural plots. DoY (Day of Year).

reasonable, since those parameters are physically related, and none of them was able to be estimated. The reason for this bad behaviour may be explained after Fig. 4, where it was observed that both were the most changing parameters among the different crops. Since the prediction is based in the average of the selected crops in the panel data, some measures may be very sensitive to the type of plant as can be seen in table A.6 for LAI with an average contribution of 0.12 for Barley (F11) whereas wheat (J12) contributes with an average of 1.16, owing their different behaviour, the average did not satisfy neither the wheat prediction (K10) nor the barley one (M9).

On the contrary, SM and PWC were the best predicted, both for wheat and barley. Interestingly, those parameters, related with the water content in soil and plants, respectively, were the worst correlated with the satellite observations in the previous work of Valcarce-Diñeiro et al. (2018) using the same dataset, but a simple regression fit. It seems that the inclusion of the temporal dimension in the model improved the estimation of the variables while capturing the dynamic changes that occur in the variables under study. Furthermore, it can be seen for instance for height (Fig. 7), that the curve of the model continues growing and it is unable to capture the end of the growing cycle at the harvesting time. Full cycle datasets are needed to improve this models or their predictions will derive in infinite plants growth.

There is not a clear pattern of under- or overestimation among the parameters. Height was underestimated using both techniques for both types of crops, but the average gap between real and predicted values was below 28%. The greatest gaps between the predicted and real values were found for FVC, being more accused for the model considering satellite factors for barley (M9). LAI was overestimated when considering satellite measures for both crop types whereas it was underestimated when considering factor satellite measures. Furthermore, the models considering factor satellite measures even predicted negative values for LAI for both crops types. Biomass, PWC and SM were overestimated when considering satellite and factor satellite measures for both crop types (only satellite for the SM). The smallest average gap between real and predicted values was found for PWC with values below the 16.5%.

6. Discussion

Prior to performing the panel data analysis, a factorial experiment with one factor was considered in order to provide a better understanding of the interactions between crop types, incidence angles and biophysical variables.

The crop type only affected the biophysical variables in three cases (Fig. 4): the height of the rye, the FVC for pasture

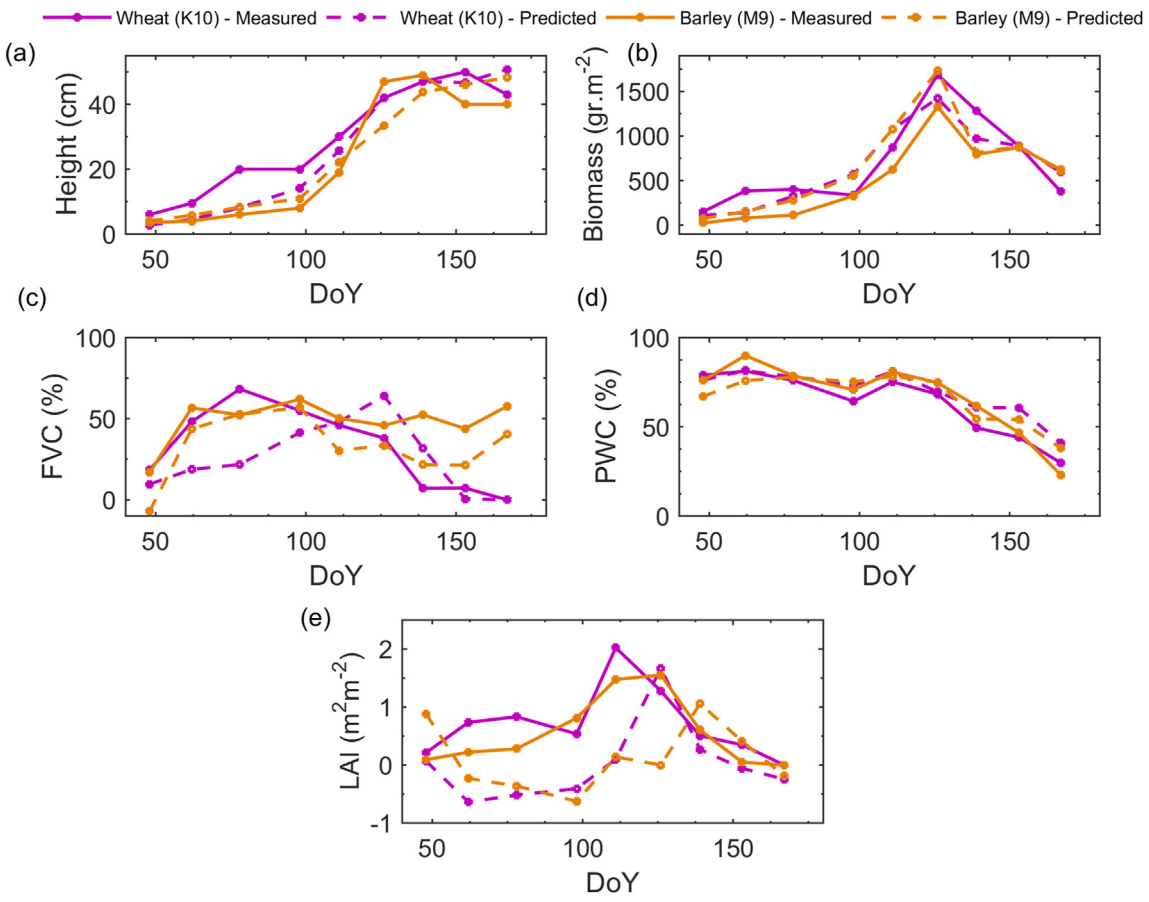

Fig. 8 – Prediction models for the original panel data for field parameters as a function of factors derived from a lineal combination of satellite parameters. (a) height; (b) biomass; (c) fraction of vegetation cover (FVC); (d) percentage of water content (PWC); (e) leaf area index (LAI). DoY (Day of Year).

showed significant differences and SM was found to be smaller for barley (F11) and higher for pasture (H9) compared to the remaining crops.

Regarding the influence of the crop type and the field measurements, several differences between crop types and plots were found. Particularly, the most changing variables among crops were LAI and FVC. Besides, pasture (FVC, LAI and SM) and rye (height and LAI) showed more differences than the rest of crops. These differences resulted in a worse prediction of LAI and FVC than the rest of variables. Therefore, the crop type seems to influence several polarimetric parameters (HV, HH/VV, HV/VV, PPD, α_1, and γ_{P1P2}). The structure of a crop canopy significantly impacts the intensity and type of scattering. Polarimetric parameters that respond to multiple or volume scattering within the crop canopy are most suitable for crop identification. Many researchers have found that the backscattering coefficient (HV) is the most important polarisation parameter for identifying the majority of crops (Lee, Grunes, & Pottier, 2001; McNairn, Champagne, Shang, Holmstrom, & Reichert, 2009; McNairn, Shang, Jiao, & Champagne, 2009). Valcarce-Diñeiro, Arias-Pérez, Lopez-Sanchez, and Sánchez (2019) also found this relationship between crop types and polarimetric parameters. They found that the dominant alpha angle (α_1), the cross-polar backscattering coefficient (HV) and the backscattering ratios (HH/VV and HV/VV) were the most important parameters for distinguishing crops.

Regarding the influence of crop type on the incidence angle and the parameters, the strongest influence occurred for the copolar channels (HH and VV), γ_{HHVV}, and H, with the differences between the incidence angle and the parameters being higher at the steepest angle (25°). The values of HH and VV were very different between wheat (L7) and wheat (J12) at 25°. A similar effect was found in (Valcarce-Diñeiro et al., 2018). Higher values of both copolar channels were observed for wheat (L7) at 25° than for wheat (J12).

Factor analysis has been demonstrated to be a useful tool for reducing the dimensionality in the number of instrumental variables in the proposed estimation models. Furthermore, linear relationships between satellite measures have been detected, allowing the construction of reference factors for field measures. Alternatives to factor analysis, such as weighted variable aggregations or network analysis have the disadvantages of not considering the specific variation in the variables involved or requiring the selection of a threshold applied to "dichotomise" the correlation matrix between variables, which is a controversial issue that could have significant effects on the structure of the resultant network.

A panel data analysis method was used to assess the potential relationships between SAR observations and field

measurements and therefore to establish a dependence model between both datasets for forecasting the biophysical variables. This approach is very new as compared with the usual methods used by the remote sensing community, e.g., regressions and physical/empirical modelling. To retrieve vegetation parameters using any type of polynomial fit between satellite/field data (Jackson et al., 2004; Kycko, Romanowska, & Zagajewski, 2019; Zhang, Hong, Qin, & Liu, 2013), the datasets need to be large and robust, which is rare owing the scarcity and cost of the field measurements. In the second case (physical or empirical approach), the modelling usually requires ancillary data describing the soil-atmosphere-plant continuum, equally difficult to gather (Bastiaanssen et al., 2005; Pu et al., 2020). Other approaches (Bellón, Bégué, Lo Seen, De Almeida, & Simões, 2017; Jiang, Liang, Wang, & Xiao, 2010) are based in time series analysis to model variables such as LAI and NDVI. Recent studies such as presented in Holloway and Mengersen (2018) review useful statistical machine learning methods. All these methods required either a large amount of data in large historical data or expensive explanatory variables difficult to measure. On the contrary, the proposed method only requires the measurements and observations of a single growing cycle.

All crop types and parameters were included in the models, with the exception of the height of rye (N9) and the SM of barley (F11) and pasture (H9) due to its different behaviour. The validating metric indicates good performance in all cases, with a determination coefficient, R^2, ranging from 0.78 to 0.96. When considering the four constructed factors, R^2 was greater than 0.88, except in the case of LAI ($R^2 = 0.29$) and FVC ($R^2 = 0.429$).

The models for the physical parameters related to water content (i.e., SM and PWC) and height were found to have higher values of R^2 (0.96, 0.95 and 0.95 respectively). More specifically, the estimation of SM using HH, VV, HV and γ_{HHVV} was found to be the best among the estimations of all of the biophysical variables. In contrast, LAI resulted in the lowest R^2, followed by FVC. It seems therefore that the model is more sensitive to the water parameters than to the volumetric parameters. Not surprisingly, these two parameters, SM and PWC were the most successfully retrieved in the validation experiment.

Among all of the biophysical variables estimated, soil moisture was the most benefited for the panel model, providing an estimation model with four polarimetric parameters: backscattering coefficients and the normalised correlation between the co-polar channels HH and VV (γ_{HHVV}). Radar backscattering coefficients have shown to be correlated with soil moisture since the dielectric constant of soil is related to the amount of water held in soil (De Roo, Du, Ulaby, & Dobson, 2001; Srivastava, Patel, Sharma, & Navalgund, 2009).

Regarding the height, it is widely recognised that radar signals tend to increase quickly with increasing crop height until a certain threshold, after which the signal increases only slightly (Baghdadi, Boyer, Todoroff, El Hajj, & Bégué, 2009; Moran et al., 2012). In this study, the coherence between the first and second channels in the Pauli basis (γ_{P1P2}) was found to be the best satellite parameter to predict this biophysical variable. The use of this parameter could be possible due to the first Pauli channel is related to the scattering from soil surface and the second channel is associated to the double-bounce between soil and stems, being sensible to the growing cycle of the crops. Liu et al. (2012) found that RADARSAT-2 Pauli basis decomposition was a good indicator of crop growth development.

The backscattering coefficient (HV) drove the estimation of biomass. Backscattering coefficients have been effective for monitoring crop conditions where changes in vegetation structure and growth were detectable (Moran et al., 2012; Valcarce-Diñeiro et al., 2018) because HH is sensitive to horizontally oriented structures, VV to vertical structures and HV to random scatters, which are often related to volume scattering. It is generally accepted that C-band polarimetric data are useful for crop biomass monitoring (Canisius et al., 2018; Wiseman et al., 2014). However, recent studies have shown the potential of using the L-band to estimate biomass. Hosseini and McNairn (2017) exploited both the C-band (RADARSAT-2) and the L-band (UAVSAR) to derive the biomass and soil moisture of wheat fields using the WCM (water cloud model) (Attema & Ulaby, 1978). Reisi-Gahrouei, Homayouni, McNairn, Hosseini, and Safari (2019) used time series of SAR polarimetric parameters from an unmanned aerial vehicle synthetic aperture radar (UAVSAR) airborne L-band to estimate crop biomass using a multiple regression model (MRM) and an artificial neural network model (ANNM).

Fraction of vegetation cover (FVC) was another biophysical variable under study. Along with leaf area index (LAI) and biomass, it is an important parameter for many agronomic, ecological and meteorological applications. In this study, the performance of FVC in the panel was relatively good, with a correlation coefficient (R^2) of 0.82. The polarimetric parameters used to estimate FVC were HV and H. Valcarce-Diñeiro et al. (2018) also found a good correlation between the cross-polar channel HV and the FVC using simple correlations. The relationship between the entropy and the FVC could be explained by assuming that at the beginning of the growing season, the radar response is characterised by a low entropy value. At this moment of the cycle, there is more bare soil than vegetation canopy (α_1 is below 15°). By the middle of the plant growth season, the value of entropy is moderate due to the presence of both soil and vegetation cover. Then, at the end of the growing season, the value of entropy decreases again due to the withering of plants. This behaviour as a function of time would be similar to the typical FVC behaviour (Fig. 4). Finally, LAI was estimated using only one polarimetric parameter (α_1). The dominant alpha angle has been reported to be a good indicator of the sensitivity of this parameter with LAI (Jiao et al., 2011; Valcarce-Diñeiro et al., 2018). The WCM is the most widely used method for estimating crop LAI from SAR data. Bériaux et al. (2015) used the semi-empirical water cloud model to estimate LAI from C-band SAR data and developed a Bayesian fusion method that improved LAI retrieval. Hosseini, McNairn, Merzouki, and Pacheco (2015) used the multipolarisation RADARSAT-2 (C-band) and a UAVSAR (L-band) to investigate the applicability of the C- and L-bands to LAI estimation.

One of the weaknesses of the present methodology for field parameter estimation and forecasting is the chance of obtaining negative values with no physical meaning. Even though this effect is not very pronounced, it occurred mainly for the field parameter that varied the most with time (LAI, Fig. 4) and for the smaller values observed (F11, Fig. 4). To improve the method in the future, truncated distributions could be used in order to avoid negative values in the estimations.

7. Conclusions

In this study, the ability of panel data and factorial techniques combined with *in situ* measurements and polarimetric parameters derived from radar imagery to predict crop biophysical variables was assessed. These statistical techniques are ideal for small datasets, such as the nine measurements in the field campaign, and, as far as we know, they have not been explored previously in this context. Although different crop types are mixed in the panel data, the resulting estimation and prediction models show a good performance when applied to validation data. These results could even be better for a single crop type in the panel data model to be retrieved and validated.

When the influence of the crop type was analysed thorough the factorial experiment, it was concluded that crop type had no effect on several biophysical variables. However, the polarimetric parameters were affected by the crop type due to the different canopy structures of the different crops, characterised by changes in the intensity and type of scattering of the radar signal. When the incidence angle was considered, large differences were found between the parameters at different incidence angles.

The panel strategy was able to model and predict many of the parameters measured in the field and to overcome the scarcity of data. The panel data analysis with no factors showed that the physical parameters related to water content (i.e., SM and PWC) were better predicted ($R^2 = 0.96$ and $R^2 = 0.95$ respectively) than those related to the structure of the crop. However, when the panel used factors to predict the biophysical variables, biomass was found to be the best predicted ($R^2 = 0.94$).

The results of this study demonstrate that the application of panel and factorial approaches to new databases, such as those provided by satellite images, is more effective than other methods, such as regressions, that are typically used in remote sensing, and can accurately predict crop biophysical variables. Owing to the increasing number of upcoming satellite SAR sensors with new capabilities in frequencies, image modes and revisit times, the panel technique may be considered a robust and convenient alternative to traditional statistical modelling.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Spanish Ministry of Science, Innovation and Universities under Projects TEC2017-85244-C2-1-P and PRODEBAT PID2019-106254RB-I00, the State Agency of Research (AEI) under Project PID2019-108311GB-I00, the European Funds for Regional Development (EFRD) under Project ESP2017-89463-C3-3-R and by the Spanish Ministry of Economy and Competitiveness, Projects MTM2016-80539-C2-2-R.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biosystemseng.2021.02.014.

REFERENCES

Adah, M. I., Kayode, O. O., & Victor, I. C. (2017). Cereal productivity in West Africa: A panel data analysis. *American Journal of Business, Economics and Management,* 5(4), 38. http://www.openscienceonline.com/journal/archive2?journalId=709&paperId=3744.

Arellano, M. (2003). Panel data econometrics. In *Advanced texts in econometrics*. Oxford University Press. https://doi:10.1093/0199245282.001.0001.

Attema, E. P. W., & Ulaby, F. T. (1978). Vegetation modeled as a water cloud. *Radio Science, 13*, 357–364. https://doi.org/10.1029/RS013i002p00357

Baghdadi, N., Boyer, N., Todoroff, P., El Hajj, M., & Bégué, A. (2009). Potential of SAR sensors TerraSAR-X, ASAR/ENVISAT and PALSAR/ALOS for monitoring sugarcane crops on Reunion Island. *Remote Sensing of Environment, 113*(8), 1724–1738. https://doi.org/10.1016/j.rse.2009.04.005

Baltagi, B. (2008). *Econometric analysis of panel data*. John Wiley & Sons.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London - Series A: Mathematical and Physical Sciences, 160*(901), 268–282. https://doi.org/10.1098/rspa.1937.0109

Bastiaanssen, W. G. M., Noordman, E. J. M., Pelgrum, H., Davids, G., Thoreson, B. P., & Allen, R. G. (2005). SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. *Journal of Irrigation and Drainage Engineering, 131*(1), 85–93. https://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9437(2005)131:1(85).

Bellón, B., Bégué, A., Lo Seen, D., De Almeida, C. A., & Simões, M. (2017). A remote sensing approach for regional-scale mapping of agricultural land-use systems based on NDVI time series. *Remote Sensing,* 9(6), 600. https://doi.org/10.3390/rs9060600

Beriaux, E., Lucau-Danila, C., Auquiere, E., & Defourny, P. (2013). Multiyear independent validation of the water cloud model for retrieving maize leaf area index from SAR time series. *International Journal of Remote Sensing, 34*(12), 4156–4181. https://doi.org/10.1080/01431161.2013.772676

Bériaux, E., Waldner, F., Collienne, F., Bogaert, P., & Defourny, P. (2015). Maize leaf area index retrieval from synthetic quad pol SAR time series using the water cloud model. *Remote Sensing, 7*(12), 16204–16225. https://doi.org/10.3390/rs71215818

Cable, J. W., Kovacs, J. M., Jiao, X., & Shang, J. (2014). Agricultural monitoring in northeastern Ontario, Canada, using multi-

temporal polarimetric RADARSAT-2 data. *Remote Sensing, 6*(3), 2343–2371. https://doi.org/10.3390/rs6032343

Canisius, F., Shang, J., Liu, J., Huang, X., Ma, B., Jiao, X., et al. (2018). Tracking crop phenological development using multi-temporal polarimetric Radarsat-2 data. *Remote Sensing of Environment, 210*, 508–518. https://doi.org/10.1016/j.rse.2017.07.031

Chauhan, S., Srivastava, H. S., & Patel, P. (2018). Wheat crop biophysical parameters retrieval using hybrid-polarized RISAT-1 SAR data. *Remote Sensing of Environment, 216*, 28–43. https://doi.org/10.1016/j.rse.2018.06.014

Choudhury, I., & Chakraborty, M. (2006). SAR signature investigation of rice crop using RADARSAT data. *International Journal of Remote Sensing, 27*(3), 519–534. https://doi.org/10.1080/01431160500239172

Colliander, A., Jackson, T. J., Bindlish, R., Chan, S., Das, N., Kim, S. B., et al. (2017). Validation of SMAP surface soil moisture products with core validation sites. *Remote Sensing of Environment, 191*, 215–231. https://doi.org/10.1016/j.rse.2017.01.021

De Roo, R. D., Du, Y., Ulaby, F. T., & Dobson, M. C. (2001). A semi-empirical backscattering model at L-band and C-band for a soybean canopy with soil moisture inversion. *IEEE Transactions on Geoscience and Remote Sensing, 39*(4), 864–872. https://doi.org/10.1109/36.917912

Engel, U., & Reinecke, J. (2012). *Analysis of change: Advanced techniques in panel data analysis*. Walter de Gruyter.

Gherboudj, I., Magagi, R., Berg, A. A., & 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), 33–43. https://doi.org/10.1016/j.rse.2010.07.011

González-Zamora, Á., Sánchez, N., Pablos, M., & Martínez-Fernández, J. (2019). CCI soil moisture assessment with SMOS soil moisture and in situ data under different environmental conditions and spatial scales in Spain. *Remote Sensing of Environment, 225*, 469–482. https://doi.org/10.1016/j.rse.2018.02.010

Gourieroux, C., & Monfort, A. (1993). Simulation-based inference: A survey with special reference to panel data models. *Journal of Econometrics, 59*(1–2), 5–33. https://doi.org/10.1016/0304-4076(93)90037-6

Holloway, J., & Mengersen, K. (2018). Statistical machine learning methods and remote sensing for sustainable development goals: A review. *Remote Sensing, 10*(9), 1365. https://doi.org/10.3390/rs10091365

Hosseini, M., & McNairn, H. (2017). Using multi-polarization C- and L-band synthetic aperture radar to estimate biomass and soil moisture of wheat fields. *International Journal of Applied Earth Observation and Geoinformation, 58*, 50–64. https://doi.org/10.1016/j.jag.2017.01.006

Hosseini, M., McNairn, H., Merzouki, A., & Pacheco, A. (2015). Estimation of Leaf Area Index (LAI) in corn and soybeans using multi-polarization C-and L-band radar data. *Remote Sensing of Environment, 170*, 77–89. https://doi.org/10.1016/j.rse.2015.09.002

Hsiao, C. (2007). Panel data analysis-advantages and challenges. *Test, 16*(1), 1–22. https://doi.org/10.1007/s11749-007-0046-x

Hu, B., & McAleer, M. (2005). Estimation of Chinese agricultural production efficiencies with panel data. *Mathematics and Computers in Simulation, 68*(5–6), 474–483. https://doi.org/10.1016/j.matcom.2005.02.002

Jackson, T. J., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., et al. (2004). Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. *Remote Sensing of Environment, 92*(4), 475–482. https://doi.org/10.1016/j.rse.2003.10.021

Jiang, B., Liang, S., Wang, J., & Xiao, Z. (2010). Modeling MODIS LAI time series using three statistical methods. *Remote Sensing of Environment, 114*(7), 1432–1444. https://doi.org/10.1016/j.rse.2010.01.026

Jiao, X., Kovacs, J. M., Shang, J., McNairn, H., Walters, D., et al.Geng, X. (2014). Object-oriented crop mapping and monitoring using multi-temporal polarimetric RADARSAT-2 data. *ISPRS Journal of Photogrammetry and Remote Sensing, 96*, 38–46. https://doi.org/10.1016/j.isprsjprs.2014.06.014

Jiao, X., McNairn, H., Shang, J., Pattey, E., Liu, J., & Champagne, C. (2011). The sensitivity of RADARSAT-2 polarimetric SAR data to corn and soybean leaf area index. *Canadian Journal of Remote Sensing, 37*(1), 69–81. https://doi.org/10.5589/m11-023

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kouser, S., & Qaim, M. (2011). Impact of Bt cotton on pesticide poisoning in smallholder agriculture: A panel data analysis. *Ecological Economics, 70*(11), 2105–2113. https://doi.org/10.1016/j.ecolecon.2011.06.008

Kycko, M., Romanowska, E., & Zagajewski, B. (2019). Lead-induced changes in fluorescence and spectral characteristics of pea leaves. *Remote Sensing, 11*(16), 1885. https://doi.org/10.3390/rs11161885

Lee, J. S., Grunes, M. R., & Pottier, E. (2001). Quantitative comparison of classification capability: Fully polarimetric versus dual and single-polarization SAR. *IEEE Transactions on Geoscience and Remote Sensing, 39*(11), 2343–2351. https://doi.org/10.1109/36.964970

Liu, C., Shang, J., Vachon, P. W., & McNairn, H. (2012). Multiyear crop monitoring using polarimetric RADARSAT-2 data. *IEEE Transactions on Geoscience and Remote Sensing, 51*(4), 2227–2240. https://doi.org/10.1109/TGRS.2012.2208649

Lopez-Sanchez, J. M., & Ballester-Berman, J. D. (2009). Potentials of polarimetric SAR interferometry for agriculture monitoring. *Radio Science, 44*(2), 1–20. https://doi.org/10.1029/2008RS004078

Mandal, D., Hosseini, M., McNairn, H., Kumar, V., Bhattacharya, A., Rao, Y. S., et al. (2019). An investigation of inversion methodologies to retrieve the leaf area index of corn from C-band SAR data. *International Journal of Applied Earth Observation and Geoinformation, 82*, 101893. https://doi.org/10.1016/j.jag.2019.06.003

Mandal, D., Kumar, V., McNairn, H., Bhattacharya, A., & Rao, Y. S. (2019). Joint estimation of Plant Area Index (PAI) and wet biomass in wheat and soybean from C-band polarimetric SAR data. *International Journal of Applied Earth Observation and Geoinformation, 79*, 24–34. https://doi.org/10.1016/j.jag.2019.02.007

Mattia, F., Le Toan, T., Picard, G., Posa, F. I., D'Alessio, A., Notarnicola, C., et al. (2003). Multitemporal C-band radar measurements on wheat fields. *IEEE Transactions on Geoscience and Remote Sensing, 41*(7), 1551–1560. https://doi.org/10.1109/TGRS.2003.813531

Mátyás, L., & Sevestre, P. (2013). *The econometrics of panel data: Handbook of theory and applications* (Vol. 28). Springer Science & Business Media.

McNairn, H., Champagne, C., Shang, J., Holmstrom, D., & 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), 434–449. https://doi.org/10.1016/j.isprsjprs.2008.07.006

McNairn, H., Ellis, J., Van Der Sanden, J. J., Hirose, T., & Brown, R. J. (2002). Providing crop information using RADARSAT-1 and satellite optical imagery. *International Journal of Remote Sensing, 23*(5), 851–870. https://doi.org/10.1080/01431160110070753

McNairn, H., Shang, J., Jiao, X., & Champagne, C. (2009b). The contribution of ALOS PALSAR multipolarization and polarimetric data to crop classification. *IEEE Transactions on*

Geoscience and Remote Sensing, 47(12), 3981–3992. https://doi.org/10.1109/TGRS.2009.2026052

Moran, M. S., Alonso, L., Moreno, J. F., Mateo, M. P. C., De La Cruz, D. F., & Montoro, A. (2012). A RADARSAT-2 quad-polarized time series for monitoring crop and soil conditions in Barrax, Spain. *IEEE Transactions on Geoscience and Remote Sensing, 50*(4), 1057–1070. https://doi.org/10.1109/TGRS.2011.2166080

Morena, L. C., James, K. V., & Beck, J. (2004). An introduction to the RADARSAT-2 mission. *Canadian Journal of Remote Sensing, 30*(3), 221–234. https://doi.org/10.5589/m04-004

Mundlak, Y., Butzer, R., & Larson, D. F. (2008). *Heterogeneous technology and panel data: The case of the agricultural production function*. The World Bank. https://doi.org/10.1596/1813-9450-4536

Pu, J., Yan, K., Zhou, G., Lei, Y., Zhu, Y., Guo, D., et al. (2020). Evaluation of the MODIS LAI/FPAR algorithm based on 3D-RTM simulations: A case study of grassland. *Remote Sensing, 12*(20), 3391. https://doi.org/10.3390/rs12203391

Reisi-Gahrouei, O., Homayouni, S., McNairn, H., Hosseini, M., & Safari, A. (2019). Crop biomass estimation using multi regression analysis and neural networks from multitemporal L-band polarimetric synthetic aperture radar data. *International Journal of Remote Sensing, 40*(17), 6822–6840. https://doi.org/10.1080/01431161.2019.1594436

Sánchez, N., Alonso-Arroyo, A., Martínez-Fernández, J., Piles, M., González-Zamora, Á., Camps, A., et al. (2015). On the synergy of airborne GNSS-R and Landsat 8 for soil moisture estimation. *Remote Sensing, 7*(8), 9954–9974. https://doi.org/10.3390/rs70809954

Sánchez, N., Martínez-Fernández, J., González-Piqueras, J., González-Dugo, M. P., Baroncini-Turrichia, G., Torres, E., et al. (2012). Water balance at plot scale for soil moisture estimation using vegetation parameters. Agricultural *and Forest Meteorology, 166*, 1–9. https://doi.org/10.1016/j.agrformet.2012.07.005

Sanchez, N., Martínez-Fernández, J., Scaini, A., & Perez-Gutierrez, C. (2012). Validation of the SMOS L2 soil moisture data in the REMEDHUS network (Spain). *IEEE Transactions on Geoscience and Remote Sensing, 50*(5), 1602–1611. https://doi.org/10.1109/TGRS.2012.2186971

Srivastava, H. S., Patel, P., Sharma, Y., & Navalgund, R. R. (2009). Large-area soil moisture estimation using multi-incidence-angle RADARSAT-1 SAR data. *IEEE Transactions on Geoscience and Remote Sensing, 47*(8), 2528–2535. https://doi.org/10.1109/TGRS.2009.2018448

Valcarce-Diñeiro, R., Arias-Pérez, B., Lopez-Sanchez, J. M., & Sánchez, N. (2019). Multi-temporal dual-and quad-polarimetric synthetic aperture radar data for crop-type mapping. *Remote Sensing, 11*(13), 1518. https://doi.org/10.3390/rs11131518

Valcarce-Diñeiro, R., Lopez-Sanchez, J. M., Sánchez, N., Arias-Pérez, B., & Martínez-Fernández, J. (2018). Influence of incidence angle in the correlation of C-band polarimetric parameters with biophysical variables of rain-fed crops. *Canadian Journal of Remote Sensing, 44*(6), 643–659. https://doi.org/10.1080/07038992.2019.1579051

Wiseman, G., McNairn, H., Homayouni, S., & 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), 4461–4471. https://doi.org/10.1109/JSTARS.2014.2322311

Wu, F., Wang, C., Zhang, H., Zhang, B., & Tang, Y. (2011). Rice crop monitoring in South China with RADARSAT-2 quad-polarization SAR data. *IEEE Geoscience and Remote Sensing Letters, 8*(2), 196–200. https://doi.org/10.1109/LGRS.2010.2055830

Xie, Q., Meng, Q., Zhang, L., Wang, C., Sun, Y., & Sun, Z. (2017). A soil moisture retrieval method based on typical polarization decomposition techniques for a maize field from full-polarization radarsat-2 data. *Remote Sensing, 9*(2), 168. https://doi.org/10.3390/rs9020168

Xu, J., Li, Z., Tian, B., Huang, L., Chen, Q., & Fu, S. (2014). Polarimetric analysis of multi-temporal RADARSAT-2 SAR images for wheat monitoring and mapping. *International Journal of Remote Sensing, 35*(10), 3840–3858. https://doi.org/10.1080/01431161.2014.919679

Yang, H., Chen, E., Li, Z., Zhao, C., Yang, G., Pignatti, S., et al. (2015). Wheat lodging monitoring using polarimetric index from RADARSAT-2 data. *International Journal of Applied Earth Observation and Geoinformation, 34*, 157–166. https://doi.org/10.1016/j.jag.2014.08.010

Yonezawa, C., Negishi, M., Azuma, K., Watanabe, M., Ishitsuka, N., Ogawa, S., et al. (2012). Growth monitoring and classification of rice fields using multitemporal RADARSAT-2 full-polarimetric data. *International Journal of Remote Sensing, 33*(18), 5696–5711. https://doi.org/10.1080/01431161.2012.665194

Zhang, N., Hong, Y., Qin, Q., & Liu, L. (2013). VSDI: A visible and shortwave infrared drought index for monitoring soil and vegetation moisture based on optical remote sensing. *International Journal of Remote Sensing, 34*(13), 4585–4609. https://doi.org/10.1080/01431161.2013.779046