



Università degli Studi della Tuscia di Viterbo
Dipartimento di Scienze Agrarie e Forestali (DAFNE)

Corso di Dottorato di Ricerca in

Scienze delle Produzioni Vegetali e Animali - XXXIII ciclo

**EXPLOITATION OF MULTI-TEMPORAL AND MULTI-SENSOR SATELLITE DATA FOR IMPROVING
BIOPHYSICAL AND AGRONOMIC VARIABLES RETRIEVAL AND YIELD PREDICTION THROUGH DATA
ASSIMILATION INTO THE AQUACROP MODEL**

(per es. **s.s.d. AGR/02**)

Tesi di dottorato di:

Dott. Deepak Upreti

Firma (Signature)

Coordinatore del corso

Prof. Roberta Bernini

Firma

.....

Tutore

Prof. Raffaele Casa

Firma

.....

A.A. 2019/20 (A.Y. 2019/20)

Dedicated to my Mother

Table of Contents

General Introduction	9
Objectives and Hypothesis.....	12
Structure of the PhD thesis	12
1. CHAPTER 1. A Comparison of Hybrid Machine Learning Algorithms for the Retrieval of Wheat Biophysical Variables from Sentinel-2	18
1. Introduction	18
2. Materials and Methods.....	20
2.1. Generation of PROSAIL Simulations.....	21
2.2. Training Machine Learning Algorithms	23
2.3. Validation with Sentinel-2 data and ground measurements.....	24
2.4. Comparative assessment of MLRAs	26
3. Results	26
4. Discussion.....	35
5. Conclusions	37
6. References	37
2. CHAPTER 2. A Comparison of Moment-Independent and Variance Based Global Sensitivity Analysis Approaches for Wheat Yield Estimation with the Aquacrop-OS Model.....	44
1. Introduction	44
2. Materials and Methods.....	47
2.1. Sensitivity Analysis	47
2.2. Crop Model: Aquacrop Open Source	49
2.3. Setting up the Parameters for AOS Model	51
2.4. Sites Description	51
2.5. Weather Data.....	55
3. Results	55
3.1. Morris Method	55
3.2. Variance-Based Extended Fourier Amplitude Sensitivity Test (EFAST).....	56
3.3. Density-Based PAWN Method	58
4. Discussion.....	63

4.1. Elementary Effects Using the Morris Method	63
4.2. Variance-Based Extended Fourier Amplitude Sensitivity Test (EFAST).....	63
4.3. Density-Based PAWN Method	64
4.4. Sensitive Parameters in Relation to Model Processes and Agronomic Conditions	64
5. Conclusions	65
References	66
3. CHAPTER 3. Bayesian Calibration of the Aquacrop-OS Model for Durum Wheat by Assimilation of Canopy Cover Retrieved from VEN μ S Satellite Data	78
1. Introduction	79
2. Materials and Methods.....	81
2.1. VEN μ S Satellite Data	81
2.2. Retrieval of Canopy Cover from VEN μ S Satellite Data.....	82
2.3. Smoothing and Fitting fvc Time Series Data Retrieved From Ven μ s Satellite Data.....	82
2.4. Study Site and Ground Validation Campaigns	82
2.5. Weather Data	84
2.6. Aquacrop-OS Model and Sensitive Parameters	85
2.7. Markov Chain Monte Carlo-Based DREAM _(KZS) Algorithm.....	85
2.8. Statistical Analysis and Validation.....	86
3. Results	87
3.1. fvc Retrieval	87
3.2. Parameters Identification	90
3.3. Calibration and Validation	91
4. Discussions	95
4.1. fvc Retrieval.....	95
4.2. Parameters Identification	96
4.3. Calibration and Validation	97
5. Conclusions	98
References	98
CHAPTER 4. Ensemble Kalman Filter based Assimilation of canopy cover retrieved from Sentinel-2 and VEN μ S satellite data into Aquacrop-OS model for improved durum wheat yield estimation in Central Italy	106
1. Introduction	106
2. Materials and Methods.....	109
2.1. Site Description, Soil and Weather Data.....	109

2.2.	Yield Data Processing	110
2.3.	Crop model: Aquacrop Open Source	110
2.4.	Sensitivity Analysis and Bayesian Calibration of the AOS crop model.....	112
2.5.	Satellite Data and Processing.....	112
3.	Results and Discussions	116
3.1.	Ensemble Kalman Filter based data assimilation.....	116
3.2.	Spatial Application of Ensemble Kalman filter based assimilation	120
4.	Conclusions	123
	References	123
	General Conclusions	127

SUMMARY

The use of remote sensing data has been shown to be very useful for field management, crop acreage estimation and yield prediction, both for field scale applications, e.g., in precision agriculture, and at regional scale. The development of new technologies and the availability of high spatial and spectral resolution satellite data allow the increasing use of these techniques. The aim of this work is to combine multi-source remote sensing data with the Aquacrop crop model, developed by the Food and Agriculture Organization (FAO), using an open-source version (Aquacrop-OS), for estimating the durum wheat yield at the field scale, using a sequential data assimilation-based ensemble Kalman filter approach.

As a first step, multiple hybrid machine learning algorithms were compared for the retrieval of wheat crop biophysical variables from Sentinel-2 satellite data. Gaussian Process Regression (GPR) performed best among all the methods of retrieval in the case of canopy cover. This method was thus further adopted, to be applied to all the Sentinel-2 and VEN μ S cloud free images available in the crop growing seasons 2017-2018 and 2018-2019 for the study site of Maccarese (Central Italy).

On the crop modeling part, initially a global sensitivity analysis approach was used to identify the most influential crop model parameters, using simulations for the study area (Maccarese Farm, Central Italy). The sensitivity analysis was performed with three different methods, Morris Screening, Extended Fourier Amplitude Sensitivity Test (EFAST) and PAWN method. These three different methods have different rationale and background. We sorted the sensitivity indices for all the parameters of the model and found a common set of most influential parameters to the target output, which was grain yield. The parameters identified as most sensitive in the analysis were further subjected to a Bayesian calibration using the observed canopy cover of the 2017-2018 season. The optimized crop model parameter values found in this calibration were validated with the observed canopy cover data of the year 2018-2019 and yield data for the year 2017-2018. A good agreement was found in the case of the canopy cover, but the yield assimilated was slightly overestimated.

After the sensitivity analysis and the Bayesian calibration, the crop model Aquacrop-OS was ready to ingest the in-season canopy cover from Sentinel- and VEN μ S satellite data. We used sequential data assimilation-based ensemble Kalman filter approach of data assimilation to combine the canopy cover time series in the crop growth model Aquacrop-Open Source. On each observation date during the crop growth season using canopy cover from the satellite data and simulated canopy cover, a new updated value was estimated, and the new model simulations start from that value, this way the simulated value gets close to the observed value. Canopy cover only from Sentinel-2, only from VEN μ S data and their combined data was ingested separately to assess the performance of the assimilation. Overall, all the three sources of canopy cover performed well, but the increased number of observations by combining both the sources of data did not improved assimilation, rather it was computationally expensive.

This study shows the possibility to estimate durum wheat production by combining multi source remote sensing data and an open-source crop growth model, Aquacrop-OS, highlighting the strengths and weakness of the assimilation procedure.

RIASSUNTO

L'uso di dati telerilevati si è dimostrato molto utile per la gestione agronomica, la stima della superficie coltivata e la previsione della resa, sia per applicazioni su scala di campo, come nell'agricoltura di precisione, sia su scala regionale. Lo sviluppo di nuove tecnologie e la disponibilità di dati satellitari ad alta risoluzione spaziale e spettrale consentono un uso crescente di queste tecniche. Lo scopo di questo lavoro è di combinare dati telerilevati satellitari provenienti da diverse fonti con il modello Aquacrop, sviluppato dalla FAO ed utilizzato in questo lavoro nella versione open-source Aquacrop-OS, al fine di stimare la produzione di grano duro su scala di campo, utilizzando l'approccio "ensemble Kalman filter" basato sull'assimilazione di dati sequenziali.

In una fase iniziale, sono stati confrontati diversi algoritmi ibridi di machine learning per ottenere le variabili biofisiche del grano duro, a partire dai dati del satellite Sentinel-2. Tra i metodi considerati, la Gaussian Process Regression (GPR) ha dato i migliori risultati per la stima della copertura fogliare. Il metodo è stato quindi applicato a tutte le immagini di Sentinel-2 e VENμS caratterizzate da ridotta nuvolosità per le stagioni colturali 2017-18 e 2018-19.

Per quanto riguarda il modello colturale, come prima cosa è stata utilizzata una "global sensitivity analysis" per identificare i parametri colturali del modello più rilevanti, con simulazioni per l'area di studio (Azienda Maccarese, Italia centrale). La sensitivity analysis è stata eseguita con tre diversi metodi: Morris Screening, Extended Fourier Amplitude Sensitivity Test (EFAST), e metodo PAWN. I tre metodi hanno diverse assunzioni e fondamenti logico-statistici. Gli indici di sensibilità di tutti i parametri del modello sono stati ordinati ed i parametri con maggiore influenza sulla produzione di granella sono stati raggruppati. Questi parametri sono stati sottoposti ad una procedura di calibrazione di tipo Bayesiano usando i dati osservati di copertura fogliare della stagione 2017-18. I valori ottimali dei parametri del modello così ottenuti sono stati validati usando i dati osservati di copertura fogliare del 2018-19 e di resa in granella del 2017-18. Un buon accordo è stato riscontrato tra dati simulati ed osservati di copertura fogliare, mentre la resa è stata leggermente sovrastimata dal modello.

Il modello colturale, tramite la sensitivity analysis e la calibrazione Bayesiana, è stato preparato per poter assimilare i dati satellitari di Sentinel-2 e VENμS. È stata utilizzata un'assimilazione sequenziale usando l'approccio "ensemble Kalman filter" per combinare le serie temporali di copertura fogliare nel modello Aquacrop-OS. Per ogni data in cui i dati satellitari di copertura fogliare erano disponibili, è stato aggiornato il valore simulato di copertura fogliare. Le nuove simulazioni sono iniziate tenendo conto di tale valore, ottenendo quindi dei valori simulati sempre più vicini a quelli osservati. I dati di copertura fogliare provenienti dal solo Sentinel-2, quelli provenienti dal solo VENμS e quelli ottenuti dalla fusione delle immagini di entrambi i satelliti sono stati combinati con il modello colturale per valutare l'efficacia dell'assimilazione. Complessivamente, l'assimilazione ha dato buoni risultati in tutti e tre i casi, ma l'aumento del numero di osservazioni ottenuto con la fusione delle immagini dei due satelliti non ha migliorato l'efficienza dell'assimilazione, portando invece ad aumento del costo computazionale.

Questo studio mostra la possibilità di stimare la produzione di grano duro combinando dati di telerilevamento provenienti da fonti diverse con un modello colturale open-source, Aquacrop-OS, ed evidenzia i punti di forza e di debolezza della procedura di assimilazione.

General Introduction

Spatial information on crop yield is valuable for both crop management and market forecast (Lobell, 2013). Currently, such data exist only at administrative units such as countries, states or provinces. For instance, the United States Department of Agriculture (USDA) releases monthly updates on crop yield predictions, weekly crop progress report for each state, and county-level yield statistics for major commodities. Such information, albeit being useful, does not indicate between and within-field variability that is essential to inform field-specific decision making for various sectors including farmers, crop insurance companies and food trading agencies (Lobell and Azzari, 2017). Therefore, robust large-scale yield estimation at the individual field level is an imperative task for agricultural research and application (Zaks and Kucharik, 2011).

Remote Sensing has played a crucial role in monitoring crop growth spatially. Using remote sensing technologies to estimate crop yields dates back to the 1970's where spectral signature from satellite imagery was associated with the crop yields using statistical models (Dorigo et al., 2007; Haun, 1974; Idso et al., 1977). Since then, the remote sensing research community developed various methods of yield estimation, which roughly fall into two categories: empirical approaches and data assimilation (Bolton and Friedl, 2013; Doraiswamy et al., 2003; Lobell et al., 2003; Thorp et al., 2012). The empirical methods rely on *in situ* or synthetic yield data, for example to train machine learning regression algorithms (Johnson, 2014; Kogan et al., 2013; Sakamoto et al., 2013). Since the training data are available only at county or state level, these studies tend to produce crop yield maps at coarse resolution (250 m -1 km) (Bolton and Friedl, 2013; Franch et al., 2015). On the other hand, data assimilation directly exploits the advantages of crop growth models in physical, chemical, and physiological simulations of canopy development (Claverie et al., 2009; Hank et al., 2015; Ines et al., 2013). There exist many strategies to integrate the crop growth simulations with the in-season crop growth variables derived usually from remote sensing platforms (Jin et al., 2018).

There are different biophysical variables that can be retrieved using remote sensing and are state variables in different crop growth models such as leaf area index (LAI), canopy cover (CC), soil moisture (SM). Biophysical variables such as LAI retrieval from remote sensing is based on the inverse modeling of the radiative transfer functions and this has shown effective when compared to the ground collected destructive measurements (Casa et al., 2019). There are different methods of biophysical variables retrieval using remote sensing mainly categorized in four different ways (Upreti et al., 2019; Verrelst et al., 2018):

- **Parametric Regression methods:** Parametric methods assume an explicit relationship between spectral observations and a specific biophysical variable. Thus, explicit parametrized expressions are built; they are usually based on some physical knowledge of absorption and scattering properties and statistical relationship between the variable and the spectral response. Typically, a band arithmetic formulation is defined (e.g., a spectral index) and then linked to the variable of interest based on a fitting function.
- **Non-Parametric Regression methods:** Nonparametric methods directly define regression functions according to information from the given spectral data and associated variable, i.e., they are data driven methods. Hence, in contrast to parametric regression methods, a non-explicit choice is to be made on spectral band relationships, transformation(s) or fitting function. Nonparametric methods can further be split into linear and non-linear methods.
- **Physically based model inversion methods:** Physically based algorithms are application of physical laws establishing photon interaction cause-effect relationships. Model variables are inferred based on specific knowledge, typically obtained with radiative transfer functions.

- **Hybrid regression methods:** A hybrid method combines elements of nonparametric statistics and physically based methods. Hybrid models rely on the generic properties of physically based methods combined with the flexibility and computational efficiency of nonparametric nonlinear regression methods.

Hybrid methods of biophysical variables retrieval are popular among the other methods, as they make combined use of both the physical laws and the training of machine learning algorithms. Various machine learning algorithms have been combined with the top of canopy reflectance data generated with the radiative transfer model such as PROSAIL. For example, Silvestro et al. (2017) combined the PROSAIL model generated top of canopy reflectance with Artificial Neural Networks to retrieve the LAI and CC from HJ1 A/B and Landsat OLI sensor. In this PhD research, we trained different linear, non-linear and kernel-based machine learning algorithms for the retrieval of LAI, Leaf Chlorophyll Content (LCC), canopy cover (CC), Fraction of Photosynthetically Absorbed Radiation (FAPAR) and Canopy Chlorophyll Content (CCC) from Sentinel-2 (Upreti et al., 2019). Validation ground data were collected in two different geographic locations: one in Central Italy and the other in Beijing, China. Among the methods compared, Gaussian Process Regression (GPR) had the advantage of providing an estimate of the uncertainty associated with the mean value of prediction.

There are different sources of uncertainty in the remote sensing-based data assimilation framework. The following figure shows the sources of the uncertainties in the remote sensing-based assimilation (Vrugt, 2016).

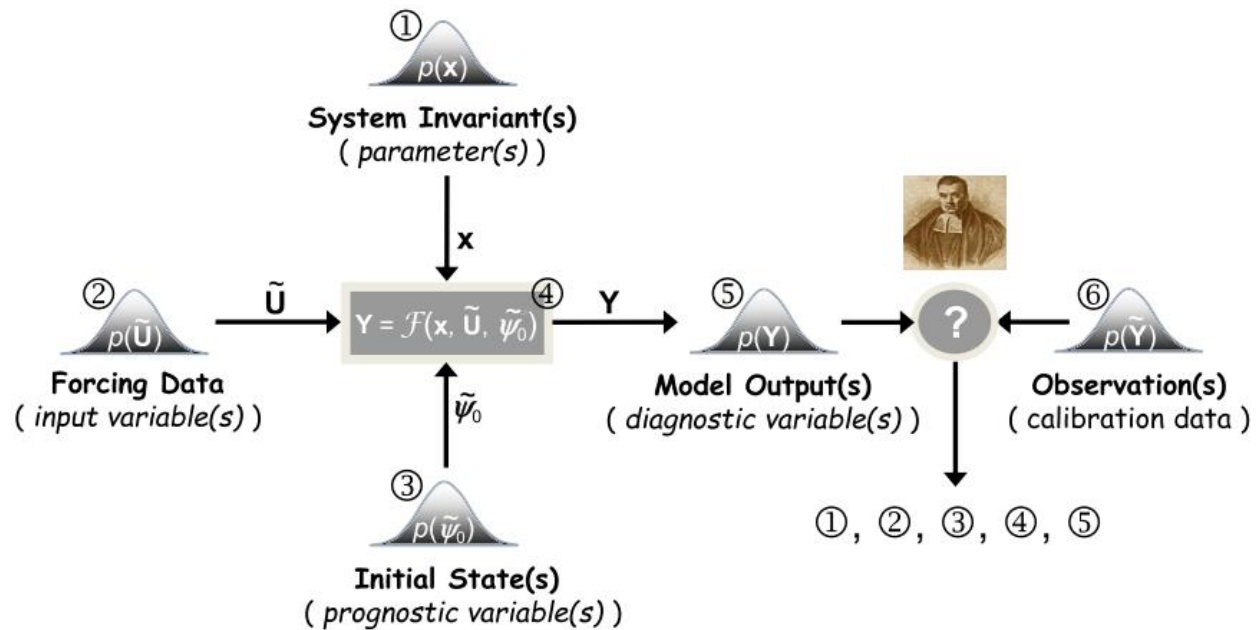


Figure 1. Schematic illustration of the most important sources of uncertainty in environmental systems modeling, including (1) parameter, (2) input data (also called forcing or boundary conditions), (3) initial state, (4) model structure, (5) Output, and (6) calibration data uncertainty. Source: (Vrugt, 2016)

Crop growth simulation models are important tools for ecology, agriculture and environmental research and management. Such models simulate the behavior of soil-crop systems in response to climate and agricultural management (de Willigen, 1991; Gervois et al., 2004; Hopmans and Bristow, 2002). In general, crop models contain many parameters and input variables, of which some cannot be measured

directly, but only can be inferred by a calibration of the responses observed from the model (Vrugt et al., 2002; Yapo et al., 1998). Identification of the most important or the highly sensitive parameters that have high impact on the output yield is crucial and computing the optimal value for these parameters at the field level can reduce the uncertainty the model output yield.

The parameters identified as the most influential to the crop yield, can be subjected to a local calibration to get the optimal values for these parameters. There are different approaches for model calibration: optimization methods that uses some kind of cost function to reduce the error between predicted and observed values are the most commonly used. Bayesian methods are based on the generation of Markov chains and can be used to transform a prior distribution function to a posterior distribution function of parameter values. The latter function can be summarized with the mean and confidence intervals of the predicted values of the parameters (Upreti et al., 2020). The advantage of these methods is that unlike in optimization methods, in which a single parameter value estimate is obtained, a distribution is produced, that can used to estimate uncertainty in the parameters.

The crop growth simulation model Aquacrop, was originally developed by the Food and Agriculture Organization (FAO) of the United Nations. Since the original model does not allow a customization, in this work we have used an open source version, developed in Matlab, of the model called Aquacrop Open Source (AOS) developed by (Foster et al., 2017). With respect to the complex crop models such as DSSAT and WOFOST, AOS model is not very complex, and has less free parameters in comparison to other models. It has been used successively to assess crop growth and yield (Jin et al., 2020; Silvestro et al., 2017; XING et al., 2017).

Data assimilation is the process of ingesting in-season crop biophysical variables into a crop growth model. This can be done in three different ways: calibration, forcing and updating. In the calibration methods the initial parameters of the crop model are adjusted to minimize the differences between the observed data and simulated data using some kind of optimization function. A major drawback of the method is that the observed data is considered true without the uncertainty contained in it (Kang and Özdoğan, 2019). Forcing methods are used to replace the simulated data with the observed data as per the availability of the observed data (Jin et al., 2018). The updating scheme is preferred over the others, as it allows to continuously update the model simulation while considering the uncertainty in both the simulated and observed data (Jin et al., 2018). The updating strategy can be implemented into two different approaches: Variational and Sequential. The former allows to update all the model simulations within a specific window, while the later recursively update model simulation using observations only from corresponding times (Jin et al., 2018, Kang and Özdoğan, 2019). Unlike statistical methods, the yield data is not required for training the model, therefore data assimilation can be applied spatially using satellite data for example from Sentinel-2 and Vegetation and Environment Monitoring Micro-Satellite (VENμS) missions.

The overall objective of the work is to retrieve the crop biophysical variables from Sentinel-2 and VENμS satellite data for the durum wheat crop growing season and integrate the time series of the canopy cover in the AOS model for the improved durum wheat yield estimates. The data assimilation framework selected for the objective is the sequential data assimilation-based Ensemble Kalman Filter, that takes into account the uncertainties contained in both the simulated and observed data to predict a new improved state. The work also identifies if increasing the number of observations by combining canopy cover from two different satellites has the effect on the yield estimate, which is the one of the assumptions in the ensemble Kaman filter for improved performance.

Objectives and Hypothesis

The work carried out in this research doctorate and illustrated in this thesis was within the ESA-MOST Dragon4 program, of scientific cooperation between the European Space Agency (ESA) and the Ministry of science and technology (MOST) of the People Republic of China. This program focuses on the use of earth observation data provided by ESA and the National Remote Sensing Center of China (NRSCC) to favor the development of applications in the terrestrial, oceanic and atmospheric fields. More specifically, the work is included in the Project ID32275 (Agricultural Monitoring) -Topic1 “Algorithm development exploiting multi-temporal and multi-sensor satellite data for improving biophysical and agronomic variables retrieval and yield prediction through data assimilation” carried out in collaboration between the University of Tuscia, The Institute of Methodology for Environmental Analysis of the National Research Council (IMAA-CNR) and National Engineering Research Center for Information Technology in Agriculture (NERCITA), Institute of Remote Sensing & Digital Earth (RADI) in Beijing. The main objective can be further split into sub-objectives as:

- Identifying the best methods to retrieve wheat crop biophysical variables from multi-source, multi-temporal satellite data
- Perform the sensitivity analysis of the crop growth model Aquacrop Open Source (AOS) version 6.0a to identify the most sensitive parameters to the output yield for the AOS crop model.
- Calibrate the crop model AOS for the local application using local ground conditions.
- Perform data assimilation by ingesting in season crop biophysical variables (canopy cover) into AOS crop model for improved yield prediction.

Structure of the PhD thesis

This Ph. D. thesis summarizes the research activities conducted at the Department of Agriculture and Forestry Sciences (DAFNE). It is divided in four chapters, each one corresponding to scientific papers published or submitted in different peer-reviewed journals (Fig. 2).

- **Chapter 1:** Upreti, D.; Huang, W.; Kong, W.; Pascucci, S.; Pignatti, S.; Zhou, X.; Ye, H.; Casa, R. A Comparison of Hybrid Machine Learning Algorithms for the Retrieval of Wheat Biophysical Variables from Sentinel-2. *Remote Sensing* **2019**, *11* (5), 481.
- **Chapter 2:** Upreti, D.; Pignatti, S.; Pascucci, S.; Tolomio, M.; Li, Z.; Huang, W.; Casa, R. A Comparison of Moment-Independent and Variance-Based Global Sensitivity Analysis Approaches for Wheat Yield Estimation with the Aquacrop-OS Model. *Agronomy* **2020**, *10* (4), 607
- **Chapter 3:** Upreti, D.; Pignatti, S.; Pascucci, S.; Tolomio, M.; Huang, W.; Casa, R. Bayesian Calibration of the Aquacrop-OS Model for Durum Wheat by Assimilation of Canopy Cover Retrieved from Venµs Satellite Data. *Remote Sensing* **2020**, *12* (16), 2666.
- **Chapter 4:** Upreti, D.; Pignatti, S.; Pascucci, S.; Tolomio, M.; Huang, W.; Casa, R. Ensemble Kalman Filter based Assimilation of canopy cover retrieved from Sentinel-2 and VENµS satellite

data into Aquacrop-OS model for improved durum wheat yield estimation in Central Italy. **(in prep.)**.

Chapter 1 is dedicated to the retrieval of the crop biophysical variables from Sentinel- satellite data. This retrieval was based on the simulations generated by the radiative transfer model PROSAIL, which were used to train different machine learning algorithms. Linear, non-linear and kernel based approaches were explored for the retrieval of the leaf area index (LAI), leaf chlorophyll content (LCC), fraction of absorbed photosynthetically active radiation (fapar), fraction of vegetation cover (fvc) and canopy chlorophyll content (CCC). The study was conducted for two different geographic locations one in Italy and the other in China. Retrieval of all the algorithms were compared with the ground data collected for the two locations.

Chapter 2 deals with the Sensitivity Analysis of the AOS crop model. For two different locations, one in Xiaotangshan, China and the other in Maccarese Farm, multiple years of weather year that shows different conditions of rainfall were used to see the impact of the model parameters to the output yield. Two different methods variance-based and moment-independent PAWN were implemented with seven years of weather data to identify the models most sensitive parameters that have influence on the yield. Both the methods were suitable for the identification of the most sensitive parameters although the ranking of the parameters varied among both the methods.

Chapter 3 is the follow up of the Chapter 2, it includes the Calibration of the most sensitive model parameters identified in the Chapter 2 for the Italian test site (Maccarese Farm). Bayesian calibration with the state-of-the-art implementation was used to get the inferences on the model parameters. The results indicate the importance of the calibration, the biomass and yield without the model calibration was overestimated and the calibration procedure get the results closer to the measured yield on the ground.

Chapter 4 is the final part of the thesis. Pixel wise data assimilation was performed using the Sequential ensemble Kalman filter approach. AOS model calibrated for the study site was used to generate the canopy cover time series and biomass and yield estimates, the state variable canopy cover was updated in the model by the canopy cover observed from remote sensing satellite data. We updated canopy cover from Sentinel-2, VEN μ S and their combination and performed the comparison to see the effects of increasing the number of observations during the crop growing season. This predicted yield was compared with the yield measured on the ground by a combine harvester equipped with a yield mapping system.

Apart from the papers published in the peer reviewed journals, the following are contributions that were made at conferences:

- Casa, R.; Upreti, D.; Palembo, A.; Pascucci, S.; Yang, H.; Yang, G.; Huang, W.; Pignatti, S. Evaluation and Exploitation of Retrieval Algorithms for Estimating Biophysical Crop Variables Using Sentinel-2, Ven μ s and Prisma Satellite Data, 2019, *Journal of Remote Sensing* **Submitted**.
- Casa, R.; Upreti, D.; Pelosi, F. Measurement and Estimation of Leaf Area Index (LAI) Using Commercial Instruments and Smartphone-Based Systems; IOP Publishing, 2019; Vol. 275, p 012006.

- Upreti, D.; Casa, R.; Pignatti, S.; Pascucci, S.; Huang, W.; Li, Z.; Yang, H.; Yang, G. Influence of Model Calibration and Sensitivity Analysis in a Data Assimilation Framework for Field-scale Crop Yield Mapping Using Biophysical Variables Retrieved from Sentinel-2 and Venus Satellite Data, 2019, (*in prep.*).

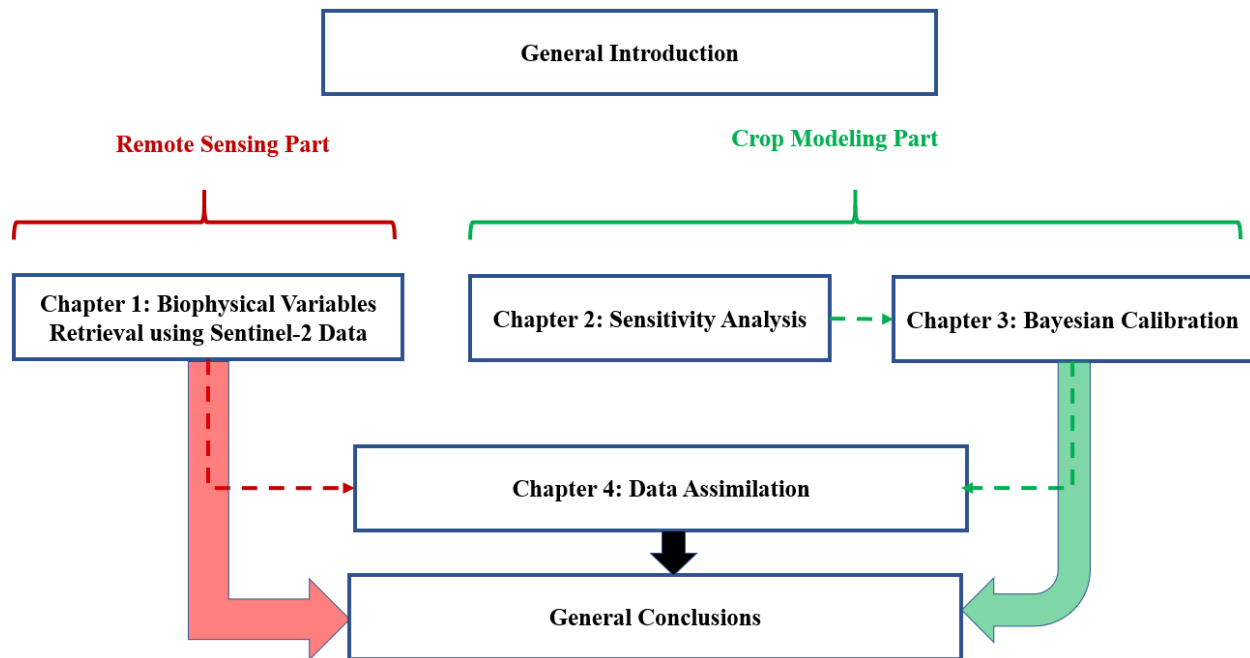


Figure 2. Thesis Structure

References of the General Introduction

Bolton, D.K., Friedl, M.A., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology* 173, 74–84.

Casa, R., Upreti, D., Pelosi, F., 2019. Measurement and estimation of leaf area index (LAI) using commercial instruments and smartphone-based systems. Presented at the IOP Conference Series: Earth and Environmental Science, IOP Publishing, p. 012006.

Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Keravec, P., Marciel, B., Ceschia, E., Dejoux, J.-F., Dedieu, G., 2009. Spatialization of crop leaf area index and biomass by combining a simple crop model SAFY and high spatial and temporal resolutions remote sensing data, in: 2009 IEEE International Geoscience and Remote Sensing Symposium. IEEE, p. III-478-III-481.

de Willigen, P., 1991. Nitrogen turnover in the soil-crop system; comparison of fourteen simulation models. *Fertilizer research* 27, 141–149. <https://doi.org/10.1007/BF01051122>

Doraiswamy, P.C., Moulin, S., Cook, P.W., Stern, A., 2003. Crop yield assessment from remote sensing. *Photogrammetric engineering & remote sensing* 69, 665–674.

- Dorigo, W.A., Zurita-Milla, R., de Wit, A.J., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International journal of applied earth observation and geoinformation* 9, 165–193.
- Foster, T., Brozović, N., Butler, A.P., Neale, C.M.U., Raes, D., Steduto, P., Fereres, E., Hsiao, T.C., 2017. AquaCrop-OS: An open source version of FAO's crop water productivity model. *Agricultural water management* 181, 18–22.
- Franch, B., Vermote, E.F., Becker-Reshef, I., Claverie, M., Huang, J., Zhang, J., Justice, C., Sobrino, J.A., 2015. Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR Growing Degree Day information. *Remote Sensing of Environment* 161, 131–148. <https://doi.org/10.1016/j.rse.2015.02.014>
- Gervois, S., de Noblet-Ducoudré, N., Viovy, N., Ciais, P., Brisson, N., Seguin, B., Perrier, A., 2004. Including croplands in a global biosphere model: methodology and evaluation at specific sites. *Earth Interactions* 8, 1–25.
- Hank, T.B., Bach, H., Mauser, W., 2015. Using a remote sensing-supported hydro-agroecological model for field-scale simulation of heterogeneous crop growth and yield: Application for wheat in central europe. *Remote Sensing* 7, 3934–3965. <https://doi.org/10.3390/rs70403934>
- Haun, J.R., 1974. Prediction of Spring Wheat Yields from Temperature and Precipitation Data 1. *Agronomy Journal* 66, 405–409.
- Hopmans, J.W., Bristow, K.L., 2002. Current capabilities and future needs of root water and nutrient uptake modeling, in: *Advances in Agronomy*. pp. 103–183.
- Idso, S.B., Jackson, R.D., Reginato, R.J., 1977. Remote-sensing of crop yields. *Science* 196, 19–25. <https://doi.org/10.1126/science.196.4285.19>
- Ines, A.V., Das, N.N., Hansen, J.W., Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment* 138, 149–164.
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., Wang, J., 2018. A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy* 92, 141–152.
- Jin, X., Li, Z., Feng, H., Ren, Z., Li, S., 2020. Estimation of maize yield by assimilating biomass and canopy cover derived from hyperspectral data into the AquaCrop model. *Agricultural Water Management* 227, 105846.
- Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment* 141, 116–128. <https://doi.org/10.1016/j.rse.2013.10.027>
- Kang, Y., Özdoğan, M., 2019. Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. *Remote Sensing of Environment* 228, 144–163.
- Kogan, F., Kussul, N., Adamenko, T., Skakun, S., Kravchenko, O., Kryvobok, O., Shelestov, A., Kolotii, A., Kussul, O., Lavrenyuk, A., 2013. Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. *International Journal of Applied Earth Observation and Geoinformation* 23, 192–203. <https://doi.org/10.1016/j.jag.2013.01.002>
- Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. *Field Crops Research* 143, 56–64.
- Lobell, D.B., Asner, G.P., Ortiz-Monasterio, J.I., Benning, T.L., 2003. Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. *Agriculture, Ecosystems & Environment* 94, 205–220.
- Lobell, D.B., Azzari, G., 2017. Satellite detection of rising maize yield heterogeneity in the US Midwest. *Environmental Research Letters* 12, 014014.

- Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2013. MODIS-based corn grain yield estimation model incorporating crop phenology information. *Remote Sensing of Environment* 131, 215–231.
- Silvestro, P., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., Huang, W., Casa, R., 2017. Estimating wheat yield in China at the field and district scale from the assimilation of satellite data into the Aquacrop and simple algorithm for yield (SAFY) models. *Remote Sensing* 9, 509.
- Thorp, K., Wang, G., West, A., Moran, M., Bronson, K., White, J., Mon, J., 2012. Estimating crop biophysical properties from remote sensing data by inverting linked radiative transfer and ecophysiological models. *Remote Sensing of Environment* 124, 224–233.
- Upreti, D., Huang, W., Kong, W., Pascucci, S., Pignatti, S., Zhou, X., Ye, H., Casa, R., 2019. A comparison of hybrid machine learning algorithms for the retrieval of wheat biophysical variables from sentinel-2. *Remote Sensing* 11, 481.
- Upreti, D., Pignatti, S., Pascucci, S., Tolomio, M., Huang, W., Casa, R., 2020. Bayesian Calibration of the Aquacrop-OS Model for Durum Wheat by Assimilation of Canopy Cover Retrieved from VENμS Satellite Data. *Remote Sensing* 12, 2666.
- Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., Moreno, J., 2018. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surveys in Geophysics* 1–41.
- Vrugt, J.A., 2016. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software* 75, 273–316.
- Vrugt, J.A., Bouten, W., Gupta, H.V., Sorooshian, S., 2002. Toward improved identifiability of hydrologic model parameters: The information content of experimental data. *Water Resources Research* 38, 48–61. <https://doi.org/10.1029/2001WR001118>
- XING, H., XU, X., LI, Z., CHEN, Y., FENG, H., YANG, G., CHEN, Z., 2017. Global sensitivity analysis of the AquaCrop model for winter wheat under different water treatments based on the extended Fourier amplitude sensitivity test. *Journal of Integrative Agriculture* 16, 2444–2458.
- Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. *Journal of Hydrology* 204, 83–97. [https://doi.org/10.1016/S0022-1694\(97\)00107-8](https://doi.org/10.1016/S0022-1694(97)00107-8)
- Zaks, D.P., Kucharik, C.J., 2011. Data and monitoring needs for a more ecological agriculture. *Environmental Research Letters* 6, 014017.

CHAPTER 1

1. CHAPTER 1. A Comparison of Hybrid Machine Learning Algorithms for the Retrieval of Wheat Biophysical Variables from Sentinel-2

(Published in Remote Sensing, 2019, 11 (5), 481)

Deepak Upreti ¹, Wenjiang Huang ², Weiping Kong ^{2,3}, Simone Pascucci ⁴, Stefano Pignatti ⁴, Xianfeng Zhou⁵, Huichun Ye ², and Raffaele Casa ^{1,*}

¹ Università della Tuscia, DAFNE, Via San Camillo de Lellis, 01100, Viterbo (Italy); dupreti@unitus.it

² Key laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, 100094 (China); huangwj@radi.ac.cn, kongwp@radi.ac.cn, yehc@radi.ac.cn

³ Key Laboratory of Quantitative Remote Sensing Information Technology, Academy of Opto-Electronics, Chinese Academy of Sciences, Beijing 100094 (China);

⁴ Consiglio Nazionale delle Ricerche, Institute of Methodologies for Environmental Analysis (CNR, IMAA), Via del Fosso del Cavaliere, 100, 00133, Rome (Italy); stefano.pignatti@cnr.it, simone.pascucci@cnr.it

⁵ College of Life Information Science and Instrument Engineering, Hangzhou Dianzi University, Hangzhou 310018 (China); zhouxu@radi.ac.cn

* Correspondence: rcasa@unitus.it; Tel.: +39 357555

Received: 8 February 2019; Accepted: 21 February 2019; Published: 26 February 2019

Abstract: This study focuses on the comparison of hybrid methods of estimation of biophysical variables Leaf Area Index (LAI), Leaf Chlorophyll Content (LCC), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Fraction of Vegetation Cover (FVC) and Canopy Chlorophyll Content (CCC), from Sentinel-2 satellite data. Different machine learning algorithms were trained with simulated spectra generated by the physically based radiative transfer model PROSAIL and subsequently applied to Sentinel-2 reflectance spectra. The algorithms were assessed against a standard operational approach, i.e. that implemented in the ESA SNAP toolbox, based on neural networks. Since kernel-based algorithms have a heavy computational cost when trained with large datasets, an active learning (AL) strategy was explored to try to alleviate this issue. Validation was carried out using ground data from two study sites, one in Shunyi (China) and the other in Maccarese (Italy). In general, the performance of the algorithms was consistent for the two study sites, though a different level of accuracy was found between the two sites, possibly due to slightly different ground sampling protocols and the range and variability of the values of the biophysical variables in the two ground datasets. For LAI estimation the best ground validation results were obtained for both sites using Least Square Linear Regression (LSLR) and Partial Least Square Regression, with the best performances values of R^2 of 0.78, RMSE of $0.68 \text{ m}^2 \text{ m}^{-2}$ and RRMSE of 19.48% obtained in the Maccarese site with LSLR. The best results for LCC were obtained using RFTB and BagT with the best performances obtained in Maccarese using RFTB ($R^2 = 0.26$, RMSE = $8.88 \mu\text{g cm}^{-2}$, RRMSE = 17.43%). GPR was the best algorithm for all variables only in the cross-validation phase, but not in the ground validation, where it ranked as the best only for FVC in Maccarese ($R^2 = 0.90$, RMSE = 0.08, RRMSE = 9.86%). It was found that the AL strategy was more efficient than the random selection of samples for training the GPR algorithm.

Keywords: LAI, LCC, FAPAR, FVC, CCC, PROSAIL, GPR, Machine Learning, Active Learning

1. Introduction

Accurate quantitative estimation of biophysical variables is of crucial importance for different agricultural and ecological applications. Such variables include Leaf Area Index (LAI), Leaf Chlorophyll Content (LCC), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Fractional Vegetation Cover (FVC) and Canopy Chlorophyll Content (CCC). Their knowledge is valuable e.g. for precision agriculture (Zhang and Kovacs, 2012), crop traits monitoring (Jay et al., 2017) and improved yield prediction and reduction of fertilizer usage (Goffart et al.,

2008; Cilia et al., 2014). The retrieval of vegetation biophysical variables from multispectral optical satellite data has been performed by many researchers in the last decades (Jay et al., 2017; Verrelst et al., 2015b). In time, there have been significant developments in the algorithms used for such tasks, with a shift from empirical to more physically based approaches. Starting from simple parametric regressions between vegetation indices and biophysical variables such as LAI and LCC (Jordan, 1969; Rouse et al., 1974; Rondeaux et al., 1996), the current state of the art relies on hybrid methods exploiting at the same time Radiative Transfer Models (RTM) and non-linear non-parametric regression algorithms (Pasolli et al., 2012; Verrelst et al., 2015b, 2016a). The methods based on the relationship of e.g., vegetation indices and biophysical variables by the means of fitting functions, typically use the information provided by two or a few spectral bands. This limits the strength of such methods in today's scenario where tens or even hundreds of spectral bands are available respectively in current super-spectral (Verrelst et al., 2015a) or forthcoming hyperspectral (Guanter et al., 2015; Pignatti et al., 2015) spaceborne sensors.

On the other side, many of the approaches that were more recently developed, for the retrieval of the biophysical variables, exploit Machine Learning Regression Algorithms (MLRA). These algorithms gained wide popularity since they allow the use of more complex models, by requiring only once a time-consuming training phase, but allowing their fast application any time thereafter for the retrieval (Durbha et al., 2007; Lázaro-Gredilla et al., 2014). Another advantage of these algorithms is the possibility of training them with full spectral information, thus overcoming issues of band selection or transformation, as in the case of parametric regression methods (Verrelst et al., 2018). These algorithms are adaptive and have the ability to cope with the strong non-linearity inherent in the remote sensing data (Knudby et al., 2010; Verrelst et al., 2012). Despite their advantages, these algorithms are computationally demanding, so for some MLRA it is difficult to carry out a training phase with a large number of samples, though there are some that perform well also when trained with small datasets (Wang et al., 2016). Some of the MLRA are also considered as black boxes, e.g. Neural Networks (NN) (Benítez et al., 1997; Verrelst et al., 2015a), in which no insight is given about the physical processes linking the spectral reflectance with the biophysical variables. Others, such as Partial Least Squares Regression (PLSR) or Gaussian Processes Regression (GPR), can instead provide some information on what spectral bands are more relevant. In hybrid methods, MLRA can be trained using physically based modelling approaches, that describe the transfer and interaction of radiation inside the canopy based on physical laws, thus providing explicit relations between the biophysical variables and canopy spectral reflectance. When used in the direct mode, biophysical variables are used as input to the physical based RTM, which in turn simulates the Top of Canopy (TOC) spectral reflectance [9]. Using turbid-medium RTMs it is easy to generate in a short time a large simulated database, with realistic ranges of variation of biophysical variables and their corresponding simulated spectra, within the limits of the assumptions of the RTM on canopy properties. These pairs of biophysical variables and reflectance spectra can be used either in numerical optimization or in Look Up Table (LUT) retrieval methods, by minimizing a cost function expressing the difference between observed (e.g., from satellite data) and RTM model simulated reflectance and looking up the corresponding values of the biophysical variables (Rivera et al., 2013). For operational applications, due to the pixel-by-pixel calculations, these algorithms are computationally very demanding. These methods are also strongly affected by noise and measurement uncertainties (Verrelst et al., 2015a). As inversion methods of physically based RTMs, they suffer from the limitation of the ill-posed nature of model inversion (Combal et al., 2003; Atzberger, 2004), by which different combinations of canopy variables lead to local minima having similar reflectance spectra (Houborg and Boegh, 2008).

In the hybrid methods of retrieval, the database obtained from RTM simulations can be used to train MLRA which are able to establish complex non-linear non-parametric models linking the biophysical variables and the spectral reflectance (Pasolli et al., 2012; Verrelst et al., 2015a). Hybrid methods have found widespread application and have reached the operational stage, in particular with the application of NN (Pozdnyakov et al., 2005; Schiller and Doerffer, 2005; Verger et al., 2008) trained with the RTM PROSAIL (Verger et al., 2011; Weiss and Baret, 2016), such as the algorithm implemented in Sentinel Application Platform (SNAP) biophysical processor tool (Weiss and Baret, 2016), developed by European Space Agency (ESA). LAI retrieval using NN trained with PROSAIL is also carried out operationally for Cyclopes and MODIS products (Chai et al., 2008). However, NN training is a delicate

phase and requires the tuning of multiple parameters, which greatly impacts the robustness of the approach (Verrelst et al., 2012). Thus, in recent years, increasing attention has been paid to alternatives to NN that are simpler to train and have better potential for retrieval accuracy, such as kernel-based methods (Camps-Valls and Bruzzone, 2009). These methods solve nonlinear regression problems by transferring the data to a higher dimensional space by the means of a kernel function (Verrelst et al., 2018). Kernel-based algorithms have been suggested to offer some advantages in comparison to NN. Kernel Ridge Regression (KRR) has proven to be simple for training and to yield competitive accuracy (Verrelst et al., 2012). Gaussian Process Regression (GPR) is partially transparent as compared to the black box nature of NN. It allows to use different kernel functions, ranging from simple to very complex, and it also provides uncertainty estimates with the mean value of prediction (Verrelst et al., 2013a, 2016a). GPR and KRR have been used to estimate successfully LAI and LCC (Verrelst et al., 2013b, 2016a). Unfortunately, some kernel-based algorithms, such as GPR and KRR are computationally very expensive if trained on large sets of simulations (Pasolli et al., 2012; Verrelst et al., 2016a). In order to have a general-purpose database, including a wide range of vegetation types, generally a huge number of simulations are performed (Verrelst et al., 2016a). However, not all pairs of the reflectance and corresponding biophysical variables will be relevant. Few attempts have been made by the researchers to optimize the simulations that are generated by the physically based models (Verrelst et al., 2016b, 2016a). Active Learning (AL) has been proposed as a useful strategy to reduce the size of the RTM generated database, to make the training of the kernel-based algorithms such as GPR more feasible [10]. AL is a sub-field of machine learning, also called optimal experimental design in statistics (Settles, 2009). It initially starts with a small subset of the samples and then, based on query strategies using either uncertainty or diversity measurement criteria, adds iteratively new samples to the initial training set of samples. In this way, the most informative samples in a dataset are selected, avoiding redundancy (Verrelst et al., 2016a). AL techniques have a great potential to optimally sample sets generated by RTM.

Very few studies (Pasolli et al., 2012; Verrelst et al., 2016a) have reported the potential of kernel-based methods using AL techniques in relation to the retrieval of biophysical variables such as LAI, LCC, FVC, FAPAR and CCC. In particular, although the estimation of these variables from hyperspectral data has been demonstrated (Verger et al., 2011), the suitability of multispectral satellites has yet to be fully proven, especially for biochemical variables such as LCC. The European Space Agency Sentinel-2 mission has been shown to provide data of a high radiometric quality (Gascon et al., 2017) and has a higher revisit frequency than what planned for hyperspectral satellites in the near future. Because of the availability of a larger number of spectral bands than other multispectral satellites, and of the inclusion of the red edge region of the spectrum, Sentinel-2 has a good potential for the estimation of these variables [9].

This work explores the application of kernel-based GPR using AL for biophysical variables retrieval from Sentinel-2, comparing the potential of this algorithm with other MLRAs and in particular of the version implemented operationally in SNAP.

The main objectives of this study are thus: (1) to compare the performances of different MLRAs, in particular with respect to the algorithm based on NN implemented in the biophysical processor tool of the ESA SNAP toolbox, by using the same database of PROSAIL simulations (Weiss and Baret, 2016); (2) to assess the accuracy of estimation of the biophysical variables LAI, LCC, FAPAR, CCC and FVC in the wheat crop, for kernel-based (GPR) and non-kernel-based MLRA hybrid methods of retrieval; (3) to explore the feasibility and potential of the use of AL strategies to optimally sample redundant PROSAIL simulations, to minimize computational time and complexity and allow the use of computationally demanding MLRAs (such as GPR) in hybrid methods.

2. Materials and Methods

The methodology adopted in this work (Figure 1), for the comparison of hybrid methods of biophysical variables retrieval, employed simulations carried out with the model PROSAIL (Jacquemoud et al., 2009) for training MLRA and finally applying them to Sentinel-2 data for assessment with ground measurements collected in two study areas,

respectively in Italy and China. The procedure involved the following steps, illustrated in detail in the following sections: 1) generation of simulations with the model PROSAIL; 2) training MLRAs with model simulations and cross-validation; 3) Sentinel-2 data pre-processing; 4) validation using ground data.

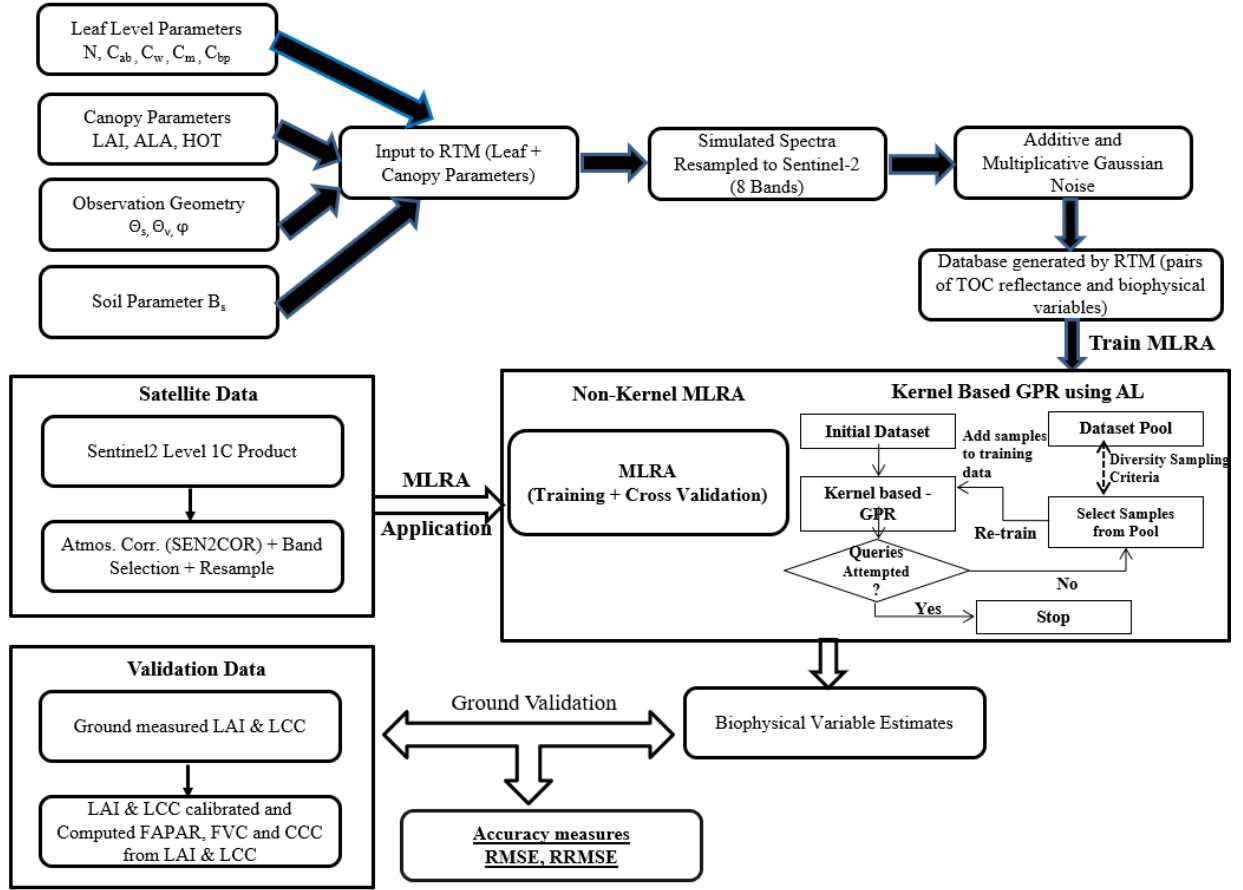


Figure 1. Flow-chart of the methodology used for the present study. Retrieval of biophysical variables: Leaf area index (LAI), Leaf chlorophyll content (LCC), Fraction of vegetation cover (FVC), Fraction of absorbed photosynthetically active radiation (FAPAR) and Canopy chlorophyll content (CCC). Radiative Transfer Model (RTM) parameters abbreviations are reported in Table 1. The RTM was used to train Machine Learning Algorithms (MLRA) including Gaussian Process Regression (GPR) using Active Learning (AL). Accuracy was assessed using the root mean squared error (RMSE) and relative RMSE (RRMSE)

2.1. Generation of PROSAIL Simulations

PROSAIL (Jacquemoud et al., 2009) is a widely used RTM for generating realizations of TOC canopy reflectances, by considering measured biophysical variables of the crop of interest. It is a combination of the leaf level model PROSPECT (Feret et al., 2008) and canopy level model SAIL (Verhoef, 1984). It has been inverted over a large range of vegetation canopies (Meroni et al., 2004; Richter et al., 2009, 2011; Palacharla et al., 2011; Le Maire et al., 2011; Si et al., 2012; Liang et al., 2016) with varying degrees of success. Reflectance and transmittance of a leaf from the PROSPECT model is used as an input for the SAIL model with the soil optical properties and illumination and observation geometry.

The benchmark against which the different hybrid methods were assessed in this study was the NN algorithm implemented in the ESA SNAP Sentinel-2 Toolbox, as biophysical processor. For this reason, PROSAIL simulations

were generated using the same input parameter values and configuration as used for SNAP (M.Weiss personal communication), by following the Sentinel-2 Toolbox level 2 products ATBD (Weiss and Baret, 2016), with the exception of the geometry of observation and illumination which was set only for the conditions of our validation dataset (section 2.3). The PROSAIL parameters were sampled using a fully orthogonal experimental plan (Bacour et al., 2002), by using a component of the BV-NNET (Biophysical Variables Neural NETwork) tool developed by (Baret et al., 2003). This procedure consists of identifying classes of values for each variable. Then all the combinations of classes are sampled once. Finally, the actual values of each variable are randomly drawn within the range of variation defined by the corresponding class, according to the distribution law specified for the variable considered. This process allows to take into account all the interactions among parameters, while having the range of variation for each variable densely and near randomly populated (Weiss and Baret, 2016).

Table 1 reports the parameter ranges, statistical distribution and number of classes of the PROSAIL model that were used in the present work, which are the same as those employed for training NNs in SNAP (M. Weiss personal communication). A total of 41472 simulations were generated, representing an extensive range of vegetation properties.

PROSAIL top of the canopy (TOC) full spectra (at 1 nm resolution) were then resampled, using Sentinel-2 spectral response functions, to the same 8 bands as used in the Sentinel-2 level 2 products ATBD (Weiss and Baret, 2016), i.e. bands 3 to 7, 8a, 11 and 12, excluding the bands most affected by atmospheric attenuation. The spectral resampling was carried out according to Eq.1:

$$\rho = \sum_i^N \left[\frac{\beta(\lambda)}{\sum_i^N \beta(\lambda)} \right] * \rho(\lambda) \quad (1)$$

where ρ and $\rho(\lambda)$ are respectively the resampled Sentinel-2 reflectance and the PROSAIL reflectance at wavelength λ . $\beta(\lambda)$ represents the weight of the band's spectral response function of the Sentinel-2 MSI sensor, whereas i and N are respectively the minimum and maximum wavelengths of the Sentinel-2 band. Similarly, a Gaussian noise model was used to better describe actual Sentinel-2 characteristics as well as the ability of the RTM used to represent actual reflectance. The noise is computed as follows (Weiss and Baret, 2016):

$$R^*(\lambda) = R(\lambda)(1 + MD(\lambda) + MI)/100 + AD(\lambda) + AI \quad (2)$$

where $R(\lambda)$ and $R^*(\lambda)$ are the raw simulated reflectance and reflectance contaminated with noise respectively. MD is the multiplicative wavelength dependent noise and MI is the multiplicative wavelength independent noise. AD and AI are the additive wavelength dependent noise and independent noise respectively. A value of 0.01 was used for AD and AI and a value of 2 % was used for MD and MI for all the bands.

Table 1. Inputs parameter values ranges, number of classes and statistical distribution used for generating the training database with the PROSAIL model

	Variable	Minimum	Maximum	Mode	Std	Nb_Classes	Law
Canopy	LAI	0	15	2	2	6	log_normal
	ALA (°)	30	80	60	20	3	gauss
	HsD	0.1	0.5	0.2	0.5	1	gauss
	N	1.2	1.8	1.5	0.3	3	gauss

Leaf	Cab ($\mu\text{g. cm}^{-2}$)	20	90	45	30	4	gauss
	Cdm (g c.m^{-2})	0.003	0.011	0.005	0.005	4	gauss
	Cw_Rel	0.6	0.85	0.75	0.08	4	uni
	Cbp	0	2	0	0.3	3	gauss
Soil	Bs	0.5	3.5	1.2	2	4	gauss

2.2. Training Machine Learning Algorithms

Some of the most widely used parametric and non-parametric regression algorithms were trained (or calibrated) with the simulations generated by the PROSAIL model, including Linear Regression, Non-linear Regression, Hierarchical Tree based approaches and kernel-based algorithms (Table 2).

The non-kernel based regression methods evaluated in this study were: Least Square Linear regression (LSLR), Partial Least Square Regression (PLSR), NN, Regression Trees (RegT), Bagging Trees (BagT), Boosting Trees (BooT), Random Forest Fit Ensemble (RFFE) and Random Forest Tree Bagger (RFTB). We also assessed the kernel-based MLRA GPR, by combining it with an AL strategy. A brief description and the references where these methods are described in detail is reported in Table 2.

Table 2. Summary of the algorithms compared in this study

Algorithm	Brief description	References
LSLR	It relies on the linear relationship between explanatory variables, parameters and the output variable.	(Bajwa et al., 2008)
PLSR	It performs the regression on the projections generated using partial least square approach.	(Geladi and Kowalski, 1986)
NN	It is an architecture composed of multiple layers of artificial neurons. For this study, hyperbolic tangent function was used and NN architecture was optimized with Levenberg-Marquardt learning algorithm with a loss function.	(Hagan and Menhaj, 1994; Verrelst et al., 2015b, 2012)
RegT	RegT, was initially a classical approach of decision trees, in which sorting and grouping methods have been added to model non-linear relationships.	(Breiman et al., 1993)
BooT	It works by sequentially applying a classification algorithm to reweighted versions of the training data and then taking a weighted majority vote of the sequence of the classifiers thus produced.	(Friedman et al., 2000; Verrelst et al., 2015b)
BagT	BagT is a method for generating multiple versions of a predictor and using them to obtain an aggregated predictor.	(Breiman, 1996)
RFFE	RFFE is an ensemble to decision tree-based approach for improving the prediction accuracy, such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest.	(Breiman, 2001)
RFTB	RFTB combines the bagging approach with the generalized approach of Random Forest of aggregating multiple decision trees.	(Breiman, 2001, 1996)
GPR	These are non-parametric kernel-based probabilistic models, based on Gaussian Processes. They are described as a collection of random variables, any finite number of which have a joint	(Williams and Rasmussen, 2006, 1996)

The simulation dataset of 41472 spectra was used for training the different algorithms. Because of the long processing time required for the training, and of the necessity of having replicates in order to carry out a statistical analysis of the results, 10 subsets of 2500 samples were extracted randomly (with replacement) from the PROSAIL simulated spectra dataset. All the MLRA were trained (and subsequently validated) ten times with the subsets of 2500 simulations each, except for GPR. Since the latter is more computationally demanding, we adopted an AL procedure for selectively further reducing the size of the training set.

AL is a technique which starts from a small set of initial training data and optimally selects samples from a larger data pool, based on different criteria, such as uncertainty or diversity indicators. Samples from the data pool are picked up and added to the initial training dataset, based on diversity criteria. Selecting samples by diversity ensures that added samples are dissimilar from those already accounted for. To identify the best method for selecting the samples from the data pool, we have used three different diversity-based criteria which performed as the best in a previous study [10]. These are Angle Based diversity (ABD), Euclidean distance-based diversity (EBD) and Cluster based diversity (CBD). In the implementation used in this work, these criteria are applied also to the label information (i.e., the biophysical variables), since this is available in the data pool (Verreslt J., personal communication). Each process in AL was started with 50 samples as initial training data and the remaining 2450 samples in a data pool. Subsequently, a maximum of 50 samples were added per iteration, with stopping criteria of 100 iterations or a RMSE decrease lower than 50%. Only when performance is improved the samples are added. The whole AL process was repeated 10 times starting from the different subset of 2500 records each.

In the comparison of the different diversity-based criteria, the full training set of 2500 samples was also used as a reference. Following this assessment, since EBD was the best performing method (see Results section), we only used EBD to further train and validate GPR with AL.

For all the algorithms described above, a cross-validation was performed with a k-fold strategy, where k is the number of folds with a value of 10.

The implementation of all the MLRAs was performed with the Matlab ARTMO Toolbox version 3.24 (Verreslt et al., 2011).

2.3. Validation with Sentinel-2 data and ground measurements

Once the regression algorithms described in the previous section had been trained using the PROSAIL simulations, they were applied to Sentinel-2 data acquired over two sites, Maccarese (Central Italy) and Shunyi (China). For both sites, cloud-free Sentinel-2 images were selected for the closest date to ground measurement collection (Table 3). Only the same 8 Sentinel-2 bands that had been used to resample PROSAIL simulations, i.e. bands 3 to 7, 8a, 11 and 12, were used. This band selection had been established as optimal for biophysical variables retrieval, by reducing the noise introduced by atmospheric effects (Camacho et al., 2013; Weiss and Baret, 2016). The images were initially processed for atmospheric correction with the Sen2cor tool implemented in SNAP (version 6.0) to obtain level 2A TOC reflectance and then resampled to a 10 m pixel size.

Maccarese (41.833° lat. N, 12.217° long. E, alt. 8 m a.s.l.) is located on the West coast of Central Italy, near Rome. It is a private farm of 3200 ha in a flat area with large fields. Field campaigns to measure biophysical variables on the durum wheat (*Triticum durum* Desf.) crop were carried out for this location from January 2018 to April 2018, for different growing conditions of the crop of interest, at dates close to Sentinel-2 acquisitions. The variables were sampled according to an elementary sampling unit (ESU) scheme, to capture the variability within and among different

fields. Each ESU consisted of a quadrat of 20 m by 20 m size, to easily accommodate the Sentinel-2 10 m pixel resolution. A total of 15 ESU, placed at different locations were employed at different sampling dates. Each of the ESU contained nine points, where LAI and LCC measurements were collected. LAI was measured using the LAI-2000 or the LAI-2200C Plant Canopy Analyzer (LI-COR, Lincoln, NE, USA). Since LAI 2000/2200C measurements were not always acquired under diffuse light conditions, data were pre-processed by applying the recommended scattering correction procedure (Li-Cor, 1992). Four readings were taken above the canopy, with and without diffused cap and six readings below the canopy with 90° cap. While processing for scattering corrections five angles of directions 7°, 23°, 38°, 53° and 68° were considered, i.e., LAI measurements including the extreme rings.

Table 3. Sentinel-2 acquisitions and ground measurements dates

Site	Date - Ground Measurements	Date - S2 Acquisition	Difference (Days)
Maccarese, Italy	31-Jan-18	29-Jan-18	2
	16-Feb-18	13-Feb-18	3
	6-Apr-18	6-Apr-18	0
	20-Apr-18	19-Apr-18	1
Shunyi, China	07 - 08 April 2016	10-Apr-16	2
	20 - 21 April 2016	23-Apr-16	2
	03 - 05 May 2016	3-May-16	0
	18 - 19 May 2016	14-May-16	4

Chlorophyll measurements were obtained using the Force-A Dualex leaf clip reader on the top-most leaves. The calibration of Dualex on durum wheat had been previously performed (Casa et al., 2015) with an estimation accuracy with root mean square error (RMSE) values ranging between 7 and 11 $\mu\text{g cm}^{-2}$. The following equation (Casa et al., 2015) was used to recalibrate and compute LCC from the Dualex measured readings

$$LCC = -3.12 + 1.55 * (\text{Dualex}) \quad (3)$$

LAI 2000/2200C measurements were also used to derive other biophysical variables, i.e. FVC and FAPAR. FVC was obtained from the LAI values measured on the ground by using the following equation (Nielsen et al., 2012):

$$FVC = (1 - e^{-0.43 \cdot \text{LAI}})^{0.52} \quad (4)$$

Previous studies have shown that the instantaneous FAPAR value at 10:00 (or 14:00) solar time is very close to the daily integrated value under clear sky conditions (Baret et al., 2003; Weiss and Baret, 2016). A Matlab script was written to compute FAPAR from the raw LAI-2200 data measured on the ground. The script extracts time dimensions, angles and gaps fraction data from raw data files, computes day of the year and calculates FAPAR at 10:00 hour solar time. CCC has been estimated by multiplying LAI with chlorophyll content at leaf level.

A dataset of ground measurements corresponding to Sentinel-2 data was also collected in Shunyi district (40.130 lat. N, 116.654 long. E), Beijing, in China. Ground campaigns were carried out in the period between April 2016 and May 2016 on the winter wheat (*Triticum aestivum* L.) crop. LAI was measured using Licor LAI 2000/2200C and chlorophyll was measured using Force-A Dualux readings and calibrated using laboratory analysis. From April 2016 to May 2016, at four different dates, with 24 observations per date, LAI and LCC were collected in different wheat fields. FVC was computed as mentioned using Eq. 4. Due to the unavailability of raw LAI data, FAPAR was not computed for this test site.

The ground validation dataset was thus composed of 45 samples for Maccarese (each one the average of an ESU) and 96 samples for Shunyi.

2.4. Comparative assessment of MLRAs

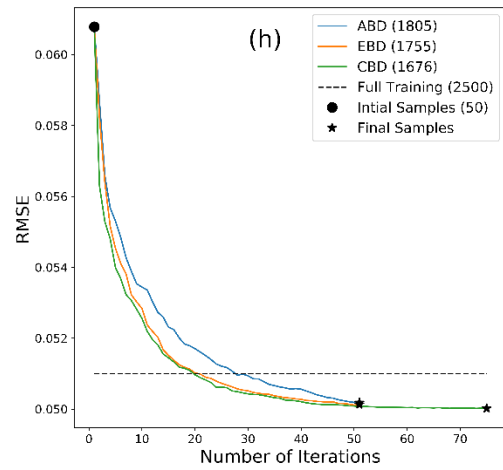
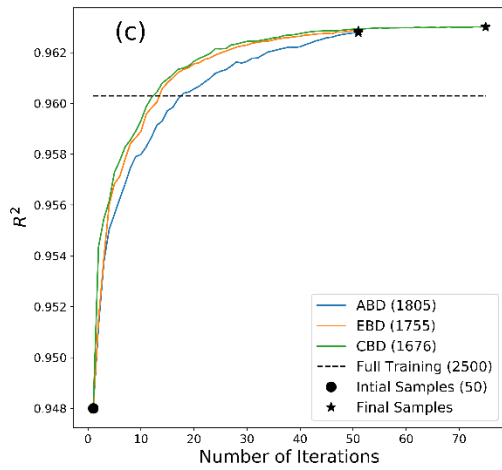
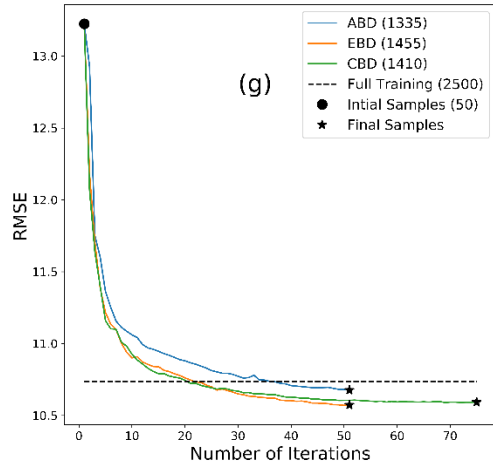
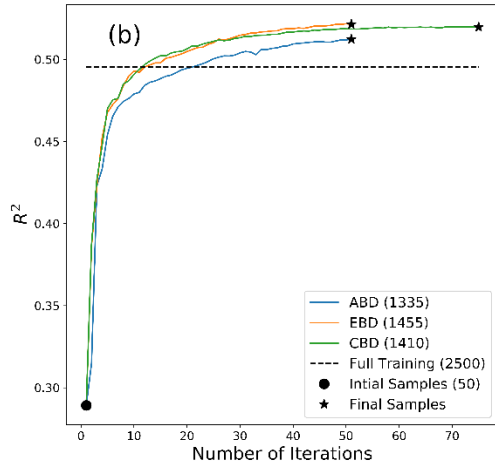
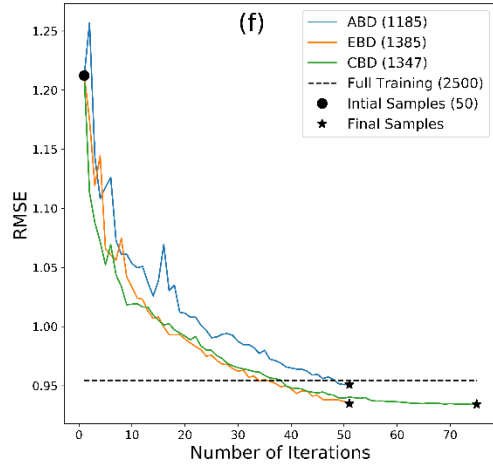
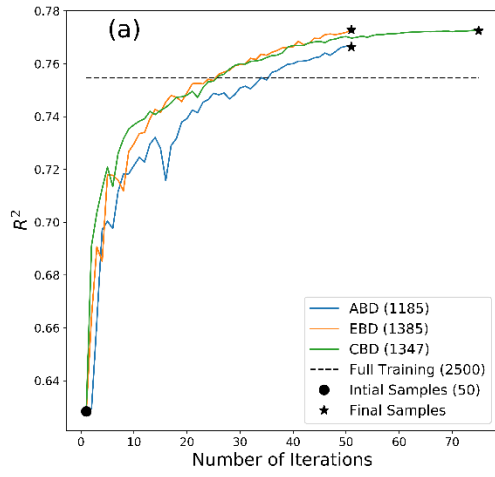
We used r^2 , root mean square error (RMSE) and relative RMSE (RRMSE %) to assess the accuracy of the retrievals for both the k-fold cross validation and the ground validation. The RRMSE was used to compare the performances across different MLRAs and variables (Verrelst et al., 2012). In general, the performances of the models were estimated by comparing the differences among RRMSE of the estimated and measured values of different biophysical variables. Lower RMSE values correspond to higher accuracy of the algorithms for the retrieval of biophysical variables from Sentinel-2.

To assess the presence of statistically significant differences among the performances of the models, we applied a Friedman test to the computed metrics (Das et al., 2018). Whenever the Friedman test revealed significant differences ($p < 0.001$) among the metric means, which indicated the presence of significant differences among the models, a Friedman's aligned ranks post-hoc test, followed by Benjamini / Hochberg (non-negative) adjusted for p-values, was performed for multiple pairwise comparisons.

3. Results

The full simulation dataset generated with PROSAIL was used as a source of subsets for training the different MLAs, as described in the methods section, except for GPR for which a data reduction procedure based on AL was employed. Preliminarily, for GPR, we compared three different diversity-based criteria used to perform AL, with respect to the full training with 2500 samples. Figure 2 shows the comparison of the different methods of sample selection in terms of R^2 and RMSE with respect to the number of iterations, for all the variables, applied to the Maccarese dataset. Only when the performance was improved the samples were added. This explains why fewer samples were added than potentially possible, i.e. less than 50 samples were added for each iteration (e.g. in Figure 2a final sample size for CBD was of 1347 at 75 iterations).

Euclidean distance-based diversity metric EBD surpasses ABD and CBD, achieving higher r^2 and lower RMSE values, i.e. higher accuracy and lower error, with a lower number of samples and iterations. EBD was thus subsequently used in all further applications of AL in the present work. It should be noted that the training of GPR with the full set of 2500 samples, i.e. without AL, provided worse performance than AL in all cases except for FAPAR, despite using a larger dataset for training, revealing some redundancy in the full set.



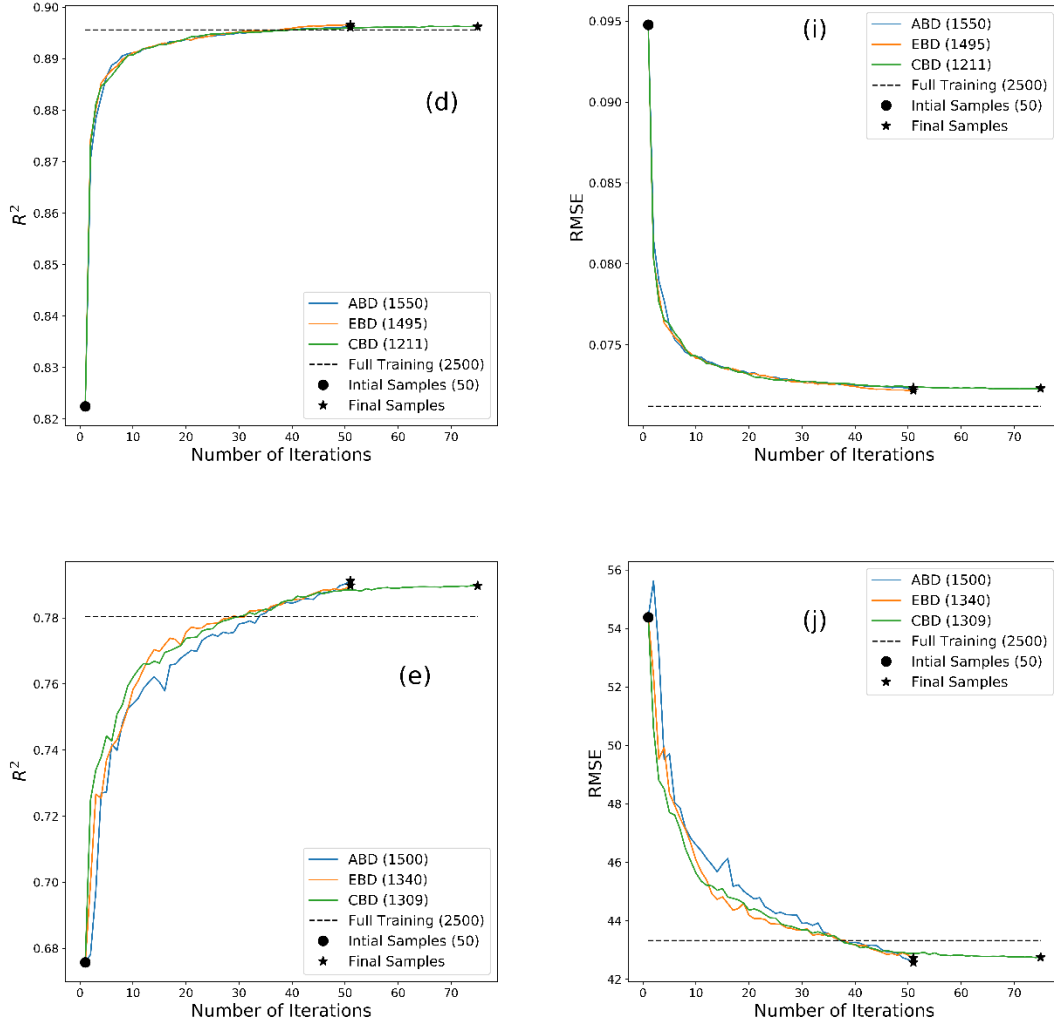


Figure 2. Comparison of three diversity criteria of data pool samples selection in AL, i.e. Angle Based diversity (ABD), Euclidean distance-based diversity (EBD) and Cluster based diversity (CBD) for the different biophysical variables for the Maccarese dataset. a), f) LAI; b), g) LCC; c), h) FVC; d), i) FAPAR; e), j) CCC. The values in parentheses in the legend indicate the final size of the training dataset after the AL process. The star symbol indicates the final samples after reaching the AL stopping criteria (see text). The horizontal dashed line indicates the performance of the GPR training with the full dataset of 2500 samples without AL.

In the comparison of all the algorithms testes, cross-validation metrics give an indication of their performance, though the actual accuracy can be more realistically assessed using ground validation results. Tables 4 and 5 report the summary of RMSE estimates for both cross-validation and ground validation for Maccarese, and Shunyi datasets, respectively, for all the algorithms tested.

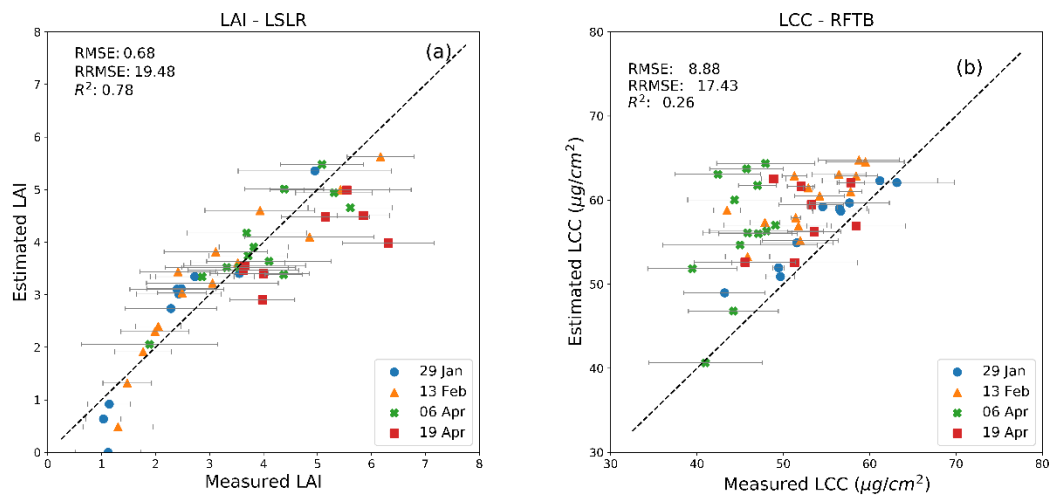
Table 4. Mean RMSE values of 10 replicates for the estimation of LAI, LCC, CCC, FVC and FAPAR using different machine learning algorithms (MLRA) for cross-validation (CV) and ground validation (GV), for the Maccarese test site. Abbreviations for the MLRA are reported in Table 2. Values sharing the same letters are not significantly different according to Friedman’s aligned ranks post-hoc test. The best results for each variable are highlighted in bold.

MLRA	LAI		LCC		CCC		FVC		FAPAR	
	CV	GV	CV	GV	CV	GV	CV	GV	CV	GV
BagT	0.99 ^b	1.01 ^d	11.4 ^c	8.90^a	45.83 ^{ab}	46.12 ^b	0.05 ^c	0.08^a	0.08 ^c	0.11 ^a
NN (ARTMO)	1.04 ^c	0.90 ^b	11.02 ^b	10.68 ^{bc}	46.53 ^b	52.1 ^c	0.05 ^b	0.09 ^{bc}	0.07 ^b	0.11 ^a
RFTB	1.00 ^b	0.93 ^{bc}	11.55 ^{cd}	8.88^a	46.47 ^b	44.71 ^b	0.06 ^d	0.08^a	0.08 ^c	0.10^a
PLSR	1.18 ^e	0.69^a	11.16 ^b	11.6 ^{cd}	53.25 ^c	40.44^a	0.06 ^e	0.13 ^d	0.09 ^e	0.12 ^b
LSLR	1.18 ^e	0.68^a	11.15 ^b	12.05 ^d	53.14 ^c	40.86^a	0.06 ^e	0.13 ^d	0.09 ^e	0.12 ^b
BooT	1.14 ^d	0.98 ^{cd}	11.61 ^d	10.94 ^{bc}	52.08 ^c	57.51 ^{cd}	0.06 ^f	0.09 ^b	0.08 ^d	0.11 ^a
RegT	1.30 ^f	1.17 ^e	14.99 ^e	11.63 ^{cd}	60.26 ^d	61.22 ^{de}	0.07 ^g	0.09 ^{bc}	0.10 ^f	0.12 ^b
RFFE	1.40 ^g	1.44 ^f	16.03 ^f	13.18 ^d	64.36 ^d	68.32 ^e	0.08 ^h	0.10 ^c	0.11 ^g	0.12 ^b
GPR + AL	0.94^a	1.31 ^f	10.55^a	10.22 ^b	43.81^a	54.56 ^c	0.05^a	0.08^a	0.07^a	0.11^a
NN (SNAP)		0.74		10.94		44.61		0.108		0.105

Table 5. Mean RMSE values of 10 replicates for the estimation of LAI, LCC, CCC, FVC and FAPAR using different machine learning algorithms (MLRA) for cross-validation (CV) and ground validation (GV), for the Shunyi test site. Abbreviations for the MLRA are reported in Table 2. Values sharing the same letters are not significantly different according to Friedman's aligned ranks post-hoc test. The best results for each variable are highlighted in bold.

MLRA	LAI		LCC		CCC		FVC	
	CV	GV	CV	GV	CV	GV	CV	GV
BagT	0.99 ^b	1.74 ^{cd}	11.67 ^d	17.28^a	46.12 ^b	56.72^a	0.05 ^c	0.19 ^c
NN(ARTMO)	1.02 ^b	1.42 ^b	11.27 ^b	24.41 ^c	46.52 ^b	98.14 ^e	0.05 ^b	0.17^a
RFTB	1.00 ^b	1.67 ^c	11.8 ^e	16.77^a	46.69 ^b	56.51^a	0.05 ^d	0.19 ^c
PLSR	1.18 ^d	1.01^a	11.43 ^c	24.77 ^c	52.94 ^c	68.11 ^c	0.06 ^f	0.18 ^b
LSLR	1.18 ^d	1.00^a	11.43 ^{bc}	25.98 ^{bc}	52.86 ^c	74.33 ^d	0.06 ^f	0.18 ^b
BooT	1.12 ^c	1.46 ^b	11.91 ^e	25.91 ^d	51.69 ^c	94.77 ^e	0.05 ^e	0.18 ^b
RegT	1.32 ^e	1.76 ^d	15.26 ^f	20.75 ^b	60.63 ^d	62.71 ^b	0.07 ^g	0.21 ^d
RFFE	1.42 ^e	2.14 ^e	16.36 ^g	25.62 ^b	64.29 ^d	76.96 ^d	0.07 ^h	0.21 ^d
GPR + AL	0.93^a	1.98 ^e	10.79^a	22.75 ^b	42.55^a	61.35 ^b	0.05^a	0.18 ^b
NN (SNAP)		1.16		23.61		123.53		0.18

For Maccaresse (Table 4), LSLR and PLSR were the best performing algorithms for LAI retrieval when validated with ground data, in terms of RMSE, with values of 0.68 ($R^2 = 0.78$ RRMSE = 19.48%) and 0.69 ($R^2 = 0.78$ RRMSE = 19.84%) respectively. These were also the fastest computing algorithms among all the MLRAs in terms of time required for training (data not shown). The cross-validation and ground validation results (Figure 3a) indicated that LSLR apparently did not show saturation, even up to LAI values of 5 or 6. LCC retrieval was best performed by RFTB (Figure 3b) and BagT, with RMSE values of $8.88 \mu\text{g cm}^{-2}$ ($R^2 = 0.26$ and RRMSE = 17.43%) and $8.90 \mu\text{g cm}^{-2}$ ($R^2 = 0.27$ and RRMSE = 17.46%). An overestimation of LCC was observed (Figure 3b). The lowest RMSE of 40.44 g cm^{-2} ($R^2 = 0.74$ RRMSE = 22.7%) (Figure 3e), was obtained for the retrieval of CCC with PLSR. The retrieval accuracy of LSLR was not significantly different from that of PLSR for CCC (RMSE = 40.86 g cm^{-2} , $R^2 = 0.74$ and RRMSE = 22.9%). Indeed, PLSR was not significantly different from LSLR in terms of RMSE for all the variables considered. The lowest RMSE for FVC was 0.08 ($R^2 = 0.9$ and RRMSE = 9.8%), obtained with the GPR algorithm (Figure 3c). RFTB and BagT were also not significantly different from GPR in terms of RMSE. FAPAR was best retrieved with a RMSE of 0.10 ($R^2 = 0.41$ and RRMSE = 12.06%) using the RFTB approach (Figure 3d), although GPR was not significantly different in terms of RMSE, which was only slightly higher. It was observed that, even in the best performing algorithms, an underestimation of low values of FVC and FAPAR was apparent (Figure 3c and 3d). It should also be noted that these latter variables show a smaller range of variation in the measured values as compared to the other biophysical variables. GPR was always the best performing algorithm in the cross-validation results, but this was not confirmed by the ground validation tests, with the exception of FVC and FAPAR estimation.



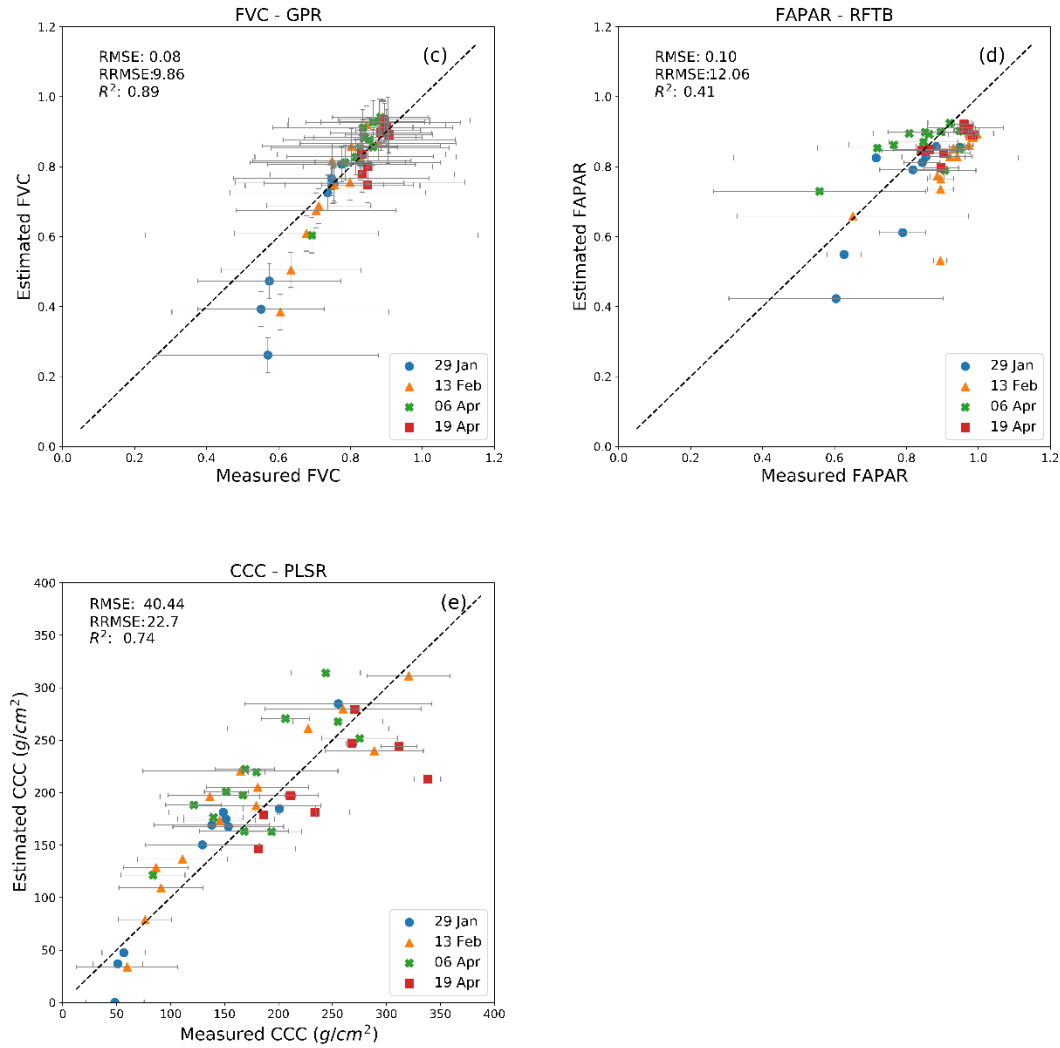


Figure 3. Best performing algorithms for the retrieval of (a) LAI, (b) LCC, (c) FVC (d) FAPAR and (e) CCC for the Maccaresse test site. The horizontal error bars represent the standard deviation of measurements inside each ESU. Vertical error bars represent uncertainty estimates using GPR for FVC (3c)

As can be observed from the horizontal error bars of Figure 3, considerable variability occurred among ground measurements inside ESUs, despite the visually apparent homogeneity of the crop canopy in the sampled area of each ESU.

For the Shunyi datasets (Table 5), the ground variability could not be reported for LAI, for which single point measurements were carried out, but an estimate of the variability for chlorophyll could be made since multiple measurements were available for each point (Figure 4b).

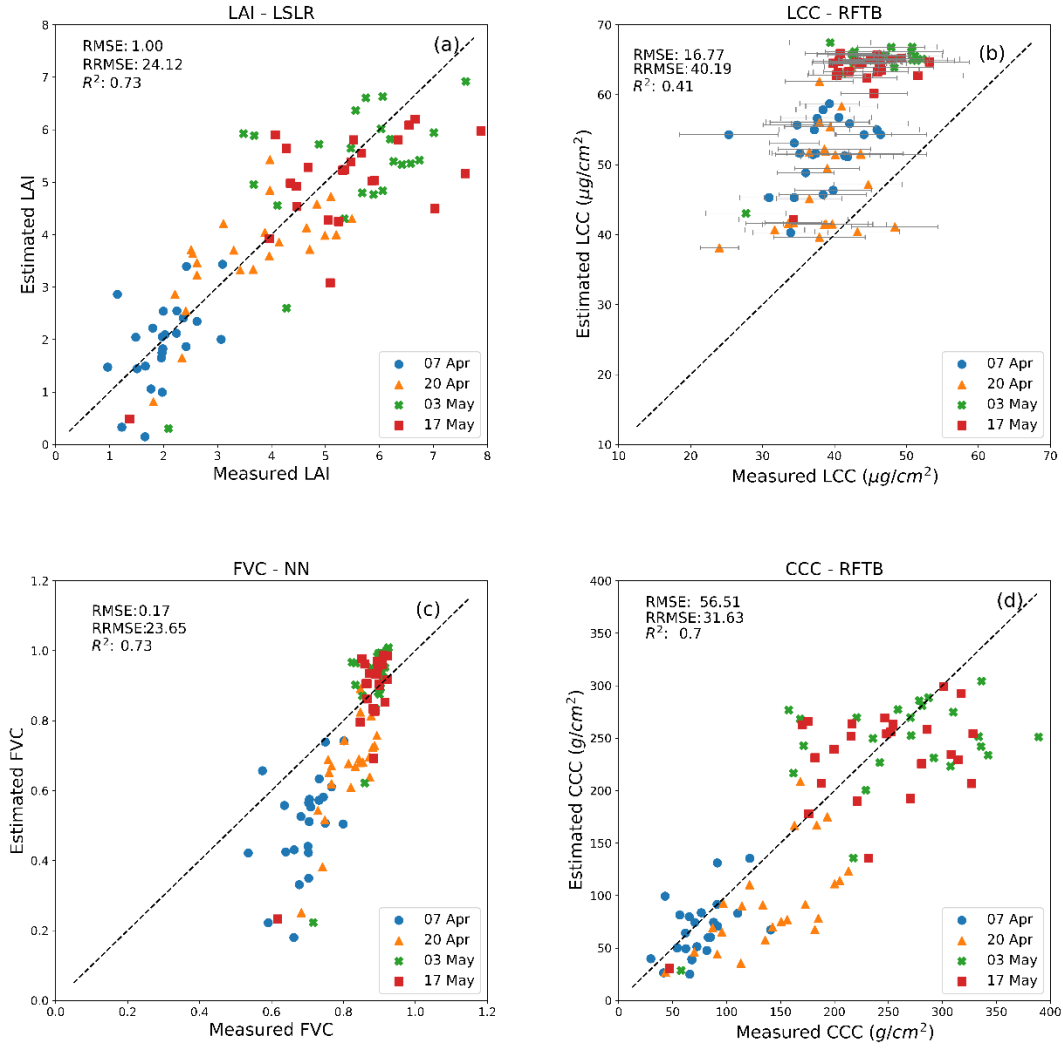
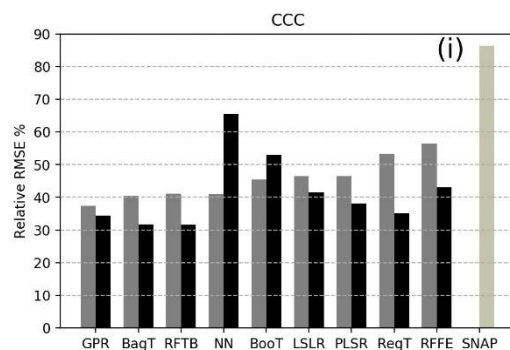
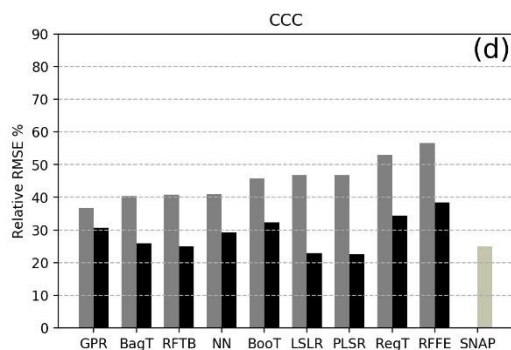
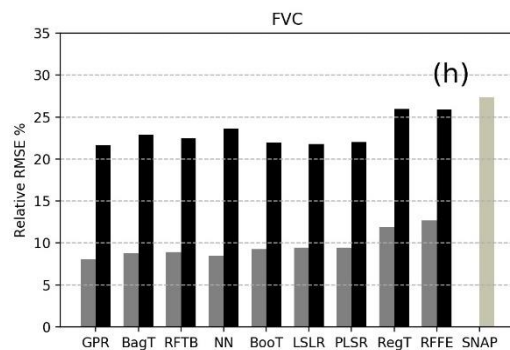
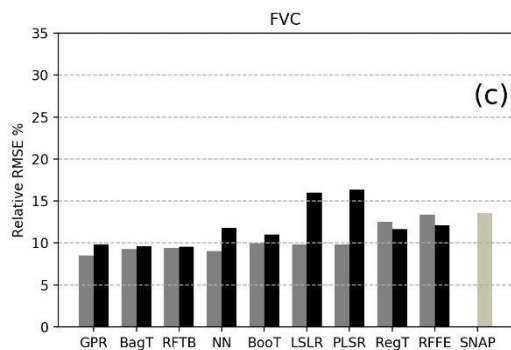
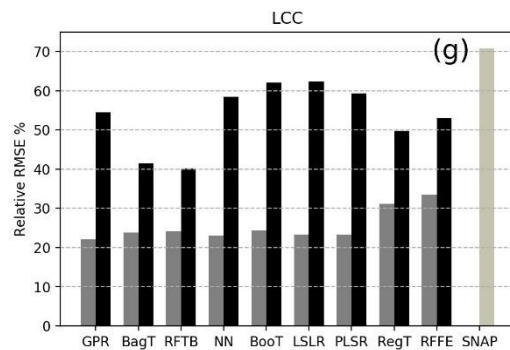
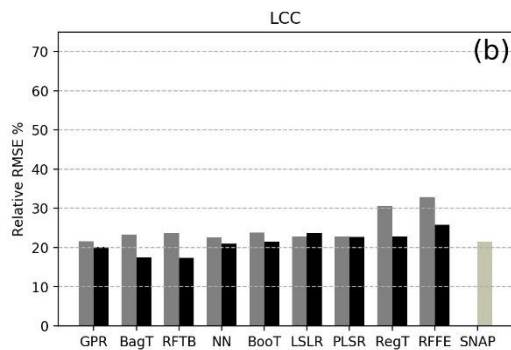
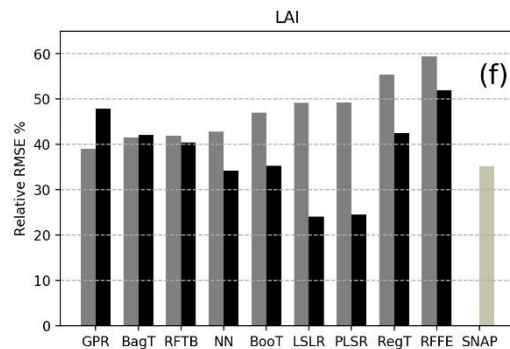
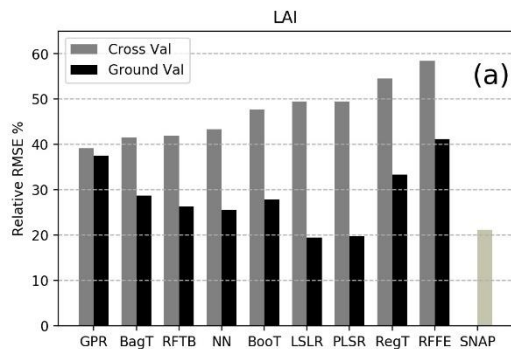


Figure 4. Best performing algorithms for (a) LAI, (b) LCC and (c) FVC d) CCC retrieval for the dataset of Shunyi, China. The horizontal error bars for LCC represent the standard deviation of single chlorophyll measurements.

The best accuracy for LAI retrieval for the Shunyi site was a RMSE of $1.00 \text{ m}^2 \text{ m}^{-2}$ ($R^2 = 0.73$ and $\text{RRMSE} = 24.12\%$) obtained with LSLR (Table 5), i.e. the same algorithm providing the best results also for Maccarese. PLSR had a similar performance as that of LSLR, with a $\text{RMSE} = 1.01 \text{ m}^2 \text{ m}^{-2}$ and ($R^2 = 0.72$ and $\text{RRMSE} = 24.50\%$), i.e. also consistently with the results of LAI retrieval for Maccarese. LCC was best retrieved with a RMSE value of $16.77 \mu\text{g cm}^{-2}$, ($R^2 = 0.41$ and $\text{RRMSE} = 40.19\%$) using RFTB (Table 5). Although RFTB, produced the lowest error for retrieving LCC, its retrieval using BagT approach was not significantly different, with $\text{RMSE} = 17.28 \mu\text{g cm}^{-2}$, $R^2 = 0.41$ and $\text{RRMSE} = 41.41\%$. For CCC retrieval the lowest RMSE, 56.51 g cm^{-2} ($R^2 = 0.7$, $\text{RRMSE} = 31.63\%$), was achieved by the RFTB method. BagT was not significantly different from RFTB in this case ($R^2 = 0.71$, $\text{RMSE} = 56.72$ and $\text{RRMSE} = 31.75\%$). FVC was retrieved with a RMSE of 0.17 ($R^2 = 0.73$ and $\text{RRMSE} = 23.65\%$) with NN (Figure 4c). Also for the Shunyi test site GPR was always the best performing algorithm in cross-validation but not in ground validation tests.

A summary of the results in terms of RRMSE for all the biophysical variables tested with different MLRAs for both sites is presented in Figure 5.



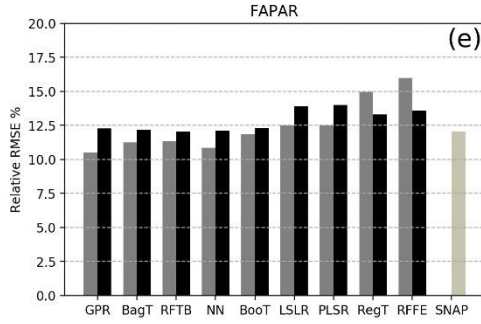


Figure 5. RRMSE (%) for LAI, LCC, FVC, CCC and FAPAR for Maccarese, Italy(left) and LAI, LCC, FVC, CCC for Shunyi (right).

It can be seen from Figure 5, that the k-fold cross-validation ($k=10$) provided in some circumstances worse results (higher RRMSE) than the validation with ground data (e.g. Figure 5a). This is particularly evident for LAI and CCC for the Maccarese site. Usually, since it is carried out with subset of the same dataset as used for the training of the MLRAs, cross-validation error is lower than the ground validation error, but this is not the case here, as it can be seen from Figure 5 a-e. The different statistical distributions of training and ground validation data might be a possible reason. It appears that the range of variation of the training sets is much larger than that of the ground validation sets, with more extreme values. This could have possibly led to more unreliable retrievals in the cross-validation when extreme values were sampled, which did not happen in the ground validation.

For the Shunyi test site, RRMSE values were generally higher than for Maccarese (Figure 5), probably because of the larger differences in the statistical distributions of the ground validation dataset from the training set, as compared for Maccarese and possibly for the larger error introduced by the sampling protocol.

In general, for LAI, algorithms such as LSLR and PLSR, seem to outperform more complex MLRAs such as NN, for example providing better results than those found by applying the biophysical processor implemented in the ESA SNAP toolbox.

For leaf chlorophyll (LCC), much worse results were obtained for Shunyi than for Maccarese, with a clear advantage of RFTB for both sites, for ground validation tests. The performance of the NN implementation of SNAP was particularly poor for Shunyi, though it should be noted that this variable was back-calculated from the CCC variable generated by SNAP.

Fractional ground cover (FVC) shows the smallest error among all the biophysical variables tested, alongside FAPAR which was only available for Maccarese. NN provided the best accuracies with the lowest values of relative RMSE for the latter variable.

Although different algorithms from different families of MLRA, perform well for different biophysical variables, GPR is considered particularly interesting, as it provides uncertainty estimates with the mean value of prediction (Figure 3c). However, despite its good cross-validation results, GPR ranked as the top algorithm only for FVC and FAPAR for the Maccarese test site (Table 4). The AL procedure only selects the most informative samples of the dataset, thus the training of GPR was performed with a smaller set of PROSAIL simulations as compared to the other algorithms, showing that it is quite efficient and robust.

The time taken for training the models ranged from 0.003 seconds for LSLR to 47.5 seconds for NN and maximum 548.3 seconds for GPR. Although, GPR took long to training, but it should be noted that it performed AL using EBD and used few samples to train and achieve high accuracy in comparison to other models.

When considering single ground sampling dates separately, generally the retrieval performances were improved (Table 6) as compared to the bulked data (Table 4). In some cases, e.g. for LAI and CCC, the worst results were obtained in the latest date, due to the fact that it is generally more difficult to estimate higher LAI values.

Table 6. Ground validation estimation results at each ground sampling date, for the best performing algorithms, for the Maccarese test site.

MLRA	29-Jan		13-Feb		6-Apr		20-Apr	
	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE
LSLR (LAI)	0.89	0.59	0.88	0.55	0.73	0.51	0.55	1.09
RFTB (LCC)	0.91	3.01	0.55	7.75	0.25	12.56	0.11	7.03
GPR (FVC)	0.90	0.12	0.89	0.08	0.90	0.05	0.71	0.05
RFTB (FAPAR)	0.74	0.10	0.45	0.14	0.44	0.09	0.56	0.06
PLSR (CCC)	0.94	26.32	0.87	33.37	0.65	44.46	0.45	55.97

4. Discussion

There are different methods for the retrieval of biophysical variables from remote sensing data and all the methods have their pros and cons (Verrelst et al., 2018). Currently, hybrid methods have the capabilities of combining physical and statistical methods and are considered state of the art. In this paper, different MLRAs (kernel based and non-kernel based) were trained and applied for the retrieval of the crop biophysical variables LAI, LCC, FAPAR, FVC and CCC with the same configuration and settings of RTM PROSAIL simulated spectra as those implemented in ESA SNAP biophysical variables retrieval toolbox, with the exception of the observation geometry. With the same of configuration of PROSAIL parameters (section 2.1), a total of 41472 simulated spectra was generated, which turned out to be rather redundant and inefficient for performing the training steps. Subsets of 2500 randomly extracted simulated spectra were used to perform the training of the algorithms and this procedure was repeated ten times with different subsets. This was done in order to allow a comparison of alternative MLRAs to a well-established operational algorithm (Weiss and Baret, 2016).

The SNAP algorithm relies on the use of PROSAIL simulations for training NNs, which are the most widely used tools for operational biophysical variables retrieval (Pozdnyakov et al., 2005; Schiller and Doerffer, 2005; Verger et al., 2011, 2008). The downside of the NNs are that they require relatively long time for training, tuning of parameters is a difficult task, they are black box in nature and can be unpredictable if training and validation data are deviating from each other even slightly (Verrelst et al., 2015b). In this paper, various alternative MLRAs outperformed NNs. For example, LAI retrieval was best performed with LSLR and PLSR, consistently for both test, although with different accuracies, i.e. RMSE of 0.68 for Maccarese and 1 for Shunyi. (Verrelst et al., 2015b) compared different retrieval strategies for LAI and reported the best performing algorithms for each category. In the case of parametric regression, the best performing algorithm was Tian 3-band formulation (RMSE = 0.615 and R² = 0.823), whereas VH-GPR performed best among non-parametric regression algorithms

(RMSE = 0.436 and $R^2 = 0.902$). However, it should be noted that accuracies found by these authors, somehow better than those of the present study, could be explained by the fact that they used cross-validation in which the training was carried out using the ground dataset, not independent model simulations such as in the present work.

LCC was best retrieved by RFTB method for both sites, revealing a general overestimation, but the error of estimation was higher for the Shunyi test site as compared with Maccarese (Figure 3b and 4b), particularly for SNAP. Also, the error of CCC estimates is very high for the SNAP tool as compared to the other algorithms tested in this work for the Shunyi ground validation (Figure 5i). It was previously shown by (Clevers et al., 2017), that Sentinel-2 bands at 10 m spatial resolution are suitable for estimating LAI, LCC and CCC. They retrieved LAI ($R^2=0.809$), LCC ($R^2=0.696$) and CCC ($R^2=0.818$) with vegetation indices approach. Multiple vegetation indices were compared for identification of potential vegetation index for the retrieval of LAI, LCC and CCC in (Frampton et al., 2013), the best correlation for LCC was with MTCI ($R^2=0.77$) and S2REP ($R^2=0.91$), for LAI, IRECI ($R^2=0.77$) and NDI45 ($R^2=0.62$), NDVI for CCC ($R^2=0.70$). These results are better than those found in the present work, but again, empirical approaches employing the ground datasets were employed by these authors for the calibration of the models.

The FVC variable was more accurately predicted than LAI or LCC, though with slightly higher error for Shunyi. With hierarchical tree-based methods such as BagT, RFTB and GPR, the error reached was even lower than 10%, which is in line with Global Monitoring for Environment and Security (GMES) goal accuracy (Drusch et al., 2010) (Figure 5c). In the case of CCC retrieval, it can be noted that the lowest RRMSE was provided by RFTB and BagT, though also GPR performs similarly to RFTB. FAPAR variable was only estimated for the Maccarese test site, due to the unavailability of the ground measurements for validation, it was not estimated for the Shunyi test site. Result show that this variable was best retrieved with RFTB method, with the lowest RMSE (Table3).

GPR has proved to be an efficient and powerful regressor for the biophysical variables retrieval in previous reports (Verrelst et al., 2012). A study conducted by (Verrelst et al., 2012), retrieved LAI, Chl and FVC using different methods, such as NN, kernel ridge regression (KRR), Support Vector Regression and GPR for different Sentinel-2 and Sentinel-3 configurations. They found that an overall good performance throughout all Sentinel configurations was provided by GPR. A study carried out by (Verrelst et al., 2016a) also found that GPR using AL techniques is an efficient and robust method for the retrieval of LAI and LCC from Sentinel-3 OLCI spectra. However, this was not the case in our study. The GPR method, coupled with AL procedure, provided the best results in terms of computational time and low error only for FVC and FAPAR ground validation tests (Figure 5c, 5h and 5e). On the other hand, GPR generally performed as the best algorithm for the retrieval of all the biophysical variables, only for the cross-validations tests (Table 4 and Table 5). Hence, of importance for further analysis to investigate how accurately GPR performs when validated against independent ground data. As (Verrelst et al., 2016a) pointed out, GPR is based on non- parametric regression in a Bayesian framework, it provides insights in bands carrying relevant information and also in theoretical uncertainty estimates, thus partially overcoming the black box problem (Verrelst et al., 2015b). These uncertainties are a useful tool for the assessment of upscaling capabilities of biophysical variables from airborne or spaceborne platforms and their respective scales (Verrelst et al., 2013a). A study conducted by (Pasolli et al., 2012) introduced the AL approach with GPR and SVM regressions to deal with the problems of training sample collection for biophysical variables estimation. Their results obtained on simulated MERIS and real SeaBAM datasets are characterized by higher performances in terms of both accuracy and stability with respect to a completely random selection strategy. The present work seems to support these results highlighting the efficiency of the AL procedure, since, when using a smaller set of selected training data, comparable results were obtained than with a random selection of larger size (Figure 2).

5. Conclusions

Hybrid methods of biophysical variables retrieval rely on the generation of simulated spectra using physically based RTM for the training of MLRAs, under the assumption of a more general applicability as compared to the training carried out using measured data, since RTMs allow to simulate a wider range of leaf and canopy properties. Operationally, so far, only NN has been implemented as a hybrid method, e.g. in the ESA SNAP toolbox. Due to complexity of parameter tuning and black box nature of the algorithm, in recent years several studies have focused on alternative MLRAs such as kernel-based methods. They offer an interesting alternative to NN in terms of performance and computational demand, sometimes partially overcoming the black box nature of NN. In the present study it was shown, using the same set of PROSAIL simulations, that some MLRAs could provide better results than NN-based algorithms implemented in SNAP. This was particularly evident for chlorophyll and fractional vegetation cover. It resulted that the best performing algorithms varied according to the biophysical variables to be estimated: e.g. LSLR and PLSR worked better for LAI whereas RFTB and BagT for LCC. This study also showed that for these variables results were consistent for two different datasets respectively collected in Italy and China on wheat, though the retrieval accuracies depended somehow on errors introduced during the ground sampling protocol. The absolute values of the retrieval accuracies are in some cases higher than the best performances reported in previous studies, but most of these previous studies did not train the regression algorithms with completely independent datasets, as was done in the present study. Further studies should be conducted to investigate alternative strategies, e.g. of active learning procedures, using different kernel functions, and validate the retrievals using a wider range of agricultural crops, also assessing their performance for canopies that depart more from the assumptions of the turbid medium used in the PROSAIL RTM.

6. References

- Atzberger, C., 2004. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. *Remote sensing of environment* 93, 53–67.
- Bacour, C., Jacquemoud, S., Tourbier, Y., Dechambre, M., Frangi, J.-P., 2002. Design and analysis of numerical experiments to compare four canopy reflectance models. *Remote Sensing of Environment* 79, 72–83.
- Bajwa, S.G., Gowda, P.H., Howell, T.A., Leh, M., 2008. COMPARING ARTIFICIAL NEURLA NETWORK WITH LEAST SQUARE REGRESSION TECHNIQUES FOR LAI RETRIEVAL FROM REMTOE SENSING DATA. Pecora.
- Baret, F., Leroy, M., Roujean, J.L., Knorr, W., Lambin, E., Linderman, M., 2003. CYCLOPES User Requirement Document. INRA-CSE, Avignon.
- Benítez, J.M., Castro, J.L., Requena, I., 1997. Are artificial neural networks black boxes? *IEEE Transactions on neural networks* 8, 1156–1164.
- Breiman, L., 2001. Random forests. *Machine learning* 45, 5–32.
- Breiman, L., 1996. Bagging predictors. *Machine learning* 24, 123–140.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1993. *Classification and Regression Trees*. Wadsworth, 1984. Google Scholar.
- Camacho, F., Baret, F., Weiss, M., Fernandes, R., Berthelot, B., Sánchez, J., Latorre, C., García-Haro, J., Duca, R., 2013. VALIDACIÓN DE ALGORITMOS PARA LA OBTENCIÓN DE VARIABLES BIOFÍSICAS CON DATOS SENTINEL2 EN LA ESA: PROYECTO VALSE-2.
- Camps-Valls, G., Bruzzone, L., 2009. *Kernel methods for remote sensing data analysis*. John Wiley & Sons.

- Casa, R., Castaldi, F., Pascucci, S., Pignatti, S., 2015. Chlorophyll estimation in field crops: an assessment of handheld leaf meters and spectral reflectance measurements. *The Journal of Agricultural Science* 153, 876–890.
- Chai, L., Qu, Y., Zhang, L., Wang, J., 2008. Lai retrieval from cyclopes and modis products using artificial neural networks, in: *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International. IEEE*, p. III-1064-III-1067.
- Cilia, C., Panigada, C., Rossini, M., Meroni, M., Busetto, L., Amaducci, S., Boschetti, M., Picchi, V., Colombo, R., 2014. Nitrogen status assessment for variable rate fertilization in maize through hyperspectral imagery. *Remote Sensing* 6, 6549–6565.
- Clevers, J.G., Kooistra, L., van den Brande, M.M., 2017. Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sensing* 9, 405.
- Combal, B., Baret, F., Weiss, M., Trubuil, A., Mace, D., Pragnere, A., Myneni, R., Knyazikhin, Y., Wang, L., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote sensing of environment* 84, 1–15.
- Das, B., Nair, B., Reddy, V.K., Venkatesh, P., 2018. Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *International journal of biometeorology* 62, 1809–1822.
- Drusch, M., Gascon, F., Berger, M., 2010. GMES Sentinel-2 mission requirements document. Earth.
- Durbha, S.S., King, R.L., Younan, N.H., 2007. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment* 107, 348–361.
- Feret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P., Ustin, S.L., Le Maire, G., Jacquemoud, S., 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote sensing of environment* 112, 3030–3043.
- Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS journal of photogrammetry and remote sensing* 82, 83–92.
- Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics* 28, 337–407.
- Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., 2017. Copernicus Sentinel-2A calibration and products validation status. *Remote Sensing* 9, 584.
- Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. *Analytica chimica acta* 185, 1–17.
- Goffart, J.P., Olivier, M., Frankinet, M., 2008. Potato crop nitrogen status assessment to improve N fertilization management and efficiency: past–present–future. *Potato Research* 51, 355–383.
- Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., 2015. The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sensing* 7, 8830–8857.

- Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the Marquardt algorithm. *IEEE transactions on Neural Networks* 5, 989–993.
- Houborg, R., Boegh, E., 2008. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote sensing of environment* 112, 186–202.
- Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT+ SAIL models: A review of use for vegetation characterization. *Remote sensing of environment* 113, S56–S66.
- Jay, S., Maupas, F., Bendoula, R., Gorretta, N., 2017. Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: Comparison of vegetation indices and PROSAIL inversion for field phenotyping. *Field Crops Research* 210, 33–46.
- Jordan, C.F., 1969. Derivation of leaf-area index from quality of light on the forest floor. *Ecology* 50, 663–666.
- Knudby, A., LeDrew, E., Brenning, A., 2010. Predictive mapping of reef fish species richness, diversity and biomass in Zanzibar using IKONOS imagery and machine-learning techniques. *Remote Sensing of Environment* 114, 1230–1241.
- Lázaro-Gredilla, M., Titsias, M.K., Verrelst, J., Camps-Valls, G., 2014. Retrieval of biophysical parameters with heteroscedastic Gaussian processes. *IEEE Geoscience and Remote Sensing Letters* 11, 838–842.
- Le Maire, G., Marsden, C., Verhoef, W., Ponzoni, F.J., Seen, D.L., Bégué, A., Stape, J.-L., Nouvellon, Y., 2011. Leaf area index estimation with MODIS reflectance time series and model inversion during full rotations of Eucalyptus plantations. *Remote Sensing of Environment* 115, 586–599.
- Liang, L., Qin, Z., Zhao, S., Di, L., Zhang, C., Deng, M., Lin, H., Zhang, L., Wang, L., Liu, Z., 2016. Estimating crop chlorophyll content with hyperspectral vegetation indices and the hybrid inversion method. *International Journal of Remote Sensing* 37, 2923–2949.
- Li-Cor, I., 1992. LAI-2000 plant canopy analyzer instruction manual. LI-COR Inc., Lincoln, Nebraska, USA.
- Meroni, M., Colombo, R., Panigada, C., 2004. Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote sensing of environment* 92, 195–206.
- Nielsen, D.C., Miceli-Garcia, J.J., Lyon, D.J., 2012. Canopy cover and leaf area index relationships for wheat, triticale, and corn. *Agronomy journal* 104, 1569–1573.
- Palacharla, P.K., Durbha, S.S., King, R.L., Gokaraju, B., Lawrence, G.W., 2011. A hyperspectral reflectance data based model inversion methodology to detect reniform nematodes in cotton, in: *Analysis of Multi-Temporal Remote Sensing Images (Multi-Temp)*, 2011 6th International Workshop on The. IEEE, pp. 249–252.
- Pasolli, E., Melgani, F., Alajlan, N., Bazi, Y., 2012. Active learning methods for biophysical parameter estimation. *IEEE Transactions on Geoscience and Remote Sensing* 50, 4071–4084.
- Pignatti, S., Acito, N., Amato, U., Casa, R., Castaldi, F., Coluzzi, R., De Bonis, R., Diani, M., Imbrenda, V., Laneve, G., Matteoli, S., Palombo, A., Pascucci, S., Santini, F., Simoniello, T., Ananasso, C., Corsini, G., Cuomo, V., 2015. Environmental products overview of the Italian hyperspectral prisma mission: The

- SAP4PRISMA project. Presented at the International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3997–4000. <https://doi.org/10.1109/IGARSS.2015.7326701>
- Pozdnyakov, D., Shuchman, R., Korosov, A., Hatt, C., 2005. Operational algorithm for the retrieval of water quality in the Great Lakes. *Remote Sensing of Environment* 97, 352–370.
- Richter, K., Atzberger, C., Vuolo, F., Weihs, P., d’Urso, G., 2009. Experimental assessment of the Sentinel-2 band setting for RTM-based LAI retrieval of sugar beet and maize. *Canadian Journal of Remote Sensing* 35, 230–247.
- Richter, R., Schlapfer, D., Muller, A., 2011. Operational atmospheric correction for imaging spectrometers accounting for the smile effect. *IEEE Transactions on Geoscience and Remote sensing* 49, 1772–1780.
- Rivera, J.P., Verrelst, J., Leonenko, G., Moreno, J., 2013. Multiple cost functions and regularization options for improved retrieval of leaf chlorophyll content and LAI through inversion of the PROSAIL model. *Remote Sensing* 5, 3280–3304.
- Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. *Remote sensing of environment* 55, 95–107.
- Rouse, J., Haas, R., Schell, J., Deering, D., Harlan, J., 1974. Monitoring the vernal advancement of retrogradation of natural vegetation, NASA/GSFC, Type III, Final Report. Greenbelt 371.
- Schiller, H., Doerffer, R., 2005. Improved determination of coastal water constituent concentrations from MERIS data. *IEEE transactions on geoscience and remote sensing* 43, 1585–1591.
- Settles, B., 2009. Active Learning Literature Survey Univ. Wisconsin-Madison, Madison, WI, 2009. CS Tech. Rep. 1648.
- Si, Y., Schlerf, M., Zurita-Milla, R., Skidmore, A., Wang, T., 2012. Mapping spatio-temporal variation of grassland quantity and quality using MERIS data and the PROSAIL model. *Remote Sensing of Environment* 121, 415–425.
- Verger, A., Baret, F., Camacho, F., 2011. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: Evaluation over an agricultural area with CHRIS/PROBA observations. *Remote Sensing of Environment* 115, 415–426.
- Verger, A., Baret, F., Weiss, M., 2008. Performances of neural networks for deriving LAI estimates from existing CYCLOPES and MODIS products. *Remote Sensing of Environment* 112, 2789–2803.
- Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. *Remote sensing of environment* 16, 125–141.
- Verrelst, J., Alonso, L., Caicedo, J.P.R., Moreno, J., Camps-Valls, G., 2013a. Gaussian process retrieval of chlorophyll content from imaging spectroscopy data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6, 867–874.
- Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G., Moreno, J., 2015a. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties—A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 108, 273–290.

- Verrelst, J., Dethier, S., Rivera, J.P., Muñoz-Marí, J., Camps-Valls, G., Moreno, J., 2016a. Active learning methods for efficient hybrid biophysical variable retrieval. *IEEE Geoscience and Remote Sensing Letters* 13, 1012–1016.
- Verrelst, J., Malenovsky, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., Moreno, J., 2018. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surveys in Geophysics* 1–41.
- Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., Moreno, J., 2012. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and-3. *Remote Sensing of Environment* 118, 127–139.
- Verrelst, J., Rivera, J., Alonso, L., Moreno, J., 2011. ARTMO: an Automated Radiative Transfer Models Operator toolbox for automated retrieval of biophysical parameters through model inversion. Presented at the Proceedings of EARSeL 7th SIG-Imaging Spectroscopy Workshop, Edinburgh, UK, pp. 11–13.
- Verrelst, J., Rivera, J.P., Gitelson, A., Delegido, J., Moreno, J., Camps-Valls, G., 2016b. Spectral band selection for vegetation properties retrieval using Gaussian processes regression. *International Journal of Applied Earth Observation and Geoinformation* 52, 554–567.
- Verrelst, J., Rivera, J.P., Moreno, J., Camps-Valls, G., 2013b. Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing* 86, 157–167.
- Verrelst, J., Rivera, J.P., Veroustraete, F., Muñoz-Marí, J., Clevers, J.G., Camps-Valls, G., Moreno, J., 2015b. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods—A comparison. *ISPRS Journal of Photogrammetry and Remote Sensing* 108, 260–272.
- Wang, L., Zhou, X., Zhu, X., Dong, Z., Guo, W., 2016. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal* 4, 212–219.
- Weiss, M., Baret, F., 2016. S2ToolBox Level 2 products: LAI, FAPAR, FCOVER.
- Williams, C.K., Rasmussen, C.E., 2006. Gaussian processes for machine learning. the MIT Press 2, 4.
- Williams, C.K., Rasmussen, C.E., 1996. Gaussian processes for regression. Presented at the Advances in neural information processing systems, pp. 514–520.
- Zhang, C., Kovacs, J.M., 2012. The application of small unmanned aerial systems for precision agriculture: a review. *Precision agriculture* 13, 693–712.

CHAPTER 2

2. CHAPTER 2. A Comparison of Moment-Independent and Variance Based Global Sensitivity Analysis Approaches for Wheat Yield Estimation with the Aquacrop-OS Model

(Published in Agronomy 2020, 10 (4), 607)

Deepak Upreti¹, Stefano Pignatti², Simone Pascucci², Massimo Tolomio¹, Zhenhai Li³, Wenjiang Huang⁴ and Raffaele Casa^{1,*}

¹ Dipartimento di Scienze Agrarie e Forestali (DAFNE), Università della Tuscia, Via San Camillo de Lellis, 01100 Viterbo, Italy; dupreti@unitus.it (D.U.)

² Consiglio Nazionale delle Ricerche, Istituto di Metodologie per l'Analisi Ambientale (CNR, IMAA), Via del Fosso del Cavaliere, 100, 00133 Rome, Italy; stefano.pignatti@cnr.it (Stefano Pignatti), simone.pascucci@cnr.it (Simone Pascucci)

³ China National Engineering Research Center for Information Technology in Agriculture (NERCITA), Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China; lizh323@126.com (Z.L.)

⁴ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; huangwj@aircas.ac.cn (W.H.)

* Correspondence: rcasa@unitus.it

Received: 22 February 2020; Revised: 18 April 2020; Accepted: 20 April 2020; Published: 24 April 2020

Abstract: The present work reports the global sensitivity analysis of the Aquacrop Open Source (AOS) model, which is the open-source version of the original Aquacrop model developed by the Food and Agriculture Organization (FAO). Analysis for identifying the most influential parameters was based on different strategies of global SA, density-based and variance-based, for the wheat crop in two different geographical locations and climates. The main objectives were to distinguish the model's influential and non-influential parameters and to examine the yield output sensitivity. We compared two different methods of global sensitivity analysis: the most commonly used variance-based method, EFAST, and the moment independent density-based PAWN method developed in recent years. We have also identified non-influential parameters using Morris screening method, so to provide an idea of the use of non-influential parameters with a dummy parameter approach. For both the study areas (located in Italy and in China) and climates, a similar set of influential parameters was found, although with varying sensitivity. When compared with different probability distribution functions, the probability distribution function of yield was found to be best approximated by a Generalized Extreme Values distribution with Kolmogorov–Smirnov statistic of 0.030 and lowest Anderson–Darling statistic of 0.164, as compared to normal distribution function with Kolmogorov–Smirnov statistic of 0.122 and Anderson–Darling statistic of 4.099. This indicates that yield output is not normally distributed but has a rather skewed distribution function. In this case, a variance-based approach was not the best choice, and the density-based method performed better. The dummy parameter approach avoids to use a threshold as it is a subjective question; it advances the approach to setting up a threshold and gives an optimal way to set up a threshold and use it to distinguish between influential and non-influential parameters. The highly sensitive parameters to crop yield were specifically canopy and phenological development parameters, parameters that govern biomass/yield production and temperature stress parameters rather than root development and water stress parameters.

Keywords: AOS; Morris; EFAST; PAWN; parameters; density-based

1. Introduction

Crop growth simulation models are important tools for ecology, agriculture and environmental research and management. Such models are used to simulate the behavior of soil-crop systems in response to climate and agriculture

management (de Willigen, 1991; Gervois et al., 2004; Hopmans and Bristow, 2002). Agricultural processes, the interaction of the crop with the environment and soil over time, are represented by mathematical equations based on theory and empirical research. These equations and parameterizations inevitably entail simplifications of reality, which leads to uncertainty and inaccuracy of output variables (Saltelli et al., 2000b). Regardless of the model used, the identification of the parameters and input variables that most affect its output is a fundamental problem for most environmental applications whenever large uncertainty on their values is expected, such as in regional-scale applications (Makowski et al., 2006; P. C. Silvestro et al., 2017; Wallach et al., 2001).

In general, crop models contain many parameters and input variables, of which some cannot be measured directly, but can only be inferred by a calibration of the responses observed from the model (Vrugt et al., 2002; Yapo et al., 1998). Manual, one-at-a-time calibration is a very tedious and time-consuming task, which is not feasible for complex models (Vrugt et al., 2003). Numerical optimization or Monte Carlo methods of calibration are widely used (Pianosi et al., 2016), but they are prone to errors, e.g., when local minima are present and their efficiency is reduced as the number of parameters increases (Duan et al., 1992). Therefore, it is often not possible to include all the parameters in the model calibration (Bekele and Nicklow, 2007; Nossent et al., 2011). In this view, Sensitivity Analysis (SA) is the set of methods that serve the purpose of selecting a limited subset of parameters or input variables, excluding those that are non-influential and have a low impact on the model output. The aim of the SA is to determine how sensitive the outputs of a crop model are with respect to the elements of the model, which are subject to uncertainty or variability (Wallach et al., 2014). The main purposes of SA are Factor Prioritization (FP) or ranking, Factor Fixing (FF) or screening and Factor Mapping (FM) (Saltelli et al., 2008). This work focuses on ranking and screening.

There are two different strategies for SA: local and global. Local SA focuses on the sensitivity of the factors at particular points in the feature space (Iooss and Lemaître, 2015), whereas global SA methods assess the sensitivity in the entire feature space and include higher-order, variance-based methods, e.g., Sobol or Fourier methods (Chen and Cournède, 2014; Saltelli et al., 1999). An extensive analysis of SA methods on crop models has been performed by different authors (Cariboni et al., 2007; Saltelli et al., 2008), which shows the inadequacy of local SA methods for crop models due to the complex nature of such models and to the inability to determine the combined effects of the parameters. For these reasons, global SA methods are currently preferred for the SA of the crop models.

Many different global SA methods have been developed (Borgonovo, 2007; Pianosi and Wagener, 2015; Saltelli et al., 2000a; Sobol', 1990; van Griensven et al., 2006). Among all these methods, variance-based methods, e.g., Sobol or Fourier, are probably the most widely used for global SA. In general, these methods measure sensitivity to an uncertain parameter or input variable by assessing the contribution of that input to the total variance of the model output variable. The advantage of the variance-based methods is their ability to quantify the contribution of the individual parameters and also quantify the contribution due to interacting effects of the parameters, independently from assumptions on the form of the input–output relation (e.g., linearity and additivity). Variance-based sensitivity indices are also easy to interpret, as they represent the fraction of the output variance caused by the variation of input (Saltelli, 2002).

Variance-based methods are moment-dependent methods and consider the second-order moment, i.e., variance as a measure of the output uncertainty. It is assumed that this moment is sufficient to describe the output variability (Saltelli, 2002). However, it has been recognized that the variance does not adequately represent output uncertainty when the model output is not represented by variance, e.g., output that is highly skewed or multi-modal (Borgonovo et al., 2011; Liu et al., 2005; Pianosi and Wagener, 2015). To overcome this limitation (moment-dependent methods), moment-independent global SA measures have been developed (Borgonovo, 2007; Liu et al., 2005; Pianosi and Wagener, 2015). These methods are also known as density-based methods. They use the entire output distribution to fully characterize the output uncertainty and to quantify the relative influence of the uncertain parameters. The main advantage of these methods, as compared to the variance-based methods, is that they do not use any specific moment

of the output distribution to measure the output variability and, therefore, are applicable regardless of its shape (e.g., multi-modal or highly skewed).

The moment-independent density-based Global Sensitivity Analysis (GSA) method, PAWN, was introduced by (Pianosi and Wagener, 2015). It measures sensitivity based on the difference between unconditional output distribution and conditional output distribution, which are obtained simultaneously when all the parameters are free to vary and when one of the parameters is fixed. The difference between PAWN and other density-based methods, e.g., the entropy-based method (Liu et al., 2005) and the δ -sensitivity measure (Borgonovo, 2007), is that the output distribution in PAWN is characterized by the cumulative distribution function (CDF), whereas in other methods the output distribution is characterized by the probability density function (PDF). The approximation of CDFs by using empirical distributions of the data sample is much easier than the approximation of their PDFs. This facilitates the analysis of the robustness and the convergence of the estimated sensitivity indices (Pianosi and Wagener, 2015).

The PAWN method was initially tested with a simple conceptual hydrological model with only five parameters. To further investigate its efficiency, it was applied to the complex environmental model - Soil and Water Assessment Tool (SWAT) (Zadeh et al., 2017) with 26 parameters and compared with variance-based method Sobol.

The Aquacrop model (Steduto et al., 2009) is a crop water productivity model developed by the Food and Agriculture Organization (FAO) of the United Nations specifically for the purpose of assessing crop response to water and increasingly used by scientists and agronomists (Foster et al., 2017; Jin et al., 2014; Saab et al., 2015; Todorovic et al., 2009; Xiangxiang et al., 2013). The Aquacrop Open Source (AOS) version was developed by (Foster et al., 2017) and made available as an open-source Matlab version. In principle, this open-source version is the same as the original developed FAO version of the Aquacrop model. Recently, a version of AOS, v6.0a, has been released. It implements all calculation procedures performed by the original FAO Aquacrop v6.0a, with the exception of soil fertility stress and soil salinity stress, as these are yet to be tested extensively for different crop types and environments worldwide (Foster, 2019). The soil water balance, stress effects, runoff and plant growth are controlled by a large number of parameters (around 100). Even when some of the parameters can be fixed a priori, calibration of the AOS remains quite challenging, given the relatively large number of parameters (54 in our case) that are typically left to be varied simultaneously. Therefore, SA is often applied prior to the calibration process to identify the most influential parameters and the non-influential ones (P. C. Silvestro et al., 2017; Vanuytrecht et al., 2014). To avoid any confusion with the model, we will hereafter refer to the FAO developed model as Aquacrop and its open-source Matlab version (The MathWorks Inc., Natick, MA, USA) as AOS that is used in this study.

There have been studies of sensitivity analyses with the FAO developed Aquacrop model (P. C. Silvestro et al., 2017; Vanuytrecht et al., 2014; XING et al., 2017), but none yet on the open-source version AOS, which nevertheless lends itself more easily to the integration into regional-scale data assimilation schemes than the standard version. The studies of Aquacrop model SA was based on the arbitrarily chosen thresholds for yield output. According to the authors' best knowledge this is the first time the dummy parameter approach (Zadeh et al., 2017) is being used to identify the most influential parameters with the recently released AOS model (v6.0).

The main objectives of this study are to (i) assess and compare the application of the Aquacrop model approach with the Extensive Fourier Amplitude Sensitivity Test (EFAST) method and the density-based PAWN method with a high number of parameters (soil and crop) in terms of parameter ranking results ; (ii) perform a sensitivity analysis of the AOS model, provided as an open-source Matlab version, that in principle is similar to the original FAO developed Aquacrop model, but differs in some parameters; (iii) compare the sensitivity indices (SI) of the EFAST and density-based PAWN methods for two different geographical locations and climates. As both methods have different backgrounds and rationale, the comparison cannot be made on the basis of the Sensitivity Index (SI); therefore, the comparison is made in terms of parameter ranking as a measure of the effectiveness with multiple years as a measure

of robustness. To assess the parameter ranking, we have used the dummy parameter approach proposed by (Zadeh et al., 2017), while the SI of the dummy parameter has used as a threshold to identify non-influential parameters.

2. Materials and Methods

2.1. Sensitivity Analysis

First, the Morris method was used to distinguish between influential and non-influential parameters of the AOS model. The Morris method is designed for an initial screening or the elementary effects that the variation of input factors (parameters and input variables) produces on model outputs. Secondly, the variance-based EFAST method and finally the density-based PAWN method were applied to all the chosen parameters considered for sensitivity analysis (see Table 1), including a dummy parameter to find the higher-order effects and the combined effect of different parameters.

2.1.1. Morris-Based Elementary Effects

The Morris method is designed for the screening of the most influential parameters (Morris, 1991). It is considered more a qualitative rather than a quantitative method as it does not quantify the sensitivity of a parameter individually, rather it represents the parameter sensitivity in terms of mean (m^*) and sigma (s). High m^* indicates higher sensitivity and higher s indicates interacting effects with other parameters. The method determines whether the effects are negligible, additive, non-linear or the factors are having interacting effects with other parameters (Campolongo et al., 2007). To sample input space, the experimental design is structured in a group of points, called trajectories. In this work, the number of trajectories has been set to 20, which is a value between the minimum suggested by (Campolongo et al., 2007) (i.e., 10) and the value used by (Vanuytrecht et al., 2014) (i.e., 25) in order to avoid an excessively high computational cost. It computes two indices for each input: the mean of the effects, which measures the total effect of input over the output, and the standard deviation of the effects, which measures the degree of interactions with the other inputs. The input factor with a mean below a threshold value can be considered as having negligible effects on the model output, thus it can be eliminated from further processing by keeping its value as fixed. Whereas, other authors have defined an arbitrary threshold (Liu et al., 2005; P. C. Silvestro et al., 2017). In this work, we chose to identify influential parameters by comparing them with a dummy parameter used as a threshold (Zadeh et al., 2017). All the parameters having a mean value greater than that of the dummy parameter are considered as influential.

2.1.2. EFAST Method

The EFAST method (Saltelli et al., 2000b) combines the FAST (Cukier et al., 1978) and the Sobol methods (Sobol, 1993). It calculates sensitivity indices using the total variance of the output of the model and the conditional variances depending on the parameters. The interaction among the factors can be quantified by calculating the main SI (Si) and the index of total sensitivity (Sti), i.e., the sum of the main effect plus the interaction between the variation of the terms of the parameters to all orders. Si and Sti range between 0 and 1, with higher values indicating more important effects (P. C. Silvestro et al., 2017).

In this study, the input parameters plus the dummy parameter were implemented using the EFAST method to identify influential parameters and to rank them. The number of sample points considered for the chosen 55 parameters, including a dummy variable, is 10,000. The required number of runs are thus $55 \times 10,000$. The number of higher harmonics M_s is 4, and the number of search curves N_{cs} was set to 1.

2.1.3. Density-Based PAWN Method

The PAWN method is a moment-independent global sensitivity analysis method that takes into account the entire model output distribution for the quantification of the sensitivity of the parameters (Pianosi and Wagener, 2015), in comparison to the other variance-based methods that consider only the moments, e.g., the variance of the model output. In general, density-based sensitivity indices measure the sensitivity to parameter x_i by the distance between unconditional PDF and conditional PDF of the model output (Borgonovo, 2007; Liu et al., 2005). Practically, PDFs are usually unknown and must be approximated using a data sample. Authors in (Pianosi and Wagener, 2015) pointed

out the difficulties and limitations of deriving empirical PDFs and suggested to use CDFs instead of PDFs, as it is much easier than using the approximation of PDFs.

In the PAWN method, sensitivity is measured by estimating the variations that are induced in the output distribution when removing the uncertainty about one or more inputs. The sensitivity to the input parameter is measured using the distance between the unconditional probability distribution and the conditional distributions that are obtained by keeping the parameter fixed using the Kolmogorov–Smirnov (KS) statistics (Kolmogorov, 1933; Smirnov, 1939).

The unconditional cumulative distribution function of the model output y is denoted by $F_y(y)$, and the conditional cumulative distribution is denoted by $F_{y|x_i}(y)$, where x_i is a parameter under consideration.

If $F_y(y)$ and $F_{y|x_i}(y)$ coincide, it can be concluded that the parameter is non-influential. If the distance between $F_y(y)$ and $F_{y|x_i}(y)$ is increased, the influence of the parameter increases as well. As a measure between unconditional and conditional CDFs, the KS statistics were used (Equation (1) (Pianosi and Wagener, 2015))

$$KS(x_i) = \max |F_y(y) - F_{y|x_i}(y)| \quad (1)$$

As KS depends on the value at which we fix x_i , the PAWN index T_i considers a statistic (e.g., the maximum or the median) over all possible values of x_i , i.e.,

$$T_i = \text{stat } |KS(x_i)| \quad (2)$$

By definition, T_i varies between 0 and 1; the lower the value of T_i the less influential x_i , and if $T_i = 0$, then x_i has no influence on y .

The total number of model evaluations to compute the sensitivity indices are $N_u + n \times N_c \times M$. M is the number of parameters, N_u and N_c are the numbers of unconditional random samples and conditional samples and n is the number of sampling points. In this work, we used $N_u = 300$, $N_c = 200$, $n = 10$ and $M = 55$.

The density-based PAWN method estimates sensitivity by implementing unconditional and conditional cumulative distribution functions. In the case of unconditional distribution functions, all the parameters are free to vary within their defined range, and in the case of the conditional distribution function, the parameter into consideration is fixed to the nominal value. The parameter with the low sensitivity or called PAWN index has the same overlapping unconditional and conditional distribution functions.

In the present study, the Morris and PAWN methods were implemented with the Matlab (The MathWorks Inc., Natick, MA, USA) - based SAFE toolbox (Pianosi et al., 2015) and EFAST based on the SA toolbox Eikos (Ekstrom, 2005). To figure out the best PDF fit for the yield output and different statistical distributions, tests including KS statistics and Anderson–Darling statistics were carried out with EasyFit Evaluation (Ver. 5.6, <http://www.mathwave.com/>).

2.1.4. Identifying Non-Influential Parameters by Using a Dummy Parameter

Theoretically, the SI of a non-influential parameter is zero. In the case of the PAWN method, SI of zero means that the conditional and unconditional CDF coincides, i.e., fixing the parameter into consideration will have no influence on the model output distribution. The value of zero in the Morris screening test indicates that the parameter has no influence on the model output. However, numerical algorithms and computations are utilized to implement SI, and small but non-zero values can be obtained for non-influential parameters. To set a threshold to identify non-

influential parameters, a dummy parameter approach was proposed by (Zadeh et al., 2017). The SI of this dummy parameter provides an indication of the approximation error of the sensitivity analysis. The parameters having SI greater than the SI of the dummy parameter are considered as the influential parameters. It should be noted that the dummy parameter is not added to the model; its SI is calculated by using the sampled data.

2.2. Crop Model: Aquacrop Open Source

Crop growth simulation model Aquacrop was originally developed by the Food and Agriculture Organization (FAO) of the United Nations. As the code of the original model is not accessible, it does not allow, for example, the implementation of data assimilation techniques such as the ensemble Kalman filter, which are increasingly employed for ingesting remote sensing observations into crop models (Steduto et al., 2009). In this work we have used the open-source Matlab version of the model called Aquacrop Open Source (AOS) developed by (Foster et al., 2017).

Aquacrop model evolved from the most commonly used empirical approach to assess crop yield response to water (Doorenbos and Kassam, 1979; Raes et al., 2009), shown as Equation (3)

$$\frac{Y_x - Y_a}{Y_x} = k_y \left(\frac{ET_x - ET_a}{ET_x} \right) \quad (3)$$

where Y_x and Y_a are the maximum and actual yield, ET_x and ET_a are the maximum and actual evapotranspiration, and k_y is the proportionality factor between relative yield loss and relative reduction in evapotranspiration. The separation of ET into Tr and E avoids the confounding effects of the nonproductive consumption use of water (E), which is important especially during incomplete ground cover, and leads to the conceptual equation at the core of the model (Raes et al., 2009) (Equation (4), derived by (Tanner and Sinclair, 1983)). Yield (Y) as a function of biomass (B) and harvest index (HI) is described as Equation (5):

$$B = WP \times \sum Tr \quad (4)$$

$$Y = B \times HI \quad (5)$$

where Tr is the crop transpiration (in mm), and WP is the water productivity (kg of biomass per m² and per mm of cumulated water transpired over the time period in which the biomass is produced), which is a constant for the given climatic condition. By normalizing it for different climatic conditions, WP becomes a conservative parameter (Steduto et al., 2007).

Reference evapotranspiration (ET_0) is provided as a model input: it can easily be estimated using an external calculator (Allen et al., 1998a) that uses the Penman–Monteith equation, following the procedures outlined in FAO Irrigation and Drainage paper no. 56 (Allen et al., 1998b). The model progressed by separating by evapotranspiration (ET) into crop transpiration (Tr) and soil evaporation (E) using canopy cover (CC). Tr is proportional to the fraction of land surface covered by the canopy, while E is proportional to the surface not covered by the canopy, with some minor adjustments. Tr in the absence of water stress is calculated using the adjusted green canopy cover (CC^*) and the crop coefficient of the full-grown crop, prior to senescence (Kcb), as in Equation (6):

$$Tr = CC^* \times Kcb \times ET_0 \quad (6)$$

A distinctive feature of Aquacrop is the use of green canopy cover instead of Leaf Area Index (LAI) to simulate canopy expansion. This allows Aquacrop to be easily used with remote sensing data (Raes et al., 2009). Soil evaporation is based on the two-stage Ritchie's approach (Ritchie, 1972), adopted also in other crop models (e.g., DSSAT crop model).

While the yield–transpiration relationship accounts for the biomass accumulation, canopy expansion and crop phenology are simulated following the thermal time concept, which uses growing degree days (GDDs) as an internal clock of crop development as described in (McMaster and Wilhelm, 1997), with minor modifications. CC development is simulated through a two-stages curve. An exponential equation is used up to half the maximum canopy cover (Equation (7)):

$$CC = CC_0 \times e^{CGC \times t} \quad (7)$$

where CC is the fractional canopy cover of soil at time t , CC_0 is the initial CC (at time 0) and CGC is the canopy growth coefficient in the percentage of existing CC at time t . The second half of canopy development follows an exponential decay equation (Equation (8)):

$$CC = CC_x - (CC_x - CC_0) \times e^{-CGC \times t} \quad (8)$$

where CC_x is the maximum canopy cover for optimal conditions. An empirical canopy decline coefficient (CDC) is used to characterize the CC fractional reduction overtime after maturity has been reached (Equation (9)):

$$CC = CC_x \left[1 - 0.05 \left(\exp^{\frac{CDC}{CC_x}} - 1 \right) \right] \quad (9)$$

The soil–water balance is simulated first by using the runoff curve number (CN) approach of the United States Department of Agriculture (USDA) Soil Conservation Service to separate runoff and infiltration, and then an exponential drainage function is used to simulate downward flow when the soil water content is above field capacity (Equation (10)):

$$\frac{\Delta\theta_i}{\Delta t} = \tau(\theta_{SAT} - \theta_{FC}) \frac{e^{\theta_i - \theta_{FC}} - 1}{e^{\theta_{SAT} - \theta_{FC}} - 1} \quad (10)$$

Where θ_i is the actual soil water content, θ_{SAT} is the soil water content at saturation θ_{FC} is the soil water content at field capacity, Δt is the time step (daily) and τ is a coefficient to account for drainage characteristics, derived from the saturated hydraulic conductivity (K_{sat}).

The effect of water-limiting conditions on crop growth is calculated with the ratio of actual available water to total available water TAW ($\theta_{SAT} - \theta_{FC}$) with an approach evolved from the FAO Irrigation and Drainage paper no. 56 (Allen et al., 1998b). Water stress coefficients (K_s) act as modifiers (with values ranging from 0 to 1) of target processes and are used to modulate transpiration, canopy expansion, canopy senescence and pollination when the readily available water (RAW) is depleted. RAW is defined as a fraction of TAW by the p_{up} parameters. Each of the K_s coefficients is defined on the basis of the fraction of actual to total available water, using three parameters: the upper threshold (p_{up} , when water stress begins: $K_s = 1$), the lower threshold (e , when water stress is complete: $K_s = 0$) and the shape factor of the curve (f_{shape_w}).

Extensive descriptions of Aquacrop processes are presented in (RAES et al., 2018; Steduto et al., 2009). AOS does not differ in the calculation procedures but is provided as an open-source Matlab version of Aquacrop (Tim Foster, 2016; Foster et al., 2017). Test simulations were conducted to verify the suitability of AOS 5.0a to reproduce accurately the calculations and outputs from the original FAO Aquacrop model, considering a range of crop types and environments: wheat production in Tunisia, rice and wheat production in Hyderabad, India, and potato production in Brussels, Belgium. The maximum RMSE and minimum r^2 for crop yields across all simulations were 0.045 t ha^{-1} and

0.993, respectively, which confirms the ability of AOS to reproduce successfully Aquacrop calculations and outputs (Tim Foster, 2016).

Recently, a version of AOS, AOS v6.0a, has been released. AOS v6.0a implements all calculation procedures performed by the original FAO Aquacrop v6.0a model, with the exception of soil fertility stress and soil salinity stress, as these are yet to be tested extensively for different crop types and environments worldwide (Foster, 2019).

There is a total of 89 parameters in the AOS model, as mentioned in the reference manual (Timothy Foster, 2016). Both Aquacrop v6.0a and AOS software include crop files with default parameter values of biomass and canopy growth, thermal time-based phenological stages, transpiration coefficients and temperature and water stresses. In AOS each of these parameters is assigned a specific abbreviation and description in a clearer way with respect to Aquacrop v6.0a. The abbreviations and descriptions of the 54 AOS parameters considered in this work for the SA are listed in Table 1, along with their Aquacrop v6.0a counterparts, as presented in two of the most recent and extensive global SA (Guo et al., 2019; Xing et al., 2017). For clarity, when referring to the previous study, all the parameter abbreviations used in this paper will be written using the AOS notation.

2.3. Setting up the Parameters for AOS Model

In this study, the EFAST and PAWN methods of SA have been applied to the AOS model and compared in terms of parameter ranking. We used all the parameters that were not fixed as mentioned in the AOS manual (Timothy Foster, 2016) to be analyzed for SA. The winter wheat grain yield was considered as the output. AOS parameters are reported in Table 1. The range of variation of the selected 54 parameters (reported in Table 1) was determined based on the AOS manual (Timothy Foster, 2016) and default values provided with the open-source AOS model database for the wheat crop and also on the previous studies of sensitivity analysis, carried out with the original FAO Aquacrop model (P. C. Silvestro et al., 2017; Vanuytrecht et al., 2014; XING et al., 2017). Since there was no prior information on parameter frequency distributions, they were assumed to have an a priori uniform distribution. It should be noted that (Vanuytrecht et al., 2014), in their study, confirms the parameter ranges are more important than their distributions in SA studies.

2.4. Sites Description

In the scenarios applied for the SA test, we used the climate and soil data from two experimental sites. The first site was located in Maccarese (41.833° N, 12.217° E; 8 m a.s.l), western coast of Central Italy, near the Fiumicino Airport, Rome. This area has a typical Mediterranean climate with wet autumn and dry hot summer falling into the Koppen–Geiger class Csa (temperate, dry summer, hot summer) (Peel et al., 2007). The average yearly maximum temperature is 27.4°C and the average minimum temperature is 5 °C. The average annual precipitation is 812 mm (Upreti et al., 2019). The soil in Maccarese is Cutanic Luvisol, with a prevailing sandy clay loam texture, becoming more clayey toward the north-east of the site.

The other site was located in Xiaotangshan (40.167° N, 116.4° E; 36 m a.s.l), Chanping district, Beijing, China. It is characterized by a continental climate, with a cold, dry winter and a hot, wet summer, belonging to the Koppen–Geiger (Peel et al., 2007) class Dwa (cold, dry winter, hot summer), i.e., monsoon-influenced, hot-summer humid continental climate. The soil in the experimental area was loam. The mean annual precipitation is approximately 602 mm. The mean temperature is approximately −4.5 °C in winter and 26.2 °C in summer. There are on average 180 frost-free days per year. The main crop is winter wheat. Table 2 shows sowing and harvest dates used for model simulations at the different test sites (XING et al., 2017), which were set on the basis of local agronomic practice. Soil information for both the test sites is presented in Table 3.

Table 1. AOS parameters considered for sensitivity analysis with the range of values adopted in the present work.

Parameters		Description	Unit	Values
Aquacrop v6.0 ^a	AOSv6.0			
Management				
den	PlantPop	Plant population	n ha ⁻¹	4,500,000
-	Planting Date	Planting date	# Date	15-Nov
-	Harvest Date	Harvest date	# Date	15-Jul
Canopy development				
ccs	SeedSize	Soil surface area covered by an individual seedling at 90% emergence	cm ²	1.5
-	CCmin	Minimum fractional canopy cover size below which yield formation does not occur	%	(0.04725, 0.08775)
mcc	CCx	Maximum fractional canopy cover size	%	0.96
cdc	CDC	Canopy decline coefficient	fraction GDD	0.004
cgc	CGC	Canopy growth coefficient	fraction GDD	0.006
Phenology				
eme	Emergence	Thermal time from sowing to emergence	GDD ^b	160
root	MaxRooting	Thermal time from sowing to maximum root development	GDD	864
sen	Senescence	Thermal time from sowing to start of canopy senescence	GDD	1670
mat	Maturity	Thermal time from sowing to physiological maturity	GDD	2372
flo	HIstart	Thermal time from sowing to start of yield formation	GDD	1243
flolen	Flowering	Duration of flowering (thermal time)	GDD	200
-	YldForm	Duration of yield formation (thermal time)	GDD	1100
Transpiration, biomass and yield				
kcb	Kcb	Maximum crop coefficient when canopy is fully developed	-	1.1
kcdcl	fage	Decline of crop coefficient due to ageing of the canopy	%d ⁻¹	0.15
wp	WP	Water productivity normalized for reference evapotranspiration and atmosphere carbon dioxide	g m ⁻²	16.5
-	Wpy	Adjustment of water productivity parameter in yield formation stage	g m ⁻²	100
-	fsink	Crop sink strength coefficient	-	(0.35,0.65)
hi	HI0	Reference harvest index	%	0.46
-	Hiini	Initial harvest index	%	(0.007, 0.013)
hinc	dHI0	Maximum possible increase in harvest index above reference value	%	0.15

<i>exc</i>	<i>exc</i>	Excess of potential fruits that is produced by the crop	%	1
Water and temperature stress				
-	<i>Tbase</i>	Base temperature below which crop growth does not occur	°C	0
-	<i>Tupp</i>	Upper temperature above which crop growth does not occur	°C	26
<i>polmx</i>	<i>Tmax_u</i> <i>p</i>	Maximum temperature above which pollination begins to fail	°C	35
-	<i>Tmax_lo</i>	Maximum temperature above which pollination fails completely	°C	40
<i>polmn</i>	<i>Tmin_up</i>	Minimum temperature below which pollination begins to fail	°C	4.5
-	<i>Tmin_lo</i>	Minimum temperature below which pollination fails completely	°C	0
<i>stbio</i>	<i>GDD_u</i> <i>p</i>	Minimum number of GDD's required for full biomass production	GDD	13.5
-	<i>GDD_lo</i>	Minimum number of GDD's required for any biomass production to occur	GDD	2.5
-	<i>fshape_b</i>	Shape factor describing the reduction in biomass production due to insufficient GDD's	-	13.81
<i>hipsflo</i>	<i>dHI_pre</i>	Possible increase of harvest index due to pre-anthesis water stress	%	(0.02, 0.05)
<i>hipsveg</i>	<i>a_HI</i>	Coefficient describing the positive impact on harvest index of restricted vegetative growth post-anthesis	-	(1, 3)
<i>hinsveg</i>	<i>b_HI</i>	Coefficient describing the negative impact on harvest index of stomatal closure post-anthesis	-	4.5
<i>puexp</i>	<i>p_up1</i>	Upper soil water depletion threshold for water stress effects on canopy expansion	fraction TAW ^c	0.2
<i>psto</i>	<i>p_up2</i>	Upper soil water depletion threshold for water stress effects on stomatal control	fraction TAW	0.65
<i>psen</i>	<i>p_up3</i>	Upper soil water depletion threshold for water stress effects on canopy senescence	fraction TAW	0.7
<i>ppol</i>	<i>p_up4</i>	Upper soil water depletion threshold for water stress effects on crop pollination	fraction TAW	0.8
<i>plexp</i>	<i>p_lo1</i>	Lower soil water depletion threshold for water stress effects on canopy expansion	fraction TAW	0.65
-	<i>p_lo2</i>	Lower soil water depletion threshold for water stress effects on stomatal control	fraction TAW	1
-	<i>p_lo3</i>	Lower soil water depletion threshold for water stress effects on canopy senescence	fraction TAW	1

-	<i>p_lo4</i>	Lower soil water depletion threshold for water stress effects on crop pollination	fraction TAW	1
<i>pexshp</i>	<i>fshape_w1</i>	Shape factor describing water stress effects on canopy expansion	-	5
<i>pstoshp</i>	<i>fshape_w2</i>	Shape factor describing water stress effects on stomatal control	-	2.5
<i>psenshp</i>	<i>fshape_w3</i>	Shape factor describing water stress effects on canopy senescence	-	2.5
-	<i>fshape_w4</i>	Shape factor describing water stress effects on crop pollination	-	1
Soil and roots				
-	<i>REW</i>	User-defined readily evaporable water depth	m	(6.3, 11.7)
-	<i>CN</i>	Curve Number	-	(65, 85)
<i>rtmx</i>	<i>Zmin</i>	Minimum effective rooting depth	m	0.3
<i>rtx</i>	<i>Zmax</i>	Maximum effective rooting depth	m	1.5
<i>rtexup</i>	<i>SxTopQ</i>	Maximum water extraction at the top of the root zone	m ³ m ⁻³ soil d ⁻¹	0.028
<i>rtexlw</i>	<i>SxBotQ</i>	Maximum water extraction at the bottom of the root zone	m ³ m ⁻³ soil d ⁻¹	0.008

^a as reported in (Guo et al., 2019; XING et al., 2017); ^b GDD: growing degree days; ^c TAW: total available soil water.

Table 2. Sowing and harvest dates for test sites.

Maccarese, Italy			Xiaotangshan, China	
Period	Sowing Date	Harvest Date	Sowing Date	Harvest Date
2014–2015	1-Nov-14	1-Jul-15	-	-
2015–2016	8-Nov-15	30-Jun-16	7-Oct-15	19-Jun-16
2016–2017	10-Nov-16	15-Jul-17	1-Oct-16	10-Jun-17
2017–2018	30-Oct-17	28-Jun-18	15-Oct-17	30-Jun-18

Table 3. Representative Soil conditions at study sites.

Maccarese, Italy									
Soil	Depth (cm)	Moisture Content (%)			Organic C (%)	Clay (%)	Silt (%)	pH	Sand (%)
		PWP	FC	Sat					
Sandy loam	0–10	8	18	45	2.4	33.3	19	6.6	47.8
Sandy loam	10–20	8	18	45	1.8	34.1	18.6	6.7	47.3
Sandy loam	20–30	8	18	45	1.8	33.6	20.4	6.7	46
Clay loam	30–40	18	23	43	1.8	34.6	19	6.7	46.4
Clay loam	40–50	18	23	43	1.5	32.4	17.8	6.8	49.8
Xiaotangshan, China									
Soil	Depth (cm)	Moisture Content (%)			Organic C (%)	Clay (%)	Silt (%)	pH	Total N (%)
		PWP	FC	Sat					
Clay loam	0–20	8.8	27.3	51.1	1.04	23.5	53.9	8.00	0.11
Clay loam	20–40	8.7	27.3	51.3	1.01	23.4	54.1	8.08	0.10
Clay loam	40–60	12.3	34.8	54.7	0.68	37.3	47.8	7.94	0.08
Clay loam	60–80	12.3	34.8	54.7	0.66	37.3	47.8	7.98	0.08
Clay loam	80–100	12.3	34.8	54.7	0.59	40.3	43	8.03	0.07

^a PWP: permanent wilting point; FC: field capacity; Sat: saturation.

2.5. Weather Data

Sensitivity analysis was carried out for four different wheat growth seasons (2014–2018) for the Maccarese site and for three seasons (2015–2018) for the Xiaotangshan site. The meteorological data used in this study to drive the model simulations for the SA were obtained from the Regione Lazio Agrometeorological Service, Italy (<http://www.arsial.it/portalearsial/agrometeo/index.asp>) and the National Meteorological Information Center of the China Meteorological Administration (<http://cdc.cma.gov.cn/>). The daily minimum and maximum temperatures, relative humidities, wind speed, rainfall data and daily solar radiation for the period were provided. The daily reference evapotranspiration was calculated based on the FAO Penman–Monteith method as described in (Allen et al., 1998a) and the ET_0 calculator (Raes and Munoz, 2009). The minimum and maximum temperature, rainfall and reference evapotranspiration for both test sites are shown in Figure 1. For both sites, the years considered in this study allowed to include wet and dry conditions, which were assessed by examining the Standardised Precipitation–Evapotranspiration Index SPEI (<https://spei.csic.es/index.html>). In particular, the spring of 2017 was severely dry in Maccarese and also in Xiaotangshan, although for the latter the drought was limited to the month of May as shown in the Supplementary Materials (Figure S1).

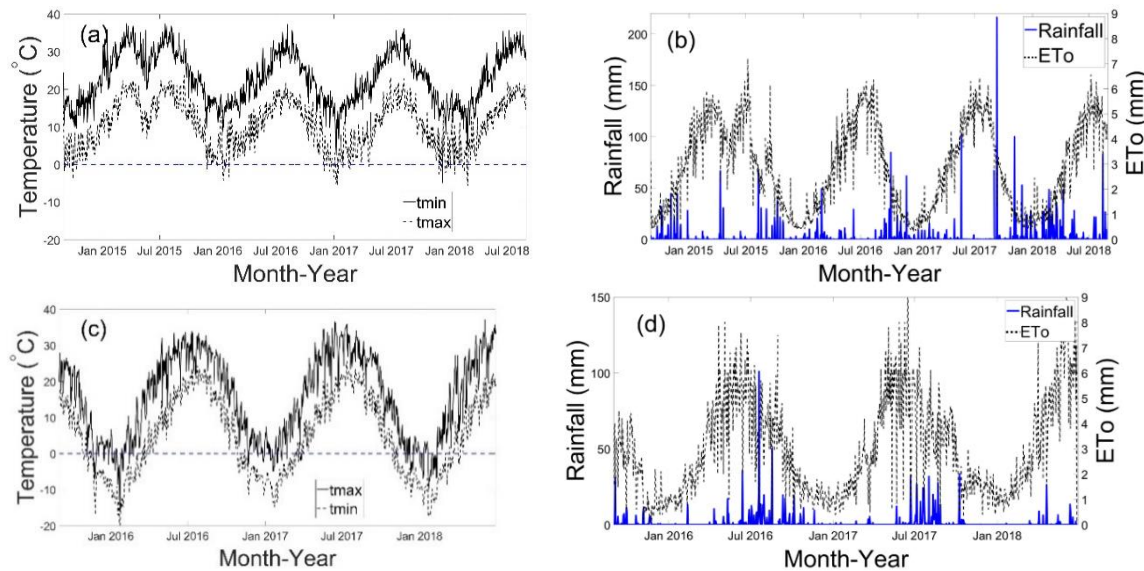


Figure 1. Maximum and minimum temperature, (b) daily precipitation and reference evapotranspiration (ET_0) for Maccarese, Central Italy. (c) Maximum and minimum temperature, (d) daily precipitation and reference evapotranspiration (ET_0) for Xiaotangshan, China.

3. Results

3.1. Morris Method

Figure 2 shows the sensitivity of the AOS model parameters obtained by applying the Morris method for the two different locations, climates and years. In Figure 2, the average value for all the years is plotted, and the error bars indicate the variation (standard deviation) of the average value.

A large magnitude of μ^* indicates the importance of the model parameter. A high standard deviation of σ implies either a non-linear relationship or interaction with other parameters. The parameters *CGC* and *H1start* were found to be of high importance for both the study sites regardless of the climate. The number of relevant parameters differs slightly for each year and study site. For the Maccarese site, 31 sensitive model parameters were identified for the years 2014–2015, 25 for 2015–2016, 23 for both the years 2016–2017 and 2017–2018. Similarly, for the Xiaotangshan site, 29 parameters were identified as most sensitive for 2014–2015, 29 for 2015–2016 and 23 for 2016–2017. When considering all years of data and study sites, 29 out of 55 (including dummy) were found important for yield output.

These 29 parameters are plotted in Figure 2. For clarity purposes, only the parameters with the mean value of or greater than 1 t ha^{-1} are labeled.

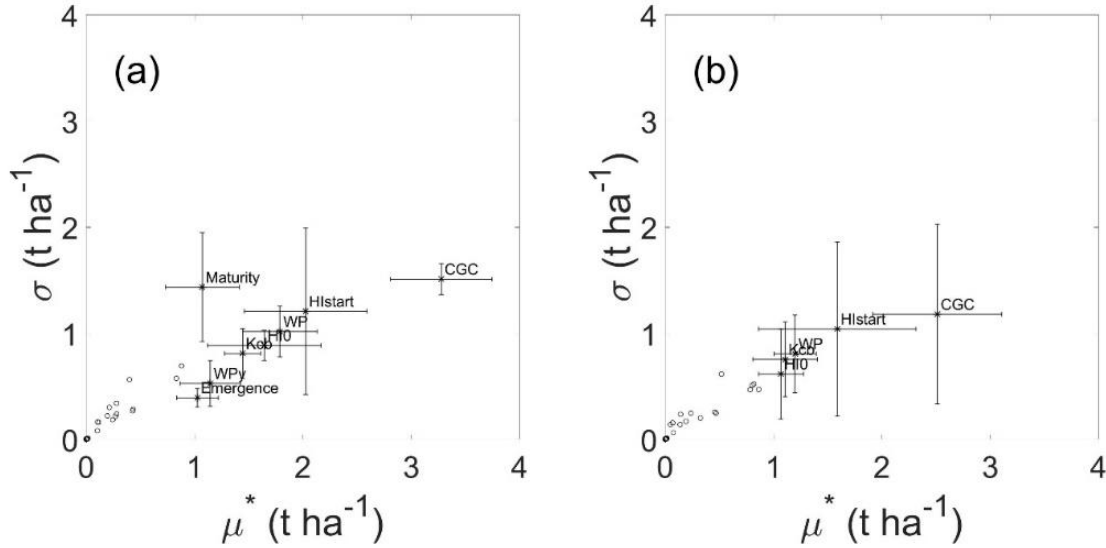


Figure 2. Morris method parameter screening results for (a) Maccarese and (b) Xiaotangshan test sites. Mean values of the Morris sensitivity indices μ^* and s are indicated by markers, and the errors bars indicate the standard deviation of index values across all the years. Only parameters with a μ^* value higher than 1 t ha^{-1} are labeled. The meanings of parameter labels are reported in Table 1.

3.2. Variance-Based Extended Fourier Amplitude Sensitivity Test (EFAST)

Figure 3 shows the EFAST indices plots for all the parameters and all the years for the Maccarese test site. Both the first-order (S_i) and the total-order (St_i) indices are calculated, as proposed in (Saltelli et al., 1999), to distinguish between the main effect and the interaction effect.

For both study sites and all the years of weather data, the most relevant parameter was *CGC*, with the exception of 2016–2017 for both test sites; thus, this parameter is responsible for most of the variance of the output (Figure 3; Figure 4). The 2016–2017 season was particularly dry in Maccarese, with a three-month accumulated standardized precipitation-evapotranspiration index (SPEI) below -1 from February to July, indicating drought conditions (Figure S1). During that season, the parameter that resulted to have the highest sensitivity was *Hlstart*. For Xiaotangshan test site (Figure S2), *Kcb*, *WP* and *HIO* were among the highest sensitive parameters after *CGC* for the years 2015–2016 and 2017–2018. During 2016–2017, *Hlstart* had the highest sensitivity followed by *CGC*, *CDC*, *Kcb*, *WP* and *HIO*.

The results of EFAST for winter wheat were quite similar for both study sites. Although, there were minor differences in the specific value of the St_i , and the general importance of the analyzed parameters was identical for both study sites. In the end, we used the average of the total sensitivity indices across all the years and study sites to find a similar set of sensitive parameters that could be adapted when calibrating AOS for regional studies (Figure 4).

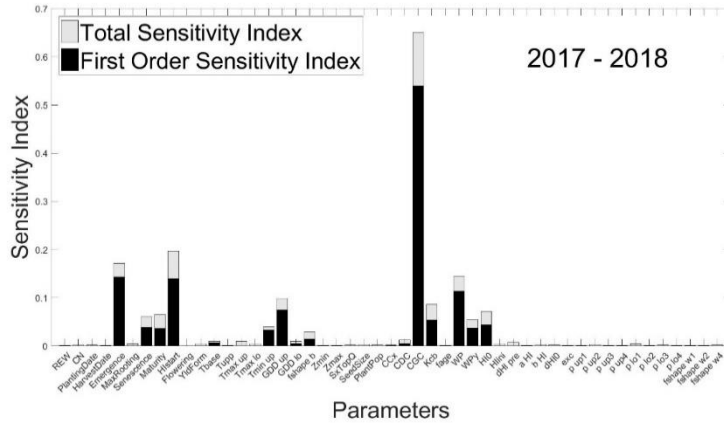


Figure 3. First-order and total-order SI for different years for Maccarese. Only the parameters with a total sensitivity index higher than the sensitivity index of the dummy parameter are shown. The details on the parameters can be found in Table 1.

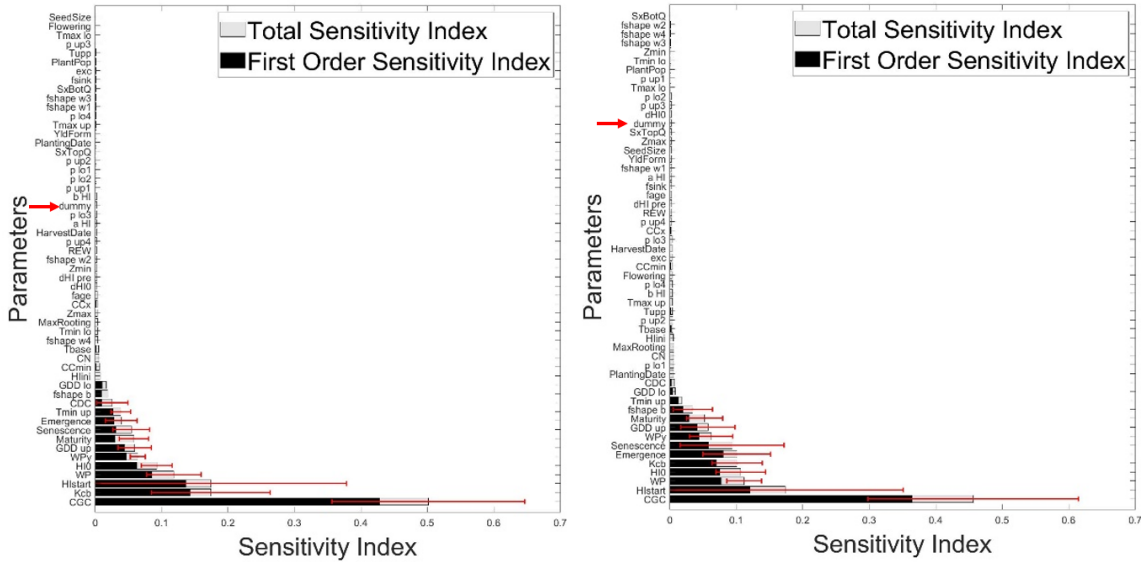


Figure 4. Average sensitivity indices using the EFAST method Maccarese (left) and Xiaotangshan (right) test sites. The error bars in red represent the extreme values of total-order sensitivity indices. The dummy parameter is highlighted with the red arrow.

The average indices for all the years are plotted in Figure 4, error bars indicating the extreme values for the total-order sensitivity indices. As expected, a small, but non-zero, value was estimated for the dummy parameter SI. For easy visualization and interpretation, the sensitivity indices are sorted in descending order. The parameters with SI values less than the SI of the dummy parameter were, thus, considered as non-influential parameters. The average SI for the Xiaotangshan location is shown in Figure 4.

3.3. Density-Based PAWN Method

Figure 5 and Figure 6 show an example of two different obtained parameters with different and similar CDFs, illustrating the behavior of non-influential and influential parameters using the PAWN method. In Figure 5 it can be seen that for the parameter *CGC* the unconditional CDF and conditional CDF are different, and the value of the KS statistics is above the critical value at 95% confidence interval. Therefore, the model output has a high sensitivity to the parameter, which can be thus considered highly influential. On the contrary, Figure 6 shows an example of another

parameter, *CDC*, with similar unconditional and conditional CDFs; the value of the KS statistic is below the critical value, and thus the parameter is non-influential and has a low SI.

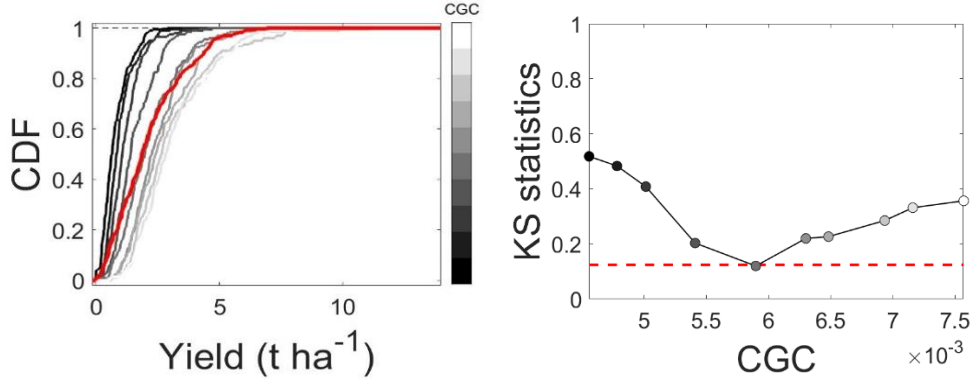


Figure 5. Cumulative Distribution Function (CDF) for the Canopy Growth Coefficient (CGC) parameter (left), red line on the left plot indicates the unconditional CDF and the grey colors represent the conditional CDF for ten different sampling points and Kolmogorov–Smirnov statistics (right), the red dashed line indicates the critical value of Kolmogorov–Smirnov (KS) statistics at 95% confidence interval.

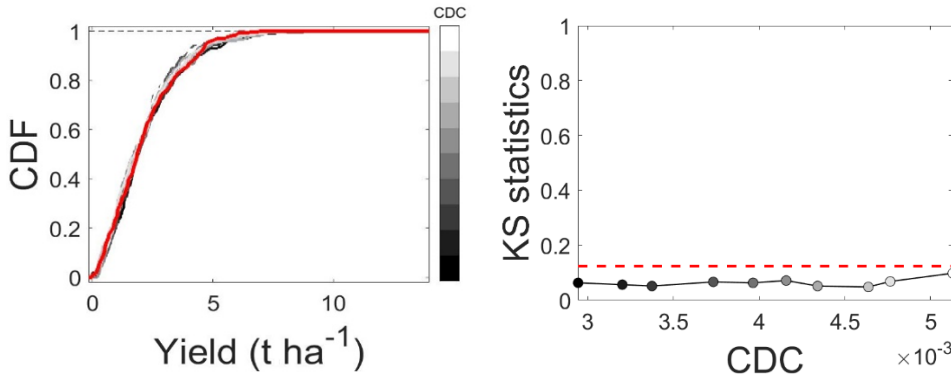


Figure 6. Cumulative Distribution Function (CDF) for the CDC parameter (left), red line on the left plot indicates the unconditional CDF and the grey colors represent the conditional CDF for ten different sampling points and Kolmogorov–Smirnov statistics (right), red dashed line indicates the critical value of Kolmogorov–Smirnov (KS) statistics at 95% confidence interval.

The PAWN index (T_i) for the different years is presented in Figure 7 for Maccarese and in Figure S3 (Supplementary Material) for Xiaotangshan. The average values of the indices for all the years are plotted in Figure 8, in which error bars represent the range of values for the PAWN sensitivity indices across the years. In comparison to EFAST, the sensitivity index appeared to be more stable. For both study sites, considering all the years of weather data, the most relevant parameter was *CGC*, also identified by EFAST method, and is also responsible for most of the variance of the model output. The second most influential parameter was *Hlstart*. However, differently from EFAST, this parameter did not have a sensitivity index higher than *CGC* even in the water-stressed conditions of 2017. Again, as expected, a small but non-zero value was estimated for the SI of the dummy parameter. The parameters with the SI less than the SI of the dummy parameter are thus non-influential parameters. The average SI for the Xiaotangshan location is shown in Figure 8.

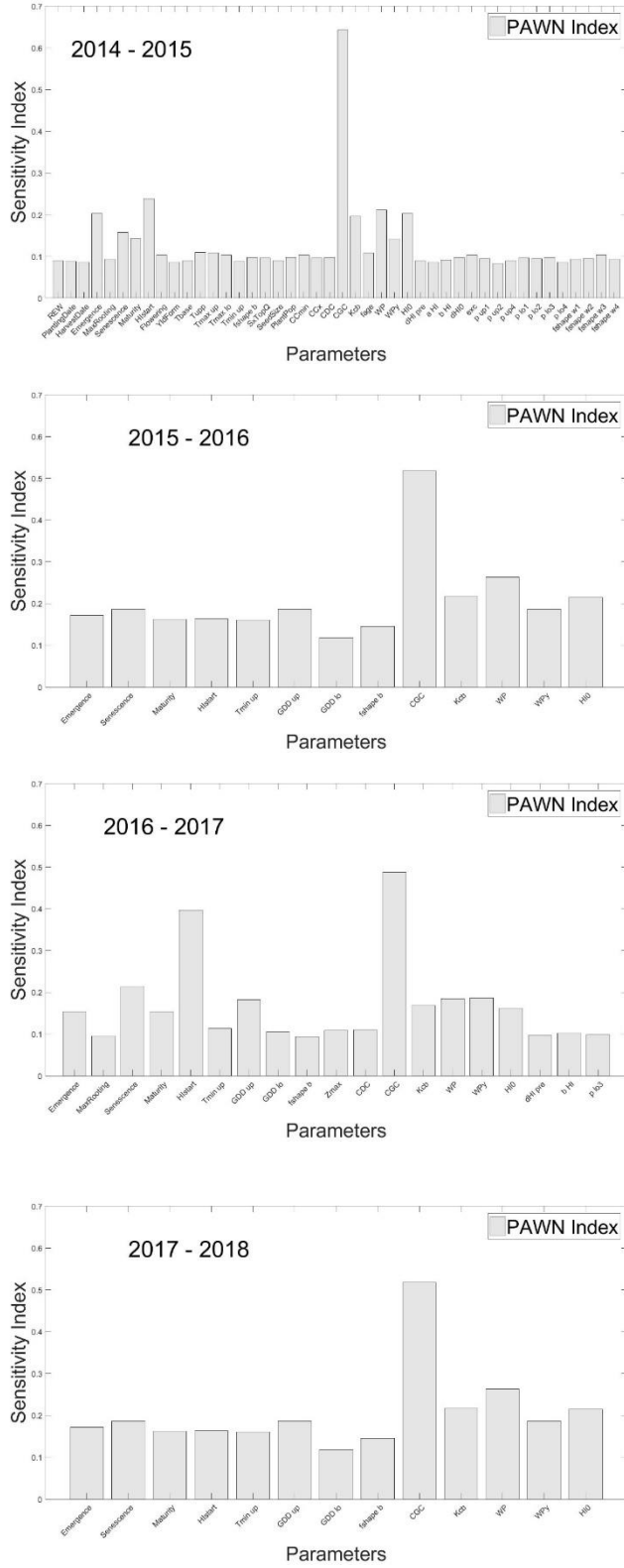


Figure 7. PAWN sensitivity index T_i for different years for the Maccarese location. Only the parameters with a total sensitivity index higher than the sensitivity index of the dummy parameter are shown. The details on the parameters can be found in Table 1.

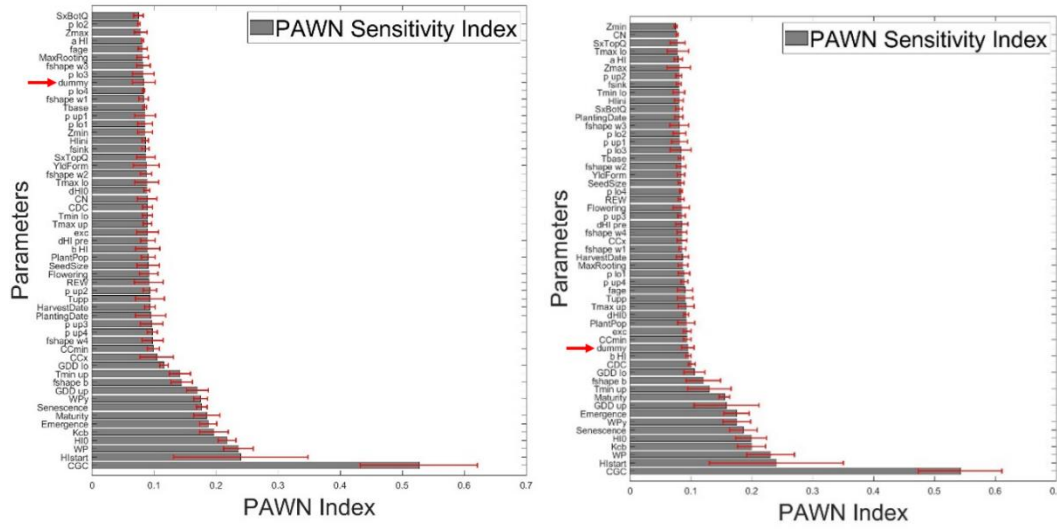


Figure 3. Average Ti using the PAWN method Maccarese test site (left) and Xiaotangshan test site (right). The error bars in red represent the extreme values of Ti. The dummy parameter is highlighted with the red arrow.

Identifying sensitive parameters using the PAWN method is suitable in the cases where the output distribution is not normal, and it follows skewed or multi-modal distribution (Pianosi and Wagener, 2015). An example of the output distribution is compared with 54 different statistical distributions. The top six best-fitted PDFs are mentioned in Table 4. The comparison is made on the basis of KS statistics and Anderson–Darling Statistics, with low values representing the best distribution. As can be seen, the Generalized Extreme Distribution has the lowest statistic value, with an Anderson–Darling statistic of 0.16401 and is second lowest in KS statistics. It is the best-suited PDF for the data, the reason for the choice of the PAWN method and is feasible to implement for identifying the most influential set of parameters for the local calibration. The Generalized Extreme Distribution is plotted in Figure 9, and statistics including the p-value and critical values at alpha (α) are presented in Table 5. The sample was from conditional CDF of the PAWN method, consisting of 200 samples, $N_c = 200$ (Section 2.1.3). It shows the statistics value and rank, indicating the raking of the distribution when compared with 54 PDFs (Table 4).

Table 7. Best probability distribution with Kolmogorov–Smirnov (KS) and Anderson–Darling statistics, ranking the best from top to bottom. Superscript in red indicates the rank of the KS and Anderson–Darling statistics.

Distribution	KS Statistics	Anderson–Darling Statistics
Gen. Extreme Values	0.03253 ⁽²⁾	0.16401 ⁽¹⁾
Johnson SB	0.03047 ⁽¹⁾	0.19169 ⁽²⁾
Gumbel Max	0.05197 ⁽³⁾	0.4357 ⁽³⁾
Logistic	0.12206 ⁽⁴⁾	3.7302 ⁽⁴⁾
Hypersecant	0.1276 ⁽⁵⁾	4.0165 ⁽⁵⁾
Normal	0.12242 ⁽⁶⁾	4.0999 ⁽⁶⁾

Similarly, for the Xiaotangshan test site, the most influential parameters are presented in Table 6, the descriptions of the parameters are mentioned in the list of AOS parameters (Table 1). The highly influential parameters for both the locations were *CGC*, *Hlstart*, *WP*, *Hl0*, *Kcb*, *Emergence*, *Senescence*, *WPy*, *GDD_up*, *Maturity*, *fshape_b*, *Tmin_up*, *GDD_lo*, and *CDC* (Figure 8 and Table 4).

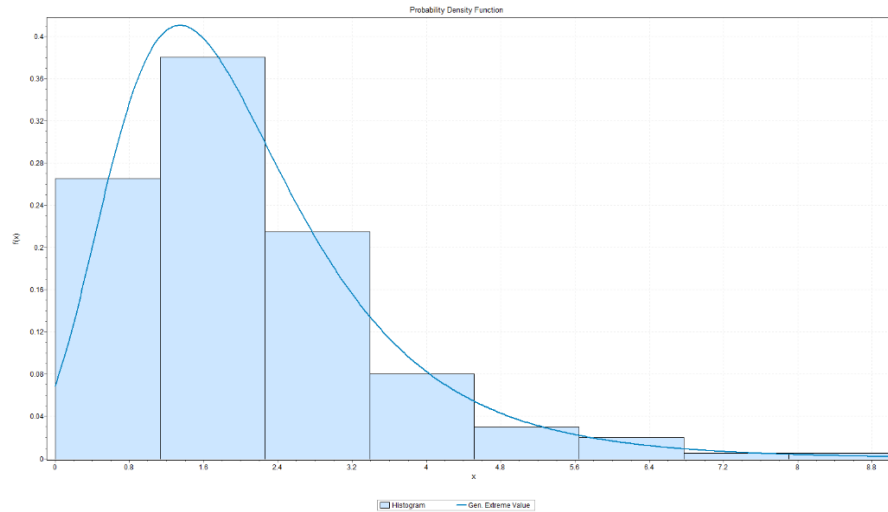


Figure 9. Generalized Extreme Values distribution fit to the output Yield from PAWN method.

Table 5. Best-fitted PDF (Generalized Extreme Values distribution) statistical summary.

Anderson–Darling Statistics					
Sample Size	200				
Statistics	0.16399				
Rank	1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Kolmogorov–Smirnov Statistics					
Sample Size	200				
Statistics	0.03251				
P-Value	0.97963				
Rank	2				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.07587	0.08648	0.09603	0.10734	0.11519

Table 68. List of sensitive parameters (based on the average SI values for all years) identified by different methods. Sorted in the decreasing order of the SI in the case of EFAST (Sti) and PAWN (Ti) method. The most influential parameters are in bold and underlined.

Study Site	Algorithm	Sensitive Parameters
Maccarese	Morris	<u>Emergence</u> , <u>Senescence</u> , <u>Maturity</u> , <u>HIstart</u> , <i>Tbase</i> , <i>Tupp</i> , <i>Tmax_up</i> , <i>Tmax_lo</i> , <i>Tmin_up</i> , <i>Tmin_lo</i> , <i>GDD_up</i> , <i>GDD_lo</i> , <i>fshape_b</i> , <i>Zmax</i> , <i>CCmin</i> , <i>CCx</i> , <i>CDC</i> , <u>CGC</u> , <u>Kcb</u> , <i>fage</i> , <u>WP</u> , <u>WPy</u> , <u>HI0</u> , <i>HIini</i> , <i>exc</i>
	EFAST	<u>CGC</u> , <u>HIstart</u> , <u>WP</u> , <u>HI0</u> , <u>Kcb</u> , <u>Emergence</u> , <u>Senescence</u> , <u>WPy</u> , <u>GDD_up</u> , <u>Maturity</u> , <u>fshape_b</u> , <u>Tmin_up</u> , <u>GDD_lo</u> , <u>CDC</u> , <u>PlantingDate</u> , <i>p_lo1</i> , <i>CN</i> , <i>MaxRooting</i> , <i>HIini</i> , <i>T_base</i> , <i>p_up2</i> , <i>Tupp</i> , <i>Tmax_up</i> , <i>b_HI</i> , <i>p_lo4</i> , <i>Flowering</i> , <i>CCmin</i> , <i>exc</i> , <i>HarvestDate</i> , <i>p_lo3</i> , <i>CCx</i> , <i>p_up4</i> , <i>REW</i> , <i>dHI_pre</i> , <i>fage</i> , <i>fsink</i> , <i>a_HI</i> , <i>fshape_w2</i> , <i>YldForm</i> , <i>SeedSize</i> , <i>Zmax</i> , <i>SxTopQ</i>
	PAWN	<u>CGC</u> , <u>HIstart</u> , <u>WP</u> , <u>Kcb</u> , <u>HI0</u> , <u>Senescence</u> , <u>WPy</u> , <u>Emergence</u> , <u>GDD_up</u> , <u>Maturity</u> , <u>Tmin_up</u> , <u>fshape_b</u> , <u>GDD_lo</u> , <u>CDC</u> , <u>b_HI</u>

Xiaotangshan	Morris	<i>REW, Emergence, Senescence, Maturity, <u>Hlstart</u>, Tbase, Tupp, Tmin_up, Tmin_lo, GDD_up, GDD_lo, fshape_b, Zmax, CCmin, CCx, CDC, <u>CGC</u>, Kcb, fage, <u>WP</u>, WPy, <u>HI0</u>, exc</i>
	EFAST	<i><u>CGC, Kcb, Hlstart, WP, HI0, WPy, GDD_up, Maturity, Senescence, Emergence, Tmin_up, CDC, fshape_b, GDD_lo, Hlini, CCmin</u>, CN, Tbase, fshape_w4, Tmin_lo, MaxRooting, Zmax, CCx, fage, HI0, dHI_pre, Zmin, fshape_w2, REW, p_up4, HarvestDate, a_HI, p_lo3</i>
	PAWN	<i><u>CGC, Hlstart, WP, HI0, Kcb, Emergence, Maturity, Senescence, WPy, GDD_up, fshape_b, Tmin_up, GDD_lo</u>, CCx, CCmin, fshape_w4, p_up4, p_up3, PlantingDate, HarvestDate, Tupp, p_up2, REW, Flowering, SeedSize, PlantPop, b_HI, dHI_pre, exc, Tmax_up, Tmin_lo, CDC, CN, dHI0, Tmax_lo, fshape_w2, YldForm, SxTopQ, fsink, Hlini, Zmin, p_lo1, p_up1, Tbase, fshape_w1, p_lo4</i>

4. Discussion

4.1. Elementary Effects Using the Morris Method

The results obtained in this study, performed using a contrasting range of climatic scenarios of winter wheat growing areas in Maccarese, Italy (Mediterranean climate) and Xiaotangshan, China (continental climate), allowed to obtain essential information on the sensitivity of the AOS model, especially required for their application within regional-scale studies (Casa et al., 2016; Jin et al., 2017; P. Silvestro et al., 2017). It is well known that the results of the sensitivity analysis studies depend on the boundary conditions chosen (Paleari and Confalonieri, 2016). In this work, these conditions are climate datasets, actual data from both sites and the range of variation of the parameters (P. C. Silvestro et al., 2017; Vanuytrecht et al., 2014; XING et al., 2017). The Morris method was used for initial screening of the parameters and to identify the influential and non-influential parameters. The influential parameters were found to be 29 for the Maccarese test site and 28 for the Xiaotangshan test site out of the total 54 model parameters considered in this study. These parameters are mentioned in Table 6, and the description of the parameters is reported in Table 1. A study (Vanuytrecht et al., 2014) on the SA of the Aquacrop model used the mean value of 0.25 as a threshold, and parameters below 0.25 t ha^{-1} were classified as non-influential parameters. It was found that the parameters (translated in AOS notation) *exc*, *dHI0*, *p_up1*, *Tmin_up*, *Tmax_up*, *p_up4*, *SxTopQ* and *SxBotQ* were non-influential ($<0.25 \text{ t ha}^{-1}$), and the other parameters *PlantPop*, *Flowering*, *HI0*, *dHI_pre*, *a_HI*, *Kcb*, *p_up2* and *fshape_w2* had low but non-negligible impacts ($0.25\text{--}1.0 \text{ t ha}^{-1}$) for all the climatic scenarios implemented. In this study, the *CGC* parameter was found to be highly influential in all the conditions for both test sites. Another study carried out by (P. C. Silvestro et al., 2017) found almost the same parameter set of the non-influential parameters as by (Vanuytrecht et al., 2014) in all the climatic conditions, with the exception of the soil curve number (*CN*) and the shape factor for water stress coefficient inducing early senescence (*fshape_w3*). It should be noted that the threshold considered in the study was 0.1 t ha^{-1} i.e., less than that used by (P. C. Silvestro et al., 2017). The tabular comparison of the sensitive parameters is presented in Table 6.

4.2. Variance-Based Extended Fourier Amplitude Sensitivity Test (EFAST)

The Morris method is only used for initial screening, and it does not quantify the higher-order interaction between the parameters. This can be assessed by using the variance-based EFAST method and the density-based PAWN method for GSA. The comparison between these algorithms is not meaningful if compared on the basis of sensitivity indices as they have different rationales and backgrounds. This work applies a comparison based on the ranking of the parameters in both methods. So, to achieve this goal, all the input parameters, i.e., 55 including a dummy parameter, are used to carry out SA in both the methods.

Figure 3 shows both the first-order and the total-order sensitivity indices for the EFAST methods for all the years considered in this study. The sensitive parameters are, in general, common in all the climate years considered, but with different values of the SI. These common parameters are *Emergence*, *Senescence*, *Maturity*, *Hlstart*, *CGC*, *Kcb*,

WP, *WPy* and *HI0*, *Tmin_up*, *GDD_up* and *fshape_b*. The average sensitivity indices plots for all the years identify the non-influential parameters having a sensitivity less than the sensitivity of the dummy parameter. As can be seen from Figure 4, *dHI0*, *p_up3*, *p_lo2*, *Tmax_lo*, *p_up1*, *PlantPop*, *Tmin_lo*, *Zmin*, *fshape_w2*, *fshape_w3*, *fshape_w4* and *SxBotQ* had low sensitivities in comparison to the dummy parameter. These parameters are clearly identified as non-influential parameters.

Similarly, for the Xiaotangshan test site, the most influential parameters are reported in Table 6. The highly influential parameters for both the locations were *CGC*, *HIstart*, *WP*, *HI0*, *Kcb*, *Emergence*, *Senescence*, *WPy*, *GDD_up*, *Maturity*, *fshape_b*, *Tmin_up*, *GDD_lo* and *CDC* (Figure 4). These parameters are in common for both the test sites, although the ranking is not the same. Authors in (Vanuytrecht et al., 2014) identified that for winter wheat in a temperate maritime climate, many parameters had an equivalent effect on the output variance, among which parameters describing canopy development were *SeedSize*, *CCx*, *CGC*, *Senescence*, *WP*, *GDD_up* and *Zmax*. In a previous study (P. C. Silvestro et al., 2017), it was found that the parameters describing the phenological cycle of the crop and the water stress were among the most highly influential. The ranking of the parameters was different from that of (Vanuytrecht et al., 2014), due to the climatic conditions employed. Also, the total-order effects were different than those of the first-order effects, indicating more complex higher-order interacting effects rather than individual effects. It should be noted that both these studies applied EFAST on the results of the Morris method, that is not the case of our study. The EFAST method for SA was directly applied to all the parameters in (XING et al., 2017) under different irrigation treatments. In the presence of rainfall, *CGC*, *CCx*, *Emergence*, *PlantPop*, *SeedSize*, *Kcb*, *Maturity*, *Zmax*, *p_up3*, *MaxRooting*, *GDD_up* and *WP* were the most sensitive to yield. In normal irrigation, *CGC*, *CCx*, *Emergence*, *PlantPop*, *SeedSize*, *Kcb*, *Maturity*, *Zmax* and *p_up3* were among the highly influential, and in over-irrigation, *CGC*, *CCx*, *Emergence*, *PlantPop*, *SeedSize*, *Kcb*, *Maturity*, *Zmax*, *p_up3* and *CDC* were the most sensitive parameters. The translations of these parameters of Aquacrop and AOS model are mentioned in Table 1.

4.3. Density-Based PAWN Method

In the PAWN method, the sensitivity of the model output (yield) to the parameter x_i is measured by the distance between the conditional CDF of output obtained by fixing x_i to a specific value and the unconditional CDF. Since all the variability related to x_i , both due to direct and interaction effects, is removed in the conditional CDF, the comparison illustrates the total effect of x_i . As a distance measure, the maximum absolute difference between two CDFs, i.e., the Kolmogorov–Smirnov statistic, is considered. The PAWN SI for the different years for Maccarese location is presented in Figure 7. The sensitive parameters are similar for all different years, but with a varying degree of the PAWN index. The average index for all the years is plotted in Figure 8. Table 6 summarizes the set of sensitive parameters based on the PAWN index.

The PAWN method is simpler and less computationally demanding in comparison to the EFAST method. Out of all the methods tested in this study, the Morris method was the simplest and required the lowest computational time, but the disadvantage with the method is that it does not quantify higher-order effects. The PAWN method, despite being simple, has two problems: first is the reliability of the CDF produced, which may become unstable as can be observed for both study sites. Fifteen parameters were identified as most influential for Maccarese study site, whereas 46 parameters appeared to have sensitivity higher than the sensitivity of the dummy parameter for the Xiaotangshan study site. However, the parameters with the highest sensitivity were in common for both the study sites. The other drawback of PAWN is the use of KS statistics to compute SI that leads to slow convergence.

4.4. Sensitive Parameters in Relation to Model Processes and Agronomic Conditions

AOS requires calibration of parameters to properly simulate crop growth, especially in new locations: SA can be used to identify the parameters that have the highest impact on crop yield, before proceeding to calibration. The interpretation of the SA results, in light of previously published literature and of the knowledge of model calculations and simulated processes, can provide useful insights for advanced as well as basic use of the model by researchers. Aquacrop parameters are divided into two main groups: conservative and non-conservative parameters. Conservative

crop parameters do not change with time, cultivar, location nor water regime and in principle should require no adjustment to local conditions. If simulations do not match observed values, the first thing to look at is the fertility regime before changing the parameters. Non-conservative parameters, on the other hand, are dependent on the cultivar (e.g., phenology parameters) and conditions.

In our case, both in Maccarese and in Xiaotangshan, both methods (EFAST and PAWN) found *CGC*, *Hlstart*, *WP*, *HlO* and *Kcb* to be the first five most sensitive parameters (Table 6), even though the order changed slightly between locations and methods. *CGC*, *WP*, *HlO* and *Kcb* are conservative parameters that showed a high impact on crop yield, while *Hlstart* is a phenological non-conservative parameter. These parameters are mainly related to crop transpiration, which is directly used by AOS to calculate biomass accumulation (AOS is a water-driven model). *CGC* describes how fast the canopy cover (*CC*) increases (Equation (7)). *CC* and *Kcb* are directly involved in the estimation of crop transpiration, which in turns determines biomass production through water productivity (*WP*). Grain yield is also highly influenced by the harvest index (*HlO*) and by the thermal time to start yield formation (*Hlstart*). High fluctuations of yield estimations can therefore be expected when adjusting these parameters, and model users should be careful in this regard. Moreover, these parameters are also used to mimic the effects of fertility stress on crop growth in a semiquantitative way, since nutrient cycling is not directly incorporated into the model (Guo et al., 2019). Testing parameter sensitivity under stress conditions is important for calibration purposes (Jin et al., 2018).

Other phenological parameters (e.g., *Emergence*, *Senescence* and *Maturity*) followed up in the SA with both locations and methods. Authors in (Zhang et al., 2019) also included *Emergence* and *Maturity* as sensitive parameters for winter wheat cultivation in China, using the EFAST method. Authors in (P. Silvestro et al., 2017) found among the most influential parameters *Senescence*, *Hlstart* and *Maturity* in winter wheat cultivated in a location in Central Italy (in the same region of Maccarese) and two locations in China (with Xiaotangshan being one of them). Similarly, authors in found *Emergency*, *Senescence* and *Maturity* to be among the highest influential parameters for winter wheat in Belgium. Authors in (XING et al., 2017) also reported *Hlstart* and *Maturity* among influential parameters for wheat cultivated near Beijing (using the EFAST method).

Water stress parameters (*p_{up}*, *p_{lo}* and *fshape_p*) are often the target of SA and calibration of Aquacrop. However, datasets for water-deficient conditions are usually needed to properly estimate the impact of these parameters on crop yield (Abdalhi and Jia, 2018). In this study, winter wheat was rainfed and not irrigated. Indeed, in the two locations considered, rainfall is frequent and evapotranspiration is low during the growing period (Figure 1), so it is not surprising that water stress parameters had lower sensitivity with respect to the canopy and phenological parameters for both locations and methods (Table 4). There are several examples that support this argument. Authors in (Vanuytrecht et al., 2014) found water stress parameters to have a higher sensitivity order for winter wheat under drier conditions and more sandy soils. Authors in (Saab et al., 2015) simulated satisfactorily barley growth in southern Italy under rainfed conditions without the need to adjust water stress parameters for each year, while other crop growth and phenological parameters were specifically changed over the years. Authors in (Salemi et al., 2011) obtained good simulations of winter wheat yield in Iran under moderate water stress without including water stress parameters in their SA and calibration. The SA carried out by (XING et al., 2017) highlighted that the parameters with higher-order sensitivity to winter wheat yield in Beijing were mainly related to phenology, biomass accumulation and canopy growth, with *p_{up3}* (the threshold for water stress-related senescence) being the only water stress-related parameter included. In another recent global SA (Guo et al., 2019), maize water stress parameters were less sensitive than crop growth and phenology parameters, under all the irrigation treatments (designed to obtain a range of water stress intensities, from no to moderate stress). These experiences confirmed our results and highlighted that water stress parameters do not have high impact on winter wheat yield when water-limiting conditions are moderate.

5. Conclusions

In this paper, we compared the application of two global SA techniques, the variance-based EFAST and the moment-independent density-based PAWN method, to the analysis of 54 parameters of the AOS model, a crop growth

simulation model widely used for canopy cover, biomass and yield simulations. With respect to the widely used variance-based sensitivity indices, the main advantage of PAWN is that it can be applied whatever the output distribution may be. The comparison was performed in terms of parameter ranking for different climates and geographical locations. In all the methods, highly influential parameters were common, and the ranking of these parameters was similar in EFAST and PAWN.

The PAWN method could be considered more adequate in our study, since the frequency distribution of the output variable was right-skewed. The use of a dummy parameter approach as a viable option to set a threshold value for parameter screening was demonstrated for all the methods considered. The results reported in this study provide insight into identifying the most influential parameters of the AOS model for two different geographic locations with two different climates.

This set of most influential parameters can be used for model simplification and calibration of the most sensitive parameters. Simplification of the model is necessary for regional-scale studies and particularly for remote sensing data assimilation. Both EFAST and PAWN methods identified the most influential parameters for all different years for both the test sites, but with a varied sensitivity and ranking. According to the PAWN method, the highly influential parameters belonged to crop parameters specifically, to the canopy and phenological development (*CGC*, *Hlstart*, *Senescence*, *maturity*), biomass/yield production (*WP*, *WPy*) and temperature stress (*Tmin_up*, *Tmin_lo*, *GDD_up*) rather than root development and water stress.

This study thus concludes that variance-based and density-based global SA methods are both useful, with the condition that the variance-based method should be preferable where a variance is an adequate proxy of the model output, i.e., where the model output distribution is normal. Although this work is limited to compare these two (EFAST and PAWN) methods on parameter ranking, for two different climates with seven years data overall, to verify the robustness of the comparison, it can be extended on the basis of efficiency, convergence or computational cost with other geographic locations, climates and crops.

References

- Abdalhi, M.A., Jia, Z., 2018. Crop yield and water saving potential for AquaCrop model under full and deficit irrigation managements. *Italian Journal of Agronomy* 13, 1288.
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998a. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome 300, D05109.
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998b. FAO Irrigation and drainage paper No. 56. Rome: Food and Agriculture Organization of the United Nations 56, e156.
- Bekele, E.G., Nicklow, J.W., 2007. Multi-objective automatic calibration of SWAT using NSGA-II. *Journal of Hydrology* 341, 165–176. <https://doi.org/10.1016/j.jhydrol.2007.05.014>
- Borgonovo, E., 2007. A new uncertainty importance measure. *Reliability Engineering & System Safety* 92, 771–784.
- Borgonovo, E., Castaings, W., Tarantola, S., 2011. Moment independent importance measures: new results and analytical test cases. *Risk Analysis: An International Journal* 31, 404–428.
- Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis of large models. *Environmental modelling & software* 22, 1509–1518.

- Cariboni, J., Gatelli, D., Liska, R., Saltelli, A., 2007. The role of sensitivity analysis in ecological modelling. *Ecological modelling* 203, 167–182.
- Casa, R., Silvestro, P.C., Yang, H., Pignatti, S., Pascucci, S., Yang, G., 2016. Assimilation of remotely sensed canopy variables into crop models for an assessment of drought-related yield losses: A comparison of models of different complexity. Presented at the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, pp. 5925–5928.
- Chen, Y., Cournède, P.-H., 2014. Data assimilation to reduce uncertainty of crop model prediction with convolution particle filtering. *Ecological Modelling* 290, 165–177.
- Cukier, R., Levine, H., Shuler, K., 1978. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of computational physics* 26, 1–42.
- de Willigen, P., 1991. Nitrogen turnover in the soil-crop system; comparison of fourteen simulation models. *Fertilizer research* 27, 141–149. <https://doi.org/10.1007/BF01051122>
- Doorenbos, J., Kassam, A., 1979. FAO irrigation and drainage paper No. 33 “Yield response to water.” FAO–Food and Agriculture Organization of the United Nations, Rome.
- Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* 28, 1015–1031. <https://doi.org/10.1029/91WR02985>
- Ekstrom, P.-A., 2005. Eikos: a simulation toolbox for sensitivity analysis in matlab. Uppsala University, Uppsala.
- Foster, T., 2019. AquaCrop-OS v6.0a Reference Manual.
- Foster, Tim, 2016. Supplementary information for ‘AquaCrop-OS: An open source version [WWW Document]’. URL <https://ars.els-cdn.com/content/image/1-s2.0-S0378377416304589-mmc1.pdf> (accessed 4.22.20).
- Foster, Timothy, 2016. AquaCrop-OS v5.0a Reference Manual.
- Foster, T., Brozović, N., Butler, A.P., Neale, C.M.U., Raes, D., Steduto, P., Fereres, E., Hsiao, T.C., 2017. AquaCrop-OS: An open source version of FAO’s crop water productivity model. *Agricultural water management* 181, 18–22.
- Gervois, S., de Noblet-Ducoudré, N., Viovy, N., Ciais, P., Brisson, N., Seguin, B., Perrier, A., 2004. Including croplands in a global biosphere model: methodology and evaluation at specific sites. *Earth Interactions* 8, 1–25.
- Guo, D., Zhao, R., Xing, X., Ma, X., 2019. Global sensitivity and uncertainty analysis of the AquaCrop model for maize under different irrigation and fertilizer management conditions. *Archives of Agronomy and Soil Science* 1–19.
- Hopmans, J.W., Bristow, K.L., 2002. Current capabilities and future needs of root water and nutrient uptake modeling, in: *Advances in Agronomy*. pp. 103–183.
- Iooss, B., Lemaître, P., 2015. A review on global sensitivity analysis methods, in: *Uncertainty Management in Simulation-Optimization of Complex Systems*. Springer, pp. 101–122.

- Jin, X., Feng, H., Zhu, X., Li, Z., Song, S., Song, X., Yang, G., Xu, X., Guo, W., 2014. Assessment of the AquaCrop model for use in simulation of irrigated winter wheat canopy cover, biomass, and grain yield in the North China Plain. *PloS one* 9, e86938.
- Jin, X., Li, Z., Nie, C., Xu, X., Feng, H., Guo, W., Wang, J., 2018. Parameter sensitivity analysis of the AquaCrop model based on extended fourier amplitude sensitivity under different agro-meteorological conditions and application. *Field crops research* 226, 1–15.
- Jin, X., Li, Z., Yang, G., Yang, H., Feng, H., Xu, X., Wang, J., Li, X., Luo, J., 2017. Winter wheat yield estimation based on multi-source medium resolution optical and radar imaging data and the AquaCrop model using the particle swarm optimization algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing* 126, 24–37.
- Kolmogorov, A., 1933. Sulla determinazione empirica di una legge di distribuzione. *Inst. Ital. Attuari, Giorn.* 4, 83–91.
- Liu, H., Chen, W., Sudjianto, A., 2005. Relative entropy based method for probabilistic sensitivity analysis in engineering design.
- Makowski, D., Hillier, J., Wallach, D., Andrieu, B., Jeuffroy, M., 2006. Parameter estimation for crop models. *Working with dynamic crop models*. Elsevier, Amsterdam 101–149.
- McMaster, G.S., Wilhelm, W., 1997. Growing degree-days: one equation, two interpretations.
- Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. *Technometrics* 33, 161–174.
- Nossent, J., Elsen, P., Bauwens, W., 2011. Sobol' sensitivity analysis of a complex environmental model. *Environmental Modelling & Software* 26, 1515–1525. <https://doi.org/10.1016/j.envsoft.2011.08.010>
- Paleari, L., Confalonieri, R., 2016. Sensitivity analysis of a sensitivity analysis: We are likely overlooking the impact of distributional assumptions. *Ecological Modelling* 340, 57–63.
- Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and earth system sciences discussions* 4, 439–473.
- Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software* 79, 214–232. <https://doi.org/10.1016/j.envsoft.2016.02.008>
- Pianosi, F., Sarrazin, F., Wagener, T., 2015. A Matlab toolbox for global sensitivity analysis. *Environmental Modelling & Software* 70, 80–85.
- Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software* 67, 1–11. <https://doi.org/10.1016/j.envsoft.2015.01.004>
- Raes, D., Munoz, G., 2009. The ETo Calculator. Reference Manual Version 3.
- RAES, D., STEDUTO, P., HSIAO, T., FERERES, E., 2018. Chapter 1: FAO crop-water productivity model to simulate yield response to water: AquaCrop: version 6.0-6.1: reference manual.
- Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2009. AquaCrop-The FAO crop model to simulate yield response to water. FAO Land and Water Division, FAO, Rome.

- Ritchie, J.T., 1972. Model for predicting evaporation from a row crop with incomplete cover. *Water resources research* 8, 1204–1213.
- Saab, M.T.A., Todorovic, M., Albrizio, R., 2015. Comparing AquaCrop and CropSyst models in simulating barley growth and yield under different water and nitrogen regimes. Does calibration year influence the performance of crop growth models? *Agricultural water management* 147, 21–33.
- Salemi, H., Soom, M.A.M., Lee, T.S., Mousavi, S.F., Ganji, A., Yusoff, M.K., 2011. Application of AquaCrop model in deficit irrigation management of winter wheat in arid region. *African Journal of Agricultural Research* 610, 2204–2215.
- Saltelli, A., 2002. Sensitivity analysis for importance assessment. *Risk analysis* 22, 579–590.
- Saltelli, A., Chan, K., Scott, M., 2000a. Sensitivity analysis. Probability and statistics series. John and Wiley & Sons, New York.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global sensitivity analysis: the primer. John Wiley & Sons.
- Saltelli, A., Tarantola, S., Campolongo, F., 2000b. Sensitivity analysis as an ingredient of modeling. *Statistical Science* 15, 377–395.
- Saltelli, A., Tarantola, S., Chan, K.-S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* 41, 39–56.
- Silvestro, P., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., Huang, W., Casa, R., 2017. Estimating wheat yield in China at the field and district scale from the assimilation of satellite data into the Aquacrop and simple algorithm for yield (SAFY) models. *Remote Sensing* 9, 509.
- Silvestro, P.C., Pignatti, S., Yang, H., Yang, G., Pascucci, S., Castaldi, F., Casa, R., 2017. Sensitivity analysis of the Aquacrop and SAFYE crop models for the assessment of water limited winter wheat yield in regional scale applications. *PloS one* 12, e0187485.
- Smirnov, N.V., 1939. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou* 2, 3–14.
- Sobol, I.M., 1993. Sensitivity estimates for nonlinear mathematical models. *Mathematical modelling and computational experiments* 1, 407–414.
- Sobol', I.M., 1990. On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe modelirovanie* 2, 112–118.
- Steduto, P., Hsiao, T.C., Fereres, E., 2007. On the conservative behavior of biomass water productivity. *Irrigation Science* 25, 189–207.
- Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop—The FAO crop model to simulate yield response to water: I. Concepts and underlying principles. *Agronomy Journal* 101, 426–437.
- Tanner, C.B., Sinclair, T.R., 1983. Efficient water use in crop production: Research or re-search? Limitations to efficient water use in crop production 1–27.
- Todorovic, M., Albrizio, R., Zivotic, L., Saab, M.-T.A., Stöckle, C., Steduto, P., 2009. Assessment of AquaCrop, CropSyst, and WOFOST models in the simulation of sunflower growth under different water regimes. *Agronomy Journal* 101, 509–521.

- Upreti, D., Huang, W., Kong, W., Pascucci, S., Pignatti, S., Zhou, X., Ye, H., Casa, R., 2019. A comparison of hybrid machine learning algorithms for the retrieval of wheat biophysical variables from sentinel-2. *Remote Sensing* 11, 481.
- van Griensven, A. van, Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., Srinivasan, R., 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of hydrology* 324, 10–23.
- Vanuytrecht, E., Raes, D., Willems, P., 2014. Global sensitivity analysis of yield output from the water productivity model. *Environmental Modelling & Software* 51, 323–332.
- Vrugt, J.A., Bouten, W., Gupta, H.V., Sorooshian, S., 2002. Toward improved identifiability of hydrologic model parameters: The information content of experimental data. *Water Resources Research* 38, 48–61. <https://doi.org/10.1029/2001WR001118>
- Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* 39, 1–14.
- Wallach, D., Goffinet, B., Bergez, J.-E., Debaeke, P., Leenhardt, D., Aubertot, J.-N., 2001. Parameter estimation for crop models. *Agronomy journal* 93, 757–766.
- Wallach, D., Makowski, D., Jones, J.W., Brun, F., 2014. Chapter 6-Parameter Estimation with Classical Methods (Model Calibration), *Working with Dynamic Crop Models*. Academic Press, San Diego.
- Xiangxiang, W., Quanjiu, W., Jun, F., Qiuping, F., 2013. Evaluation of the AquaCrop model for simulating the impact of water deficits and different irrigation regimes on the biomass and yield of winter wheat grown on China's Loess Plateau. *Agricultural Water Management* 129, 95–104.
- Xing, H., Li, Z., Xu, X., Feng, H., Yang, G., Chen, Z., 2017. Multi-assimilation methods based on AquaCrop model and remote sensing data. *Transactions of the Chinese Society of Agricultural Engineering* 33, 183–192.
- XING, H., XU, X., LI, Z., CHEN, Y., FENG, H., YANG, G., CHEN, Z., 2017. Global sensitivity analysis of the AquaCrop model for winter wheat under different water treatments based on the extended Fourier amplitude sensitivity test. *Journal of Integrative Agriculture* 16, 2444–2458.
- Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. *Journal of Hydrology* 204, 83–97. [https://doi.org/10.1016/S0022-1694\(97\)00107-8](https://doi.org/10.1016/S0022-1694(97)00107-8)
- Zadeh, F.K., Nossent, J., Sarrazin, F., Pianosi, F., Griensven, A. van, Wagener, T., Bauwens, W., 2017. Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model. *Environmental Modelling & Software* 91, 210–222. <https://doi.org/10.1016/j.envsoft.2017.02.001>
- Zhang, T., Su, J., Liu, C., Chen, W.-H., 2019. Bayesian calibration of AquaCrop model for winter wheat by assimilating UAV multi-spectral images. *Computers and Electronics in Agriculture* 167, 105052.

Figure S1 (Supplementary Materials)

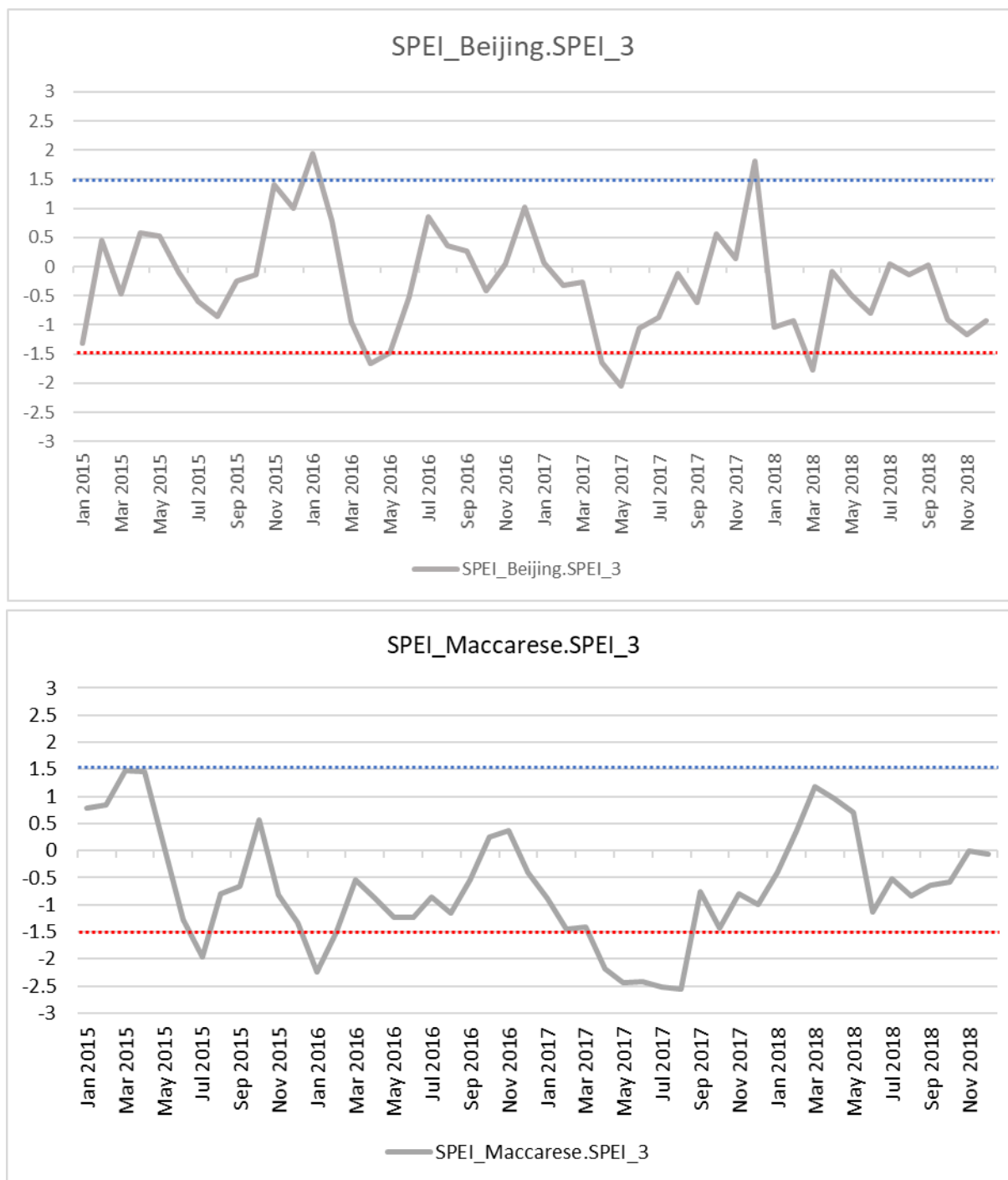
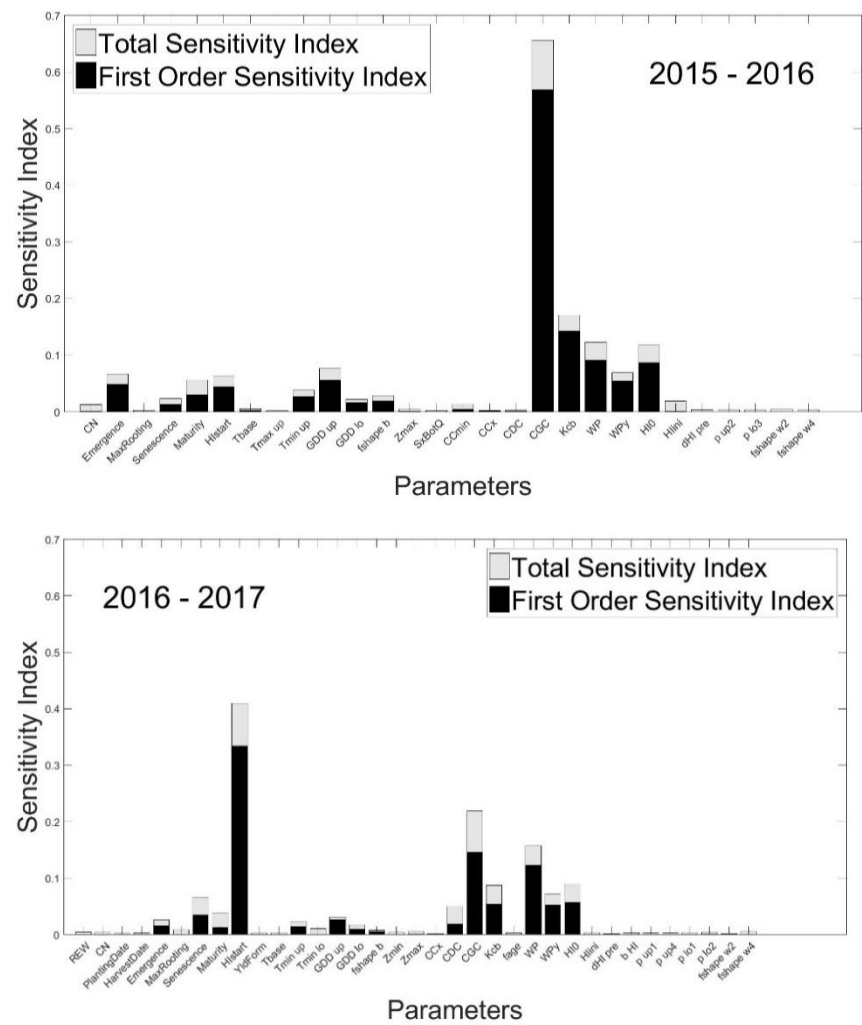


Figure S1. Standardised Precipitation-Evapotranspiration Index 3-months cumulative values trends for the Maccaresse and Xiaotangshan sites for the periods used in the study (downloaded from <https://spei.csic.es/index.html>). Dotted lines show thresholds for Extremely Wet $SPEI > 1.5$ (blue) or Severely Dry $SPEI < -1.5$ (red) conditions.

Figure S2



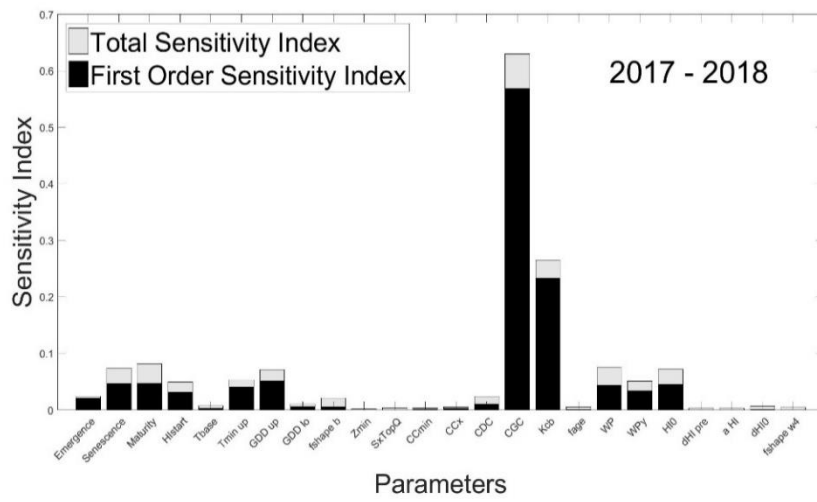
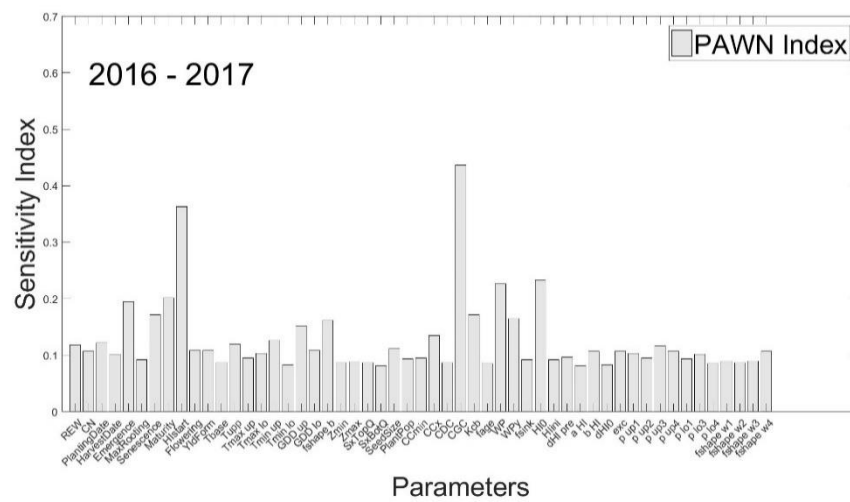
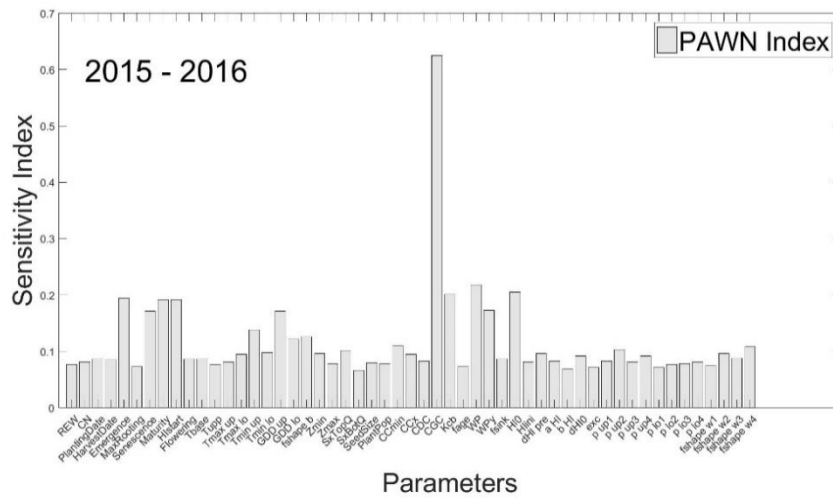


Figure S2. First order and total order EFAST SI for different years for Xiaotansghan. Only the parameters with a total sensitivity index higher than the sensitivity index of the dummy parameter are shown. The details on the parameters can be found in Table 1.

Figure S3



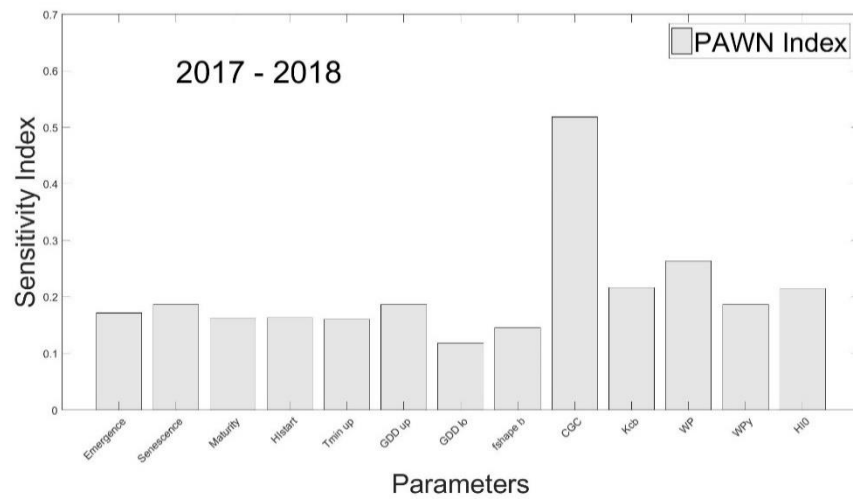


Figure S3. PAWN sensitivity index T_i for different years for Xiaotansghan. Only the parameters with a total sensitivity index higher than the sensitivity index of the dummy parameter are shown. The details on the parameters can be found in Table 1.

CHAPTER 3

3. CHAPTER 3. Bayesian Calibration of the Aquacrop-OS Model for Durum Wheat by Assimilation of Canopy Cover Retrieved from VENμS Satellite Data

(Published in Remote Sensing 2020, 12 (16), 2666)

Deepak Upreti ^{1,*}, Stefano Pignatti ², Simone Pascucci ², Massimo Tolomio ¹, Wenjiang Huang ³ and Raffaele Casa ¹

¹ Università della Tuscia, Via San Camillo de Lellis, 01100, Viterbo (Italy); massimo.tolomio@unitus.it(M.T.); rcasa@unitus.it(R.C.)

² Consiglio Nazionale delle Ricerche, Institute of Methodologies for Environmental Analysis (CNR, IMAA), Via del Fosso del Cavaliere, 100, 00133, Rome (Italy); stefano.pignatti@cnr.it(Stefano Pignatti), simone.pascucci@cnr.it(Simone Pascucci)

³ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; huangwj@aircas.ac.cn(W.H.)

* Correspondence: dupreti@unitus.it; Tel.: +39-0761-357555

Received: 22 July 2020; Revised: 13 August 2020; Accepted: 17 August 2020; Published: 18 August 2020

ABSTRACT

Crop growth models play an important role in agriculture management, allowing, for example, the spatialized estimation of crop yield information. However, crop model parameter calibration is a mandatory step for their application. The present work focused on the regional calibration of the Aquacrop-OS (AOS) model for durum wheat by assimilating high spatial and temporal resolution canopy cover data retrieved from VENμS satellite images. The assimilation procedure was implemented using the Bayesian approach with the recent implementation of the Markov chain Monte Carlo (MCMC)-based Differential Evolution Adaptive Metropolis (DREAM) algorithm DREAM(KZS). The fraction of vegetation cover (*fvc*) was retrieved from the VENμS satellite images for two years, during the durum wheat growing seasons of 2018 and 2019 in Central Italy. The retrieval was based on a hybrid method using PROSAIL Radiative Transfer Model (RTM) simulations for training a Gaussian Process Regression (GPR) algorithm, combined with Active Learning to reduce the computational cost. The AOS model was calibrated with the *fvc* data of 2017–2018 for the Maccarese farm in Central Italy and validated with the 2018–2019 data. The retrieval accuracy of the *fvc* from the VENμS images were the Coefficient of Determination (R^2) = 0.76, Root Mean Square Error (RMSE) = 0.09, and Relative Root Mean Square Error (RRMSE) = 11.6%, when compared with the ground-measured *fvc*. The MCMC results are presented in terms of Gelman–Rubin R statistics and MR statistics, Markov chains, and marginal posterior distribution functions, which are summarized with the mean values for the most sensitive crop parameters of the AOS model subjected to calibration. When validating for the *fvc*, the R^2 of the model for year (2018–2019) ranged from 0.69 to 0.86. The RMSE, Relative Error (RE), Relative Variability (α), and Relative Bias (β) ranged from 0.15 to 0.44, 0.19 to 2.79, 0.84 to 1.45, and 0.91 to 1.95, respectively. The present work shows the importance of the calibration of the AOS crop water productivity model for durum wheat by assimilating remote sensing information from VENμS satellite data.

Keywords: Aquacrop, AOS, Bayesian calibration, Markov chain, Monte Carlo, DREAM, DREAM_{KZS}

1. Introduction

Spatial crop yield information is valuable both for crop management and for market forecasting (Lobell, 2013). These data only reside in the administrative units. The data is useful but does not provide information on the scale of the field nor on the variability within the field, which is essential to inform field-specific decision-making for the different sectors, including farmers, crop insurance companies, and food trading agencies (Azzari et al., 2017). Clear, large-scale yield estimates at the individual field level are therefore an important activity for research and application in agriculture (Zaks and Kucharik, 2011).

Remote sensing-based crop yield estimation methods roughly fall into two categories: empirical or based on crop models, the latter including data assimilation approaches (Bolton and Friedl, 2013; Doraiswamy et al., 2005; Lobell et al., 2003; Thorp et al., 2012). Empirical methods generally require training of parametric or non-parametric regression algorithms with in situ data, which are generally available at a coarse resolution. Thus, these methods provide yield data at a coarse resolution (Azzari et al., 2017; X. Jin et al., 2017). A few studies (Azzari et al., 2017; Z. Jin et al., 2017; Lobell et al., 2015) attempted to produce fine-resolution yield maps using Landsat satellite data by training machine learning regression algorithms with the synthetic yield data; these methods were computationally efficient but suffered from potential bias, especially when the synthetic samples were not adequate representatives of the real-world conditions.

Simulation of crop growth models is an important tool for research and management in ecology, agriculture, and the environment. These models are used to simulate soil-cultivation system actions in response to climate and agricultural management (de Willigen, 1991; Gervois et al., 2004; Hopmans and Bristow, 2002). Agricultural processes, crop interaction with the atmosphere, and soil over time are described by theoretically based mathematical equations and empirical studies. Such equations and parametrizations necessarily require truth assumptions, leading to confusion and inaccuracy of the output variables (Saltelli et al., 2000). Regardless of the model used, the identification of the parameters that most affect the output is a fundamental problem for most of the environmental applications when great uncertainty is expected about their values, for example in regional applications (Makowski et al., 2006; P. C. Silvestro et al., 2017; Wallach et al., 2001). The system parameters, therefore, need to be calibrated before they refer to regional scenarios. To this end, the integration of crop growth models and remote sensing data is attracting even greater interest in research (Jin et al., 2018). Noteworthy, higher spatial and temporal resolution remote sensing data improves the predictive power of the calibrated model.

The most common approach for model calibration is based on optimization techniques. Various optimization methods have been used and the crop model parameters are typically obtained after the cost function between the simulations and the measured data has been minimized (Jin et al., 2018). There have been different algorithms implemented for model calibration, e.g., particle swarm optimization, simplex, least square, genetic algorithms and shuffled complex evolution, and four-dimensional variational data assimilation (4D-Var) (He et al., 2015; Jeon et al., 2014; Jin et al., 2016, 2018; Jasper A. Vrugt et al., 2003). One of the drawbacks of these algorithms is that they can get trapped into local optima due to the complexity of the optimization problem. Additionally, they only provide point estimates of the parameters, without indications of their uncertainty, e.g., not providing confidence intervals. Calibration techniques that follow a Bayesian formalism, as an alternative optimization method, include a posterior distribution of the model parameters. These can be derived by conditioning the model's spatio-temporal behavior on measurements of the system response observed (Vrugt, 2016). If the prior distribution and the probability function have been established, the posterior distribution can be summarized by, for example, the mean, covariance, or percentiles of individual parameters. Summarizing the posterior distribution is not analytically feasible for complex dynamic system models, such as crop models, which are highly non-linear, and Monte Carlo (MC) simulation methods are needed to generate a sample of the posterior distribution (Vrugt, 2016).

The earliest Markov chain Monte Carlo (MCMC) method, i.e., the random walk Metropolis (RWM) algorithm, was developed by (Metropolis et al., 1953). In this method, a candidate parameter value is initially sampled from a

distribution of asymmetric proposals, centered around the current state of the Markov chain. The Metropolis probability is then computed for this candidate. When the candidate is approved, the chain must switch to the new candidate otherwise the chain must remain in the current state. By repeating this procedure, the Markov chain results in an equilibrium distribution of the parameter. In (Hastings, 1970), the RWM algorithm was generalized by implementing an unequal probability density of the forward and backward jumps; it has become one of the most revolutionary algorithms in computational statistics and is called the Metropolis–Hastings (MH) algorithm (Beichl and Sullivan, 2000).

In recent decades, much work has been dedicated to enhancing the performance of the MCMC methods and enhancing the original RWM and MH algorithms. These are multi-chain and single-chain approaches. Single-chain methods work a single trajectory and multi-chain methods use various parallel trajectories to investigate the posterior distributions of the targets. The most common single-chain approaches are adaptive proposals (Haario et al., 1999), adaptive Metropolis (Haario et al., 2001), and delayed rejection adaptive Metropolis (Haario et al., 2006). The use of multiple chains provides robust protection against premature convergence and provides a wide range of statistical measures to test whether convergence has been achieved to a restrictive distribution (Gelman and Rubin, 1992). The most widely used multiple-chain methods are the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA) (Jasper A Vrugt et al., 2003) and Differential Evolution Markov Chain (DE-MC) (Ter Braak, 2006). Using adaptive randomized subspace sampling, multiple chain pairs for proposal development, and explicit consideration of aberrant trajectories, the efficiency of the DE-MC approach has been increased. This new method called Differential Evolution Adaptive Metropolis (DREAM) (Vrugt et al., 2009) maintains the detailed balance and ergodicity and has demonstrated excellent performance on a wide range of issues, including nonlinearity, high dimensionality, and multimodality (Laloy and Vrugt, 2012).

Some extensions were also developed based on DREAM, e.g., $\text{DREAM}_{(ZS)}$ and $\text{MT-DREAM}_{(ZS)}$, suitable for high-dimensional problems (Laloy et al., 2013; Laloy and Vrugt, 2012), as well as $\text{DREAM}_{(D)}$ for discrete and combinatorial posterior distribution problems (Vrugt and Ter Braak, 2011) and $\text{DREAM}_{(ABC)}$ for diagnostic model evaluation (Sadegh and Vrugt, 2014), etc.

We use $\text{DREAM}_{(KZS)}$ in the present work, which is an extension to $\text{DREAM}_{(ZS)}$ and explicitly designed to accelerate the convergence of $\text{DREAM}_{(ZS)}$ in exploring the high-dimensional target distributions. The new proposal distribution is based on the simpler ensemble updating scheme (Evensen, 2009), which only updates the model parameters by assimilating all historical measurements. The information found in the model parameter covariance structure, the measurements, and model performance is used in the Kalman jump. So, the Kalman jump can generate a more directional update of the model parameters than the parallel direction jump and the snooker jump in $\text{DREAM}_{(KZS)}$ (Zhang et al., 2017).

Durum wheat is one of Italy’s key crops and therefore improving crop yield simulation efficiency is critical in addressing the challenges of precision farming, such as managing fertilization and predicting yield before harvest. Previous studies have mainly focused on optimization methods with satellite data or ground data (Castaldi et al., 2015). Integration with crop models, e.g., with the Aquacrop model (Steduto et al., 2009), originally developed by the Food and Agriculture Organization (FAO) of the United Nations to simulate crop yield responses to water, has been shown to provide promising results (Silvestro et al., 2014). Aquacrop simulates daily canopy cover, biomass, and yield as a function of water productivity. The open-source version of the Aquacrop model (AOS) was developed by (Foster, 2016), and recently a new version of the AOS, AOS v6.0a (Foster, 2019), has been made available.

Given the importance of model calibration, our interest was that of assessing the value of high temporal and spatial frequency remotely sensed data for the calibration of the AOS model on durum wheat. The aim of this study was thus to find the optimal values of the AOS model parameters for the study site and assess the Aquacrop-OS model performances by exploiting the fvc retrieved from high spatial and temporal resolution VEN μ S satellite images. We

implemented the state-of-the-art Markov chain Monte Carlo-based DREAM algorithm, $DREAM_{(KZS)}$, to get a Bayesian inference on the crop model parameters for durum wheat, which is the predominant crop in our study region of the Maccarese farm in Central Italy. Calibration and validation were carried out separately using the fvc from the VEN μ S images: The calibration was developed in 2017–2018 for the fvc while the validation was carried out in 2018–2019 for the fvc for the different fields of the farm.

2. Materials and Methods

The overall workflow is presented in Figure 1.

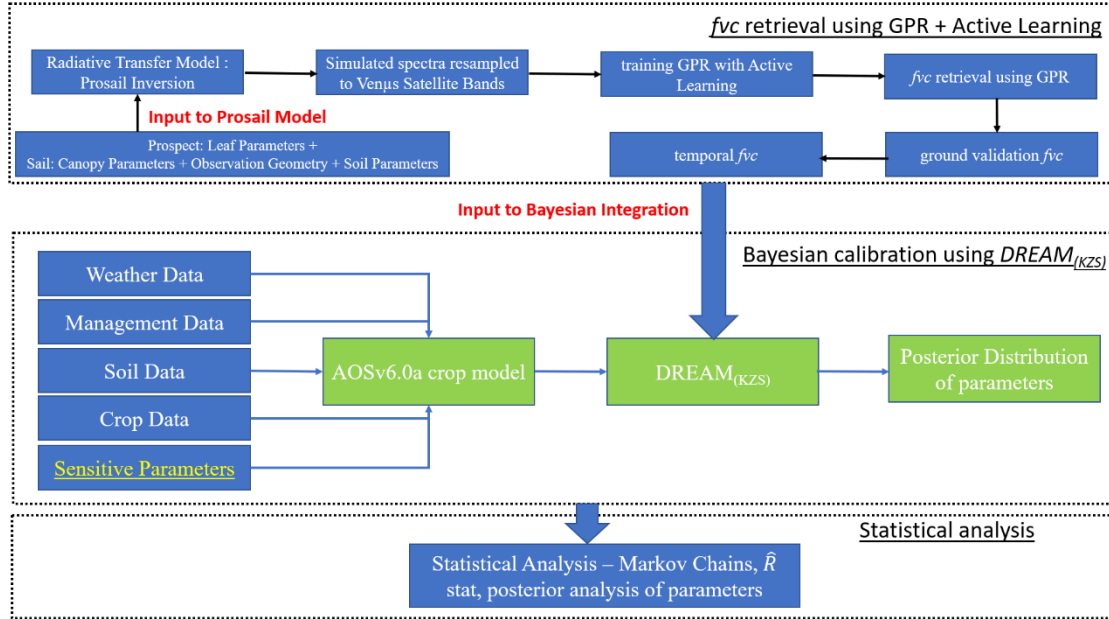


Figure 1. Bayesian calibration framework for the Aquacrop-OS (AOS) model using VEN μ S satellite data.

2.1. VEN μ S Satellite Data

VEN μ S (Vegetation and Environment monitoring micro-Satellite) is a joint Earth observation mission of Israel and France’s space agencies. It was launched in August 2017 with an expected life span of 3 years. VEN μ S satellite data is of high temporal (2 days) and spatial resolution (5.3 m) and is acquired on about 100 sites all around the world (Ferrier et al., 2010). It operates in 12 spectral bands from visible to near-infrared, suitable for vegetation studies. Due to the signal-to-noise ratio, the Level 2A data, i.e., after atmospheric corrections, were provided at 10 m (<https://theia.cnes.fr/>). In this work, VEN μ S 10 m atmospherically corrected surface reflectance data were used to carry out the processing. This 10 m data for the period November 2017–June 2018 and November 2018–June 2019 were used for the retrieval of the fvc for our study site (see Section 2.2). Currently, the full archive is being re-processed to provide the data at 5 m. The VEN μ S satellite bands are reported in Table 1.

Table 1. VEN μ S satellite bands information.

Bands	Central Wavelength (nm)	Bandwidth (nm)	Main Applications
B1	420	25	Atmospheric Correction Water
B2	443	40	Aerosols, Clouds
B3	490	20	Atmospheric Correction, Water

B4	555	20	Land
B5	638	24	Vegetation Indices
B6	638	24	DEM, Image Quality
B7	672	16	Red Edge
B8	702	24	Red Edge
B9	742	16	Red Edge
B10	782	16	Red Edge
B11	865	20	Vegetation Indices
B12	910	20	Water Vapor

2.2. Retrieval of Canopy Cover from VENμS Satellite Data

The PROSAIL model (Jacquemoud et al., 2009) is a widely used Radiative Transfer Model (RTM) for generating realizations of top of canopy reflectance by considering the measured biophysical variables of the crop of interest. The PROSAIL model's parameter ranges, statistical distribution, and the number of classes were similar to those used in previous research on the biophysical variables for the same study area (Upreti et al., 2019). A total of 7 out of the 12 bands (band 4,5,7-11) were selected from the VENμS satellite data that best approximates the Sentinel-2 8 bands suitable for vegetation studies (Weiss and Baret, 2016). Resampling of the PROSAIL simulations of the VENμS satellite selected bands and the addition of Gaussian noise was performed to better represent the actual reflectance of the canopy, simulated by the RTM. The detailed process is described by (Weiss and Baret, 2016), to which interested readers are referred.

In a previous comparative study of multiple hybrid machine learning algorithms based on PROSAIL simulations, for the biophysical variables retrieval from Sentinel-2, Gaussian Process Regression (GPR) with Active Learning (AL) was found to be the best algorithm (Upreti et al., 2019). We used here the same implementation with the sampling changed from Sentinel-2 to the VENμS bands for the retrieval of the *fvc* from VENμS satellite data.

2.3. Smoothing and Fitting *fvc* Time Series Data Retrieved From VenμS Satellite Data

The *fvc* time series was retrieved from the VENμS data for two years of the durum wheat growing seasons. Initially, the *fvc* time series were interpolated into daily time intervals using a 4253H twice filter (Velleman and Hoaglin, 1981), and then a double logistic function was used to fit the data (Jönsson and Eklundh, 2004). The 4253H twice is a median filter that takes the running median of 4, then 2, then 5, and then 3 data points over the *fvc* time series followed by a running weighted average. This procedure is repeated with the residuals and then both the results are added together (Kang and Özdoğan, 2019). Finally, the peak values were retained by scaling the *fvc* time series to the maximum of the un-smoothed *fvc* time series (see Results section).

2.4. Study Site and Ground Validation Campaigns

The test site of Maccarese (Figure 2) was used in the present study. Maccarese (41.833° lat. N, 12.217° long. E, 8 m a.s.l.) is a private farm of 3200 ha in a flat area with large fields, located on the west coast of Central Italy, near Rome. Field campaigns to measure the LAI of the durum wheat (*Triticum durum* Desf.) crops were carried out for this location from January 2018 to April 2018 at dates close to the VENμS acquisitions. The variables were sampled according to an Elementary Sampling Unit (ESU) scheme, to capture the variability within and among the different fields. Each ESU consisted of a quadrat 20 m by 20 m in size, to easily accommodate the VENμS 10 m pixel resolution. A total of 15 ESUs, placed at different locations, were employed at the different sampling dates. Each of the ESUs contained nine points, where the LAI measurements were collected. The LAI was measured using a LAI-2000 or LAI-2200C Plant Canopy Analyzer (LI-COR, Lincoln, NE, USA). Since LAI 2000/220C measurements were not always

acquired under diffuse light conditions, data were pre-processed by applying the recommended scattering correction procedure (Biosciences, 1992).

LAI 2000/2200C measurements were used to derive the f_{vc} , by using the following equation (Nielsen et al., 2012):

$$f_{vc} = 0.94 \times (1 - e^{-0.43 \times LAI})^{0.52} \quad (1)$$

Table 2 presents the calibration and the validation datasets and Table 3 shows the dates of the VEN μ s satellite data acquisitions close to the ground data collection dates.

The aboveground biomass was collected on two different dates, 16 February and 20 April in the year 2018, in eight different ESUs in different fields (see Figure 2a). The samples were collected destructively at the center point of each ESU within an area of 1 m² and were subsequently oven-dried at 72° C for three days. Finally, the dried samples were weighted and the biomass data was used to compare the results of the calibration. The yield data was obtained from a New Holland CX860 combine harvester, equipped with a yield mapping system. The data were pre-processed by removing the spatial outliers and points where the speed of the machine was less than 0.5 km h⁻¹. All points having a yield greater than 15 t ha⁻¹ and less than 0.2 t ha⁻¹ were also removed from the processing. Further, very low yielding values were removed among the adjacent high-yield passes, which indicated that the operator was not working at full cutting width. As a next step, block kriging was carried out using a local variogram option; this processing was completed in the open-source software Variogram Estimation and Spatial Prediction with Error (Vesper), developed by the Australian Center for Precision Agriculture (Whelan et al., 2002). The final yield map produced after corrections and processing was used for pixel-wise validation of the yield predictions obtained after calibration on the same year (2017–2018 cropping season).

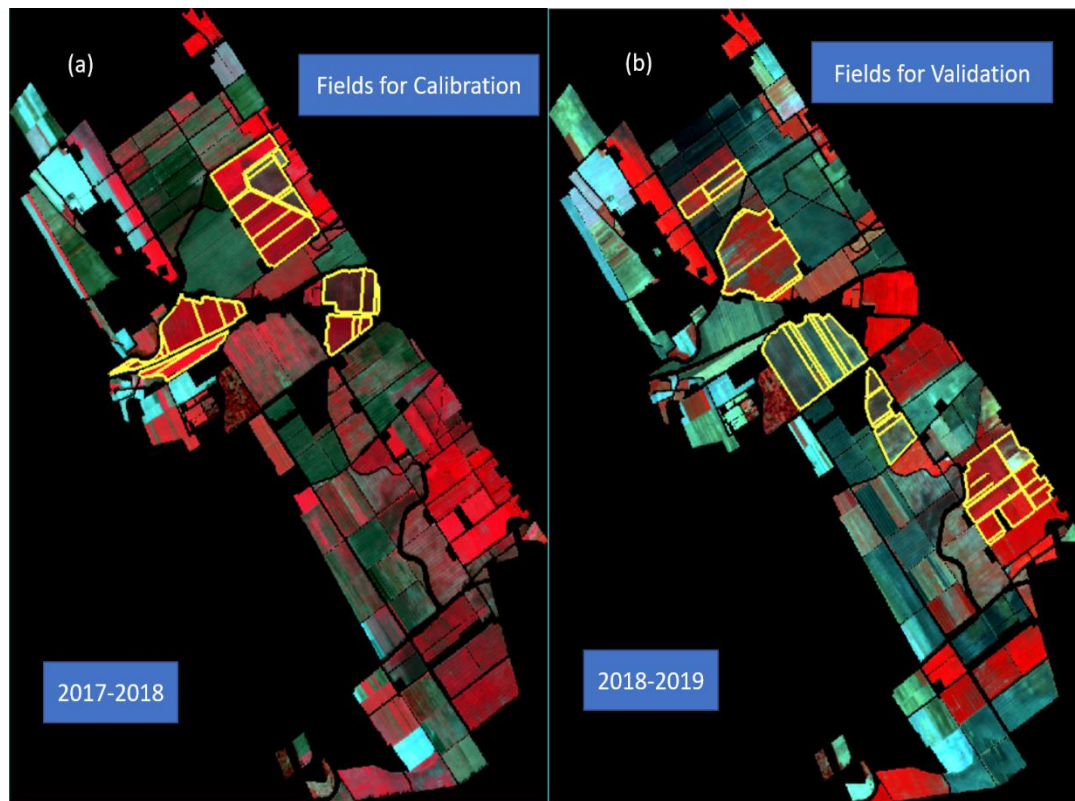


Figure 2. Durum wheat fields considered for calibration (a) and validation (b) in the Maccarese Farm (Central Italy). The backgrounds are the VEN μ S images of 24 January 2018 (a) and 12 March 2019 (b) (red = 865 nm, green = 672 nm, and blue = 555 nm; see Table 1).

Table 2. Description of the calibration and validation datasets.

AOS Model Calibration Data	AOS Model Validation Data
<i>fvc</i> time series of 2017–2018 retrieved from VEN μ S data	<i>fvc</i> time series of 2018–2019 retrieved from VEN μ S data
(20 pixels that represents yield variability)	biomass measured on the ground for the years 2017–2018
	yield observed from the combine for the years 2017–2018

Table 3. VEN μ S acquisitions and ground measurements dates.

Date—Ground Measurements	Date—VEN μ S Acquisition	Difference (Days)
January 31, 2018	January 28, 2018	3
February 16, 2018	February 13, 2018	3
April 6, 2018	April 8, 2018	2
April 20, 2018	April 20, 2018	0

2.5. Weather Data

The meteorological data used in this study to drive the model simulations were obtained from a weather station at about 3 km from the farm, from the Agrometeorological Service of the Lazio Region. For the period 2017–2019, the daily minimum and maximum temperature, relative humidity, wind speed, rainfall data, and daily solar radiation

were provided. The measurement of the regular reference evapotranspiration was based on the FAO Penman–Monteith method, as described in (Allen et al., 1998), and the ETo calculator (Raes, 2009). Figure 3 shows the minimum and maximum temperature, rainfall, and reference evapotranspiration for the 2017–18 and 2018–19 durum wheat growing seasons (October to July) at the Maccarese site.

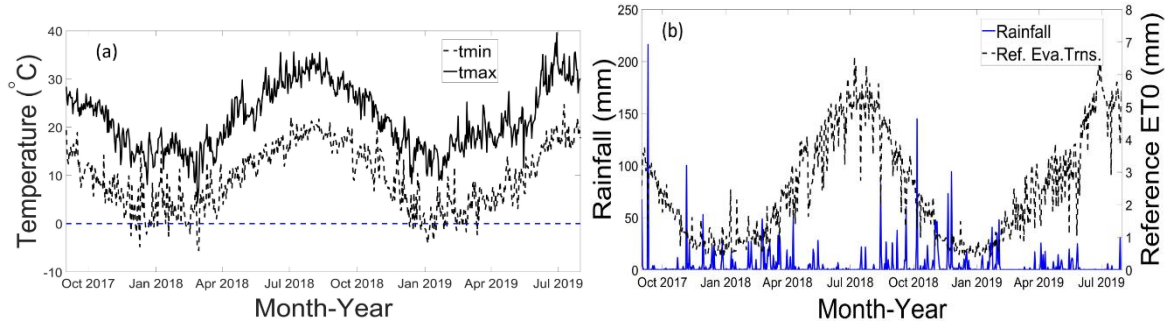


Figure 3.(a) Maximum and minimum temperature and (b) daily precipitation and reference evapotranspiration for two durum wheat growing seasons recorded in Maccarese (Rome, Italy).

2.6. Aquacrop-OS Model and Sensitive Parameters

The Aquacrop crop water productivity model was introduced by (Steduto et al., 2009). It simulates the relationship between crop yield and crop transpiration under different conditions. Aboveground biomass is simulated daily by using the normalized crop water productivity (NCWP) concept. Biomass (*bio*) is obtained as the product of the NCWP to the ratio of crop transpiration (*CT*) and reference evapotranspiration (*ET_o*; Equation (2)):

$$bio = NCWP \times \sum \frac{CT}{ET_o} \quad (2)$$

Grain yield (*yld*) is obtained by multiplying the harvest index (*HI*) by *bio* (Equation (3)):

$$yld = bio \times HI \quad (3)$$

The Aquacrop model originally developed by FAO includes a graphical user interface to run the model, but there have been different versions developed, including a command-line plug-in version and a GIS version (see <http://www.fao.org/aquacrop>). This model has been tested on different climates, geographical locations, and crops under different irrigation and field management practices (Andarzian et al., 2011; Jin et al., 2020; Saab et al., 2015; Todorovic et al., 2009; Xiangxiang et al., 2013). To allow the customization of the model and to implement the model for different data demanding applications, an open-source Matlab version, Aquacrop-OS (OS), was developed by (Foster et al., 2017). The latest version of AOS, Aquacrop-OS v6.0a, is used in the present study. Since AOS is slightly different from the original FAO version in terms of parameters, specific sensitivity analyses and calibrations should be carried out since the studies carried out on the standard FAO Aquacrop model cannot be directly applied to AOS.

In previous work on the AOS sensitivity analysis, (Upreti et al., 2020) applied two different methods: the moment independent density-based PAWN and variance-based Extended Fourier Amplitude Sensitivity Test (EFAST). A common set of the most influential parameters was found with both methods, although the rankings of the parameter varied for the same study site. In the present work, we considered the same set of parameters and parameter ranges for the Bayesian calibration, using the DREAM_(KZS) version of the DREAM algorithm originally developed by (Vrugt, 2016).

2.7. Markov Chain Monte Carlo-Based DREAM_(KZS) Algorithm

The Bayesian estimation aims to update the distribution of the critical parameters by combining observations and prior values (Van Oijen et al., 2005). Bayes theorem states that, given the new observations, the posterior probability of a hypothesis is proportional to the product of the prior probability of the hypothesis and likelihood of the same hypothesis. By adopting a Bayesian formalism, the posterior distribution of the model's parameters can be derived by conditioning the model's spatio-temporal behavior on measurements of the system response observed:

$$p(x|\tilde{Y}) = \frac{p(x)p(\tilde{Y}|x)}{p(\tilde{Y})} \quad (4)$$

where $p(x)$ and $p(x|\tilde{Y})$ signify the prior and posterior parameter distribution, respectively, and $L(x|\tilde{Y})=p(\tilde{Y}|x)$ denotes the likelihood function. If the parameters are well defined in the prior distribution, the key issue is the interpretation of the likelihood function. $L(x|\tilde{Y})$ is used to summarize the difference between the simulations of the model and the data of the measurements. If the error residuals are assumed to be uncorrelated then the likelihood of the n-vector of residual error can be written as follows:

$$L(x|\tilde{Y}) = f_{\tilde{y}_1}(y_1(x)) \times f_{\tilde{y}_2}(y_2(x)) \times \dots \times f_{\tilde{y}_n}(y_n(x)) = \prod_{t=1}^n f_{\tilde{y}_t}(y_t(x)) \quad (5)$$

where $f_a(b)$ signifies the probability density function of a evaluated at b, given that a and b represent the calculation of the likelihood function at each measurement. If we assume further that the residuals of the error are normally distributed, then Equation (5) can be written as

$$L(x|\tilde{Y}, \hat{\sigma}^2) = \prod_{t=1}^n \left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp \left[-\frac{1}{2} \left(\frac{\tilde{y}_t - y_t(x)}{\hat{\sigma}_t} \right)^2 \right] \right) \quad (6)$$

where an n-vector with standard deviations of the measurement error of the observations is $\hat{\sigma}_t = \{\hat{\sigma}_1, \dots, \hat{\sigma}_n\}$. This formulation allows for homoscedastic and heteroscedastic measurement errors.

The posterior distribution is often high dimensional and analytically intractable for complex crop models, and sampling methods are required to estimate the target, and then the desired description of the posterior distribution can be obtained from the sample (Wallach et al., 2018). To generate samples from the posterior distribution, a lot of iterative methods have been developed. In a way, all of these methods rely on Monte Carlo simulations. As already stated, we implemented the DREAM_(KZS) algorithm (Zhang et al., 2017) for parameter estimation in this work.

2.8. Statistical Analysis and Validation

$\hat{R}$ statistics, as proposed by (Gelman and Rubin, 1992), was used to determine when the convergence of the sample chains to a limiting distribution has been achieved. This analysis compares the within and between chain variance for each parameter using

$$\hat{R}_j = \sqrt{\frac{N+1}{N} \frac{\hat{\sigma}_+^{2(j)}}{W_j} - \frac{T-2}{NT}} \quad (7)$$

where N is the number of chains, T implies the number of samples in each chain, W_j is the within chain variance, and $\hat{\sigma}_+^{2j}$ is an estimate of the variance of the jth parameter of the target distribution

$$\hat{\sigma}_+^{2(j)} = \frac{T-2}{T} W_j + \frac{2}{T} B_j \quad (8)$$

For each parameter $j=\{1, \dots, d\}$, to declare convergence, $\hat{R}_j$ should be ≤ 1.2 , or otherwise the value of T should be increased and the chains should run longer. Since the various N chains are launched from different starting points, the $\hat{R}$ diagnostic is a relatively robust estimator (Metropolis et al., 1953) (see Results section).

The accuracy of the *fvc* retrieval using VEN μ S satellite data is computed in terms of Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Relative Root Mean Square Error (RRMSE). Bayesian calibration of the AOS model was carried out using the *fvc* time series data for multiple pixels (center pixels of ESU) from eight fields for the years 2017–2018 (Figure 2a) and validation was carried out for seven fields for the years 2018–2019 (Figure 2b). Different metrics, namely, the RMSE, Relative Error (RE), Relative Variability (a), Relative Bias (b), Linear

Correlation Coefficient (r), and Kling–Gupta Efficiency (KGE), were used to assess the performance of the validation. As indicated by (Gupta et al., 2009), the significant metric Normalized Standard Error (NSE) can be divided down into three distinctive components: a linear correlation between simulations and observations, b normalized by the standard deviation in the observed values, and a as a measure of the relative variation in the simulated and observed values. In (Gupta et al., 2009), KGE is defined as

$$\alpha = \frac{\sigma_s}{\sigma_o} \quad (9)$$

$$\beta = \frac{\mu_s}{\mu_o} \quad (10)$$

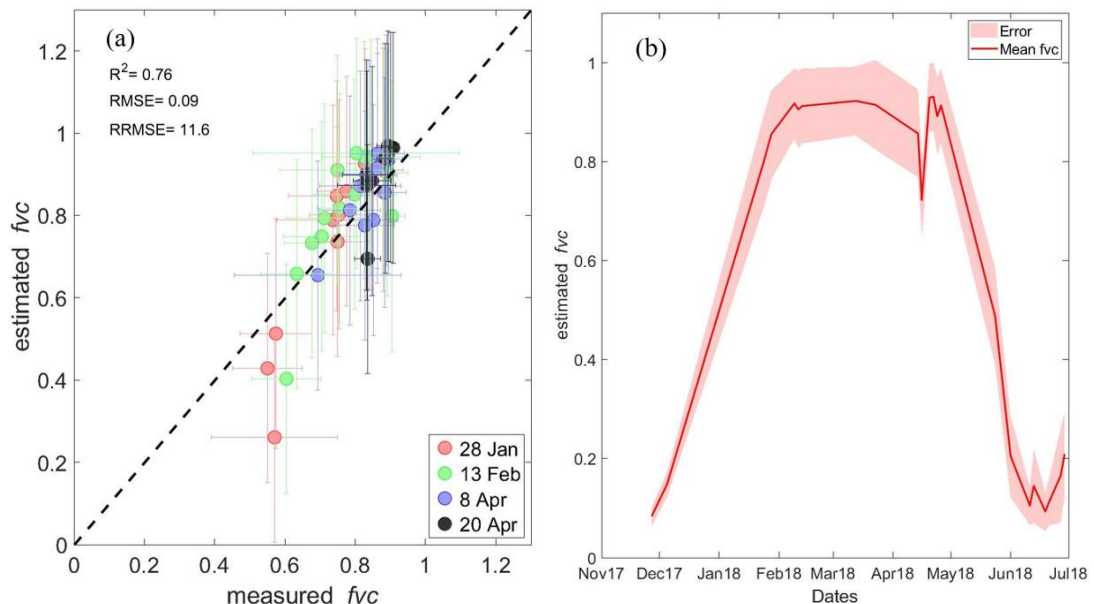
$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (11)$$

where σ is the standard deviation and μ is the mean value (with the subscript “s” for simulations and “o” for observations), a is the relative variability, and b is the relative bias. For calibration, we have used the fvc time series of the years 2017–2018 and one simulation for each pixel and applied the process to multiple pixels in each field presented in Figure 2a. The pixels selected were the center point of each ESU where the biomass sample was collection on the ground. Parameter values obtained by the calibration process were used for fvc validation for the years 2018–2019 and we have a single simulation for validation (Figure 2b). The simulated biomass and yield for the years 2017–2018 were also validated with the ground-measured biomass and the yield data measured at the end of the season.

3. Results

3.1. fvc Retrieval

The advantage of using the GPR algorithm is that it also estimates associated uncertainty along with the mean values of predictions (fvc). The fvc time series of a pixel as an example is shown in Figure 4b; the shaded region represents the uncertainty in the estimation. It can also be visualized that the uncertainty is higher at the peak in comparison to the start or end of the growing season. The sudden dip in the fvc value (Figure 4b) at the peak represents the partial cloudy date and so it makes the smoothing and the fitting of the fvc time series necessary before further processing of assimilation. Figure 4c shows an example of the fvc time series of a pixel with the fvc retrieved from VEN μ S satellite data, which has partially the effects of the cloud; this original fvc time series was processed using a 4253H twice filter followed by fitting a double logistic function to this data to obtain smooth fvc time series data, which will be ingested in the assimilation procedure.



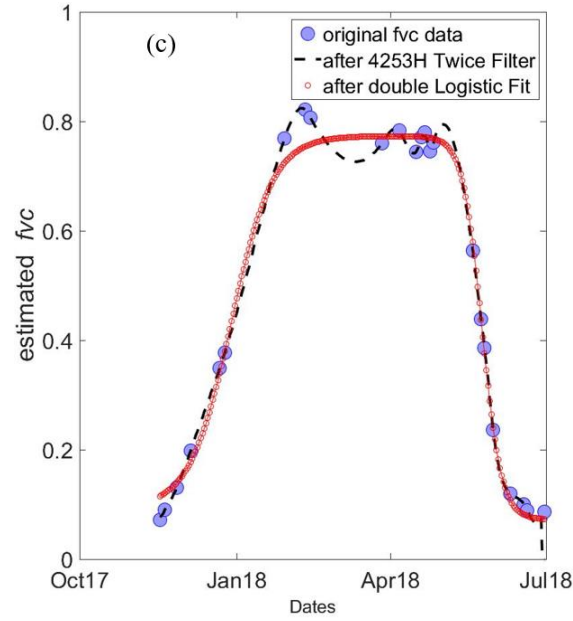
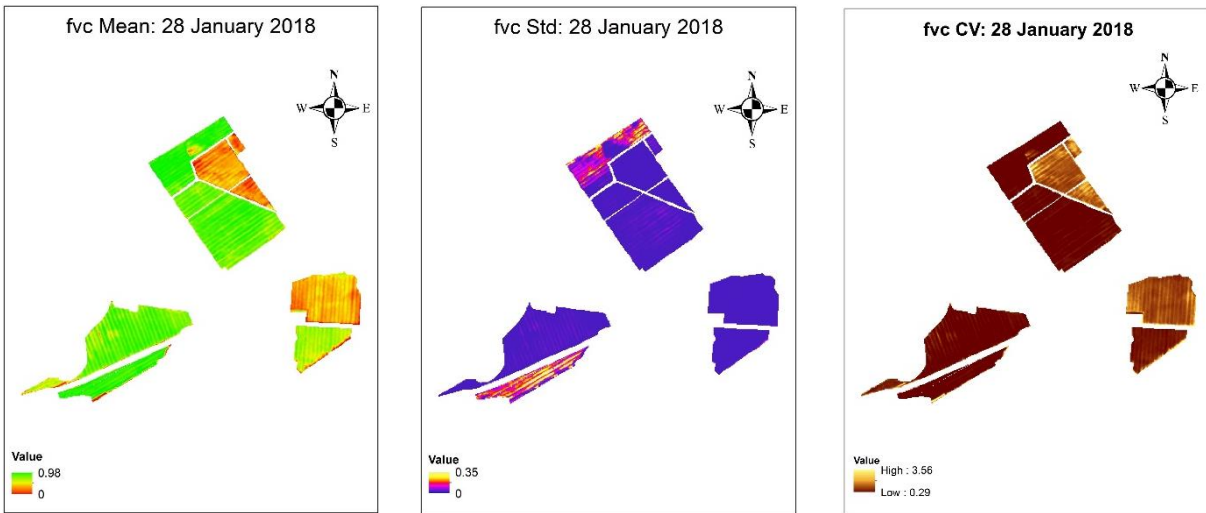


Figure 4. (a) Ground-measured fvc and estimated fvc from VEN μ S data; horizontal error bars represent variation within an ESU and vertical error bars represent uncertainty estimates using the GPR algorithm. (b) Example of a seasonal trend of fvc in the 2018 season with the associated uncertainties (shaded region) and with the mean value of the fvc predictions. (c) Example of an fvc time series estimation from VEN μ S data before and after the application of a filter and smoothers.

An example of *fvc* mean estimation maps with the associated uncertainty expressed as the standard deviation (SD) around the mean and the relative uncertainty expressed as the coefficient of variation (SD/mean) for four dates for the durum wheat fields, as shown in Figure 5a (mean *fvc*), Figure 5b (associated uncertainty), and Figure 5c (relative uncertainty).



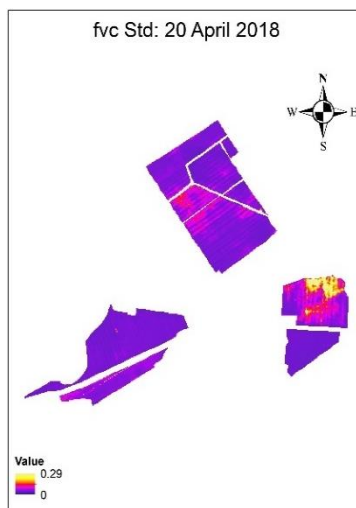
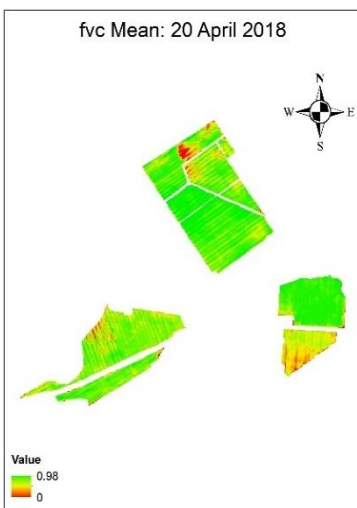
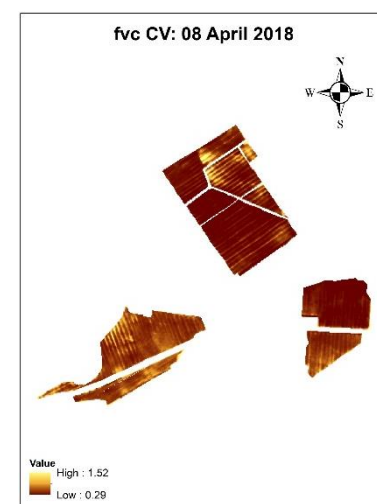
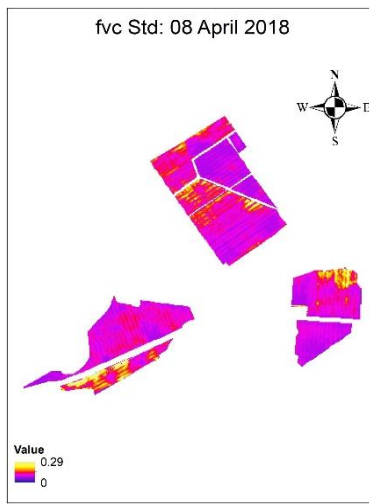
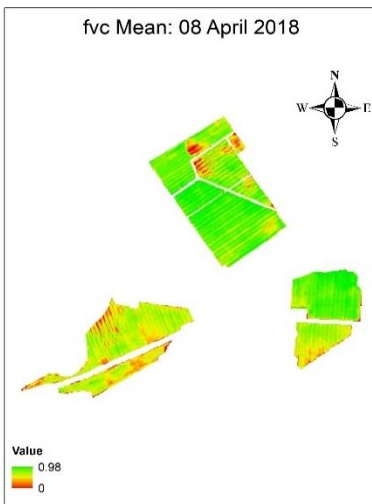
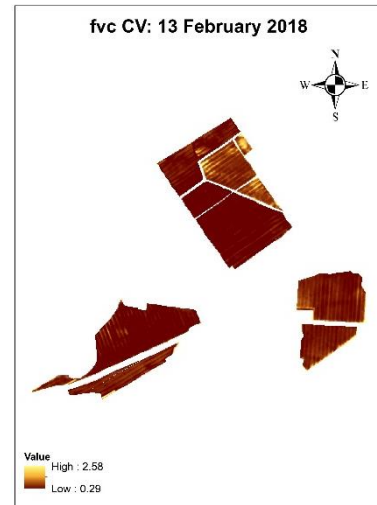
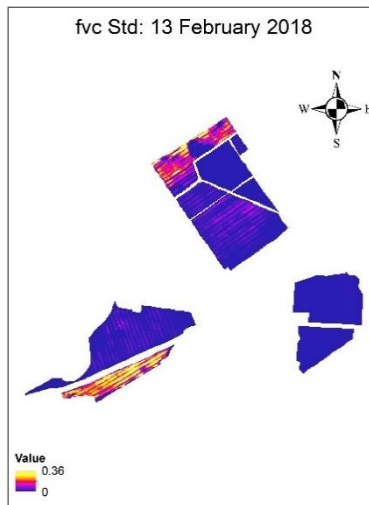
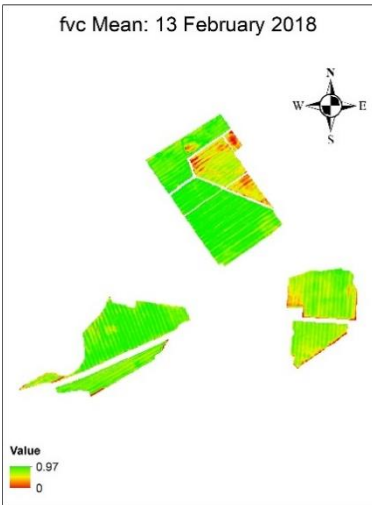


Figure 5. The fvc mean predictions (a), associated uncertainties (b), and relative uncertainties (c) for four dates, for durum wheat fields in the 2018 season in Maccarese.

It can be seen from Figure 5a, b that the uncertainty increases during the peak fvc mean estimates, which can also be assessed in Figure 4b. Relative uncertainty (Figure 5c) may provide a more meaningful interpretation. It can be observed that the fallow areas or bare soil retrievals have a rather high relative uncertainty. By applying a threshold, the more uncertain retrievals can be masked out. Hence, uncertainty can function as a spatial mask that enables displaying only pixels with a high certainty (Verrelst et al., 2015).

3.2. Parameters Identification

Figure 6a,b represents the evolution of the univariate and multivariate $\hat{R}$ statistics of the 15 crop model parameters, as proposed by (Gelman and Rubin, 1992) for the analysis of the convergence. The threshold of 1.2 for convergence diagnosis is represented by the black dashed line.

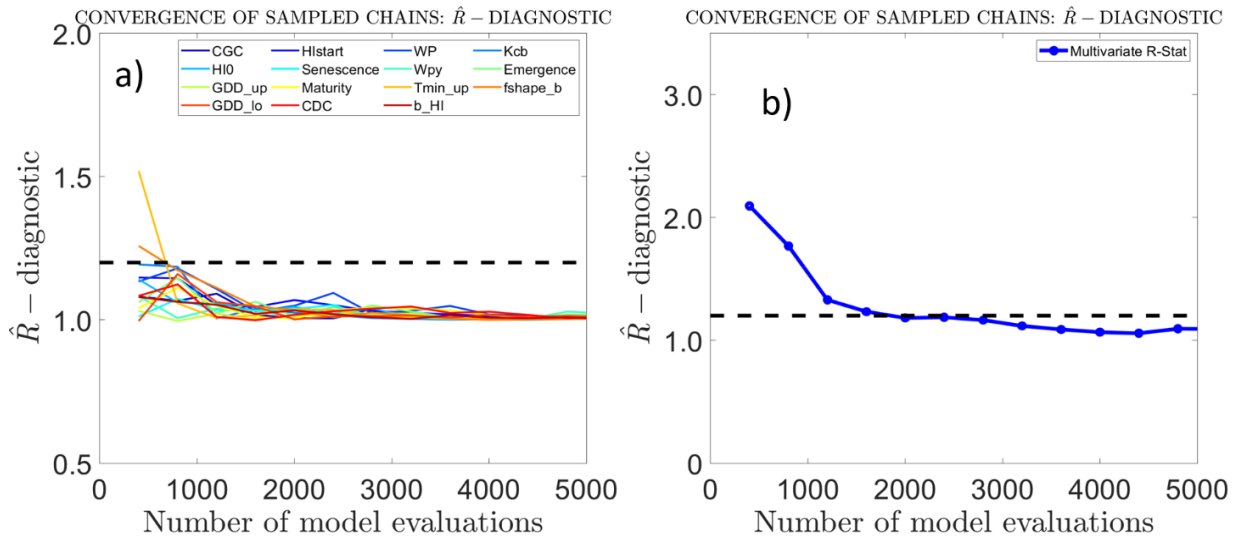


Figure 6. Evolution of the univariate (a) and multivariate (b) Gelman–Rubin $\hat{R}$ statistics of the 15 model parameters. The threshold of 1.2 for a convergence diagnosis is represented by the black dashed line (see Section 2.8).

To better explore the parameter space, we run $N = 4$ parallel chains in $\text{DREAM}_{(KZS)}$. The length of the chain was set as 5000, which means that the total number of model evaluations became 20,000 (5000×4). We restrict the Kalman jump to the initial 20% of the simulation time as it cannot maintain a detailed balance and set the probabilities of using the Kalman jump, the parallel direction jump, and the snooker jump to 20%, 72%, and 8%, respectively, while in the remaining 80% simulation, these are set 0%, 90%, and 10% to sample the posterior, which can satisfy a detailed balance with diminishing adaptation (Zhang et al., 2017). When the $\hat{R}$ statistic of all the 15 parameters is below 1.2, it can be concluded that the Markov chains converge to the equilibrium distribution. The number of chains and the number of generations for each chain is sufficient for the mixing of the chains and the exploration of the entire parameter space.

Figure 7 presents the marginal posterior probability density functions (PPDF) of the parameter estimates when the sampling process had achieved a stationary distribution at the end of the process. It can be visualized from Figure 7 that the prior distribution function has been transformed to the posterior distribution function for all the AOS model parameters subjected to the calibration.

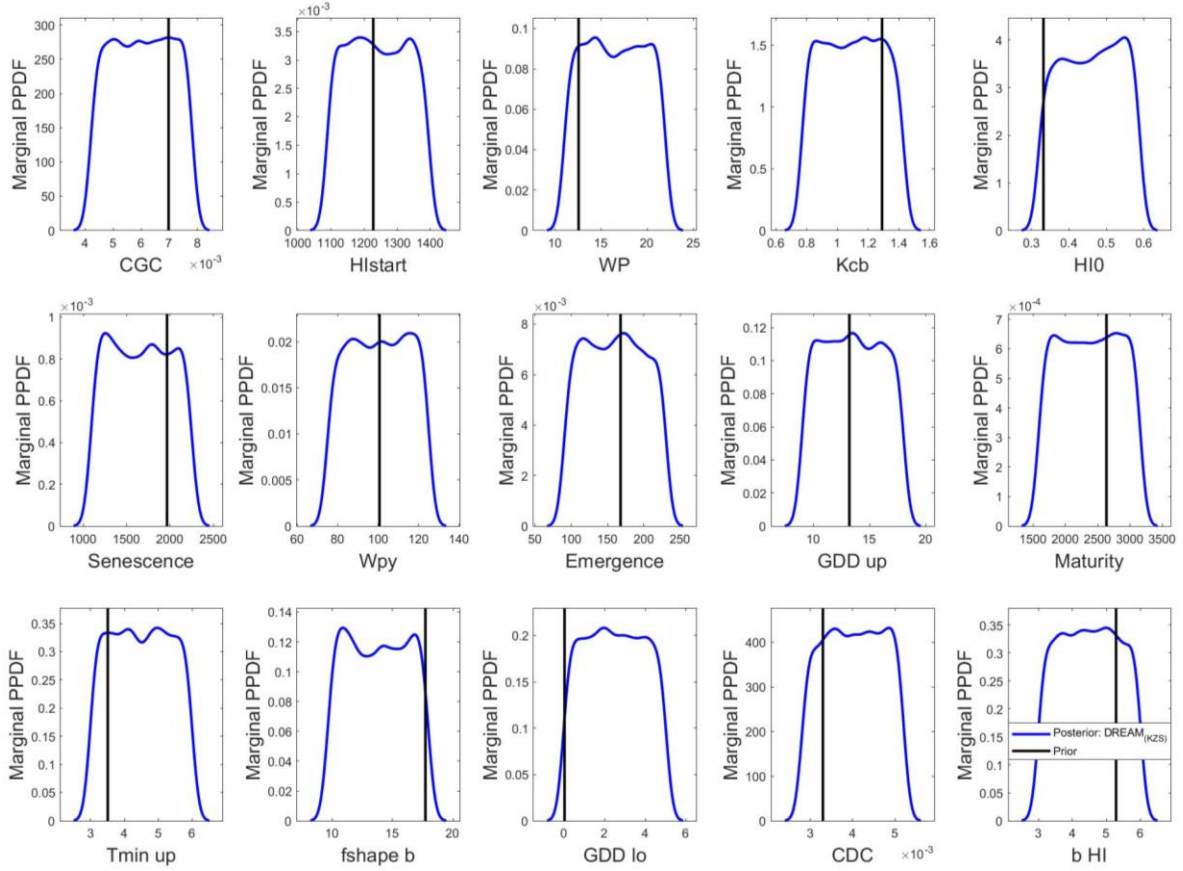


Figure 7. Marginal PPDF of the AOS parameters. The vertical black line indicates the prior values of the parameters.

Prior ranges of the model parameters (uniform distribution) and the final value after processing of the $DREAM_{(KZS)}$ algorithm are presented in Table 4. Since the output of the $DREAM_{(KZS)}$ algorithm used for calibration is a posterior distribution, rather than a single point estimate, the mean value of all the 15 crop model parameters was extracted from the marginal posterior distribution function, as is most commonly used. Although the calibration was carried out using *fvc* data from eight different durum wheat fields for the years 2017–2018, the parameter values do not differ much among each field. The mean values of the posterior distribution across fields are thus presented in Table 4.

3.3. Calibration and Validation

3.3.1. *fvc*

Error metrics for the estimation of the *fvc* for the validation fields presented in Figure 2b, for the years 2018–2019, are presented in Table 5. The calibration of the Aquacrop model was previously performed by (Trombetta et al., 2016) with two years of canopy cover data retrieved from MODIS LAI images and validated on one year of data for two farms in Italy; the error metrics have more or less similar values, but the linear correlation coefficient is higher in this work (Table 5).

Table 4. Prior ranges and parameter values computed from the posterior distribution (mean) of the 15 parameters of the AOS model used for calibration.

Parameter	Unit	AOS Standard	Prior Range	Estimated Parameter
CGC	fraction GDD	0.0050	(0.0042, 0.0078)	0.0060

<i>HIstart</i>	GDD	1250	(1090, 1395)	1243
<i>WP</i>	gm ⁻²	15	(11, 22)	16.50
<i>Kcb</i>	-	1.10	(0.77, 1.43)	1.10
<i>HI0</i>	%	0.48	(0.32, 0.59)	0.46
<i>Senescence</i>	GDD	1700	(1090, 2250)	1670
<i>Wpy</i>	gm ⁻²	100.00	(75, 125)	100.00
<i>Emergence</i>	GDD	150	(90, 230)	160
<i>GDD_up</i>	GDD	14	(9, 18)	13.50
<i>Maturity</i>	GDD	2400	(1590, 3150)	2370
<i>Tmin_up</i>	°C	5	(3, 6)	4.50
<i>fshape_b</i>	-	13.81	(9.6694, 17.9575)	13.81
<i>GDD_lo</i>	GDD	0	(0, 5)	2.5
<i>CDC</i>	fraction GDD	0.0040	(0.0028, 0.0052)	0.0040
<i>b_HI</i>	-	7	(3, 6)	4.50

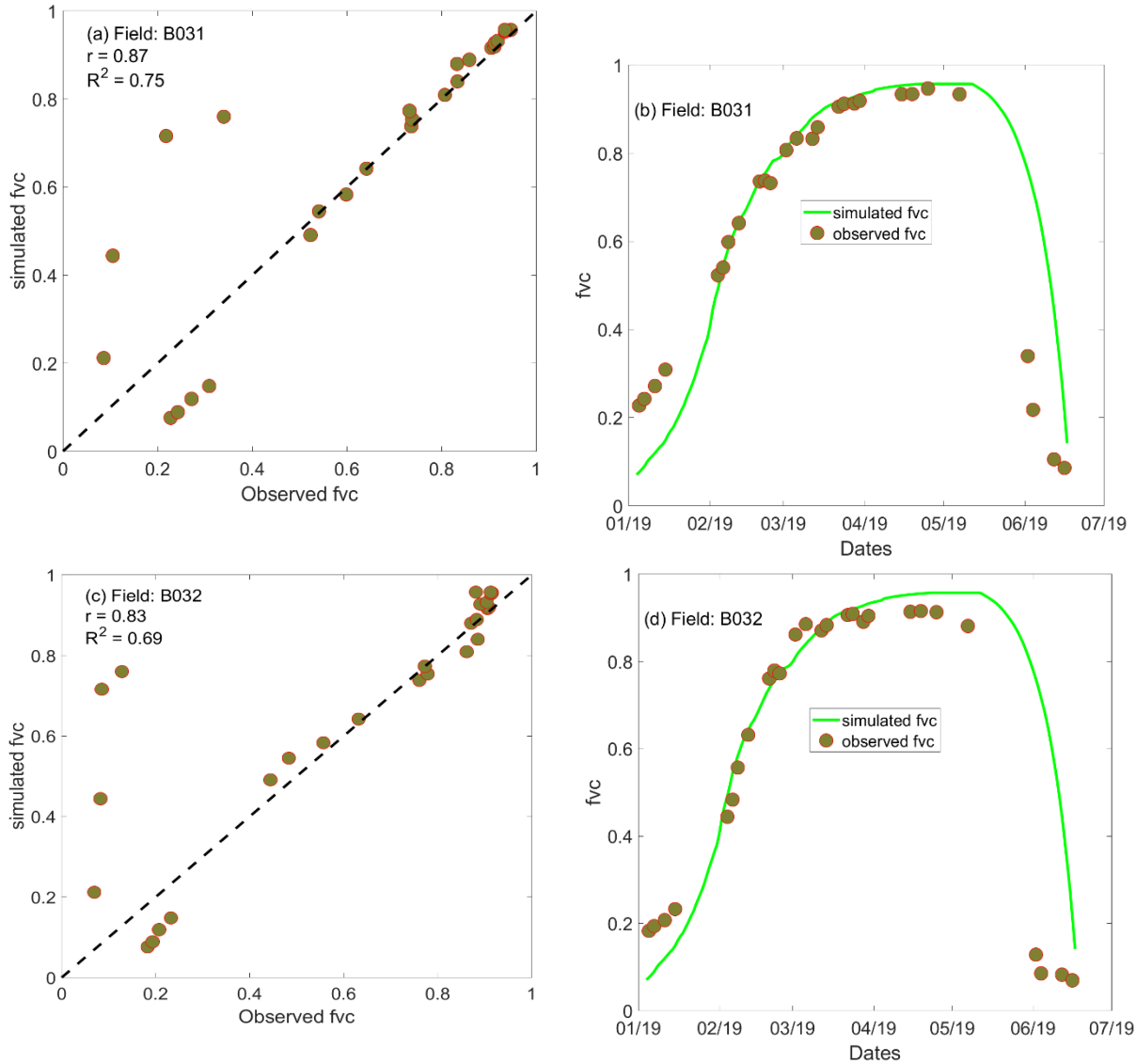
Table 5. RMSE, RE, α , β , r , and KGE for validation of the *fv*c year 2018–2019.

Field code	RMSE	RE	α	β	r	KGE
B031	0.16	0.25	1.11	1.03	0.87	0.18
B032	0.19	0.60	0.96	1.07	0.83	0.15
B044	0.20	0.44	1.11	1.13	0.87	0.22
B062	0.41	2.79	0.84	1.9	0.61	0.93
B069	0.44	2.29	0.88	1.95	0.52	0.97
B085	0.23	0.65	0.96	1.30	0.86	0.33
B109	0.15	0.19	1.45	0.91	0.93	0.47

From Table 5 it can be seen that r is close to 1, except for the two fields: Field B062 and Field B069, where it has a value of 0.61 and 0.52. This shows that the linear relation between the observed and simulated *fv*c time series is strong. Figure 2b shows the background image of the VEN μ S data on 12 March 2019, where the fields B062 and B069 can be seen clearly as fallow lands. The not so strong correlation for these two fields may be the difference between the sowing and planting dates in comparison to the other fields. This can also be understood from the RMSE and RE data from Table 5, where it has the highest error for these two fields on validation, i.e., 0.41 and 0.44 for RMSE and 2.79 and 2.29 for RE. α and β are the representative for the relative variability and relative bias, and their values range from 0.84 to 1.45 and 0.91 to 1.95, respectively. The value of α is close to 1 for almost all the fields, which is the ideal value for it; similarly, β also has values close to 1, to its ideal value except for the fields mentioned earlier. KGE, which combines α , β , and r in an equation to find out the efficiency of the validation, ranges from 0.15 to 0.93, which is higher than during the calibration (metrics not shown).

The calibrated values of the 15 AOS model parameters from Table 4 were used with the other model parameters set as default to generate the *fv*c, biomass, and yield time series data for the years 2018–2019. This simulated *fv*c time

series is represented as a green curve in Figure 8. As can be seen, the curve is the same for all the fields; this is because we generated a single set of simulations for the years 2018–2019 that is representative of the average *fvc* in the study area (Figure 2b). The field-wise average *fvc* time series derived from the VEN μ S satellite images for the years 2018–2019 were plotted with this simulated *fvc* time series. The results are mentioned in Table 5; it can also be visually assessed from the Figure 8—the R^2 ranges from 0.69 to 0.86 and r from 0.83 to 0.93. There is a good agreement between the simulated *fvc* time series and VEN μ S satellite image-derived *fvc* time series.



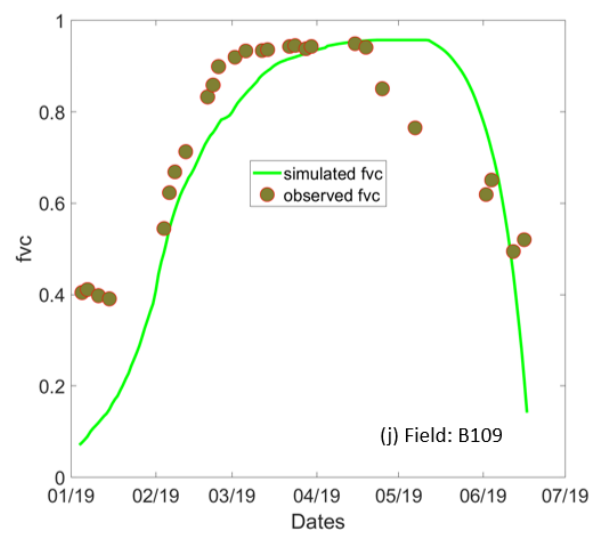
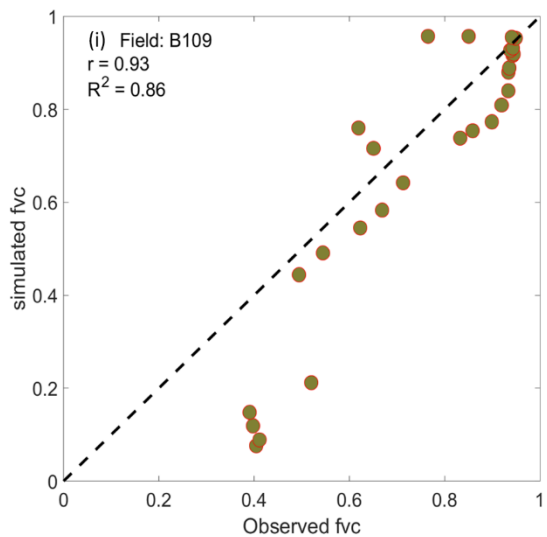
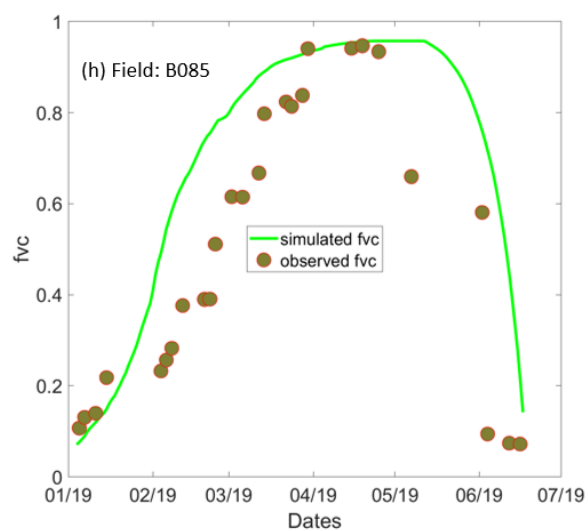
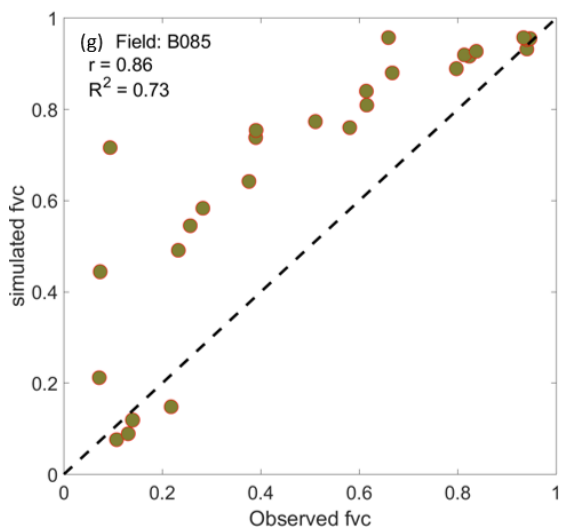
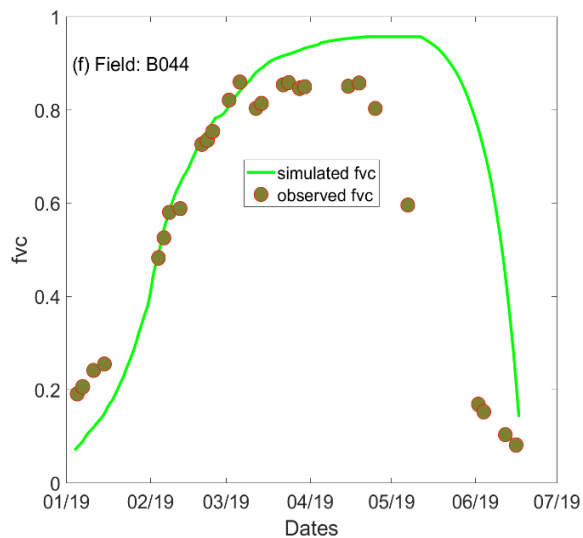
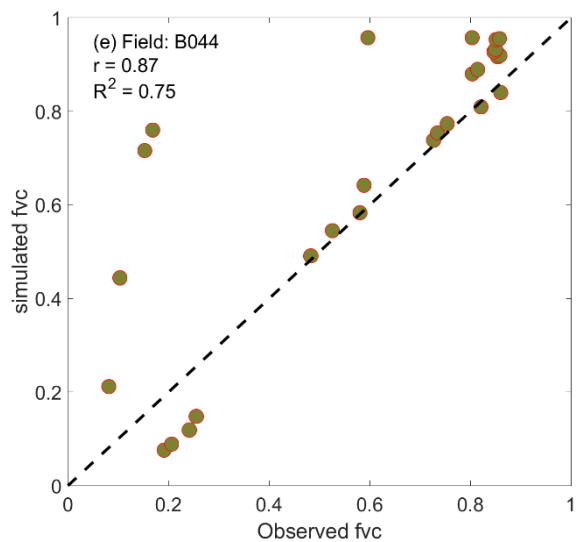


Figure 8. Validation—simulated and observed *fvc* of the durum wheat in different fields in the Maccarese farm for the years 2018–2019. (a,b) Field: B031; (c,d) Field: B032; (e,f) Field: B044; (g,h) Field: B085; (i,j) Field: B109.3.3.2. Biomass and Yield

The biomass collected inside the ESU on two dates (16 February 2018 and 20 April 2018) and yield data collected by the combine on multiple pixels in different fields in 2018 were compared with the simulation predictions obtained after the pixelwise calibration carried out using VEN μ S *fvc* (Figure 9a,b). As can be seen visually, the simulated biomass is overestimated in comparison to the ground, and the pixelwise yield predictions shows low error when compared to the measured yield (RMSE of 1.51 t ha⁻¹), with a tendency towards overestimation that is less pronounced than for biomass. Although AOS overestimated biomass and to a lesser extent yield as compared to the ground measurements, it is in good agreement with the *fvc* (Figure 8), confirming the importance of the calibration.

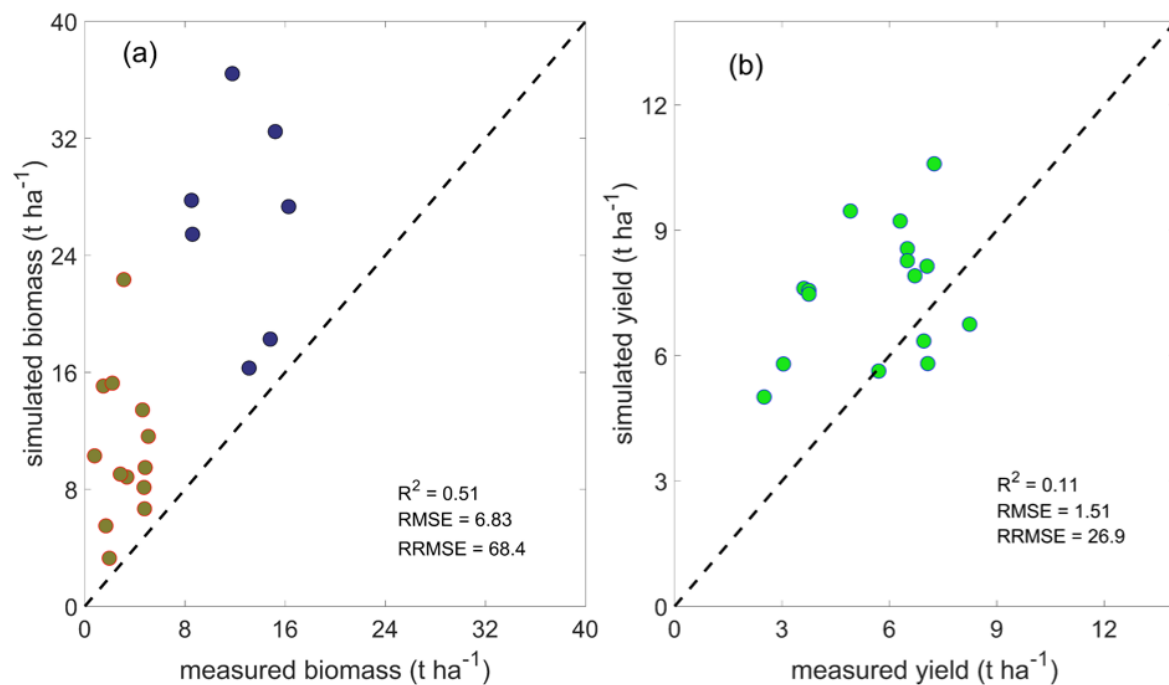


Figure 9. Measured and simulated biomass and (b) measured and simulated yield after pixelwise calibration for the years 2017–2018.

4. Discussions

4.1. *fvc* Retrieval

The R^2 obtained was 0.76, the RMSE 0.09, and the Relative RMSE (RRMSE) was 11.6 (%), with an apparent overall slight overestimation of high values and underestimation of low values in the retrieval. Due to the Bayesian backbone of the GPR family algorithms, the advantage of using the GPR algorithm for the estimation of the *fvc* is that it provides uncertainty intervals associated with the mean predictions, which are reported as vertical bars in Figure 4a, whereas the horizontal bars show the variability (in terms of standard deviation) within single ESUs.

The original *fvc* data retrieved from the VEN μ S data using the GPR algorithm, which sometimes showed an erratic seasonal pattern (Figure 4c), were smoothed using the 4253H twice filter, fitted using a double logistic function (Jönsson and Eklundh, 2004) and then rescaled to the minimum and maximum of the original *fvc* time series retrieved using GPR. Since the goal of the present work was to get the Bayesian inference on the AOS crop model parameters subjected to calibration, we considered this retrieval of *fvc* from the VEN μ S satellite 10 m data to be acceptable. Maps of the *fvc* for the Maccarese farm were obtained for 15 cloud-free VEN μ S images acquired from 2017 to 2018 and 27

images acquired for the 2018–2019 wheat growing seasons. An example *fvc* mean estimation maps with associated uncertainty and relative uncertainty for four dates for the durum wheat fields is shown in Figure 5. The related uncertainty images (Figure 5b) indicate higher retrieval certainties from the trained model in the low-intensity blue color. When applied to any image, the generation of uncertainty estimates allows insight into the pixel-by-pixel approach. It thus enables the definition for which land covers are correlated with a high degree of certainty and also those areas in which additional sampling would gain land cover. (Verrelst et al., 2015).

Uncertainty (σ) is also correlated with the mean estimate (μ). Relative uncertainties (σ/μ) can give a more meaningful interpretation (Figure 5c). It can be observed that there is a relatively high relative uncertainty regarding fallow areas or bare soil retrievals. The application of a threshold will cover some more questionable retrievals. Uncertainty can, therefore, act as a spatial mask that only display pixels with a high degree of certainty (Verrelst et al., 2015).

A similar approach of biophysical variables retrieval utilizing Sentinel-2 data was implemented by (Upreti et al., 2019). The accuracy that they found was somewhat higher than in this work, possibly because in (Upreti et al., 2019) a choice of optimal bands was made following the analysis discussed in (Weiss and Baret, 2016). Three different sources of data, namely LIDAR, aerial imagery, and spaceborne imagery, were used for the retrieval of canopy cover by (Ma et al., 2017), among which the accuracy of the retrieval was highest with spaceborne WorldView2 data with an $R^2 = 0.58$, followed by aerial imagery with an $R^2 = 0.50$, and then LIDAR data with an $R^2 = 0.33$. The accuracies were low in comparison to the present work; however, the data concerned forest vegetation. The authors in (P. Silvestro et al., 2017) used the Chinese satellite Huan Jing (HJ1A/B) and Landsat-8 Operational Land Imager (OLI) images to retrieve the canopy cover in wheat fields. The authors initially generated simulations on the PROSAIL model and then the simulated data was used to train an Artificial Neural Network (ANN). The retrieval was performed for three years: 2013, 2014, and 2015, and compared with the ground data of that year with the overall RMSE = 7.17 and RRMSE = 9%. The RRMSE is somewhat higher in the present work than reported in (P. Silvestro et al., 2017); there could be two possible reasons for this, the first one may be the non-optimal band selection of the VEN μ S satellite data for *fvc* retrieval and the other could be the equation that was used to convert the measurements of LAI into *fvc*. The solution to the first issue can be the identification of the optimal bands of the VEN μ S satellite data for *fvc* retrieval, for example using GPR band analysis (Revill et al., 2019), and for the second issue, the conversion equation can be tested with multiple-year canopy cover data. Although, in the present work, the RRMSE is a little higher than the Global Monitoring for Environment and Security (GMES) goal accuracy, i.e., 10%. We consider this *fvc* retrieval further as the aim of the work is to calibrate the AOS crop model for the localized application. The method of retrieving biophysical variables from multispectral satellite images have proven effective to train machine learning models with the PROSAIL simulated data (Silvestro et al., 2014). However, other machine learning methods do not provide the uncertainty estimates with the mean value of predictions as in the case of GPR.

4.2. Parameters Identification

It can be seen that the initial uniform prior distribution has been transformed into a posterior distribution by integrating the observations into the AOS crop growth model. From the distributions, the parameter estimates (mean values) have been derived and the confidence intervals of the parameters estimated can be quantified. Figure 7 shows the marginal pdf of all the 15 crop model parameters subjected to the calibration of the model. It can be visualized from Figure 7 that the prior values of the crop model parameters differ from the mean values of the posterior distribution, especially in the case of CGC, WP, Kcb, HI0, Sencence, Tmin_up, fshape_b, GDD_lo CDC, and b_HI, which also signifies the importance of the Bayesian calibration. In (Trombetta et al., 2016), the calibration of the Aquacrop model was performed by integrating the MODIS LAI images for wheat, considering a similar set of parameters as that in the present work. The study carried out by (Zhang et al., 2019) on the Bayesian calibration of the Aquacrop model also find WP, CGC, Kcb, and CDC important for the calibration, with the high errors reported by using the default parameter values. The authors only considered 10 model parameters for the local calibration of the model, unlike the present study that evaluated 15 crop model parameters for calibration, which identified the most sensitive crop model parameters to yield from previous work on global sensitivity analyses (Upreti et al., 2020).

4.3. Calibration and Validation

The *fvc* time series simulated for the years 2018–2019, using the calibrated parameter values, was compared with the *fvc* time series of 2018–2019, generated using VEN μ S images. A good fit was observed in all the simulations (Figure 8), in agreement with the comments by (Andarzian et al., 2011). As can be seen from Table 5, in the present study the linear correlation is in the range of 0.81 to 0.93, with the exception of 0.52 and 0.61 for the fields B062 and B069; because of the phenological shift in the *fvc*, these metrics are better than as reported in (Trombetta et al., 2016). The AOS model parameter values obtained by the calibration were used for the validation and Figure 8 shows the validation on the years 2018–2019 for the *fvc*. The R^2 for the validation of the *fvc* for the years 2018–2019 also ranges from 0.69 to 0.86 for the different fields, while the retrieval of *fvc* from VEN μ S satellite images when compared to the ground-measured *fvc* data was 0.76 (Figure 4a).

As can be visualized from Figure 9a, the RRMSE for the simulated biomass is much higher than the usually acceptable range of 0–30%. This is the clear interpretation that the calibrated parameters did not work well for the estimation of the biomass. Although, there are uncertainties in the observed biomass and its processing, from the destructured samples collected to the oven dried samples; this could be one reason for the uncertainty accounted. While simulating yield, the RRMSE was 26.9%, which is fair and acceptable. The accuracy may be improved if the model is spatially applied to the entire field. In the present study, our focus was to find the optimal crop model parameter for the study area, so the calibration was performed on a few pixels. We are convinced that if the model is applied spatially to the study area, the accuracy would be higher than reported in the present work.

It should also be noted that the calibration was performed using *fvc* as a target variable, so the results for the yield simulations might be affected by other factors not taken into account in the calibration. In [74], it was argued that the estimating parameters on one response may not necessarily improve the predictions for a different response, but rather points at reducing the overall error of the model. Moreover, AOS is a new recent version of Aquacrop, coded on a different platform (MATLAB). Even if its performances have been checked (Foster et al., 2017), the newly coded versions of Aquacrop may still require refinement and extensive testing (Camargo Rodriguez and Ober, 2019).

In general, simulations performed after calibration showed lower values of biomass and yield with respect to values obtained with the standard AOS parameters, decreasing the simulation bias with respect to the measured values. For the years 2018–2019, the grain yield simulated using the standard parameters was 15.4 t ha⁻¹, while the yield obtained with the calibrated parameters (average of the pixelwise parameters estimated for the previous year) was 11.9 t ha⁻¹, closer to the yield range usually recorded by their combination (2–10 t ha⁻¹).

The advantages of calibrating the model using a Bayesian approach over classical methods is that the parameter estimation is a posterior probability distribution that can be used to implement uncertainty analysis. This distribution can be summarized with the mean estimate along with confidence intervals unlike a single point estimate as in the classical methods. These approaches are becoming increasingly popular for parameter estimation of complex mathematical models, such as crop models, as it provides a coherent framework for dealing with uncertainty (Makowski et al., 2006). A comparative study of Bayesian calibration and optimization method-based calibration for the Aquacrop model utilizing UAV multispectral images shows the improved performance with the Bayesian method (Zhang et al., 2019). The main reason of the poor performance in the optimization method was that it aimed at minimizing the error between the measurement data and model output data, when inappropriate observations are chosen. On the other side, in the case of Bayesian approach, even if no sufficient dataset is available, one can easily observe this by inspecting the parameter estimation confidence (Zhang et al., 2019).

All this considered, the significant improvements in the estimation of *fvc* and yield and the calibration workflow that integrates the VEN μ S satellite imagery and AOS offer encouraging results for the combination of remote sensing and crop modeling to improve yield estimations.

5. Conclusions

In this work, we have used VEN μ S satellite data to retrieve the *fvc* using state-of-the-art hybrid machine learning algorithms, specifically, GPR trained on PROSAIL model simulations of the canopy reflectance. The GPR algorithm was then applied to the VEN μ S satellite data time series to obtain the *fvc* data with their mean value of prediction and uncertainty and these were used to calibrate the AOS model in a Bayesian framework.

Bayesian inferences were preferred for the calibration as they transform the initial prior distribution to the posterior distribution from which the mean and median values can be computed, which reflects the optimal value of the parameter computed from its prior distribution. We implemented the Monte Carlo Markov chain-based DREAM extension, DREAM_(KZS), which involves three jumps: a parallel direction jump, snooker jump, and also a Kalman jump to speed up the convergence of the algorithm; we found that the algorithm is efficient in computing the parameter values in a time-constrained manner.

The results show that the AOS model gives good estimations of the canopy cover development on durum wheat. VEN μ S satellite images were useful to integrate the in-season *fvc* data for durum wheat, due to the revisit frequency of the satellite. Although the *fvc* provided good estimations of the canopy cover development, biomass and yield were overestimated in comparison to the ground data.

This study, therefore, concludes that the AOS crop model calibration improved the *fvc* simulated though it overestimated biomass and yield. Overestimation of the yield was not completely minimized, but it was reduced and thus it would be important to calibrate the model before yield estimation applications. The *fvc* derived from the VEN μ S satellite images was useful to perform such a calibration, specifically with Bayesian methods.

References

- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome 300, D05109.
- Andarzian, B., Bannayan, M., Steduto, P., Mazraeh, H., Barati, ME, Barati, MA, Rahnama, A., 2011. Validation and testing of the AquaCrop model under full and deficit irrigated wheat production in Iran. *Agricultural Water Management* 100, 1–8.
- Azzari, G., Jain, M., Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sensing of Environment* 202, 129–141.
- Beichl, I., Sullivan, F., 2000. The metropolis algorithm. *Computing in Science & Engineering* 2, 65–69.
- Biosciences, L., 1992. Lai 2000 Plant Canopy Analyzer. Operating Manual. Lincoln, USA.
- Bolton, D.K., Friedl, M.A., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology* 173, 74–84.
- Camargo Rodriguez, A.V., Ober, E.S., 2019. AquaCropR: Crop Growth Model for R. *Agronomy* 9, 378.

- Castaldi, F., Casa, R., Pelosi, F., Yang, H., 2015. Influence of acquisition time and resolution on wheat yield estimation at the field scale from canopy biophysical variables retrieved from SPOT satellite data. *International Journal of Remote Sensing* 36, 2438–2459.
- de Willigen, P., 1991. Nitrogen turnover in the soil-crop system; comparison of fourteen simulation models. *Fertilizer research* 27, 141–149. <https://doi.org/10.1007/BF01051122>
- Doraiswamy, P.C., Sinclair, T.R., Hollinger, S., Akhmedov, B., Stern, A., Prueger, J., 2005. Application of MODIS derived parameters for regional crop yield assessment. *Remote sensing of environment* 97, 192–202.
- Evensen, G., 2009. Data assimilation: the ensemble Kalman filter. Springer Science & Business Media.
- Ferrier, P., Crebassol, P., Dedieu, G., Hagolle, O., Meygret, A., Tinto, F., Yaniv, Y., Herscovitz, J., 2010. VEN μ S (Vegetation and environment monitoring on a new micro satellite), in: 2010 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 3736–3739.
- Foster, T., 2019. AquaCrop-OS v6.0a Reference Manual.
- Foster, T., 2016. AquaCrop-OS v5.0a Reference Manual.
- Foster, T., Brozovic, N., Butler, A.P., Neale, C.M., Raes, D., Steduto, P., Fereres, E., Hsiao, T.C., 2017. AquaCrop-OS: A tool for resilient management of land and water resources in agriculture, in: EGU General Assembly Conference Abstracts. p. 2842.
- Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical science* 7, 457–472.
- Gervois, S., de Noblet-Ducoudré, N., Viovy, N., Ciais, P., Brisson, N., Seguin, B., Perrier, A., 2004. Including croplands in a global biosphere model: methodology and evaluation at specific sites. *Earth Interactions* 8, 1–25.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology* 377, 80–91.
- Haario, H., Laine, M., Mira, A., Saksman, E., 2006. DRAM: efficient adaptive MCMC. *Statistics and computing* 16, 339–354.
- Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. *Bernoulli* 7, 223–242.
- Haario, H., Saksman, E., Tamminen, J., 1999. Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics* 14, 375–396.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109. <https://doi.org/10.1093/biomet/57.1.97>
- He, B., Li, X., Quan, X., Qiu, S., 2015. Estimating the Aboveground Dry Biomass of Grass by Assimilation of Retrieved LAI Into a Crop Growth Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8, 550–561. <https://doi.org/10.1109/JSTARS.2014.2360676>
- Hopmans, J.W., Bristow, K.L., 2002. Current capabilities and future needs of root water and nutrient uptake modeling, in: *Advances in Agronomy*. pp. 103–183.

- Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT+ SAIL models: A review of use for vegetation characterization. *Remote sensing of environment* 113, S56–S66.
- Jeon, J.-H., Park, C.-G., Engel, B.A., 2014. Comparison of performance between genetic algorithm and SCE-UA for calibration of SCS-CN surface runoff simulation. *Water* 6, 3433–3456.
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., Wang, J., 2018. A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy* 92, 141–152.
- Jin, X., Kumar, L., Li, Z., Xu, X., Yang, G., Wang, J., 2016. Estimation of winter wheat biomass and yield by combining the aquacrop model and field hyperspectral data. *Remote Sensing* 8, 972.
- Jin, X., Li, Z., Feng, H., Ren, Z., Li, S., 2020. Estimation of maize yield by assimilating biomass and canopy cover derived from hyperspectral data into the AquaCrop model. *Agricultural Water Management* 227, 105846.
- Jin, X., Li, Z., Yang, G., Yang, H., Feng, H., Xu, X., Wang, J., Li, X., Luo, J., 2017. Winter wheat yield estimation based on multi-source medium resolution optical and radar imaging data and the AquaCrop model using the particle swarm optimization algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing* 126, 24–37.
- Jin, Z., Azzari, G., Lobell, D.B., 2017. Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches. *Agricultural and forest meteorology* 247, 207–220.
- Jönsson, P., Eklundh, L., 2004. TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers & geosciences* 30, 833–845.
- Kang, Y., Özdoğan, M., 2019. Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. *Remote Sensing of Environment* 228, 144–163.
- Laloy, E., Rogiers, B., Vrugt, J.A., Mallants, D., Jacques, D., 2013. Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. *Water Resources Research* 49, 2664–2682.
- Laloy, E., Vrugt, J.A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-trace DREAM (ZS) and high-performance computing. *Water Resources Research* 48.
- Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. *Field Crops Research* 143, 56–64.
- Lobell, D.B., Asner, G.P., Ortiz-Monasterio, J.I., Benning, T.L., 2003. Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. *Agriculture, Ecosystems & Environment* 94, 205–220.
- Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. *Remote Sensing of Environment* 164, 324–333.
- Ma, Q., Su, Y., Guo, Q., 2017. Comparison of canopy cover estimations from airborne LiDAR, aerial imagery, and satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10, 4225–4236.
- Makowski, D., Hillier, J., Wallach, D., Andrieu, B., Jeuffroy, M., 2006. Parameter estimation for crop models. *Working with dynamic crop models*. Elsevier, Amsterdam 101–149.

- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. *The journal of chemical physics* 21, 1087–1092.
- Nielsen, D.C., Miceli-Garcia, J.J., Lyon, D.J., 2012. Canopy cover and leaf area index relationships for wheat, triticale, and corn. *Agronomy journal* 104, 1569–1573.
- Raes, D., 2009. The ETo calculator, evapotranspiration from a reference surface. Reference manual version 3.1, January. Food and Agricultural Organization of the United Nations.
- Revill, A., Florence, A., MacArthur, A., Hoad, S.P., Rees, R.M., Williams, M., 2019. The value of Sentinel-2 spectral bands for the assessment of winter wheat growth and development. *Remote Sensing* 11, 2050.
- Saab, M.T.A., Todorovic, M., Albrizio, R., 2015. Comparing AquaCrop and CropSyst models in simulating barley growth and yield under different water and nitrogen regimes. Does calibration year influence the performance of crop growth models? *Agricultural water management* 147, 21–33.
- Sadegh, M., Vrugt, J.A., 2014. Approximate bayesian computation using Markov chain Monte Carlo simulation: DREAM (ABC). *Water Resources Research* 50, 6767–6787.
- Saltelli, A., Tarantola, S., Campolongo, F., 2000. Sensitivity analysis as an ingredient of modeling. *Statistical Science* 15, 377–395.
- Silvestro, P., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., Huang, W., Casa, R., 2017. Estimating wheat yield in China at the field and district scale from the assimilation of satellite data into the Aquacrop and simple algorithm for yield (SAFY) models. *Remote Sensing* 9, 509.
- Silvestro, P.C., Casa, R., Pignatti, S., 2014. Development of an Assimilation Scheme for the Estimation of Drought-Induced Yield Losses Based on Multi-Source Remote Sensing and the AcquaCrop Model, in: *Dragon 3 Mid-Term Results Symposium*, Chengdu, China, 26–29 May 2014. ESA Communications.
- Silvestro, P.C., Pignatti, S., Yang, H., Yang, G., Pascucci, S., Castaldi, F., Casa, R., 2017. Sensitivity analysis of the Aquacrop and SAFYE crop models for the assessment of water limited winter wheat yield in regional scale applications. *PloS one* 12, e0187485.
- Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop—The FAO crop model to simulate yield response to water: I. Concepts and underlying principles. *Agronomy Journal* 101, 426–437.
- Ter Braak, C.J., 2006. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing* 16, 239–249.
- Thorp, K., Wang, G., West, A., Moran, M., Bronson, K., White, J., Mon, J., 2012. Estimating crop biophysical properties from remote sensing data by inverting linked radiative transfer and ecophysiological models. *Remote Sensing of Environment* 124, 224–233.
- Todorovic, M., Albrizio, R., Zivotic, L., Saab, M.-T.A., Stöckle, C., Steduto, P., 2009. Assessment of AquaCrop, CropSyst, and WOFOST models in the simulation of sunflower growth under different water regimes. *Agronomy Journal* 101, 509–521.
- Trombetta, A., Iacobellis, V., Tarantino, E., Gentile, F., 2016. Calibration of the AquaCrop model for winter wheat using MODIS LAI images. *Agricultural Water Management* 164, 304–316.

- Upreti, D., Huang, W., Kong, W., Pascucci, S., Pignatti, S., Zhou, X., Ye, H., Casa, R., 2019. A comparison of hybrid machine learning algorithms for the retrieval of wheat biophysical variables from sentinel-2. *Remote Sensing* 11, 481.
- Upreti, D., Pignatti, S., Pascucci, S., Tolomio, M., Li, Z., Huang, W., Casa, R., 2020. A comparison of moment-independent and variance-based global sensitivity analysis approaches for wheat yield estimation with the Aquacrop-OS model. *Agronomy* 10, 607.
- Van Oijen, M., Rougier, J., Smith, R., 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiology* 25, 915–927.
- Velleman, P.F., Hoaglin, D.C., 1981. Applications, basics, and computing of exploratory data analysis. Duxbury Press.
- Verrelst, J., Rivera, J.P., Veroustraete, F., Muñoz-Marí, J., Clevers, J.G., Camps-Valls, G., Moreno, J., 2015. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods—A comparison. *ISPRS Journal of Photogrammetry and Remote Sensing* 108, 260–272.
- Vrugt, J.A., 2016. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software* 75, 273–316.
- Vrugt, Jasper A., Gupta, H.V., Bastidas, L.A., Bouten, W., Sorooshian, S., 2003. Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water resources research* 39.
- Vrugt, Jasper A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* 39, 1–14.
- Vrugt, J.A., Ter Braak, C.J., 2011. DREAM (D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrology and Earth System Sciences* 15, 3701–3713.
- Vrugt, J.A., Ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation* 10, 273–290.
- Wallach, D., Goffinet, B., Bergez, J.-E., Debaeke, P., Leenhardt, D., Aubertot, J.-N., 2001. Parameter estimation for crop models: a new approach and application to a corn model. *Agronomy journal* 93, 757–766.
- Wallach, D., Makowski, D., Jones, J.W., Brun, F., 2018. Working with dynamic crop models: methods, tools and examples for agriculture and environment. Academic Press.
- Weiss, M., Baret, F., 2016. S2ToolBox Level 2 products: LAI, FAPAR, FCOVER.
- Whelan, B.M., McBratney, A.B., Minasny, B., 2002. Vesper 1.5—spatial prediction software for precision agriculture, in: *Proceedings of the 6th International Conference on Precision Agriculture*. (Ed. PC Robert)(ASA-CSSA-SSSA: Madison, WI).
- Xiangxiang, W., Quanjiu, W., Jun, F., Qiuping, F., 2013. Evaluation of the AquaCrop model for simulating the impact of water deficits and different irrigation regimes on the biomass and yield of winter wheat grown on China's Loess Plateau. *Agricultural Water Management* 129, 95–104.

Zaks, D.P., Kucharik, C.J., 2011. Data and monitoring needs for a more ecological agriculture. *Environmental Research Letters* 6, 014017.

Zhang, J., Vrugt, J.A., Shi, X., Lin, G., Zeng, L., Wu, L., 2017. Speed-up of posterior inference of highly-parameterized environmental models from a Kalman proposal distribution: DREAM (KZS). *arXiv preprint arXiv:1707.05431*.

Zhang, T., Su, J., Liu, C., Chen, W.-H., 2019. Bayesian calibration of AquaCrop model for winter wheat by assimilating UAV multi-spectral images. *Computers and Electronics in Agriculture* 167, 105052.

CHAPTER 4

CHAPTER 4. Ensemble Kalman Filter based Assimilation of canopy cover retrieved from Sentinel-2 and VENμS satellite data into Aquacrop-OS model for improved durum wheat yield estimation in Central Italy

(paper in preparation)

Deepak Upreti ^{1,*}, Stefano Pignatti ², Simone Pascucci ², Massimo Tolomio ¹, Zhenhai Li ³, Wenjiang Huang ⁴, Raffaele Casa ^{1,*}

¹ Department of Agriculture and Forestry Sciences, University of Tuscia, Via San Camillo de Lellis, 01100, Viterbo, Italy

² Consiglio Nazionale delle Ricerche, Institute of Methodologies for Environmental Analysis (CNR, IMAA), Via del Fosso del Cavaliere, 100, 00133, Rome, Italy

³ Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

⁴ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

* Corresponding authors: dupreti@unitus.it (D.U.); rcasa@unitus.it (R.C.)

Abstract: Accurate yield estimation at the field scale is essential for precision agriculture applications. In this paper, Aquacrop Open Source crop model was employed with the canopy cover time series retrieved from the Sentinel-2 and VENμS satellite data in the ensemble Kalman filter based data assimilation framework. The efficiency of the assimilation was compared by assimilating the canopy cover from only Sentinel-2, only VENμS and by combining both the Sentinel-2 and VENμS satellite data canopy cover. The assimilation procedure was applied at the field scale for two different years in the Maccarese Farm, Central Italy for the durum wheat crop growth season. Overall, all the three scenarios were able to simulate yield, but an overestimation was clearly observed in the assimilated yield specifically with the VENμS satellite derived canopy cover. The best approximation of the assimilated yield was with Sentinel-2 derived canopy cover with RMSE ranging from 0.98 tha⁻¹ – 2.29 tha⁻¹ for different fields. Similarly, RMSE ranged from 1.16 tha⁻¹ – 2.73 tha⁻¹ with the canopy cover assimilated from VENμS satellite data and 1.34 tha⁻¹ – 2.64 tha⁻¹ when assimilating canopy cover from both the satellite data. More frequent and increased number of observations from two different sources of satellite data did not improved the efficiency of the assimilation and simulated yield, 15-20 number of observations were sufficient to have the best approximation of the yield during the crop growth season.

Keywords: Aquacrop, AOS, canopy cover, ensemble Kalman filter, data assimilation, Sentinel-2, VENμS

1. Introduction

Spatial information on the crop yield is valuable for both crop management and market forecast (Lobell, 2013). This information is only available at the national or state level. Although being useful, this information does not capture within-field variability that is essential to inform field specific decision making for various sectors including farmers and crop insurance companies (Lobell and Azzari, 2017).

Therefore, robust methods for large scale field level crop yield estimation are important for agriculture research and applications (Zaks and Kucharik, 2011).

Remote Sensing data can provide spatially and temporally consistent information over large agricultural areas (Yuan et al., 2016) to monitor the seasonal patterns related to the biological lifecycle of the crop (Atzberger, 2013). Crop yield predictions can be obtained by extracting some variables from remote sensing data. Broadly, these methods are categorized as statistical methods and data assimilation methods (Bolton and Friedl, 2013; Doraiswamy et al., 2003; Lobell et al., 2003; Thorp et al., 2012). Statistical methods uses the *in situ* or synthetic data to train regression models or Machine Learning algorithms (Azzari et al., 2017; Jin et al., 2017; Sakamoto et al., 2013; Sibley et al., 2014). Despite being fast, simple and computationally efficient, these methods can suffer potential bias, if for example the synthetic data generated by the crop models are not an adequate proxy of the real data, or if *in situ* data are limited in terms of spatial resolution.

Remote Sensing can be used to derive in-season crop biophysical variables such as leaf area index (LAI) and canopy cover (CC) for monitoring crop growth. These variables can have good agreement upon validation with the corresponding ground measured values (Casa et al., 2019). The predictions of the models can be adjusted by updating in-season biophysical variables, that are state variables in the crop models, and can be extracted using remote sensing data. Potentially, such measurements could lead to a large improvement in model predictions (Makowski et al., 2006). How good will predictions be after correction with measurements? This depends on a number of factors, how closely the measured variables and predicted variables are related, the number of observations, early or late season observations etc... (Makowski et al., 2006).

There are three different strategies for integrating observed in-season biophysical variables into crop models: calibration, forcing and updating. In the calibration methods, the initial parameters of the crop model are adjusted to minimize the differences between the observed data and the simulated data using an optimization algorithm and some kind of cost function. A major drawback of the method is that the observed data is considered true, without taking into account the uncertainty contained in it (Kang and Özdoğan, 2019). Forcing methods are used to simply replace the simulated data with the observed data, and has the drawback that the latter should have the same time step as the former, e.g. daily observations typically in case a crop model is used (Casa et al., 2012). In the updating strategy, simulated data are updated with the observed data whenever observed data are available (Jin et al., 2018). The updating scheme is preferred over the others as it allows to consider the uncertainty in both the simulated and observed data (Jin et al., 2018). The updating strategy can be implemented using two different groups of techniques: variational and sequential. The former allows to update all the model simulations within a specific window, while the later recursively updates model simulation using observations only from corresponding times. Unlike in the statistical methods, yield data are not required for training the model in data assimilation, therefore this approach can be more generally applied, using satellite data for example from Sentinel-2, Vegetation and Environment Monitoring Micro-Satellite (VEN μ S) or Landsat missions.

Ensemble Kalman filter is the most commonly used sequential data assimilation method for crop models (Kang and Özdoğan, 2019). It updates the model state variable sequentially by remote sensing observation of derived biophysical/ agronomic variables such as leaf area index (Silvestro et al., 2017), canopy cover (Upreti et al., 2020a) or soil moisture (Ines et al., 2013). Ensemble Kalman filter considers the uncertainty in both the remotely sensed observed variables and the uncertainties in the model simulations,

calculating the Kalman gain factor from the current state and the uncertainties. The Kalman gain factor is designed in such a way that it is biased towards the less uncertain state, may it be simulated or observed. In this way the future prediction is improved. Mathematically it is simple and has only two steps of forecast and analysis at each observation date. There have been various studies of the application of ensemble Kalman filter data assimilation in earth science and crop modeling studies. Silvestro et al. (2017) assimilated leaf area index derived from Chinese satellite HJ1A/B and Landsat 8 Operational Land Imager into the SAFYE crop model for field and district scale yield prediction and found low error between simulated yield and observed yield. Huang et al. (2016, 2015) assimilated leaf area index into the WOFOST crop model and found low errors on yield prediction for smaller regions. However, the model was initially calibrated with the ground leaf area index and yield. The local weather conditions, soil profile, sowing and emergence dates greatly reduced model uncertainty and potential bias. A study (Ines et al., 2013) shows better performance of the ensemble Kalman filter for estimating yield at county level by assimilating MODIS derived leaf area index into the DSSAT model.

The performance of the ensemble Kalman filter is affected by the quality of the remote sensing observations and the quantification of the uncertainties in model and observations (Casa et al., 2012). The development of the gaussian processes regression algorithms for the retrieval of the in-season crop biophysical variables from remote sensing data allows the quantification of the associated uncertainties with the mean value of the prediction due to their Bayesian backbone (Upreti et al., 2019, 2020a; Verrelst et al., 2013). This associated uncertainty with the mean value of the biophysical variable can be used in the ensemble Kalman filter data assimilation for the uncertainty in the observed variables. Some studies (Dorigo et al., 2007; Huang et al., 2013; Ines et al., 2013; Wu et al., 2011) focused on the number of the observations during the crop growing season also affects the efficiency of the assimilation. Very limited observations may not have improved the prediction and reduce errors in the estimation. This can be overcome by combining satellite data from different sources.

The availability of remote sensing data with high temporal resolution would potentially offer advantages in monitoring crop phenological development variability during different development stages. However, satellite data with high temporal resolution generally tend to have a coarse spatial resolution and consistent time series of cloud-free images at medium spatial resolution are seldom available due to the frequent cloud cover during the crop growing season (Huang et al., 2016). With the advent of European Copernicus program, the Sentinel-2 data is available at 10m with a revisit of 5 days (Sentinel-2A & Sentinel-2B). However, the cloud cover in crop growing season is still a problem. The frequency of the data can be increased by combining different sources of satellite data.

VENμS is a joint Earth observation mission of the Israeli and French Space Agencies. It operates in 12 spectral bands with a revisit time of 2 days and a spatial resolution of 10m (Ferrier et al., 2010).

In the present study, we aim to understand the effects of combining data and their uncertainties from two different data sources and increase the temporal dimension which is essential for monitoring different development stages. Sentinel-2 and VENμS satellite derived canopy cover with their uncertainties in the retrieval are assimilated using ensemble Kalman filter into the recently developed Aquacrop-Open Source (AOS v6.0a) (Foster, 2019) crop water productivity model for the estimation of durum wheat yield for two different crop growing season 2017-2018 and 2018-2019 in Maccarese Farm, Central Italy.

2. Materials and Methods

2.1. Site Description, Soil and Weather Data

The experimental site is located in Maccarese (41.833° lat. N, 12.217° long. E, 8m a.s.l), western coast of Central Italy, near Fiumicino Airport, Rome (Figure 1). This area has a typical Mediterranean climate with wet autumn and dry hot summer. The average maximum temperature is 27.4°C in the summer, and the average minimum temperature is 5.0°C in the winter. The average annual precipitation is 812.9 mm. The soil in Maccarese is Cutanic Luvisol, with a prevailing sandy clay loam texture, becoming more clayey toward the North-East of the site.

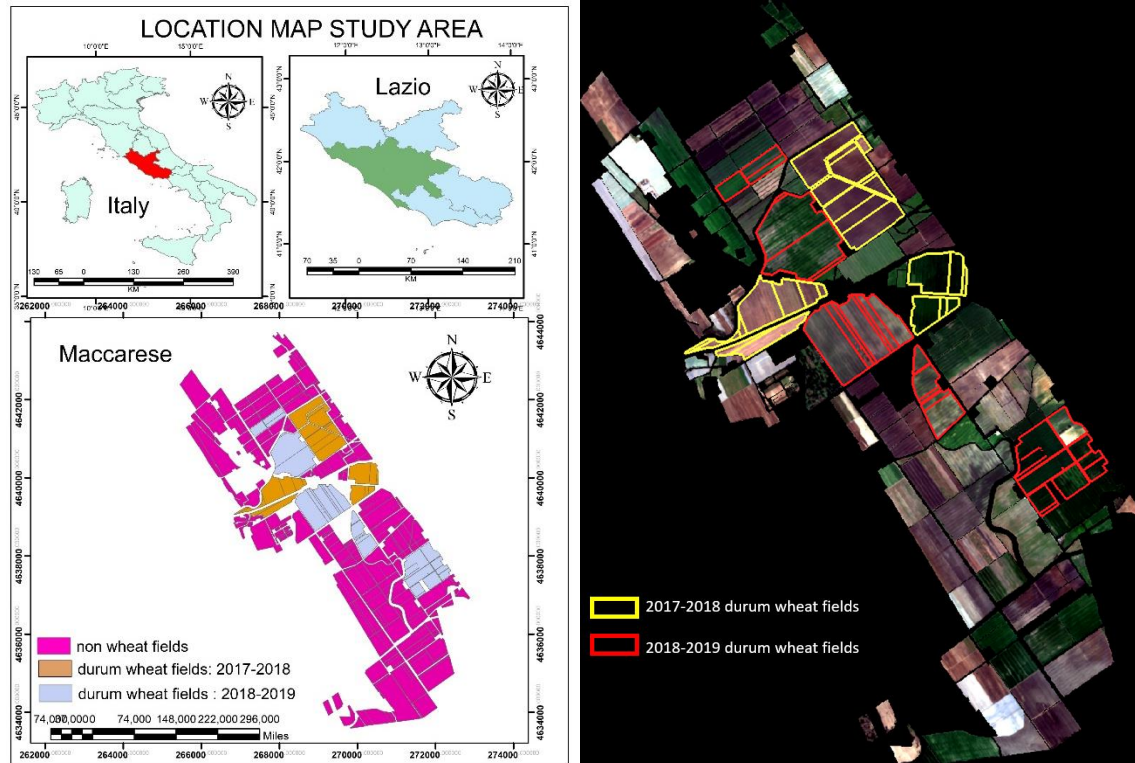


Figure 1. Study Site - Maccarese Farm, Rome, Central Italy (Left), VENμS, 6 March 2019 in true color composite (right) for durum wheat fields in Maccarese Farm

The sowing dates and the harvest dates for durum wheat crop for both the years 2018 and 2019 are presented in Table 1. Representative average soil conditions are presented in Table 2.

Table 1. Sowing and harvest dates for test sites

Period	Sowing Date	Harvest Date
2017-2018	30-Oct-17	28-Jun-18
2018-2019	10-Nov-18	15-July-19

Table 2. Representative Soil Conditions at the test site

Soil	Depth (cm)	Moisture Content (%)			Organic C (%)	Clay (%)	Silt (%)	pH	Sand (%)
		PWP	FC	Sat					

Sandy loam	0-10	8.0	18.0	45.0	2.4	33.3	19.0	6.6	47.8
Sandy loam	10-20	8.0	18.0	45.0	1.8	34.1	18.6	6.7	47.3
Sandy loam	20-30	8.0	18.0	45.0	1.8	33.6	20.4	6.7	46.0
Clay loam	30-40	18.0	23.0	43.0	1.8	34.6	19.0	6.7	46.4
Clay loam	40-50	18.0	23.0	43.0	1.5	32.4	17.8	6.8	49.8

The meteorological data used in this study to drive the model simulations were obtained from the Regione Lazio Agrometrological Service, Italy. The daily minimum and maximum temperature, relative humidity, wind speed, rainfall data and daily solar radiation for the period (2017 - 2019) were provided. The daily reference evapotranspiration was calculated based on the FAO Penman-Monteith method as described in (Allen et al., 1998) and the ET₀ calculator (Raes, 2009). The temperature minimum and maximum, rainfall and reference evapotranspiration for the test site is shown in Figure 2.

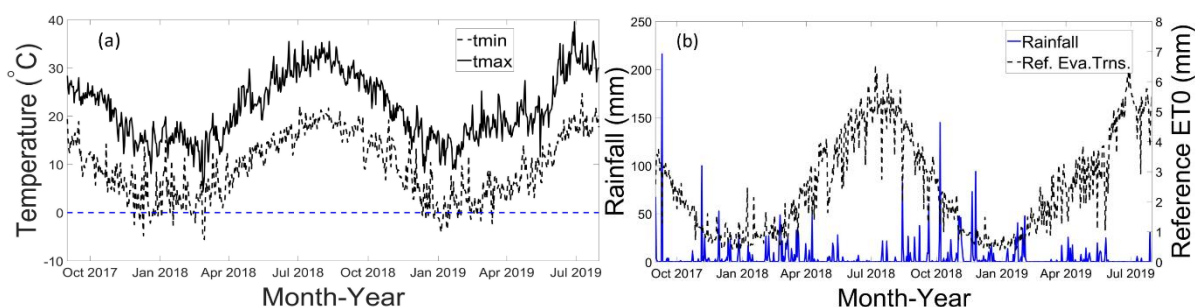


Figure 2. Maximum and minimum temperature (a), daily precipitation and reference evapotranspiration (b) for Maccarese, Central Italy

2.2. Yield Data Processing

The yield data obtained from the Maccarese Farm CNH 860x combine harvester yield mapping system was used for the validation of yield estimates. The raw data was previously cleaned and pre-processed. The pre-processing includes the removal of the points where the speed of the machine was less than 0.5 km h⁻¹. Also, all the points having yield greater than 15 t ha⁻¹ and less than 0.2 t ha⁻¹ were removed from the processing, as these were deemed unrealistic values. Further, very low yielding values were removed among adjacent high yield passes, which means that the operator was working not at full cutting width. As a next step, block kriging was carried out using local variogram option, this processing was completed in the open-source software Variogram Estimation and Spatial Prediction with Error (Vesper), developed by Australian Center for Precision Agriculture (Whelan et al., 2002).

2.3. Crop model: Aquacrop Open Source

The crop growth simulation model Aquacrop, was originally developed by Food and Agriculture Organization (FAO) of the United Nations. Since the original model does not allow the customization and the implementation of the Ensemble Kalman Filter (EnKF) data assimilation approach, In this work we have used the open source Matlab version of the model called as Aquacrop Open Source (AOS) developed by Foster et al. (2017). With respect to complex crop models such as DSSAT and WOFOST, Aquacrop is simpler, and has less free parameters in comparison to the other models. It has been used successfully to assess crop growth and yield (Jin et al., 2020, 2016; Silvestro et al., 2017; Xing et al., 2017).

Aquacrop model evolves from the most commonly used empirical approach to assess crop yield response to water (Doorenbos and Kassam, 1979; Raes et al., 2009). Eq.1

$$\left(\frac{Y_x - Y_a}{Y_x}\right) = k_y \left(\frac{ET_x - ET_a}{ET_x}\right) \quad (1)$$

Where Y_x and Y_a are the maximum and actual yield, ET_x and ET_a are the maximum and actual evapotranspiration, and k_y is the proportionality factor between relative yield loss and relative reduction in evapotranspiration. The model progressed by separating by evapotranspiration (ET) into crop transpiration (Tr) and soil evaporation (E) and Yield (Y) as a function of biomass (B) and harvest index (HI) (Raes et al., 2009). The separation of ET into Tr and E avoids the confounding effects of the non-productive consumption use of water (E), which is important especially during incomplete ground cover, and led to the conceptual equation at the core of the model (Raes et al., 2009) Eq.2

$$B = WP \times \Sigma Tr \quad (2)$$

$$Y = B \times HI \quad (3)$$

Where Tr is the crop transpiration (in mm) and WP is the water productivity (kg of biomass per m² and per mm of cumulated water transpired over the time period in which the biomass is produced) which is a constant for given climatic condition. By normalizing it for different climatic conditions, WP becomes a conservative parameter (Steduto et al., 2007). The other main advantage of Aquacrop is the time scale used, it simulate crop growth in daily time steps, a period closer to and approaching the time scale of crop responses to water deficits (Acevedo et al., 1971).

Aquacrop structures its soil-crop-atmosphere continuum by including (i) the soil, with its water balance, (ii) the crop, with its growth, development and yield processes and (iii) the atmosphere, with its thermal regime, rainfall, evaporative demand and annual carbon dioxide concentration. Some management practices are included, with emphasis on irrigation, but also on the level of soil fertility, since they affect crop development, water productivity and crop adjustment to stresses and therefor final yield (Raes et al., 2009).

AOS (T. Foster et al., 2017; Timothy Foster et al., 2017) is the open source Matlab based version of the original Aquacrop model developed by the FAO. Simulations were conducted to verify the suitability of AOS 5.0a to reproduce accurately the calculations and outputs from the original FAO Aquacrop model. Test simulations considered a range of crop types and environments: wheat production in Tunisia, rice and wheat production in Hyderabad, India and potato production in Brussels, Belgium. The maximum RMSE and minimum r^2 for crop yields across all simulations were 0.045 t ha⁻¹ and 0.993 respectively, that confirms the ability of AOS to reproduce successfully Aquacrop calculations and outputs (Foster, 2016).

Recently, a new version of AOS, AOS v6.0a has been released (Foster, 2019). AOS v6.0a implements all calculation procedures performed by the original FAO Aquacrop v6.0a model, with the exception of soil fertility stress, soil salinity stress and weed management as these are yet to be tested extensively for different crop types and environment worldwide. There are some 89 parameters in the AOS model as mentioned in the reference manual (Foster, 2016). The list of AOS crop model parameters are presented in Table 3.

2.4. Sensitivity Analysis and Bayesian Calibration of the AOS crop model

In a previous work on sensitivity analysis, Upreti et al. (2020b) applied two different methods: moment independent density-based PAWN and variance-based Extended Fourier Amplitude Sensitivity Test (EFAST) for the identification of the most sensitive AOS crop model parameters. A common set of the most influential parameters was found with both methods, although the ranking of the parameters varied, for the same study site. Upreti et al. (2020a) calibrated the same set of parameters by assimilating canopy cover retrieved from VEN μ S satellite images with Markov Chain Monte Carlo (MCMC) based Differential Evolution Adaptive Metropolis (DREAM) extension DREAM_(KZS). Table 3 presents the free parameters of the AOS crop model used in the present work after performing sensitivity analysis (Upreti et al., 2020b) and Bayesian calibration (Upreti et al., 2020a) for the same study site of Maccarese Farm.

Table 3. Free parameters of the AOS crop model parameters, descriptions and ranges used in the present work.

Parameter	Unit	Description	Range
<i>CGC</i>	fraction GDD	Canopy growth coefficient	(0.0042, 0.0078)
<i>Hstart</i>	GDD	Thermal time from sowing to start of yield formation	(1090, 1395)
<i>WP</i>	gm ⁻²	Water productivity normalized for reference evapotranspiration and atmospheric carbon dioxide	(11, 22)
<i>HIO</i>	%	Reference harvest index	(0.32, 0.59)
<i>Senescence</i>	GDD	Thermal time from sowing to start of canopy senescence	(1090, 2250)
<i>Emergence</i>	GDD	Thermal time from sowing to emergence	(90, 230)
<i>GDD_up</i>	GDD	Minimum number of GDD's required for full biomass production	(9, 18)
<i>Maturity</i>	GDD	Thermal time from sowing to physiological maturity	(1590, 3150)
<i>Tmin_up</i>	°C	Minimum temperature below which pollination begins to fail	(3, 6)
<i>GDD_lo</i>	GDD	Minimum number of GDD's required for any biomass production to occur	(0, 5)
<i>b_HI</i>		Coefficient describing the negative impact on harvest index of stomatal closure post-anthesis	(3, 6)

GDD: growing degree days

2.5. Satellite Data and Processing

2.5.1. Sentinel-2 and VEN μ S satellite data

Sentinel-2 level 2A atmospherically corrected satellite data were downloaded for two years 2017-2018 and 2018-2019 of durum wheat crop growing seasons for the study site. Overall, 16 cloud free images for the period 2017-2018 and 22 cloud free images were available for the period 2018-2019 for our study site. The images were downloaded from Copernicus Open Access Hub (<https://scihub.copernicus.eu/>).

VEN μ S, a joint earth observation mission of Israel and France space agencies was launched in August 2017 with an expected life span of 3 years. VEN μ S satellite data has a high temporal (2 days) and spatial resolution (5.3 m) for around 100 sites all around the world (Ferrier et al., 2010). It operates in 12 spectral bands from visible to near-infrared, suitable for vegetation studies. Due to the signal to noise ratio, the data after atmospheric corrections were initially provided at 10 m spatial resolution. This 10 m data for the two-durum wheat crop growing seasons November 2017-June 2018 and November 2018 – June 2019 was downloaded from Theia Land Data Center (<https://theia.cnes.fr/>). Overall, 15 images were available

for the year 2017-2018 and 33 for 2018-2019. It should be noted that, at present, the entire archive is being re-processed to deliver the satellite data at 5m.

On combining canopy cover from both the satellite data, i.e., to increase the number of observations during the crop growing season, we could obtain 24 images for the year 2017-2018 and 47 images for the year 2018-2019. A comparison of Sentinel-2 and VEN μ S satellite bands and details are presented in Table 4.

Table 4. Sentinel-2 and VEN μ S satellite bands information.

Sentinel-2				VEN μ S			
Bands	Center (nm)	Width (nm)	Resolution (m)	Bands	Center (nm)	Width (nm)	Applications
B1	443	20	60	B1	420	25	Atmospheric Correction
B2	490	65	10	B2	443	40	Aerosols, Clouds
B3	560	35	10	B3	490	20	Atmospheric Correction
B4	665	30	10	B4	555	20	Land
B5	705	15	20	B5	638	24	Vegetation Indices
B6	740	15	20	B6	638	24	DEM, Image Quality
B7	783	20	20	B7	672	16	Red Edge
B8	842	115	10	B8	702	24	Red Edge
B8A	865	20	20	B9	742	16	Red Edge
B9	945	20	60	B10	782	16	Red Edge
B10	1375	30	60	B11	865	20	Vegetation Indices
B11	1610	90	20	B12	910	20	Water Vapor
B12	2190	180	20	-	-	-	-

2.5.2. Canopy cover retrieval from Sentinel-2 and VEN μ S data

We have used canopy cover retrieved from Sentinel-2 and VEN μ S satellite to update the crop model simulations in the data assimilation framework. Canopy cover time series of a single field / pixel during the growing season typically from November to June, were generated using the Gaussian Process Regression algorithm (GPR) algorithm. Multiple hybrid machine learning regressors were compared for the retrieval of biophysical variables from Sentinel-2 data for the test site. GPR was best performing for the retrieval of canopy cover (Upreti et al., 2019). When compared with the ground measured canopy cover data, the values of R^2 , RMSE and RRMSE were 0.89, 0.08 and 9.86% respectively. The canopy cover retrieved using VEN μ S satellite data was also in good agreement when compared to the ground measured ($R^2 = 0.76$, RMSE=0.09, RRMSE=11.6%) (Upreti et al., 2020a), however, the accuracy of retrieval with Sentinel-2 was higher than on VEN μ S data. The other advantage of the GPR algorithm is that it also provides the uncertainty estimates with the mean value of prediction. This error and uncertainty was also used by Upreti et al. (2020a) in his study to get the Bayesian inferences on the AOS crop model parameters. The same algorithm was also employed to retrieve canopy cover from both Sentinel-2 and VEN μ S satellite data for the same period, i.e. November - June for two years 2018 and 2019. For the details on the retrieval of the canopy cover from Sentinel-2 and VEN μ S satellite data interested readers are referred to Upreti et al. (2020a, 2019).

2.5.3. Smoothing and gap-filling time series canopy cover data

The canopy cover time series retrieved from Sentinel-2 and VEN μ S satellite data was then screened and smoothed to remove outliers and increase observation frequency. We implemented 4253H twice filter (Hoaglin and Velleman, 1949) for smoothing of the canopy cover time series, the smoothing process first interpolated the canopy cover time series into daily interval and then fit a double logistic function to approximate phenological development. The 4253H twice algorithm first takes a running median of 4, then 2, and then 5, and then 3 data points over the raw time series, followed by a running weighted average. This procedure is then applied to the residuals and the two results are added together. Finally, the entire time series was scaled to the maxima of the original (un-smoothed) to retain the peak values (Upreti et al., 2020a). Figure 3 shows an example of the smoothing and the double logistic fitting for a single pixel from a field for the test site Maccarese, in 2018.

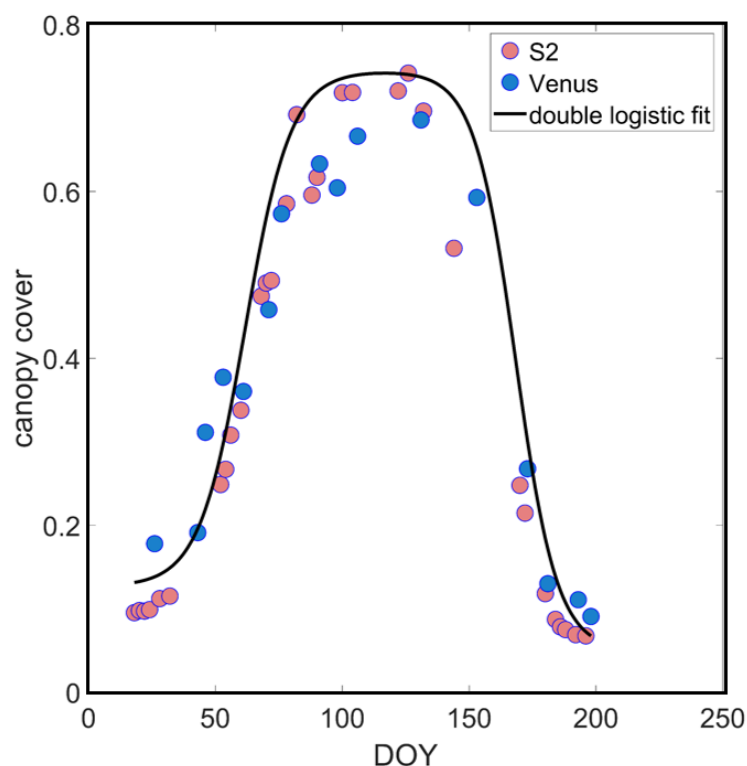


Figure 3. Sentinel-2 and VEN μ S satellite retrieved canopy cover and after smoothing and fitting double logistic fit

2.5.4. Data Assimilation Strategy

Kalman filter is best known for updating the state variables sequentially in time i.e., on the availability of the observations for linear models (Kalman, 1960). Due to the modeling of the physical processes as mathematical equations, the crop models are complex and highly non-linear in nature and updating the state variables using Kalman filter can be performed by transforming non-linear crop models to the linear models the method popularly known as extended Kalman filter. The main drawback of the algorithm is

that linearization of the non-linear models has been shown to be a poor approximation in a number of applications (Makowski et al., 2006).

Unlike the extended Kalman filter, the methodology of assimilation using ensemble Kalman filter does not require the linear approximation of the model. The methodology of assimilation using ensemble Kalman filter is presented in Figure 4. The ensemble Kalman filter was described by Burgers et al. (1998). The principle is to approximate the probability distributions using random samples of state variable and parameter values. First, an ensemble of N values of $\phi_t, \phi_t^1, \dots, \phi_t^j, \dots, \phi_t^N$ and an ensemble of N values of Measurements $M_t + \tau_t, M_t^1, \dots, M_t^j, \dots, M_t^N$, are randomly generated. Second, the Kalman filter equation is applied to each ensemble element as follows (Makowski et al., 2006):

$$\phi_{t,K}^j = \phi_t^j + K_t^e (M_t^j - R_t \phi_t^j) \quad (4)$$

where K_t^e is a $(n+p) \times q$ matrix defined by $K_t^e = \Sigma_t^e R_t^T [R_t \Sigma_t^e R_t^T + \text{var}(\tau_t)]^{-1}$, $\text{var}(\tau_t)$ is the variance – covariance matrix of the measurement error, and Σ_t^e is the $(n+p) \times (n+p)$ variance – covariance matrix of N vectors $\phi_t^j, j = 1, \dots, N$. The ensemble of state variables $\phi_t^j, j = 1, \dots, N$, describes the uncertainty in the state variable and parameter values before using the measurements M_t .

In the present work canopy cover time series derived from Sentinel-2 and VEN μ S satellite data were updated in the assimilation framework. GPR algorithm was implemented for the retrieval of the canopy cover time series from satellite data, due to the Bayesian backbone of the algorithm the associated uncertainties with the mean value of the prediction was estimated (Upreti et al., 2020a, 2019).

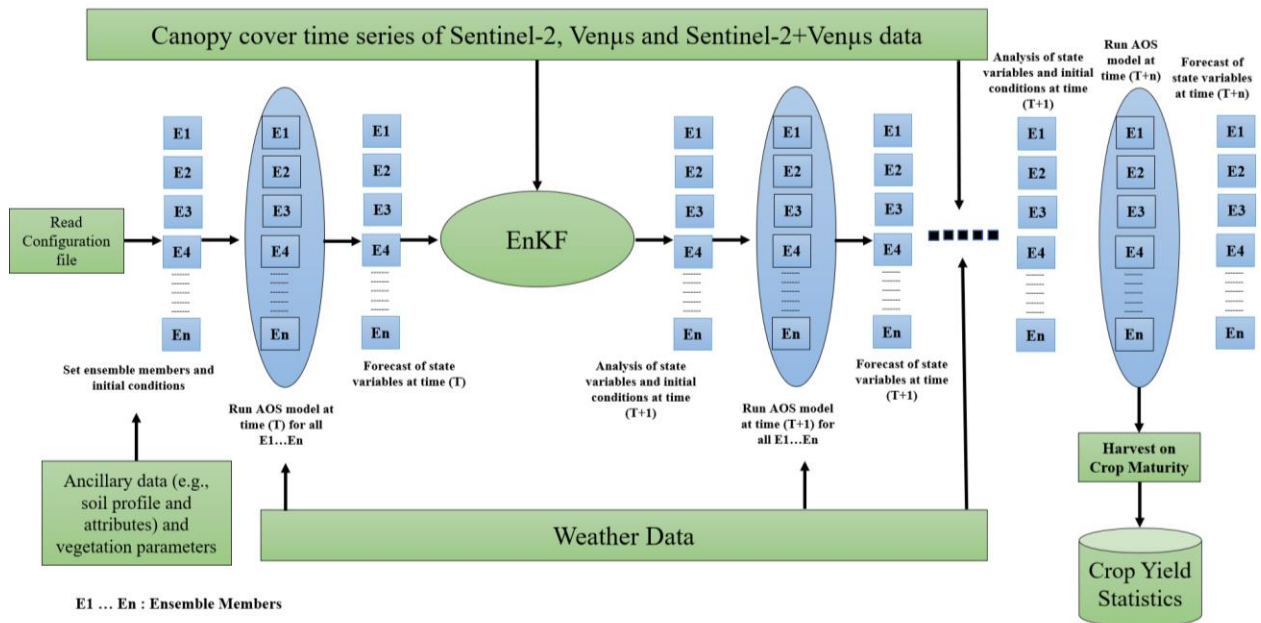


Figure 4. Data Assimilation procedure, when an observation of canopy cover is available, the ensemble stops propagating and performs analysis that updates simulated canopy cover for time T in daily time steps. Assimilation performed with canopy cover from only Sentinel-2 data, only VEN μ S data and combining both Sentinel-2 and VEN μ S satellite data.

2.5.5. Evaluation and Validation

The performance of data assimilation was assessed by comparing pixel-wise estimated grain yield values with measured values from the Maccaresse farm combine harvester yield mapping system, using the standard error metrics including Root Mean Square Error (RMSE) and Relative RMSE (RRMSE). Figure 5 and Figure 6 present the visual interpretation of the uncertainties in both the assimilated yield and the observed yield data processed using Vesper for two different fields for two different crop growing seasons of 2018 and 2019.

3. Results and Discussions

3.1. Ensemble Kalman Filter based data assimilation

The yield is computed in four steps in the AOS model. Initially the crop development is computed, followed by the crop transpiration estimation. Biomass is then computed and finally with harvest index the crop yield is simulated. The canopy cover time series was retrieved for both Sentinel-2 and VEN μ S for two years 2017-2018 and 2018-2019. The canopy cover was retrieved with $R^2=0.9$, RMSE=0.08 and RRMSE=9.86 (Upreti et al., 2020b) from Sentinel-2 data and with $R^2=0.8$, RMSE=0.09 and RRMSE=11.6 (Upreti et al., 2020a) from VEN μ S data. The estimated yield was compared to the observed yield. Table 5 presents the assimilation results. Figure 5 and Figure 6, presents the scatter plots of observed and simulated yield, observed uncertainties and uncertainties in Sentinel-2, VEN μ S and their combination.

As can be seen from the Figure 5, the assimilation results with Sentinel-2 data derived canopy cover are better than the assimilation of the canopy cover derived from VEN μ S data and their combination. The uncertainties in the observed yield data is lower than then uncertainties in the assimilation of the canopy cover derived from the satellite data.

The canopy cover time series during the crop growth season was assimilated for 15 different fields in the study area, 8 fields of durum wheat crop for the year 2017-2018 and 7 different fields of durum wheat crop for the year 2018-2019 (Figure 1). The assimilation results for all the 15 fields for both years are presented in Table 5. Figure 5 presents the results for the field B041 and Figure 6 shows the results for the field B069. It can be seen from both Figure 5 and Figure 6 that the observed yield uncertainty is lower than the assimilated uncertainty.

The simulated yield is comparatively in good agreement for Sentinel-2 data derived canopy cover, then with the canopy cover of VEN μ S data and their combination. The simulated yield by assimilating with the VEN μ S data derived canopy cover and Sentinel-2, VEN μ S combined canopy cover is overestimated. Also, the assimilated uncertainties are higher with the VEN μ S and combined data. The point to be noted is that the assimilation results did not simulated low yield values. In Figure 5, we see that the observed yield on the ground is ranging from 3 t ha⁻¹ to 9 t ha⁻¹. When we assimilate the Sentinel-2 data, the simulated yield is ranging from 5 t ha⁻¹ to 10 t ha⁻¹, and the simulated yield assimilating VEN μ S data is ranging from 6 t ha⁻¹ to 11 t ha⁻¹. When we combine the canopy cover from both the satellite data, the simulated yield ranges from 6 t ha⁻¹ to 10 t ha⁻¹. In all cases, it is clear that the simulated yield is overestimated. But the simulated yields using Sentinel-2 data was closer to the observed yield on the ground.

Table 5. RMSE and RRMSE between predicted yield and observed yield for different fields

Year	Field Code	Sentinel-2		VEN μ S		Sentinel-2 + VEN μ S	
		RMSE	RRMSE	RMSE	RRMSE	RMSE	RRMSE
2017-2018	B037	0.98	21.15	1.30	26.85	1.45	30.37
	B040	1.07	17.84	1.16	21.58	1.46	34.44
	B041	1.35	20.30	2.10	31.57	1.34	19.83
	B043	1.16	15.19	1.67	23.92	1.87	30.32
	B050	1.73	26.97	2.23	40.28	2.64	56.27
	B057	2.10	33.42	2.56	35.67	2.80	35.85
	B064	2.29	25.55	2.73	30.59	2.54	28.39
	B065	1.94	23.81	1.91	23.96	2.62	32.20
2018-2019	B031	1.83	24.65	2.04	27.49	2.40	30.09
	B032	1.99	30.37	2.09	31.69	2.39	36.00
	B044	1.87	40.27	2.10	42.53	2.41	53.10
	B062	1.97	29.56	2.39	31.26	1.82	29.34
	B069	1.79	23.58	2.02	26.62	2.18	28.31
	B085	1.32	41.23	1.76	49.31	2.04	48.13
	B109	1.78	35.46	2.15	31.43	1.74	30.86

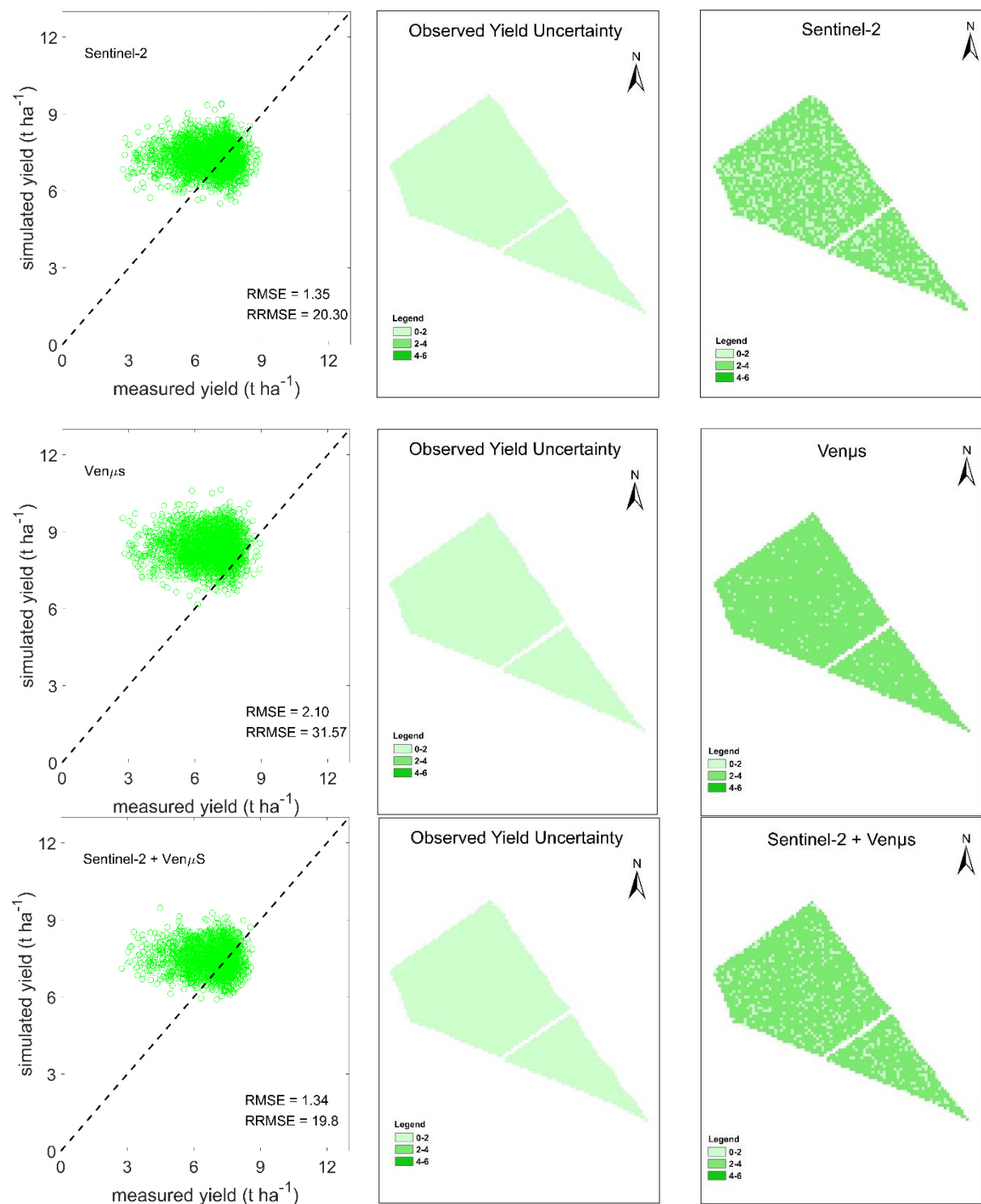


Figure 5. Observed and Predicted Yield, observed uncertainty and assimilated uncertainty for field B041 for the year 2017-2018

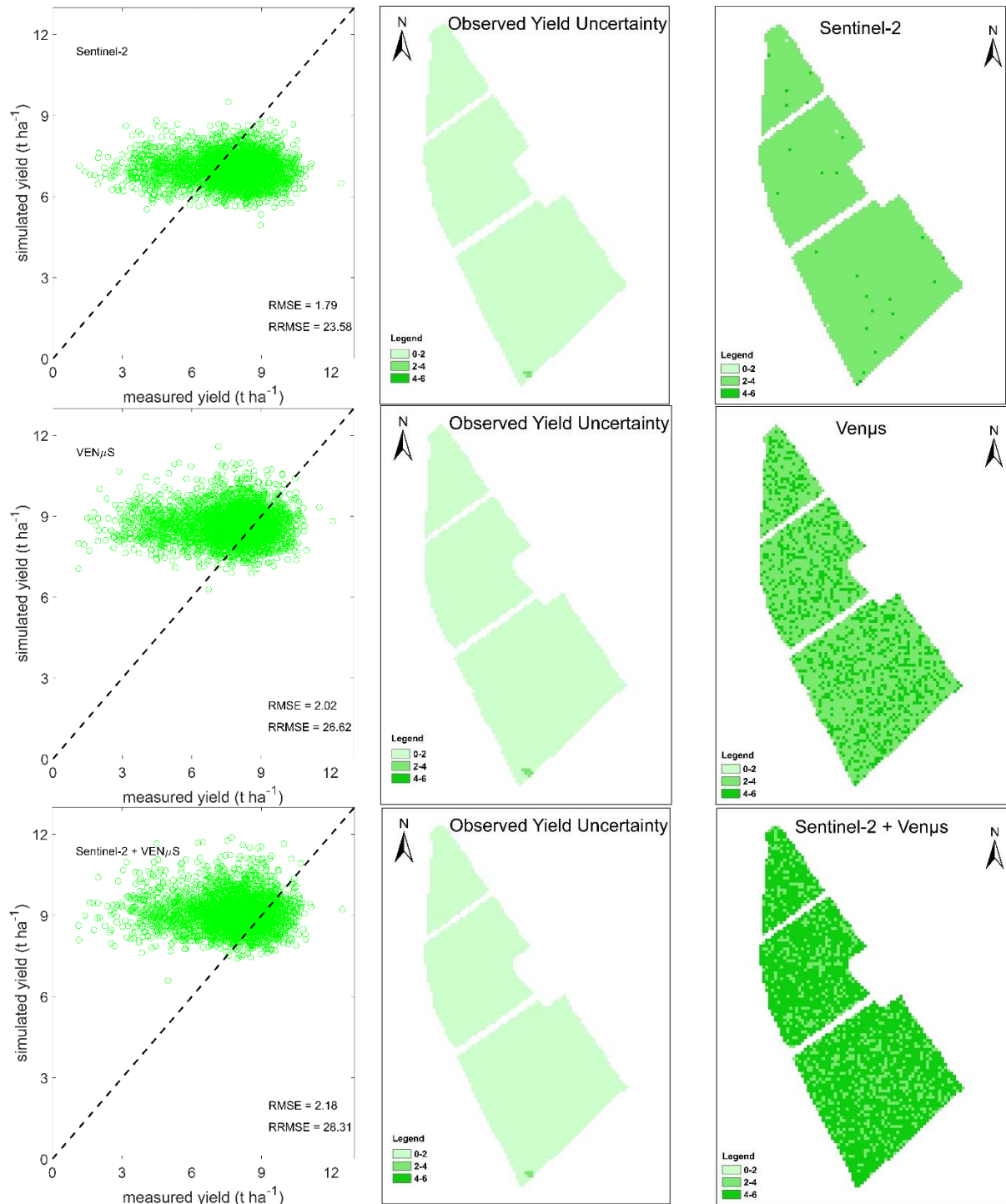


Figure 6. Observed and Predicted Yield, observed uncertainty and assimilated uncertainty for field B069 for the year 2018-2019

In Figure 6, we see that the observed yield on the ground is ranging from 2 t ha⁻¹ to 11 t ha⁻¹. When we assimilate the Sentinel-2 data, the simulated yield is ranging from 4 t ha⁻¹ to 10 t ha⁻¹, the simulated yield assimilating VENμS data is ranging from 6 t ha⁻¹ - 12 t ha⁻¹ and when we combine the canopy cover from both the satellite data, the simulated yield ranges from 6 t ha⁻¹ - 12 t ha⁻¹. In all the cases, it is clear that the

simulated yield is overestimated. But the simulated yields using Sentinel-2 data was closer to the observed yield on the ground.

Over all, the RMSE between the simulated yield and the observed yield for all the 15 fields for both the year's crop growing season ranges from 0.98 t ha^{-1} – 2.29 t ha^{-1} by assimilating Sentinel-2 data derived canopy cover, 1.16 t ha^{-1} – 2.73 t ha^{-1} by assimilating VEN μ S derived canopy cover and 1.34 t ha^{-1} – 2.64 t ha^{-1} by assimilating the canopy cover combining both Sentinel-2 and VEN μ S data (Table 5).

The Sentinel-2 data is available every 5 days and VEN μ S data in every 2 days in cloud free conditions and correct image processed and their availability. In our work we could obtain, during the crop growing season of 2017-2018, 16 cloud free images for Sentinel-2, 15 images of VEN μ S data and 24 images when we combined both satellite data. The increased number of observations is expected to improve the yield estimation in the assimilation process (Huang et al., 2013; Ines et al., 2013; Wu et al., 2011), however in our study we found that the assimilated yield was better approximated with 16 in-season observations of canopy cover from Sentinel-2, and similar approximation was also found by assimilating 24 observations of in season canopy cover from Sentinel-2 and VEN μ S combined. The minimum number of observations used in assimilation should be 6 in order to improve the efficiency of the assimilation (Silvestro et al., 2017). We found that the assimilation results with 16 observations from Sentinel-2 was similar to the assimilation of increased number to assimilated observations, however assimilation with the increased number of observations increased the computational time. Also, the point to note is that the canopy cover derived from the VEN μ S satellite led to the overestimation of the of the assimilated yield, however with the 15 observations during the crop growing season. The reason may be the canopy cover derived from VEN μ S satellite data was somewhat overestimated in comparison to the canopy cover derived from the Sentinel-2 data (Upreti et al., 2020a, 2019). The canopy cover retrieved from Sentinel-2, VEN μ S and their combination was processed with the smoothing using 4253H twice filter and double logistic fit for the best approximation of the phenological cycle and reduce the uncertainties (Figure 3). The uncertainties were minimized by taking into account the effect of the most influential crop model parameters by applying sensitivity analysis (Upreti et al., 2020b) and performing the model calibration (Upreti et al., 2020a) for the study area, this allowed us to implement the optimal value of the most influential crop model parameters into the assimilation procedure.

3.2. Spatial Application of Ensemble Kalman filter based assimilation

Figure 7 and Figure 8 shows the spatialized application of the observed yield and the assimilated yield. The assimilation procedure was applied to the all the pixels in the field. In Figure 7, an example of a single field B041 for the year 2017-2018 durum wheat crop growing season is presented. The observed and assimilated yield ranges from 2 t ha^{-1} to 12 t ha^{-1} . As we can see, the observed yield prevalingly lies between 2 t ha^{-1} and 8 t ha^{-1} . The yield after assimilating Sentinel-2 data, ranges from 4 t ha^{-1} to 8 t ha^{-1} . By assimilating VEN μ S satellite derived canopy cover the yield is overestimated, ranging from 6 t ha^{-1} to 10 t ha^{-1} . By combining Sentinel-2 and VEN μ S derived canopy cover, the assimilation error is somewhat reduced and is closer to the assimilated yield using only Sentinel-2.

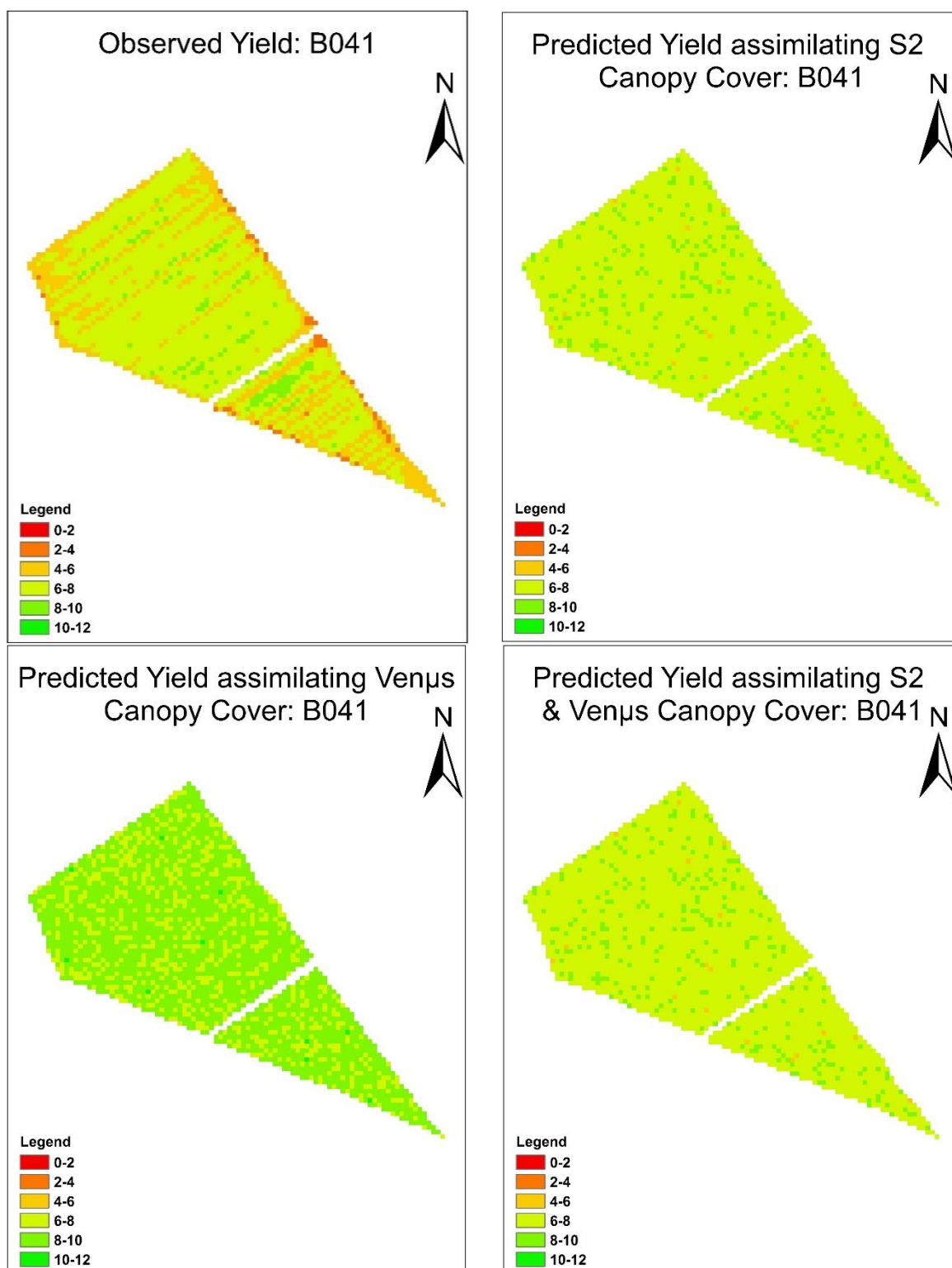


Figure 7. Pixel level map of observed yield and assimilated yield with Sentinel-2, VENµS and S2+VENµS derived canopy cover for field B041 for the year 2017-2018

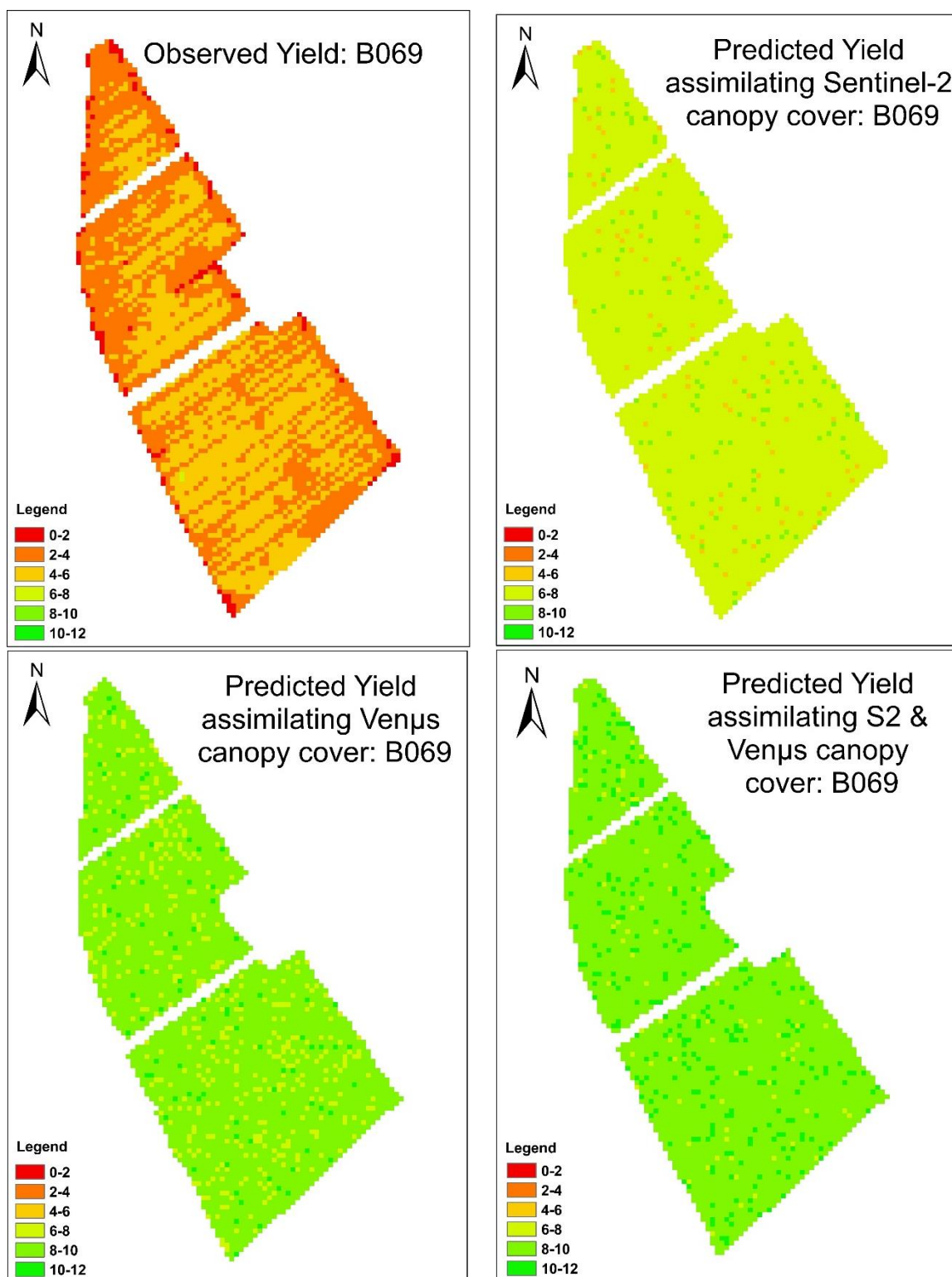


Figure 8. Pixel level map of observed yield and assimilated yield with Sentinel-2, VENµS and S2+VENµS derived canopy cover for field B069 for the year 2018-2019

Similarly, for the durum wheat crop growing season of 2018-2019, an example of assimilated and observed yield for the field B069 is presented in Figure 8. As for the field in Figure 7, field B069 in the Figure

8 shows that the assimilated yield using Sentinel-2 is $4 \text{ t ha}^{-1} - 8 \text{ t ha}^{-1}$, which is close to the observed yield $2 \text{ t ha}^{-1} - 8 \text{ t ha}^{-1}$. On the other hand, the assimilated yield using VEN μ S satellite derived canopy cover lies in between 6 t ha^{-1} and 10 t ha^{-1} , and that of the combination also ranges between 6 t ha^{-1} and 10 t ha^{-1} .

Overall, the assimilation results are in good agreement with the in-season canopy cover derived from the Sentinel-2 satellite data with respect to the canopy cover derived from the VEN μ S satellite data and by their combination in terms of average values. However, it should be noted that assimilation results failed to reproduce the within-field spatial variability observed in the measured yield data. The more frequent number of observations in the assimilation procedure did not improve results. It may be due to the fact that the observations were assimilated during a saturation period of maximum canopy cover, which is nearly 50 days. The availability of the satellite data in the first half of the season is more important than the availability of the satellite data in the saturation period.

4. Conclusions

In this work, we derived canopy cover time series from two different satellite Sentinel-2, VEN μ S for two different durum wheat crop growing season in Maccarese Farm in Central Italy. The main purpose of the work is to assimilate the canopy cover into the Aquacrop-OS model for durum wheat yield at field scale and to analyze the efficiency of the assimilation procedure by combining canopy cover only from Sentinel-2, only from VEN μ S and by combining both of them into the crop model.

The canopy cover retrieved from Sentinel-2 was in better agreement than the canopy cover retrieved from the VEN μ S satellite data. The time series of the canopy cover was somehow noisy, which was improved by performing the smoothing and fitting of the temporal data. The sensitivity analysis and the calibration of the crop model gave us the optimal values for the crop model parameters, which in turn reduced uncertainties in the model and the temporal canopy cover was assimilated into the model.

As sequential data assimilation-based ensemble Kalman filter approach is generally assumed to be improved by ingesting an increased number of observations, but this did not match with our results. The minimum number of observations is generally considered to be 6 in the previous studies. We found that a number of 15-20 observations assimilated provided the best approximation of the yield in the comparison to the observed data. By further increasing the number of observations during the season did not improve the assimilation results, it rather increases the computational time of the assimilation.

References

- Acevedo, E., Hsiao, T.C., Henderson, D., 1971. Immediate and subsequent growth responses of maize leaves to changes in water status. *Plant Physiology* 48, 631–636.
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome 300, D05109.
- Atzberger, C., 2013. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sensing* 5, 949–981.
- Azzari, G., Jain, M., Lobell, D.B., 2017. Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sensing of Environment* 202, 129–141.

- Bolton, D.K., Friedl, M.A., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology* 173, 74–84.
- Burgers, G., Jan van Leeuwen, P., Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. *Monthly weather review* 126, 1719–1724.
- Casa, R., Upreti, D., Pelosi, F., 2019. Measurement and estimation of leaf area index (LAI) using commercial instruments and smartphone-based systems. Presented at the IOP Conference Series: Earth and Environmental Science, IOP Publishing, p. 012006.
- Casa, R., Varella, H., Buis, S., Guérif, M., De Solan, B., Baret, F., 2012. Forcing a wheat crop model with LAI data to access agronomic variables: Evaluation of the impact of model and LAI uncertainties and comparison with an empirical approach. *European journal of agronomy* 37, 1–10.
- Doorenbos, J., Kassam, A., 1979. FAO irrigation and drainage paper No. 33 “Yield response to water.” FAO–Food and Agriculture Organization of the United Nations, Rome.
- Doraiswamy, P.C., Moulin, S., Cook, P.W., Stern, A., 2003. Crop yield assessment from remote sensing. *Photogrammetric engineering & remote sensing* 69, 665–674.
- Dorigo, W.A., Zurita-Milla, R., de Wit, A.J., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International journal of applied earth observation and geoinformation* 9, 165–193.
- Ferrier, P., Crebassol, P., Dedieu, G., Hagolle, O., Meygret, A., Tinto, F., Yaniv, Y., Herscovitz, J., 2010. VENμS (Vegetation and environment monitoring on a new micro satellite), in: 2010 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 3736–3739.
- Foster, T., 2019. AquaCrop-OS v6.0a Reference Manual.
- Foster, Tim, 2016. Supplementary information for ‘AquaCrop-OS: An open source version [WWW Document]’. URL <https://ars.els-cdn.com/content/image/1-s2.0-S0378377416304589-mmc1.pdf> (accessed 4.22.20).
- Foster, Timothy, 2016. AquaCrop-OS v5.0a Reference Manual.
- Foster, Timothy, Brozovic, N., Butler, A.P., Neale, C.M., Raes, D., Steduto, P., Fereres, E., Hsiao, T.C., 2017. AquaCrop-OS: A tool for resilient management of land and water resources in agriculture, in: EGU General Assembly Conference Abstracts. p. 2842.
- Foster, T., Brozović, N., Butler, A.P., Neale, C.M.U., Raes, D., Steduto, P., Fereres, E., Hsiao, T.C., 2017. AquaCrop-OS: An open source version of FAO’s crop water productivity model. *Agricultural water management* 181, 18–22.
- Hoaglin, D.C., Velleman, P.F.A., 1949. Applications, basics and computing of exploratory data analysis.
- Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., Tian, L., Zhang, X., Fan, J., Wu, W., 2016. Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. *Agricultural and forest meteorology* 216, 188–202.
- Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., Su, W., Zhang, X., Zhu, D., Wu, W., 2015. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agricultural and Forest Meteorology* 204, 106–121.

- Huang, Y., Zhu, Y., Li, W., Cao, W., Tian, Y., 2013. Assimilating remotely sensed information with the wheatgrow model based on the ensemble square root filter for improving regional wheat yield forecasts. *Plant Production Science* 16, 352–364.
- Ines, A.V., Das, N.N., Hansen, J.W., Njoku, E.G., 2013. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment* 138, 149–164.
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., Wang, J., 2018. A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy* 92, 141–152.
- Jin, X., Kumar, L., Li, Z., Xu, X., Yang, G., Wang, J., 2016. Estimation of winter wheat biomass and yield by combining the AquaCrop model and field hyperspectral data. *Remote Sensing* 8, 972.
- Jin, X., Li, Z., Feng, H., Ren, Z., Li, S., 2020. Estimation of maize yield by assimilating biomass and canopy cover derived from hyperspectral data into the AquaCrop model. *Agricultural Water Management* 227, 105846.
- Jin, Z., Azzari, G., Lobell, D.B., 2017. Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches. *Agricultural and forest meteorology* 247, 207–220.
- Kalman, R.E., 1960. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering* 82, 35–45. <https://doi.org/10.1115/1.3662552>
- Kang, Y., Özdoğan, M., 2019. Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. *Remote Sensing of Environment* 228, 144–163.
- Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. *Field Crops Research* 143, 56–64.
- Lobell, D.B., Asner, G.P., Ortiz-Monasterio, J.I., Benning, T.L., 2003. Remote sensing of regional crop production in the Yaqui Valley, Mexico: estimates and uncertainties. *Agriculture, Ecosystems & Environment* 94, 205–220.
- Lobell, D.B., Azzari, G., 2017. Satellite detection of rising maize yield heterogeneity in the US Midwest. *Environmental Research Letters* 12, 014014.
- Makowski, D., Guérif, M., Jones, J., Graham, W., 2006. Data assimilation with crop models. *Working with Dynamic Crop Models* 151–172.
- Raes, D., 2009. The ETo Calculator: Evapotranspiration from a Reference Surface, Reference Manual, Version 3.2. FAO Publication.
- Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2009. AquaCrop-The FAO crop model to simulate yield response to water. FAO Land and Water Division, FAO, Rome.
- Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2013. MODIS-based corn grain yield estimation model incorporating crop phenology information. *Remote Sensing of Environment* 131, 215–231.
- Sibley, A.M., Grassini, P., Thomas, N.E., Cassman, K.G., 2014. Testing remote sensing approaches for assessing yield variability among maize fields. *Agronomy Journal* 106, 24–32.
- Silvestro, P., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., Huang, W., Casa, R., 2017. Estimating wheat yield in China at the field and district scale from the assimilation of satellite data into the Aquacrop and simple algorithm for yield (SAFY) models. *Remote Sensing* 9, 509.

- Steduto, P., Hsiao, T.C., Fereres, E., 2007. On the conservative behavior of biomass water productivity. *Irrigation Science* 25, 189–207.
- Thorp, K., Wang, G., West, A., Moran, M., Bronson, K., White, J., Mon, J., 2012. Estimating crop biophysical properties from remote sensing data by inverting linked radiative transfer and ecophysiological models. *Remote Sensing of Environment* 124, 224–233.
- Upreti, D., Huang, W., Kong, W., Pascucci, S., Pignatti, S., Zhou, X., Ye, H., Casa, R., 2019. A Comparison of Hybrid Machine Learning Algorithms for the Retrieval of Wheat Biophysical Variables from Sentinel-2. *Remote Sensing* 11, 481.
- Upreti, D., Pignatti, S., Pascucci, S., Tolomio, M., Huang, W., Casa, R., 2020a. Bayesian Calibration of the Aquacrop-OS Model for Durum Wheat by Assimilation of Canopy Cover Retrieved from VEN μ S Satellite Data. *Remote Sensing* 12, 2666.
- Upreti, D., Pignatti, S., Pascucci, S., Tolomio, M., Li, Z., Huang, W., Casa, R., 2020b. A comparison of moment-independent and variance-based global sensitivity analysis approaches for wheat yield estimation with the Aquacrop-OS model. *Agronomy* 10, 607.
- Verrelst, J., Rivera, J.P., Moreno, J., Camps-Valls, G., 2013. Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing* 86, 157–167.
- Whelan, B.M., McBratney, A.B., Minasny, B., 2002. Vesper 1.5—spatial prediction software for precision agriculture, in: *Proceedings of the 6th International Conference on Precision Agriculture*. (Ed. PC Robert)(ASA-CSSA-SSSA: Madison, WI).
- Wu, S., Huang, J., Liu, X., Fan, J., Ma, G., Zou, J., 2011. Assimilating MODIS-LAI into crop growth model with EnKF to predict regional crop yield, in: *International Conference on Computer and Computing Technologies in Agriculture*. Springer, pp. 410–418.
- Xing, H., Li, Z., Xu, X., Feng, H., Yang, G., Chen, Z., 2017. Multi-assimilation methods based on AquaCrop model and remote sensing data. *Transactions of the Chinese Society of Agricultural Engineering* 33, 183–192.
- Yuan, W., Chen, Y., Xia, J., Dong, W., Magliulo, V., Moors, E., Olesen, J.E., Zhang, H., 2016. Estimating crop yield using a satellite-based light use efficiency model. *Ecological indicators* 60, 702–709.
- Zaks, D.P., Kucharik, C.J., 2011. Data and monitoring needs for a more ecological agriculture. *Environmental Research Letters* 6, 014017.

General Conclusions

Crop models can be a key component in addressing issues of global food security as they can be utilized to monitor and improve crop production. However, their use, particularly in isolated and rural areas, is commonly limited by the lack of sufficient and reliable field data. This data scarcity increases uncertainties in crop simulations, therefore, the outputs produced can be of limited value to agricultural advisors. However, some of this uncertainty can be mitigated by using remotely sensed data. As part of the continuing efforts to explore the potential of using remote sensing data to calibrate and validate crop models, as well as to integrate the two to improve model simulations, especially in data-limited areas, the study assessed how well a crop model assimilated with remotely sensed data performed. More specifically we focused on the sequential assimilation of remote sensing derived canopy cover time series from Sentinel-2 and VEN μ S satellite into the Aquacrop-OS model in the framework of Ensemble Kalman filter. This work also compared the efficiency of the assimilation by ingesting the in-season number of observations from only Sentinel-2, VEN μ S and by combining both of them.

This approach is feasible to address the issue of within-field variability but it also has various sources of the uncertainties that need to be considered for the improved results. These sources include model parameters, input variables, initial conditions, model outputs, remote sensing data, assimilation. We have focused on the uncertainties in the remote sensing data, model parameters, initial conditions and assimilation.

In the first part of the work, we explored the state-of-the-art hybrid methods for the retrieval of the crop biophysical variables. The advantages of these methods are that they combined both the physical laws of the radiative transfer model with the state-of-the-art machine learning regression algorithms. Different machine learning regression algorithms including linear and non-linear and the kernel-based methods were tested for the retrieval of the crop biophysical variables. The kernel-based algorithms are computationally inefficient when trained on the large sets and so, to reduce the computational complexity, we also explored active learning strategies and combined them with the kernel-based gaussian regression algorithm. Different methods perform well with different crop biophysical variables.

In the second part, we dealt with the crop model. The dynamic crop growth models are the efficient tools for yield prediction and several studies indicate that the estimation of the yield could be improved by accounting the uncertainties in the crop model parameters. For the uncertainties in the crop model parameters, we focused on the identification of the most sensitive parameters of the crop model for two different geographic conditions and multiple years of weather data. For this, we compared three different methods from different rationale and background and we found a similar set of most influential parameters to the output yield, although the ranking of the parameters varied among different methods.

The set of the parameters identified as most influential in the sensitivity analysis, were subjected to the local calibration using the Bayesian approach. We adapted the recent developments of the Markov Chains methods and found the optimal value of the parameters for our field level application. The advantage of the Bayesian methods was that it not only provided the mean estimate, it also provided the confidence intervals of the parameter values and it considers the uncertainty in the remote sensing data that was used for the calibration purpose.

Finally, with the reduced uncertainties in the model parameters and initial conditions, we assigned a random error to the Aquacrop-OS model and the uncertainties in the retrieval of Sentinel-2 and VEN μ S canopy cover using gaussian process regression were considered for the sequential based data assimilation using

Ensemble Kalman Filter approach. The approach is useful as it considers uncertainties in both the model simulations and the remote sensing data for the new state estimation calculated using Kalman gain factor. The new state is an improved state estimation with reduced error. In general, the ensemble Kalman filter based data assimilation provided the assimilated yield values close to the observed yield, however the increased use of combined satellite data did not improve the estimation. The assimilation with the VEN μ S canopy cover was clearly indicating overestimation, but Sentinel-2 and combined use was somewhat close to the observed yield and showed low RMSE. Increased use of in-season observations did not improve the results also on the other end it increased processing time. Additionally, the spatialized application of the assimilation revealed that this technique was unable to capture the spatial variability observed in yield data. On the other side, low values of the yield were not simulated correctly, estimated yield had a narrow range then the observed yield on the ground, excluding the low yield areas.

The study acknowledges the complexity of the model used, the cropping system being modelled, as well as the possible errors in the remote sensing data and the approach of the assimilation implemented. Thus, to provide more vigorous and useful information to the agricultural policymakers, the study acknowledges that the current results could be improved in multiple ways. For example, one of the limitations of this study was that the canopy cover on the ground was computed using the equation that transforms leaf area index into the canopy cover. On the comparison of the measured canopy cover with the estimated canopy cover, the underestimation of the canopy cover at low values was present. However, leaf area index raw gap fraction data was also explored for the estimation of the canopy cover, but we continued processed leaf area index data as it was better correlated to the estimated canopy cover. Therefore, we recommend that similar future studies should focus on this aspect. Another limitation of this study is that it only considered one state variable canopy cover for the purpose of the calibration and validation. The validation showed better performance with the canopy cover but not with the biomass and yield. Hence, we recommend to consider the parameters related to the biomass and yield accumulation for recalibration. This work did not consider the uncertainties in the weather data. One of the assumptions in Ensemble Kalman filter-based assimilation is that the increasing number of observations during the crop growth season increases the efficiency of the assimilation. But this was not the case in our study. The reason maybe the assimilation of very frequent observations when it is not required as the saturation period in the canopy cover lies for 50-60 days and the observation in this phase was not important rather the increase number of observations increased the computational cost. Further research can focus on the uncertainties and on the assimilation of different state variables, and reducing the number of observations in the saturation period and focus only on observations that shows a change in the canopy cover. An equal interval observation where the changes in the canopy cover can be observed for example, during the crop growing and the senescence would make more sense in assimilation and it needs further investigation.

Overall, the PhD activity provided a complete evaluation the algorithms for the retrieval of biophysical variables from remote sensing data, improved the understanding of the most influential crop model parameters for the output yield and the state-of-the-art calibration methods for the crop models. The work assimilated remote sensing derived canopy cover from different satellites and evaluated if the increased number of in season observations can improve the yield estimates in the ensemble Kalman filter.

The use of crop models is very important to better understand the local cropping systems and how they are affected by environmental conditions as well as management strategies. Currently, crop models are immensely useful and do very well in cases whereby they deal with western or commercial studies but could also benefit so much from being to handle small-holder farming systems and process. Thus, integrating these models with remote sensing data can attract and improve the use cop models by holder farmers, extension offers, and researchers in data limited area such as rural and remote areas.

Acknowledgements

I would like to thank my research supervisor, Prof. Raffaele Casa for his scientific and human support for my research activities.

I thank to Prof. Giovanni Laneve for providing me research support and stay in his lab in Sapienza University of Rome.

I also thank Dr. Stefano Pignatti and Dr. Simone Pascucci for the research supervision, guidance and my research stay in Rome.

My special thanks to Prof. Wenjiang Huang for his support with the project administration and funding and for my research stay at Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China.

I like to thank Dr. Zhenhai Li, for providing the support in the research work and for accepting the evaluation of the thesis work.

I also to thank Anna Dalla Marta for accepting the evaluation of the thesis work.

I would like to thank my colleagues Massimo Tolomio, Nada Mzid and Biyao Zhang for providing the friendly environment for completing the research in Italy and Beijing, China.

I thank to Davide Tahani and Fabiola Fontana for helping in the data collection during the crop growing season.

Lastly, I thank all other people that either directly or indirectly supported the research work.