



Bader Oulaid

**Investigating the impact of coupling process-based and
data-driven models on wheat crops in arid and semi-arid
regions**

School of Water, Energy and Environment

PhD

Academic Year: 2020 - 2024

Supervisor: Professor Ron Corstanje, Dr Alice E. Milne
Associate Supervisor: Dr Toby W. Waine, Mr Rafiq El Alami

July 2024



School of Water, Energy and Environment

PhD

Academic Year 2020 – 2024

Bader Oulaid

**Investigating the impact of coupling process-based and
data-driven models on wheat crops in arid and semi-arid
regions**

Supervisor: Professor Ron Corstanje, Dr Alice E. Milne
Associate Supervisor: Dr Toby W. Waine, Mr Rafiq El Alami

July 2024

This thesis is submitted in partial fulfilment of the requirements for
the degree of PhD

© Cranfield University 2024. All rights reserved. No part of this
publication may be reproduced without the written permission of the
copyright owner.

Academic integrity declaration

I declare that:

- the thesis submitted has been written by me alone.
- the thesis submitted has not been previously submitted to this university or any other.
- that all content, including primary and/or secondary data, is true to the best of my knowledge.
- that all quotations and references have been duly acknowledged according to the requirements of academic research.

I understand that to knowingly submit work in violation of the above statement will be considered by examiners as academic misconduct.

Abstract

Accurate prediction of wheat yields in arid and semi-arid regions is challenging due to water scarcity, varying environmental conditions, and the dynamic nature of factors influencing crop growth. This thesis aims to enhance scalable wheat yield prediction by integrating remote sensing (RS) data into process-based and data-driven models for more precise and accurate yield prediction in these regions, supporting both tactical and strategic decision-making in agriculture. AquaCrop was chosen for its robust simulation of crop yield response to water. Four interlinked research questions are addressed in this study. First, I identify key factors impacting wheat yield prediction based on sensitivity and SHAP analysis for process-based and data-driven models, respectively. Second, I compare the trade-offs between calibrating process-based models using ground-based hemispherical data and freely available remotely sensed data, highlighting the trade-offs between accuracy and practicality. Third, I evaluate the effectiveness of early-season data-driven yield prediction models across two geographic regions, emphasising the need for region-specific calibrations to maintain accuracy and quantifying accuracy loss due to model transferability. Model performance improved as the season progressed, with Support Vector Regressors achieving an RMSE of 0.23 t ha⁻¹ in the arid regions and Random Forests achieving 0.50 t ha⁻¹ and 0.46 t ha⁻¹ in semi-arid and global models. Fourth, I examine the integration of data-driven models into process-based models through data assimilation techniques, demonstrating how Bayesian assimilation and high-temporal resolution data improve yield prediction accuracy. Bayesian assimilation reduced the prediction errors, decreasing RMSE and MAPE by 25% and 76.5%, respectively, compared to no assimilation approach. This research contributes to the body of knowledge by providing a comprehensive framework for integrating remote sensing data into yield prediction models, supporting precise and timely agricultural decision-making to optimise productivity in water-limited environments.

Keywords:

Data assimilation, particle swarm optimisation, remote sensing, in-season yield prediction, AquaCrop, sensitivity analysis.

Acknowledgements

In February 2019, I faced a tough decision: accept a well-paid pre-employment internship with a multinational company in the vibrant capital of Morocco or take an unpaid internship at UM6P. Most of my friends and family thought I was crazy for turning down the financial security and excitement of Rabat for something so uncertain. Honestly, I couldn't explain it myself—something deep inside told me to go to Benguerir. Six months later, while at UM6P, I came across a PhD opportunity in the UK. It felt like fate. I applied, was endorsed, and accepted. That decision opened the door to an amazing PhD journey that I wouldn't trade for anything. A journey that has been transformative, both academically and personally. Along the way, I have been incredibly fortunate to work with inspiring supervisors and meet wonderful new colleagues who became more than friends.

To Alice, who provided constant support, not just in my research but in all life matters. Our weekend meetings, her patience with code debugging and her thoughtful treats that made the tough days manageable. To Ron, who taught me how to think like a scientist. His guidance on science communicating has been invaluable, pushing me to articulate my ideas clearly and confidently. To Toby, who shared his expertise in remote sensing, providing me with guidance and skills that were essential for my research. To Rafiq, who gave me the opportunity to pursue this PhD and believed in my potential from the beginning.

Special thanks to my colleagues at the NetZero and Resilient Farming (NZRF) group, especially Kev and Vasthi, thank you for your help, mentoring and incredible presentations on soil science. To my friends Imane, Musa, Tadesse, Sophia, Latifa and Parul, your support during tough times, especially through the pandemic, kept me going.

To my wonderful wife, thank you for leaving all the open doors in Morocco and coming to the UK to support your 'PhD Student' husband. Your support and understanding, especially during my mood swings, made all the difference. To my little daughter Leena, who decided to arrive just one month before my thesis

submission. To my parents, your unwavering belief in me have been my foundation, you are my greatest joy and motivation.

Thank you all for being part of this wonderful journey.

Table of Contents

Academic integrity declaration.....	i
Abstract	ii
Acknowledgements	iv
List of Figures.....	ix
List of Tables	xii
List of Equations.....	xiv
List of Abbreviations	xvi
1 Introduction.....	17
1.1 The challenge	17
1.2 Background to yield prediction	20
1.2.1 Process-based models.....	20
1.2.2 Data-driven models	28
1.2.3 Integrating process-based models and data-driven models through data assimilation techniques	32
1.3 Aim and objectives of the research	34
1.4 Thesis structure	38
1.5 Project context	41
1.6 References	42
2 Stepwise model parametrisation using satellite imagery and hemispherical photography: Tuning AquaCrop sensitive parameters for improved winter wheat yield predictions in semi-arid and arid regions	82
Abstract.....	82
2.1 Introduction	83
2.2 Material and methods	87
2.2.1 AquaCrop model	87
2.2.2 Study and data description	88
2.2.3 Sensitivity analysis methods	95
2.2.4 Stepwise model parameter optimisation.....	106
2.2.5 Model evaluation	108
2.3 Results.....	109
2.3.1 Sensitivity analysis	109
2.3.2 Optimisation	117
2.4 Discussion	121
2.4.1 Insights from Morris and FAST sensitivity analysis of AquaCrop parameters	121
2.4.2 Impact of the source of canopy cover on the predictive performance of AquaCrop	122
2.5 Conclusions	126
2.6 References	128

3 Towards a better understanding of rainfed wheat yield predictions in arid and semi-arid regions: How early, precise, and transferable can these models be?	168
Abstract.....	168
3.1 Introduction	169
3.2 Study area and data.....	173
3.2.1 Study area.....	173
3.2.2 Yield and associated management data	174
3.2.3 Satellite data	179
3.3 Methodology	187
3.3.1 Feature selection.....	187
3.3.2 Yield modelling techniques.....	187
3.3.3 Hyperparameter tuning.....	190
3.3.4 Model Evaluation.....	190
3.3.5 Percentage feature contribution	191
3.3.6 Transferability and generalisation	192
3.4 Results.....	192
3.4.1 Feature selection.....	192
3.4.2 Model evaluation	197
3.4.3 Feature contribution	199
3.4.4 Transferability and generalisation analysis.....	201
3.5 Discussion	203
3.5.1 Key factors contributing to yield prediction across varying climates	203
3.5.2 Model performance across growth stages and climate zone suitability	205
3.5.3 Model transferability	206
3.5.4 Model generalisation	208
3.6 Conclusion	209
3.7 References	211
4 Enhancing wheat yield prediction through data assimilation of satellite-derived canopy cover into AquaCrop: A comparative study of assimilation techniques and timings.....	251
Abstract.....	251
4.1 Introduction	252
4.2 Material and methods	257
4.2.1 Study site	257
4.2.2 Crop and management data.....	258
4.2.3 Crop growth model.....	259
4.2.4 Remote sensing data	260
4.2.5 Data assimilation framework	261

4.2.6 Model evaluation	266
4.3 Results.....	267
4.3.1 Canopy cover estimation from remote sensed data	267
4.3.2 Canopy cover assimilation	271
4.3.3 Yield predictions after assimilations	275
4.4 Discussion	278
4.4.1 Remotely sensed state variables in crop models	278
4.4.2 Effectiveness of data assimilation techniques and periods	280
4.5 Conclusions	283
4.6 References	285
5 Discussion	294
5.1 What are the key factors that significantly impact wheat yield prediction in arid and semi-arid regions and how does that change over the season?	294
5.1.1 Key findings.....	294
5.1.2 Limitations and recommendations for further research	298
5.2 What are the trade-offs associated with calibrating process-based models for yield prediction using ground-based data in comparison to remotely sensed data?.....	299
5.2.1 Key findings.....	299
5.2.2 Limitations and recommendations for further research	301
5.3 How effective are early season data-driven yield prediction models across geographic regions?.....	302
5.3.1 Key findings.....	302
5.3.2 Limitations and recommendations for further research	305
5.4 To what extent does the integration of data-driven models into process- based models through data assimilation techniques enhance the precision of wheat yield predictions?.....	306
5.4.1 Key findings.....	306
5.4.2 Limitations and recommendations for further research	308
5.5 What are the most promising approaches for wheat yield prediction in arid and semi-arid regions for strategic and tactical decision-making?	309
5.6 References	312
6 Conclusion.....	316
Appendices.....	318
Appendix A Ethical approval letter	318
Appendix B – Supplementary Material to Chapter 2	319
Appendix C – Supplementary Material to Chapter 3.....	322
Appendix D – Supplementary Material to Chapter 4.....	329
Appendix E – Supplementary Material to Chapter 5	332

List of Figures

Figure 1 Graphical representation of an agricultural process-based model.	23
Figure 2 Schematic representation of AquaCrop.....	24
Figure 3 Schematic representation of forcing methods in crop modelling. The forced state variable is extracted from remote sensing data and will replace the modelled state variable.....	26
Figure 4 Schematic representation of calibration methods in crop modelling. The calibrated state variable will replace the modelled state variable.	27
Figure 5 Graphical representation of a data-driven model implementation process.....	30
Figure 6 Schematic representation of an example of updating methods in crop modelling. The updated state variable is taking into consideration uncertainties in the remotely sensed and the modelled state variable.....	32
Figure 7 Location of the study area showing the study fields in which irrigated wheat was grown. The red parcels are from season 2003–2004 and the grey parcels are related to 2002–2003 growing season.	89
Figure 8 Conceptual framework of the sensitivity analysis and the stepwise parametrisation of AquaCrop.....	107
Figure 9 FAST analysis results for the canopy cover time series. Parcels B1 and B2, sown earlier, exhibited lower water stress due to additional rainfall. .	113
Figure 10 FAST analysis results for the aboveground biomass time series. Parcels B1 and B2, sown earlier, exhibited lower water stress due to additional rainfall.....	114
Figure 11 FAST analysis results for the yield with cut-off value (horizontal red dashed line).	116
Figure 12 Comparison of model simulated canopy cover using hemispherical photography, Landsat7 and field observed canopy cover using hemispherical photography and Landsat7.	118
Figure 13 Comparison of model simulated yield using hemispherical photography, Landsat7 and field observed yield in the calibration sites (left) and the validation sites (right).	120
Figure 14 Morocco study area with (a) the land-cover distribution and (b) the climatic zones with wheat yield distribution for investigated sites for the 2020–2021 cropping season.	174
Figure 15 Wheat yield distribution and box plots for arid and semi-arid regions (2020–2021 season).....	175

Figure 16 Distribution of applied fertilizer (NPK) and sowing dates in arid and semi-arid sites (2020–2021 season).....	178
Figure 17 Vegetation index smoothing and phenology extraction methodology	182
Figure 18 Correlation with the yield and correlation based selected features in arid, semiarid, and global datasets.	194
Figure 19 Stage-specific recursive feature elimination for yield prediction in the arid, semi-arid and global datasets.....	195
Figure 20 Venn diagrams showing the unique and shared features between arid and semi-arid climate at the four prediction stages.	196
Figure 21 Stage-specific evaluation of the global (a) and localised models (i.e., b and c).....	198
Figure 22 Percentage contribution of each feature to the yield prediction at 4 stages in arid (dark blue),semiarid (brown) and global (light blue) models based on SHAP values. Features are ranked from highest to lowest contributors.....	200
Figure 23 Transferability and generalisation evaluation metrics for MLR, SVR and RF trained to predict the yield at four different prediction stages. SA2SA and SA2A represent models trained in semi-arid and tested in semi-arid and arid regions, respectively.	202
Figure 24 Location of the study area in Morocco showing the soil texture distribution of the studies sites.....	257
Figure 25 Histogram and spatial distribution of the observed yield ($t\ ha^{-1}$) across the studied sites.....	258
Figure 26 Spatial distribution of planting and key phonological dates for the studies sites.....	259
Figure 27 Histograms of AquaCrop sampled parameters.	264
Figure 28 Analysis of the Lasso regression model for canopy prediction using Landsat 7 data.	269
Figure 29 Regression lines with 95% confidence and prediction intervals for NDVI, NDMI and SAVI transformation between Landsat7 and Sentinel 2.	270
Figure 30 Temporal distribution of satellite imagery and canopy cover distribution from Landsat7 and Sentinel2 used for data assimilation into AquaCrop.	273
Figure 31 Comparison of data assimilation techniques for simulated canopy cover using Sentinel 2 observations. ENM shows AquaCrop canopy cover ensemble with the sampled parameters. EnKF, ETKF and Bayes denote the	

Ensemble Kalman Filter, Transform Kalman Filters and Bayesian assimilations, respectively. The black line and shaded area indicate the mean and standard deviation of the assimilated canopy cover. The red dots with errors represent the sentinel2-based canopy cover with its standard deviation.	274
Figure 32 Performance evaluation of data assimilation techniques using Sentinel2, Landsat 7 and a combined dataset of Landsat 7 and Sentinel2 for yield prediction.....	276
Figure 33 Performance evaluation of data assimilation techniques using Sentinel 2 for yield prediction under different assimilation timings.....	277
Figure_Apx B.1-1 Daily meteorological data on the experimental site for the 2002–2003 and 2002–2003 growing seasons.	319
Figure_Apx C.1-1 Correlation heatmap in the global dataset.	322
Figure_Apx C.1-2 Correlation heatmap in the semi-arid climate dataset.	323
Figure_Apx C.1-3 Correlation heatmap in the arid climate dataset.	324
Figure_Apx C.3-1 Hyperparameter tuning using grid-search and cross validation H represent the set of hyperparameters, P_i is the performance of the model using hyperparameter combination H_i	326
Figure_Apx C.4-1 Stage-specific Mean Absolute Percentage Error (MAPE) of the global (a) and localised models (i.e., b and c).	327
Figure_Apx C.4-2 Percentage of change in RMSE of transferred and generalised models across the four prediction stages.	328
Figure_Apx D.2-1 In-season assimilation accuracy metrics for Landsat 7	330
Figure_Apx D.2-2 n-season assimilation accuracy metrics for the combined dataset.....	331

List of Tables

Table 1 Thesis structure and status of paper submission with links to specific research questions.	39
Table 2 Sowing and irrigation schedule for the studied fields (taken from Toumi et.al 2016). Fields A1, A2, A3 were irrigated with 30 mm per event, while fields B1, B2, B3, B4, B5, B6 received 60 mm per irrigation event.	90
Table 3 Landsat 7 band specifications.	92
Table 4 Average aboveground biomass and yield in the study area with the range of measured values given in brackets.....	94
Table 5 List of the parameters from AquaCrop that were investigated in the sensitivity analysis with published lower and upper bounds where available and associated references.....	98
Table 6 The rank order of the mean of elementary effects (μ^*) for the Morris sensitivity analysis. Increasing ranks imply reduced sensitivity with the parameter ranked one the most sensitive (See table 5 for definitions). ...	109
Table 7 Non-influential parameters that are common for all studied fields according to the Morris Analysis. Min and Max represent the minimum and maximum Morris mean of elementary effects (μ^*) in the studied parcels.	111
Table 8 Model performance indicators comparing observed and simulated canopy cover.	119
Table 9 Model performance indicators comparing observed and simulated yield.	120
Table 10 Summary statistics of wheat yield and applied fertilizer and sowing dates in 2020–2021 cropping season	177
Table 11 Sentinel-2 MSI L2A band characteristics for both S2A and S2B sensors The spectral bands used in this study are in bold.....	179
Table 12 Vegetation indices used in this study, where NIR, RED, GREEN and SWIR1 correspond to Sentinel’s B8, B4, B3 and B11, respectively.....	181
Table 13 Detailed list of features used for model development.	184
Table 14 Sentinel-2 MSI and Landsat 7 ETM+ bands used in this study.	260
Table 15 Fitted model and performance metrics for Lasso model trained and validated using Landsat 7 data.	267
Table 16 Performance metrics for vegetation indices transformation.....	268

Table_Apx B.2-1 AquaCrop main sensitive model parameters values obtained using the two studied approaches	320
Table_Apx C.2-1 Detailed list of hyperparameters used for model development. Default values ate based on the python Scikit-learn (version 0.22)	325
Table_Apx D.1-1 Full list of vegetation indices used to fit the Lasso model ...	329

List of Equations

(2-1)..... 88

(2-2)..... 88

(2-3)..... 88

(2-4)..... 92

(2-5)..... 95

(2-6)..... 95

(2-7)..... 95

(2-8)..... 96

(2-9)..... 96

(2-10)..... 106

(2-11)..... 106

(2-12)..... 108

(2-13)..... 108

(2-14)..... 108

(2-15)..... 108

(3-1)..... 180

(3-2)..... 183

(3-3)..... 188

(3-4)..... 189

(3-5)..... 190

(3-6)..... 190

(3-7)..... 190

(3-8)..... 191

(3-9)..... 191

(3-10)..... 191

(3-11)..... 192

(4-1)..... 261

(4-2).....	261
(4-3).....	261
(4-4).....	262
(4-5).....	262
(4-6).....	263
(4-7).....	263
(4-8).....	263
(4-9).....	263
(4-10).....	264
(4-11).....	264
(4-12).....	265
(4-13).....	265
(4-14).....	265
(4-15).....	265
(4-16).....	265
(4-17).....	266
(4-18).....	266
(4-19).....	266
(4-20).....	267

List of Abbreviations

CC	Canopy Cover
CGDD	Cumulative Growing Degree Days
ENKF	Ensemble Kalman Filter
ETKF	Ensemble Transform Kalman Filter
FAST	Fourier Amplitude Sensitivity Test
GEE	Google Earth Engine
GNDVI	Green Normalized Difference Vegetation Index
HP	Hemispherical Photograph
LASSO	Least Absolute Shrinkage and Selection Operator
L7	Landsat 7
MLR	Multiple Linear Regression
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
PSO	Particle Swarm Optimization
RF	Random Forests
PSSR	Pigment Specific Simple Ratio
SAVI	Soil Adjusted Vegetation Index
SVR	Support Vector Regression
SHAP	Shapley Additive Explanations
S2	Sentinel 2

1 Introduction

1.1 The challenge

Climate change is escalating pressures on agricultural production system, thereby resulting in lower yields and poorer crop health (Abdelmajid et al., 2021; N. Amiri et al., 2021). This challenges the sustainability of food production and the assurance of food security. Consequently, stakeholders face significant uncertainties and must adapt to evolving environmental conditions. To manage these challenges, advancing yield prediction methodologies has become critical (Chang et al., 2023a; Shuai & Basso, 2022). Specifically, predictive models are essential for informing decision-making processes and improving agricultural practices (Castañeda-Vera et al., 2015). For instance, policymakers use yield predictions to develop food security policies and manage agricultural subsidies (Yu et al., 2018), ensuring that resources are allocated efficiently (Linker, 2020). Moreover, agronomists and farmers rely on these predictions to recommend best practices for crop management and to optimize input use, such as water and fertilizers (Z. Jin et al., 2017; Linker, 2020). However, despite the clear benefits, there are technical limitations and barriers to the wider use of yield prediction models. These include the need for high-quality data, the requirement for model calibration to specific local conditions and the right choice of the needed yield modelling approaches.

In regions with arid and semi-arid climates, particularly in Morocco, where weather conditions are extreme, sustaining agricultural productivity becomes even more difficult (Bouras et al., 2019). Wheat, being an essential crop both locally and worldwide, is especially susceptible to the impact of these environmental pressures. Consequently, ensuring food security in Morocco heavily relies on overcoming the challenges facing wheat production. These constraints include unpredictable rainfall with recent levels around 50% of the 10-year average (Meddi & Eslamian, 2021). This has led to a heavy reliance on wheat imports, currently estimated at 7.7 million tonnes, which is a 52% increase from Morocco's 10-year import average (Belmahi et al., 2023). These imports primarily come from Russia and in fact the recent Russia-Ukraine conflict has

greatly impacted these supply chains, resulting in a substantial price increases and supply shortages (Bertassello et al., 2023; Lin et al., 2023). Thus, climatic factors combined with global crisis have aggravated the vulnerability of wheat supply in Morocco making accurate yield prediction even more important to manage local production and anticipate import needs. My research aims to tackle this challenge by developing methods to improve both the accuracy and timeliness of wheat yield prediction models for arid and semi-arid regions. By customising models to local conditions, we can improve their predictive precision and provide earlier, more reliable information to farming stakeholders. This, in turn, will enhance the productivity of wheat farming in these challenging environments. The integration of such models into agricultural practices will provide critical support for decision-making at both tactical and strategic levels.

Prediction timing and scale are key components of how effective and useful a crop yield model could be to various stakeholders (Leukel et al., 2023). Forecasts can be made at different stages of the crop growth cycle, with early season predictions being useful for tactical decisions and end-of-season predictions aiding in strategic planning. Strategic support involves long-term planning and policy formulation, such as developing national food security policies and managing agricultural subsidies. Tactical support focuses on short-term decisions and immediate actions, including real-time adjustments to crop management practices. Yield predictions can also be generated at different spatial scales, from individual fields to regional or national levels. Fine-scale predictions assist farmers in making precise management decisions, while large-scale predictions help policymakers plan and distribute resources efficiently. By tailoring yield prediction models to meet the specific needs of different stakeholders, their practical applicability and effectiveness can be significantly enhanced (Yli-Heikkilä et al., 2022). To achieve this, it is essential to understand the limitations of the different yield prediction models and determine how readily available data, such as data from satellite, can mitigate these challenges.

To develop accurate yield prediction models, leveraging diverse, high-quality data sources from both ground-based observations and remote sensing technologies is essential, as they provide detailed environmental data that are critical for this purpose as well as data for calibration (E. Cheng et al., 2022). Ground-based hemispherical photographs for instance offer information about canopy structure and light conditions, which are necessary for calibration decision support systems. Courault et al., (2021) used hemispherical photography to generate leaf area index maps to parametrise STICS model for rice yield prediction in France. While the results are accurate, they are labour and computationally intensive. In contrast, remote sensing (RS) sensors such Landsat7 and Sentinel2, deliver valuable data on vegetation indices, land cover, and surface reflectance, which help infer model parameters and variables essential for yield prediction (Ali et al., 2022a; S. Wu et al., 2021). Sentinel2's high resolution optical imagery is particularly useful for monitoring crop health and assessing crop biomass, providing essential inputs for yield predictions. Meanwhile, Landsat7's long-term historical data supports trend analysis and change detection in agricultural landscapes, enhancing the temporal aspect of yield modelling. These comprehensive data inputs enhance the model precision and reduce the need for extensive ground-based data collection. By balancing the trade-offs between data quality, effort and accuracy, my research aims to develop models that provide reliable yield prediction.

Historical yield data averaging, process-based models and data-driven models are commonly acknowledged as main tools for estimating crop yields. Still, there are limitations associated with each method (Di Paola et al., 2016). Historical data averages involve using past yield data to calculate averages for future predictions. This method, while simple, fails to account for yearly variability and lacks the adaptability to changing conditions. In contrast, process-based models provide a comprehensive understanding of crop development, but they often need time-consuming calibrations specific to each area (Roberts et al., 2017). In this regard, RS can reduce the need for extensive ground data collection by providing accurate, high-resolution environmental data, such as soil

properties, temperature, and moisture levels. This not only speeds up the calibration process but also enhances the spatial applicability of these models to different regions without extensive localised calibrations (G. Ma et al., 2013; S. Wu et al., 2021). Conversely, data-driven models are proficient in handling large datasets, but they are challenged in terms of adaptability to many scenarios and comprehensibility (Chang et al., 2023a). To overcome these issues, RS has been used to enhance model training and validation by supplying essential input features, such as spectral indices and thermal imagery, that are essential for developing robust machine learning algorithms (Ali et al., 2022b). Moreover, the large volume of data provided by RS technologies enable the exploration of complex interactions between different factors affecting crop yields early in the season, which might not be evident through traditional data sources alone (Chen & Tao, 2022). Consequently, RS is regarded as a scalable and efficient tool to enhance the capabilities of each modelling approach (Kasampalis et al., 2018).

1.2 Background to yield prediction

1.2.1 Process-based models

1.2.1.1 Overview and limitations

Process-based models (PBMs) are mathematical representations of natural systems that capture the underlying biological, physical, and chemical processes that determine a system behaviour such as crop growth, water movement, nutrient cycling, and soil erosion. PBMs are categorized based on the underlying balances they simulate. These primarily include crop physiological process models (PPMs), soil water and nutrient balance models (SWAN) and energy balance models (EBM). Physiological process models, such as Crop Environment Resource Synthesis (CERES), the Rothamsted Landscape model (RLM) and Agricultural Production Systems Simulator (APSIM), simulate metabolic processes such as photosynthesis and nutrient uptake in crops (Adnan et al., 2017; Coleman et al., 2017a; Kumar et al., 2023). Soil water and nutrient balance models such as AquaCrop and the Soil and Water Assessment Tool

(SWAT) simulate water and nutrient movements within the soil-plant system, considering processes such as evapotranspiration and root uptake (Rostami et al., 2022; Steduto et al., 2009a). Finally, energy balance models quantify the energy transfer between the atmosphere, vegetation, and soil, considering latent heat fluxes and radiation. Process-based models provide substantial advantages due to their capacity to i) clarify the complex relationships associated with natural systems (Pasquel et al., 2022), ii) investigate multiple scenarios (Kephe et al., 2021a), iii) assist informed decision-making (Beveridge et al., 2018), and iv) estimate the outcomes of various agricultural strategies (Gavasso-Rita et al., 2024). These models play a crucial role in enhancing our understanding of agricultural ecosystems by offering useful perspectives on the fundamental processes that drive crop development and ecosystem dynamics.

Yield prediction by process-based models (PBMs) originated in the mid-20th century and was led by De Wit. (1958) who highlighted the relationship between transpiration and crop yields, particularly in environments with limited water resources. Advances in computer technology and multidisciplinary collaboration has expanded the use of PBMs for estimating crop production. Methodical data gathering on climate, soil properties and crop management practices is the first step in using process-based models for yield prediction. The PBM then uses this information to replicate the biological, physical, or chemical processes—such as photosynthesis, nutrient absorption, and biomass accumulation—that control crop development. Model accuracy is improved by calibrating simulations to observed data in various climatic and management situations. Finally, the model outputs provide farmers and policymakers comprehensive information to plan sowing dates (Bassu et al., 2009), enhance crop management (Jing et al., 2021), and evaluate risks to improve agricultural productivity (Wellens et al., 2022), sustainability (Javan & Darestani, 2024), and resilience (Zampieri et al., 2020).

Process-based models (Figure 1) are important tools in agriculture and environmental management; however, they are not without limitations (Boote et al., 1996). One significant limitation is the uncertainty associated with input data

(Chapagain et al., 2022), which fundamentally shapes model outputs. The variability or lack of precision in climatic data (X. Feng et al., 2023), soil properties (Folberth et al., 2016), and crop genetics (Parent & Tardieu, 2014) can introduce significant errors into model simulations. Additionally, the resolution of data, both spatially and temporally, often fails to capture the inherent variability of natural systems, which can limit the model's ability to deliver precise predictions at local scales or over prolonged periods, thereby affecting strategic decision-making in agricultural management and environmental conservation (Al-Shammari et al., 2021; Constantin et al., 2019; J. Zhang et al., 2023). The complexity inherent in PBMs presents further challenges. These models require the calibration of numerous parameters that govern multiple simulated processes, making the parameterization a demanding and often imprecise task, especially when model data or parameters are not directly observable (Silva & Giller, 2020; Wallach et al., 2021). Additionally, assumptions required to develop these models may not universally apply under all conditions (Hoogenboom & White, 2003), potentially leading to divergences between predicted and actual system outcomes. PBMs also exhibit sensitivity to initial conditions, where small changes in the starting state of the system lead to significantly different outcomes over time (Manschadi et al., 2021; Müller et al., 2024). These challenges necessitate ongoing calibration and validation of PBMs to maintain their accuracy and applicability in practical scenarios.

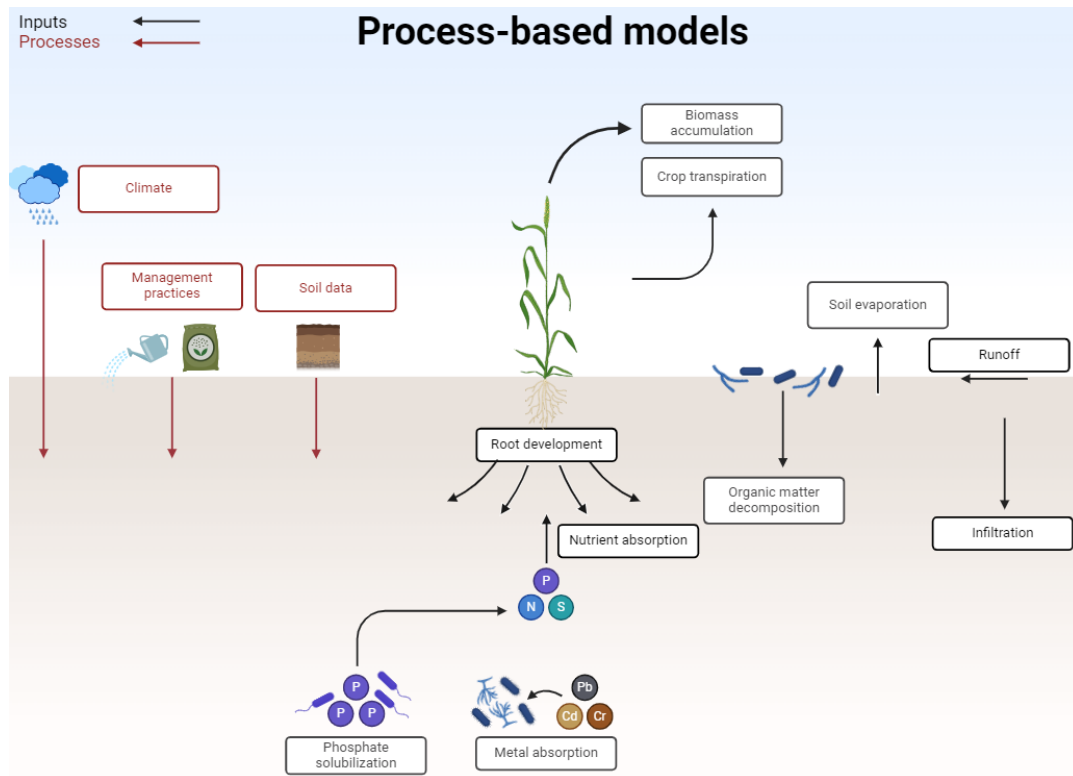


Figure 1 Graphical representation of an agricultural process-based model.

Given the challenges associated with process-based models, it is important to opt for a model that balances practicality with reliability. AquaCrop, developed by the Food and Agriculture Organization (FAO), is designed specifically to simulate crop response to water availability while prioritizing simplicity and effectiveness. The model focuses on key processes essential to plant growth, including soil moisture dynamics, crop expansion, crop transpiration and biomass production. A unique feature of AquaCrop is its partitioning of evapotranspiration into soil evaporation and crop transpiration, which provides a more precise assessment of water use efficiency. This allows the model to represent different crop growth stages (from emergence to flowering and maturity) while considering the effects of water stress on yield formation. AquaCrop operates by integrating different environmental inputs, such as daily weather data (i.e., temperature, rainfall, solar radiation) and soil characteristics such as texture, hydraulic properties, and water retention capacity. These inputs

drive the simulation of processes related to water balance, such as infiltration, runoff and root zone water uptake. The model further incorporates crop-specific parameters, including canopy growth rate, rooting depth, and harvest index to estimate growth stages and potential yield. Management practices, such as irrigation schedules can also be specified to assess their impact on crop performance. By integrating environmental inputs, soil properties and crop-specific parameters, AquaCrop simulates the processes from water balance and canopy development to transpiration and biomass accumulation, providing a detailed representation of yield formation under varying water availability and management conditions (Figure 2).

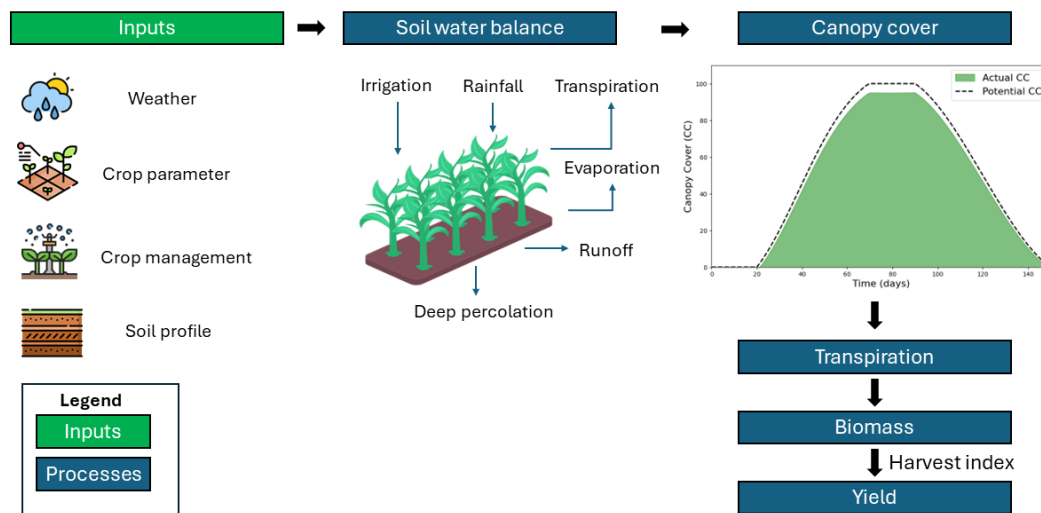


Figure 2 Schematic representation of AquaCrop

1.2.1.2 Parametrisation of process-based models through remote sensing

Remote sensing has transformed the traditional approach to parameterizing process-based models (PBMs) by providing an efficient way to collect data across large geographic areas. This capability potentially reduces the dependency on expensive and time-consuming fieldwork, enabling dynamic adjustments and fine-tuning of model parameters to adapt to changing environmental conditions.

The strategies for incorporating remote sensing data into PBMs can be categorised into forcing and calibration methods.

Forcing methods involve directly substituting simulation data with RS data, thus supplying the PBM with observed values rather than modelled estimates (Figure 3). This approach is particularly beneficial in decision support systems where factors such as soil moisture (J. Zeng et al., 2021), climate (Ruane et al., 2015) and vegetation cover (Casa et al., 2012a) are crucial for accurate simulations. For instance, Bai et al. (2019) enhanced the accuracy of yield predictions for jujube orchards by forcing leaf area index (LAI) data from Landsat 8 into the WOFOST model. This application resulted in improved prediction accuracy compared to unforced simulations. In addition, Morel et al. (2014) significantly enhanced the efficiency of MOSICAS sugarcane model by forcing fraction intercepted photosynthetically active radiation (fIPAR) data derived from SPOT satellite images. Although this method improves the accuracy of crop models based on real-world data, it is not without its limitations. Key challenges include reliance on the quality and accessibility of remote sensing data, both of which are susceptible to variation due to elements such as cloud cover and sensor defects (Breunig et al., 2023; Dubovik et al., 2021). Additionally, simulation errors can occur if the spatial and temporal dimensions of the data are misaligned with the requirements of the model (Wu & Li, 2009). Thus, the overreliance of this method on precise and frequent RS data inhibits model's ability to predict outcomes under unobserved conditions, potentially reducing its robustness and adaptability in varying scenarios.

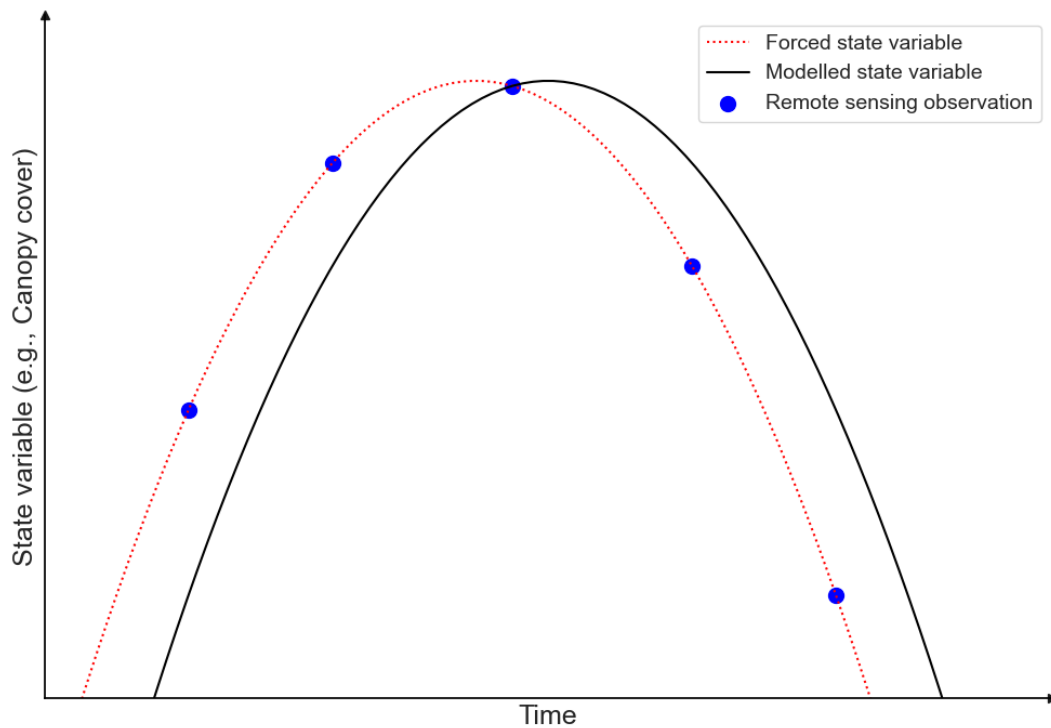


Figure 3 Schematic representation of forcing methods in crop modelling. The forced state variable is extracted from remote sensing data and will replace the modelled state variable.

To avoid the overdependence on RS observations, the calibration method adjusts the parameters of the model to minimize the differences between remote sensing and model simulated data (Figure 4). This method uses advanced algorithms to minimise the objective function, including simulated annealing (Guo et al., 2021; J. Wang et al., 2013), differential evolution (Bandara et al., 2021; Martínez-Ruiz et al., 2021), and particle swarm optimization (Z. Li et al., 2015; Oulaid et al., 2024). However, these methods can be computationally intensive, requiring significant time and expertise. The process also introduces subjectivity, as decisions must be made on which parameters to adjust. Thus, while it is a valuable tool to parametrise crop models, It demands careful implementation to ensure its effectiveness.

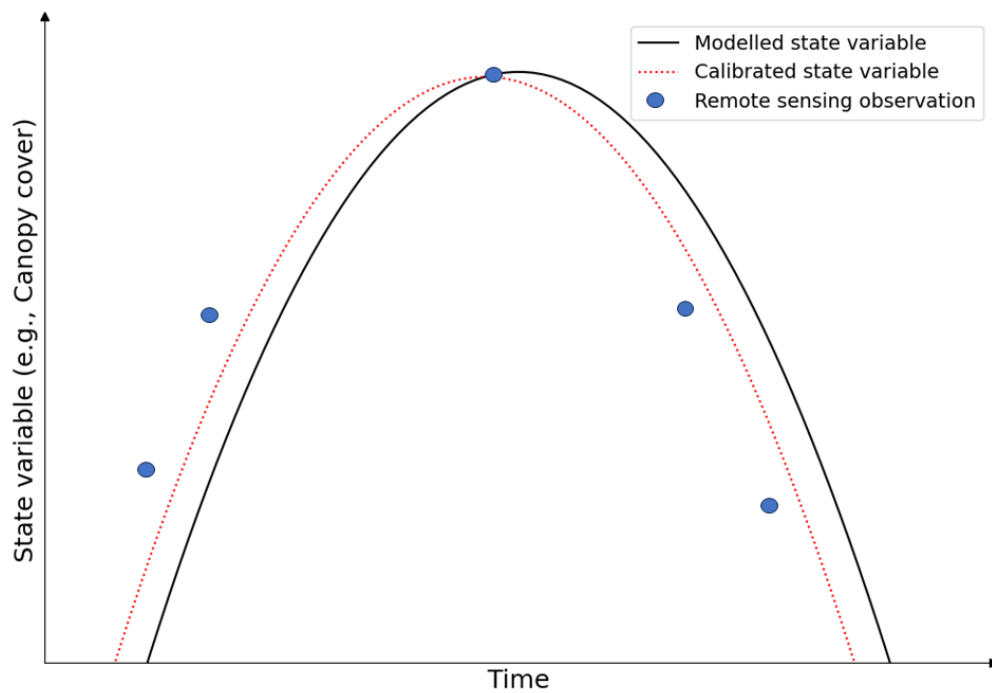


Figure 4 Schematic representation of calibration methods in crop modelling. The calibrated state variable will replace the modelled state variable.

The effectiveness of remote sensing data in process-based models parametrisation heavily depends on the resolution and source of the data (Dlamini et al., 2023). The resolution, both spatial and temporal, determines the granularity and frequency of observations available for model parametrisation. High-resolution data allows for detailed and precise assessments of environmental conditions and emerging patterns, enhancing model accuracy in capturing complex landscapes and rapid crop phenological changes (Diao et al., 2024; Rivas-Fandiño et al., 2023). Conversely, low-resolution data could lead to oversimplifications that misrepresent the dynamics of the area being modelled (Han et al., 2024).

The source of RS data is equally crucial in determining the effectiveness of process-based model parametrisation (Niu et al., 2021). Multispectral data from satellites such as MODIS and Landsat provide essential details into crop health (Doraiswamy et al., 2004), soil moisture (ZHANG et al., 2014), and land cover changes (Ait El Haj et al., 2023) due to its balanced spatial resolution and

coverage frequency. Hyperspectral sensors capture hundreds of contiguous spectral bands, allowing for precise identification of crop types and nutrient status, which are essential for detailed crop modelling (U. Patel et al., 2023; Peng et al., 2020). SAR sensors, unaffected by weather conditions and daylight, offer high-resolution imagery for soil moisture monitoring and crop structure assessment, with platforms such as Sentinel-1 and RADARSAT providing critical data in cloud-covered regions (Khabbazan et al., 2019; Wood et al., 2002).

Integrating data from multiple sources might help mitigate the limitations of relying on a single source, particularly when advanced data fusion techniques are used. This approach aims to create a more robust and comprehensive dataset, potentially reducing biases and improving model reliability and accuracy (Barbedo, 2022). However, it also increases complexity in data processing and analysis, necessitating sophisticated algorithms and significant computational resources. Consequently, while the choice of RS data source and resolution can profoundly impact PBM accuracy and applicability, it remains uncertain whether the integration of different RS-based data sources consistently enhances model performance.

1.2.2 Data-driven models

1.2.2.1 Overview and limitations

Data-driven models (DDMs) are derived using machine learning or statistical analysis to detect patterns, relationships, or trends in large datasets (Figure 5). Unlike PBMs that rely on simulating physical, biological, or chemical processes, DDMs use empirical data to generate their predictions (Lee et al., 2023). These models are especially useful in areas with abundant data. DDMs can be categorized into several subtypes depending on their learning strategies. Supervised learning, for example, involves training models on labelled data using algorithms like linear regression and neural networks to learn input-output mappings (Dhanapal et al., 2021; K. Patel & Patel, 2021). Conversely, unsupervised learning detects patterns within unlabelled data based on

techniques like clustering and principal component analysis (Xavier et al., 2017). Semi-supervised learning combines labelled and unlabelled data to enhance learning accuracy, particularly advantageous when labelled data is sparse (Bhullar et al., 2023). Finally, reinforcement learning models, which are often used in robotics and game theory, learn to make decisions through feedback from their environmental interactions (R. Tao et al., 2023).

The use of data-driven models for yield prediction emerged prominently with the advancements in digital computing and data analytics capabilities. It begins with systematically gathering important data, such as climatic factors, soil qualities, and information on crop management strategies. The large dataset allows for the use of statistical techniques and machine learning algorithms such as decision trees, random forests, and deep learning networks in DDMs to accurately analyse and forecast crop production (Elbasi et al., 2023a; Jeong et al., 2016; Oikonomidis et al., 2023). The precision of DDMs continually improves as they learn from new data inputs and validate their predictions against actual observed outcomes. DDMs have been used to assist farmers and agricultural managers in making well-informed decisions about optimal planting times (Khaki et al., 2020), efficient resource allocation (Lorenz, 2013), strategic crop rotation (Puntel et al., 2016), and timely market management (Arunraja et al., 2021), enhancing both sustainability and economic returns.

Despite their benefits in agricultural decision-making, data-driven models (DDMs) have several limitations (Rozenstein et al., 2024). A primary issue is their reliance on high-quality, comprehensive datasets that cover a range of variables and extended historical data as incomplete, noisy, or biased data can result in inaccurate predictions and poor decision-making. Furthermore, while DDMs can detect patterns and correlations, they often do not explain the causal relationships behind these patterns, limiting their usefulness in understanding complex agricultural systems. Overfitting is another problem, where models become too specific to the training data and fail to generalize to new unseen data (Nansen et al., 2013). Implementing DDMs also demands substantial computational resources and expertise. Additionally, the interpretability of DDMs is a concern,

as many machine learning algorithms, particularly deep learning networks, function as “black boxes” with limited transparency about their decision-making processes (Hu et al., 2023). This lack of clarity could limit the adoption of these models by stakeholders who need clear and understandable decision-making tools. Finally, integrating DDMs into current agricultural practices and systems presents practical challenges, including the necessity for continuous updates and validation to maintain model accuracy and relevance when faced with evolving environmental conditions and agricultural practices.

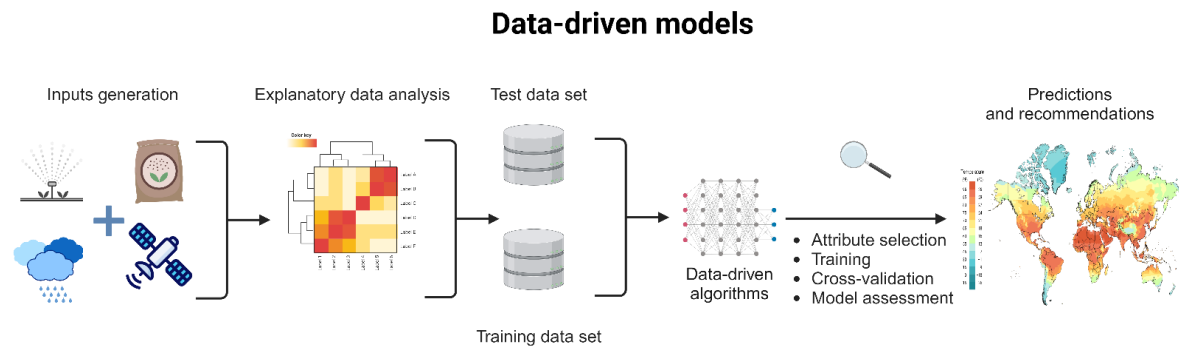


Figure 5 Graphical representation of a data-driven model implementation process.

1.2.2.2 The role of remote sensing in assisting data-driven models

Remote sensing can significantly enhance the capabilities of data-driven models (DDMs) by providing critical phenology metrics and capturing the spatial and temporal variability of key factors that influence crop yields (Zeng et al., 2020). These contributions are important for improving the accuracy and reliability of DDMs in agricultural applications. RS plays a crucial role in generating phenology metrics, which are vital for understanding the timing of life cycle events in plants. These metrics can be derived using three main methods:

(i) Threshold-based methods, which identify key phenological stages by setting specific thresholds for remote sensing data, such as vegetation indices. For

example, thresholds for Normalized Difference Vegetation Index (NDVI) values can be used to determine the onset of greening or senescence in crops (Li & Chen, 2018; Tang et al., 2015).

(ii) Amplitude-based methods which rely on the amplitude of seasonal changes in remote sensing signals to identify phenological events. This approach is particularly useful for detecting peak growth periods and assessing overall vegetation health (Gerard et al., 2020).

(iii) Finally, derivative-based methods which apply mathematical derivatives to remote sensing time series data to detect changes in growth rates that indicate phenological stages. Consequently, critical growth phases such as flowering and maturation can be derived by analysing the first or second derivatives of vegetation indices (Xin et al., 2020; Ye et al., 2023).

In addition to phenology metrics, remote sensing can capture the spatial and temporal variability of key factors influencing yield. Soil moisture is a critical factor that can be monitored using remote sensing to understand water availability and stress conditions in crops. Accurate soil moisture data helps in optimizing irrigation schedules and ensuring efficient water use (Mohanty et al., 2017; Wang & Qu, 2009). Vegetation health is another important variable monitored through indices like NDVI, which assesses plant health (Petersen, 2018), detects diseases (H. Zhao et al., 2020), nutrient deficiencies (Marzukhi et al., 2016) and therefore enables early intervention and targeted management practices. The comprehensive and high-resolution data provided by remote sensing also allows DDMS to account for the variability and complexity of agricultural systems, leading to more reliable predictions and actionable recommendations for farmers and agricultural managers. This integration supports informed decision-making, enhances resource management, and ultimately contributes to sustainable and productive agricultural practices.

1.2.3 Integrating process-based models and data-driven models through data assimilation techniques

Data assimilation methods advance the integration of real-time remote sensing data with process-based models by continuously refining model outputs. Unlike the forcing method, which directly substitutes simulation data with RS observations, and the calibration method, which adjusts initial parameters to reduce discrepancies between observed and simulated data, updating methods dynamically adjust model parameters as new RS data becomes available (Figure 6). This approach leverages algorithms such as Kalman filter variations (J. Huang et al., 2016; Koo et al., 2003; S. Wu et al., 2011) and Bayesian methods (Wu et al., 2022; T. Zhang et al., 2019) to align model outputs with observed conditions. This not only enhances the model's precision and reliability under varied environmental conditions but also considers uncertainties in both model parameters and RS data.

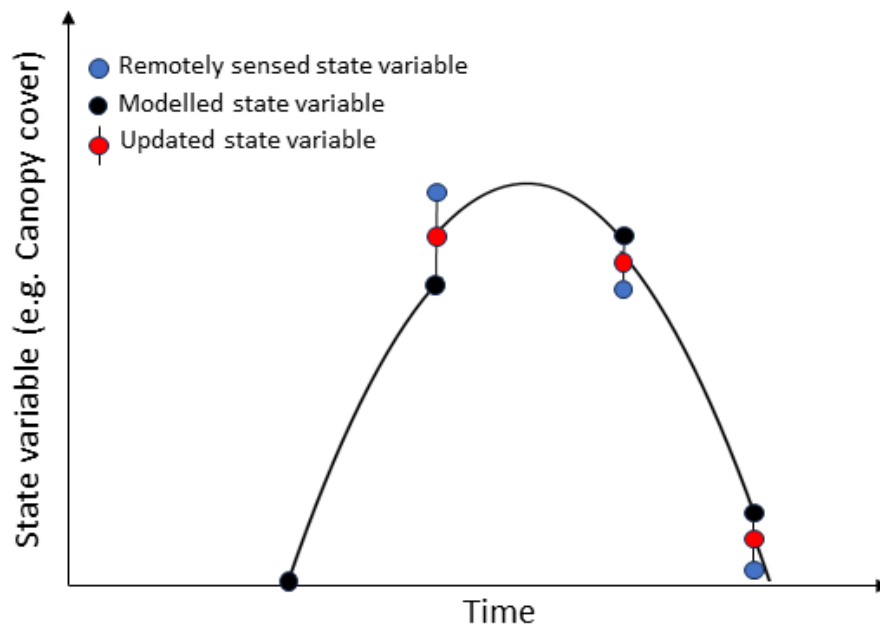


Figure 6 Schematic representation of an example of updating methods in crop modelling. The updated state variable is taking into consideration uncertainties in the remotely sensed and the modelled state variable.

Integrating process-based models and data-driven models through data assimilation techniques combines the strengths of both approaches to improve agricultural decision-making. PBMs offer a detailed understanding of the biological, physical, and chemical processes that govern system behaviour, while DDMS provide robust predictions derived from large datasets using machine learning and statistical methods. In this integration, state variables are initially predicted using data-driven models, leveraging their capability to make use of large and complex datasets. These predicted state variables are then assimilated into PBMs. Data assimilation merges observational data with model predictions, ensuring that models remain accurate and reflective of real-world conditions. This hybrid approach can enhance model reliability, providing more comprehensive and precise recommendations for agricultural management (Maestrini et al., 2022a).

Data assimilation techniques can be broadly categorized into sequential-based and variational-based methods. Sequential-based methods, such as the Kalman Filter and its variants, update model predictions step by step as new data becomes available. These methods are particularly effective in incorporating real-time data, continuously adjusting model outputs to reflect the latest observations (Bertino et al., 2003; R. Li et al., 2010). The Kalman Filter, for example, uses a recursive algorithm to process incoming data, providing estimates of unknown variables by assigning case-specific weight to new observations, ensuring that the integrated model remains accurate and responsive to changing conditions. In agricultural applications, state variables such as canopy cover and soil moisture are updated as new data is received, improving the model's accuracy and robustness. Additionally, these methods take into account the standard deviations of both model predictions and observations, enhancing the reliability of the forecasts by quantifying uncertainties. This capability makes the sequential assimilation methods suitable for dynamic agricultural environments where real-time decision-making is needed. Variational-based methods in contrast, involve optimizing the entire model state by minimizing the differences between model predictions and observations over a specified period. These methods use

sophisticated optimization techniques to find the best fit for the observed data within the constraints of the model (Fischer et al., 2005). One prominent approach is Four-Dimensional Variational Data Assimilation (4D-Var), which adjusts the model's initial conditions to minimize discrepancies with observations over a time window. Variational methods are particularly efficient for incorporating large datasets and are widely used in weather forecasting. In the context of crop yield prediction, variational methods have proven to be highly effective. For instance, the 4D-Var method has been applied to integrate satellite-derived features into crop yield models, significantly enhancing their predictive performance (Huang et al., 2020). Sequential-based and variational-based methods are both essential for improving model predictions, yet they operate differently and have distinct strengths. While sequential methods are better for real-time applications, variational methods provide comprehensive optimizations, making both approaches complementary in supporting effective agricultural management and decision-making.

1.3 Aim and objectives of the research

My overall thesis aim is to advance scalable wheat yield prediction by exploring how remote sensing (RS) data can support the integration of process-based and data-driven models for precise and accurate yield prediction in arid and semi-arid regions. AquaCrop, specifically, has been chosen for this research due to its robust simulation of crop yield response to water, balancing simplicity, and robustness in process-based models. Despite significant research on data-driven models for yield prediction, existing studies have primarily focused on late yield prediction at the end of the season when crop management is minimal. Moreover, these studies have not considered a large number of sites, limiting their applicability and scalability. To achieve the overall aim, my thesis is structured around four interlinked research questions.

Research Question 1: What are the key factors that significantly impact wheat yield prediction in arid and semi-arid regions and how does that change over the season?

Previous studies have successfully calibrated AquaCrop to simulate yield in arid and semi-arid regions (M. Cheng et al., 2022; Mubvuma et al., 2021a). However, these calibrations have been done either using the full set of parameters or using local sensitivity analysis methods that miss interactions between parameters. This has led to complex model behaviour and a lack of understanding of the factors influencing yield predictions. To address these gaps, I used a stepwise methodology that first screens the factors with marginal effects and then quantifies the main effect-sensitive index of factors influencing the yield (Chapter 2). This approach allowed for a more comprehensive exploration of parameter interactions and their impacts on yield prediction accuracy. Additionally, I investigated the temporal sensitivity of AquaCrop outputs prior to yield formation, focusing on variables such as canopy cover and above ground biomass. This temporal analysis provided insights into how model outputs vary throughout the growing season and under different planting dates and irrigation managements. Unlike process-based models that use mechanistic ways to predict yield, key factors that influence yield prediction in data-driven models are elusive due to the black box nature of these models. Prior studies lacked the in-depth analysis of feature contributions to models' predictions across different prediction stages (Elbasi et al., 2023b; Khaki & Wang, 2019; Sharma et al., 2023). I investigated the climate specific factors influencing yield prediction for data-driven models in two climatic zones of Morocco (Chapter 3) and assessed their impact on models' accuracy and transferability. By addressing this research question, I increase our understanding of how factors (parameters and features) influencing yield prediction differ between process-based and data-driven models under varying climates and crop management conditions.

Research Question 2: What are the trade-offs associated with calibrating process-based models for yield prediction using ground-based data in comparison to remotely sensed data?

Previous research has predominantly relied on ground-based measurements and manual trial-and error methods for AquaCrop parametrization in semi-arid and arid regions (Amiri et al., 2024; Mubvuma et al., 2021b; Toumi et al., 2016). This traditional approach while highly accurate and specific to local conditions, is labour-intensive, time-consuming, and often limited in spatial coverage. However, the advancements in remote sensing technologies present new opportunities for AquaCrop parametrization. Recognizing the limitations of traditional methods, while the use of hemispherical photography and satellite imagery represents a significant shift from localised, manual techniques to more extensive, automated approaches, it introduces a set of trade-offs. Hemispherical photography offers high-resolution, direct measurements of canopy cover, providing detailed insights into crop growth dynamics. Yet, its practicality is limited by the need for frequent field visits. On the other hand, satellite imagery offers broad spatial coverage and temporal consistency, facilitating large-scale analysis and reducing the reliance on ground measurements. However, it suffers from spectral saturation issues, particularly in regions with dense vegetation cover, leading to inaccuracies in canopy cover estimation and subsequent yield predictions (Joshi et al., 2017; P. Zhao et al., 2016).

To address these considerations, I assessed the practical implications and trade-offs associated with each data source (Chapter 2). By systematically comparing the accuracy, reliability, and resource requirements for these parametrisation approaches, I aimed to assess the most suitable approach for AquaCrop parametrisation based on earth observation data. Furthermore, I investigated how the choice of data source influences model performance, particularly in terms of canopy cover and yield simulation accuracy. By highlighting these factors, my research contributes to enhancing the effectiveness and applicability of AquaCrop calibrations in regions facing water scarcity and climate variability.

Research Question 3: How effective are early season data-driven yield prediction models across geographic regions?

The exponential increase in data availability from varying sources, including satellites, sensors, and drones, led to an increase in the research on data-driven models for yield prediction (Darra et al., 2023). Most of these have focused on yield prediction at the end of the season, when crop management practices are minimal. Because these models are driven by training data, their efficiency is diminished when used outside areas of their implementation. However, there has been limited research focusing on understanding the errors related to early prediction and model transferability.

In Chapter 3, I explored the trade-offs between model's accuracy and earliness by implementing climate-specific yield prediction models that forecast yield at various growth stages. To be useful, the model predictions need to be made at a time appropriate for the decision of interest, allowing timely agricultural management practices. Additionally, I quantified the errors associated with the transferability of these models across different climatic zones. This analysis also included a comparison between site-specific models and generalized models to determine their relative effectiveness in diverse environments. By addressing this research question, I increased our understanding of how climate-specific early yield prediction models operate across different growth stages and climatic zones, demonstrating that for models to be effective, they need to maintain high accuracy across different environments. This will provide valuable perspectives regarding the optimization of agricultural management practices and resource allocation to mitigate risks associated with climate fluctuations and water shortage.

Research Question 4: To what extent does the integration of data-driven models into process-based models through data assimilation techniques enhance the precision of wheat yield predictions?

Process-based models like AquaCrop have been valuable tools for yield prediction, but they often rely on simplifications, assumptions and site-specific calibrations that could limit their scalability. As discussed in research question 2, remote sensed data offers a means to calibrate models like AquaCrop but given the timeliness with which these data are available there is opportunity to use them to update within season predictions through data assimilation methods. By leveraging data-driven models to predict key AquaCrop responses based on satellite imagery data and integrating them into AquaCrop through methods based on Kalman Filtering variations and Bayesian algorithms, I aim to advance our understanding of how data assimilation techniques can enhance the precision of regional yield predictions in AquaCrop, while accounting for the inherent errors in both the model and the remote sensing data. More specifically, in Chapter 4, I incorporated machine learning-predicted canopy cover into AquaCrop, evaluating the impact of different assimilation methods, sensor types and assimilation timings on the accuracy of wheat yield predictions. Ultimately, addressing this research question will contribute to the development of more reliable tools for agricultural planning and management in water-limited environments.

1.4 Thesis structure

This thesis is organized into six chapters, each specifically designed to address the research questions which are related to the thesis overall aim (Table 1). Chapter 1 establishes the foundation of the thesis, outlining the critical background and rationale for the research. It sets the context by emphasizing the importance of yield prediction in agricultural management and decision-making processes and the role that remote sensing plays in coupling different crop modelling approaches. Chapter 2 presents the results of a study on the trade-offs associated to different sources of remote sensing data used to assist automatic

calibrations. In Chapter 3, the focus shifts to assessing the accuracy, early prediction capabilities, and transferability of yield prediction models across different climatic conditions. This analysis is pivotal for understanding how these models perform under varied environmental stresses, aligning with the first and third objectives of the thesis. Chapter 4 explores the integration of data-driven models with process-based models using data assimilation techniques. Chapter 5, synthesizes the research findings from the preceding chapters, discusses their limitations, and evaluates the broader implications of the research in the field of agricultural modelling. Finally, Chapter 6 concludes the thesis by summarizing the key findings and reflecting on the different contributions and practical applications of the research.

Table 1 Thesis structure and status of paper submission with links to specific research questions.

Chapter	Paper	Linked to research question	Title	Journal and status
1	-	-	Introduction	-
2	1	1 and 2	Stepwise model parametrisation using satellite imagery and hemispherical photography: Tuning AquaCrop sensitive parameters for improved winter wheat yield	Field Crops Research, Volume 309, 1 April 2024, 109327 Received 18 May 2023 Revised 21 February 2024

			predictions in semi-arid regions	Accepted 27 February 2024 Available online 8 March 2024
3	2	1 and 3	Towards a better understanding of rainfed wheat yield predictions in arid and semi-arid regions: How early, precise, and transferable can these models be?	Agriculture and Forest Meteorology Submitted 10 June 2024
4	3	4	Enhancing wheat yield prediction through data assimilation of satellite-derived canopy cover into AquaCrop: A Comparative study of assimilation techniques and timings	Agricultural systems With Co-authors
5	-	-	Discussions	-
6	-	-	Conclusions	-

1.5 Project context

This work is part of the Dynamic Landscapes project, funded by OCP which aims to develop a GIG-based modelling framework (GIS-MF) to integrate process-based, data-driven, and agent-based models to assess agricultural systems and their impact on dynamic landscapes in Morocco. My PhD constitutes Work Package 4, which focuses on investigating the impact of coupling process-based and data-driven models on wheat crops. Specifically, my research involves assisting crop models with remotely sensed data and developing early prediction models for tactical and strategic decisions. To disseminate my findings, I have presented posters and delivered talks at different academic conferences and symposia in Morocco, Austria, and the UK. Workshops with different agricultural stakeholders were organised to validate the models and to explore ways to enhance their practical application. These engagements have been essential in ensuring the models' relevance and accuracy, while also facilitating knowledge transfer.

1.6 References

- A, A., E, H., A, A., & S, D. P. (2021). E-Farm Management and Yield Prediction. *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 1–6. <https://doi.org/10.1109/ICIRCA51532.2021.9544759>
- Abdelmajid, S., Mukhtar, A., Baig, M. B., & Reed, M. R. (2021). Climate change, agricultural policy and food security in Morocco. In *Emerging Challenges to Food Production and Security in Asia, Middle East, and Africa: Climate Risks and Resource Scarcity* (pp. 171–196). Springer International Publishing. https://doi.org/10.1007/978-3-030-72987-5_7
- Abi Saab, M. T., el Alam, R., Jomaa, I., Skaf, S., Fahed, S., Albrizio, R., & Todorovic, M. (2021). Coupling remote sensing data and aquacrop model for simulation of winter wheat growth under rainfed and irrigated conditions in a mediterranean environment. *Agronomy*, 11(11). <https://doi.org/10.3390/agronomy11112265>
- Adnan, A. A., Jibrin, J. M., Kamara, A. Y., Abdulrahman, B. L., Shaibu, A. S., & Garba, I. I. (2017). CERES–Maize model for determining the optimum planting dates of early maturing maize varieties in Northern Nigeria. *Frontiers in Plant Science*, 8. <https://doi.org/10.3389/fpls.2017.01118>
- Agapiou, A., Hadjimitsis, D. G., Papoutsas, C., Alexakis, D. D., & Papadavid, G. (2011). The Importance of accounting for atmospheric effects in the application of NDVI and interpretation of satellite imagery supporting archaeological research: The case studies of Palaepaphos and Nea Paphos sites in Cyprus. *Remote Sensing*, 3(12), 2605–2629. <https://doi.org/10.3390/rs3122605>
- Ait El Haj, F., Ouadif, L., & Akhssas, A. (2023). Monitoring Land Use and Land Cover Changes Using Remote Sensing Techniques and the Precipitation-Vegetation Indexes in Morocco. *Ecological Engineering and Environmental Technology*, 24(1), 272–286. <https://doi.org/10.12912/27197050/154937>

- Ali, A. M., Abouelghar, M., Belal, A. A., Saleh, N., Yones, M., Selim, A. I., Amin, M. E. S., Elwesemy, A., Kucher, D. E., Maginan, S., & Savin, I. (2022a). Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). *The Egyptian Journal of Remote Sensing and Space Science*, 25(3), 711–716. <https://doi.org/https://doi.org/10.1016/j.ejrs.2022.04.006>
- Ali, A. M., Abouelghar, M., Belal, A. A., Saleh, N., Yones, M., Selim, A. I., Amin, M. E. S., Elwesemy, A., Kucher, D. E., Maginan, S., & Savin, I. (2022b). Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). *The Egyptian Journal of Remote Sensing and Space Science*, 25(3), 711–716. <https://doi.org/https://doi.org/10.1016/j.ejrs.2022.04.006>
- Al-Shammari, D., Whelan, B. M., Wang, C., Bramley, R. G. V., Fajardo, M., & Bishop, T. F. A. (2021). Impact of spatial resolution on the quality of crop yield predictions for site-specific crop management. *Agricultural and Forest Meteorology*, 310. <https://doi.org/10.1016/j.agrformet.2021.108622>
- Amiri, E., Irmak, S., & Tari, D. B. (2024). Comparison of the AquaCrop and CERES-Maize models for simulating maize phenology, grain yield, evapotranspiration and water productivity under different irrigation and nitrogen levels. *Irrigation and Drainage*. <https://doi.org/10.1002/ird.2917>
- Amiri, N., Lahlali, R., Amiri, S., Jarroudi, M. E. L., Khebiza, M. Y., & Messouli, M. (2021). Development of an integrated model to assess the impact of agricultural practices and land use on agricultural production in morocco under climate stress over the next twenty years. *Sustainability (Switzerland)*, 13(21). <https://doi.org/10.3390/su132111943>
- Ashourloo, D., Nematollahi, H., Huete, A., Aghighi, H., Azadbakht, M., Shahrabi, H. S., & Goodarzashti, S. (2022). A new phenology-based method for mapping wheat and barley using time-series of Sentinel-2 images. *Remote Sensing of Environment*, 280, 113206. <https://doi.org/https://doi.org/10.1016/j.rse.2022.113206>

- Bai, T., Zhang, N., Mercatoris, B., & Chen, Y. (2019). Improving jujube fruit tree yield estimation at the field scale by assimilating a single Landsat remotely-sensed LAI into the WOFOST model. *Remote Sensing*, 11(9). <https://doi.org/10.3390/rs11091119>
- Bandara, W. B. M. A. C., Sakai, K., Nakandakari, T., Kapetch, P., Anan, M., Nakamura, S., Setouchi, H., & Rathnappriya, R. H. K. (2021). Global optimization of cultivar trait parameters in the simulation of sugarcane phenology using gaussian process emulation. *Agronomy*, 11(7). <https://doi.org/10.3390/agronomy11071379>
- Bandaru, V., Yaramasu, R., PNVR, K., He, J., Fernando, S., Sahajpal, R., Wardlow, B. D., Suyker, A., & Justice, C. (2020). PhenoCrop: An integrated satellite-based framework to estimate physiological growth stages of corn and soybeans. *International Journal of Applied Earth Observation and Geoinformation*, 92. <https://doi.org/10.1016/j.jag.2020.102188>
- Barbedo, J. G. A. (2022). Data Fusion in Agriculture: Resolving Ambiguities and Closing Data Gaps. In *Sensors* (Vol. 22, Issue 6). MDPI. <https://doi.org/10.3390/s22062285>
- Bassu, S., Asseng, S., Motzo, R., & Giunta, F. (2009). Optimising sowing date of durum wheat in a variable Mediterranean environment. *Field Crops Research*, 111(1–2), 109–118. <https://doi.org/10.1016/j.fcr.2008.11.002>
- Belmahi, M., Hanchane, M., Mahjoub, A., Najjari, F., Khayati, A., & Kessabi, R. (2023). Sustainability assessment of the main cereals market in Morocco: Evaluating production and import. *European Journal of Sustainable Development*. <https://api.semanticscholar.org/CorpusID:259255834>
- Bertassello, L., Winters, P., & Müller, M. F. (2023). Access to global wheat reserves determines country-level vulnerability to conflict-induced Ukrainian wheat supply disruption. *Nature Food*, 4(8), 673–676. <https://doi.org/10.1038/s43016-023-00806-w>

- Bertino, L., Evensen, G., & Wackernagel, H. (2003). Sequential Data Assimilation Techniques in Oceanography. *International Statistical Review*, 71(2), 223–241. <https://doi.org/https://doi.org/10.1111/j.1751-5823.2003.tb00194.x>
- Beveridge, L., Whitfield, S., & Challinor, A. (2018). Crop modelling: towards locally relevant and climate-informed adaptation. *Climatic Change*, 147(3–4), 475–489. <https://doi.org/10.1007/s10584-018-2160-z>
- Bhullar, A., Nadeem, K., & Ali, R. A. (2023). Simultaneous multi-crop land suitability prediction from remote sensing data using semi-supervised learning. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-33840-6>
- Bischiniotis, K., van den Hurk, B., Coughlan de Perez, E., Veldkamp, T., Nobre, G. G., & Aerts, J. (2019). Assessing time, cost and quality trade-offs in forecast-based action for floods. *International Journal of Disaster Risk Reduction*, 40. <https://doi.org/10.1016/j.ijdr.2019.101252>
- Boote, K. J., Jones, J. W., & Pickering, N. B. (1996). Potential Uses and Limitations of Crop Models. *Agronomy Journal*, 88(5), 704–716. <https://doi.org/https://doi.org/10.2134/agronj1996.00021962008800050005x>
- Bouras, E., Jarlan, L., Khabba, S., Er-Raki, S., Dezetter, A., Sghir, F., & Tramblay, Y. (2019). Assessing the impact of global climate changes on irrigated wheat yields and water requirements in a semi-arid environment of Morocco. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-55251-2>
- Breunig, M., Kuper, P., Reitze, F., Landgraf, S., Al-Doori, M., Stefanakis, E., Abdulmuttalib, H., & Kugler, Z. (2023). IMPROVING DATA QUALITY AND MANAGEMENT FOR REMOTE SENSING ANALYSIS: USE-CASES AND EMERGING RESEARCH QUESTIONS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10(1-

W1-2023), 41–49. <https://doi.org/10.5194/isprs-annals-X-1-W1-2023-41-2023>

Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, 22(10), 1509–1518. <https://doi.org/10.1016/j.envsoft.2006.10.004>

Casa, R., Varella, H., Buis, S., Guérif, M., De Solan, B., & Baret, F. (2012a). 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), 1–10. <https://doi.org/10.1016/j.eja.2011.09.004>

Casa, R., Varella, H., Buis, S., Guérif, M., De Solan, B., & Baret, F. (2012b). 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), 1–10. <https://doi.org/10.1016/j.eja.2011.09.004>

Castañeda-Vera, A., Leffelaar, P. A., Álvaro-Fuentes, J., Cantero-Martínez, C., & Mínguez, M. I. (2015). Selecting crop models for decision making in wheat insurance. *European Journal of Agronomy*, 68, 97–116. <https://doi.org/https://doi.org/10.1016/j.eja.2015.04.008>

Chang, Y., Latham, J., Licht, M., & Wang, L. (2023a). A data-driven crop model for maize yield prediction. *Communications Biology*, 6(1). <https://doi.org/10.1038/s42003-023-04833-y>

Chang, Y., Latham, J., Licht, M., & Wang, L. (2023b). A data-driven crop model for maize yield prediction. *Communications Biology*, 6(1). <https://doi.org/10.1038/s42003-023-04833-y>

Chapagain, R., Remenyi, T. A., Harris, R. M. B., Mohammed, C. L., Huth, N., Wallach, D., Rezaei, E. E., & Ojeda, J. J. (2022). Decomposing crop model

- uncertainty: A systematic review. In *Field Crops Research* (Vol. 279). Elsevier B.V. <https://doi.org/10.1016/j.fcr.2022.108448>
- Chen, Y., & Tao, F. (2022). Potential of remote sensing data-crop model assimilation and seasonal weather forecasts for early-season crop yield forecasting over a large area. *Field Crops Research*, 276, 108398. <https://doi.org/https://doi.org/10.1016/j.fcr.2021.108398>
- Cheng, E., Zhang, B., Peng, D., Zhong, L., Yu, L., Liu, Y., Xiao, C., Li, C., Li, X., Chen, Y., Ye, H., Wang, H., Yu, R., Hu, J., & Yang, S. (2022). Wheat yield estimation using remote sensing data based on machine learning approaches. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.1090970>
- Cheng, M., Wang, H., Fan, J., Xiang, Y., Liu, X., Liao, Z., Abdelghany, A. E., Zhang, F., & Li, Z. (2022). Evaluation of AquaCrop model for greenhouse cherry tomato with plastic film mulch under various water and nitrogen supplies. *Agricultural Water Management*, 274. <https://doi.org/10.1016/j.agwat.2022.107949>
- Choruma, D. J., Balkovic, J., Pietsch, S. A., & Odume, O. N. (2021). Using EPIC to simulate the effects of different irrigation and fertilizer levels on maize yield in the Eastern Cape, South Africa. *Agricultural Water Management*, 254. <https://doi.org/10.1016/j.agwat.2021.106974>
- Choudhury, A. K., Molla, M. S. H., Zahan, T., Sen, R., Biswas, J. C., Akhter, S., Ishtiaque, S., Ahmed, F., Maniruzaman, M., Hossain, M. B., Sarker, P. C., Dessoky, E. S., Hassan, M. M., & Hossain, A. (2021). Optimum sowing window and yield forecasting for maize in northern and western bangladesh using ceres maize model. *Agronomy*, 11(4). <https://doi.org/10.3390/AGRONOMY11040635>
- Ciric, C., Ciffroy, P., & Charles, S. (2012). Use of sensitivity analysis to identify influential and non-influential parameters within an aquatic ecosystem

- model. *Ecological Modelling*, 246(C), 119–130.
<https://doi.org/10.1016/j.ecolmodel.2012.06.024>
- Coleman, K., Muhammed, S. E., Milne, A. E., Todman, L. C., Dailey, A. G., Glendining, M. J., & Whitmore, A. P. (2017a). The landscape model: A model for exploring trade-offs between agricultural production and the environment. *Science of the Total Environment*, 609, 1483–1499.
<https://doi.org/10.1016/j.scitotenv.2017.07.193>
- Coleman, K., Muhammed, S. E., Milne, A. E., Todman, L. C., Dailey, A. G., Glendining, M. J., & Whitmore, A. P. (2017b). The landscape model: A model for exploring trade-offs between agricultural production and the environment. *Science of the Total Environment*, 609, 1483–1499.
<https://doi.org/10.1016/j.scitotenv.2017.07.193>
- Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., & Acutis, M. (2010). Comparison of sensitivity analysis techniques: A case study with the rice model WARM. *Ecological Modelling*, 221(16), 1897–1906.
<https://doi.org/10.1016/j.ecolmodel.2010.04.021>
- Constantin, J., Raynal, H., Casellas, E., Hoffmann, H., Bindi, M., Doro, L., Eckersten, H., Gaiser, T., Grosz, B., Haas, E., Kersebaum, K., Klatt, S., Kuhnert, M., Lewan, E., Maharjan, G., Moriondo, M., Nendel, C., Roggero, P. P., Specka, X., & Bergez, J.-E. (2019). Management and spatial resolution effects on yield and water balance at regional scale in crop models. *Agricultural and Forest Meteorology*, 275, 184–195.
<https://doi.org/10.1016/j.agrformet.2019.05.013>
- Corstanje, R., Kirk, G. J. D., & Lark, R. M. (2008). The behaviour of soil process models of ammonia volatilization at contrasting spatial scales. *European Journal of Soil Science*, 59(6), 1271–1283. <https://doi.org/10.1111/j.1365-2389.2008.01086.x>
- Corstanje, R., Mercer, T., Rickson, R., Lynda, K., Newell Price, P., Holman, I., Kechavarsi, C., & Waine, T. (2017). Physical soil quality indicators for

- monitoring British soils. *Solid Earth*, 8, 1003–1016.
<https://doi.org/10.5194/se-8-1003-2017>
- Courault, D., Hossard, L., Demarez, V., Dechatre, H., Irfan, K., Baghdadi, N., Flamain, F., & Ruget, F. (n.d.). *STICS crop model and Sentinel-2 images for monitoring rice growth and yield in the Camargue region*.
<https://doi.org/10.1007/s13593-021-00697-w/Published>
- Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., & Schaibly, J. H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics*, 59(8), 3873–3878. <https://doi.org/10.1063/1.1680571>
- Darra, N., Anastasiou, E., Kriezi, O., Lazarou, E., Kalivas, D., & Fountas, S. (2023). Can Yield Prediction Be Fully Digitized? A Systematic Review. In *Agronomy* (Vol. 13, Issue 9). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/agronomy13092441>
- De Wit, C. T. (1958). *Transpiration and Crop Yields*. Institute of Biological and Chemical Research in Field Crops and Herbage.
<https://books.google.co.uk/books?id=3YolHQAACAAJ>
- DeJonge, K. C., Ascough, J. C., Ahmadi, M., Andales, A. A., & Arabi, M. (2012). Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. *Ecological Modelling*, 231, 113–125. <https://doi.org/10.1016/j.ecolmodel.2012.01.024>
- Dhanapal, R., Ajanraj, A., Balavinayagapragathish, S., & Balaji, J. (2021). Retraction: Crop price prediction using supervised machine learning algorithms. In *Journal of Physics: Conference Series* (Vol. 1916, Issue 1). IOP Publishing Ltd. <https://doi.org/10.1088/1742-6596/1916/1/012042>
- Dhillon, M. S., Dahms, T., Kuebert-Flock, C., Borg, E., Conrad, C., & Ullmann, T. (2020). Modelling crop biomass from synthetic remote sensing time series:

- Example for the DEMMIN test site, Germany. *Remote Sensing*, 12(11). <https://doi.org/10.3390/rs12111819>
- Di Paola, A., Valentini, R., & Santini, M. (2016). An overview of available crop growth and yield models for studies and assessments in agriculture. *Journal of the Science of Food and Agriculture*, 96(3), 709–714. <https://doi.org/https://doi.org/10.1002/jsfa.7359>
- Diao, C., Augspurger, C. K., Zhao, Y., & Salk, C. F. (2024). A satellite-field phenological bridging framework for characterizing community-level spring forest phenology using multi-scale satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211, 83–103. <https://doi.org/10.1016/j.isprsjprs.2024.03.018>
- Dirwai, T. L., Senzanje, A., & Mabhaudhi, T. (2021). Calibration and evaluation of the fao aquacrop model for canola (*Brassica napus*) under varied moistube irrigation regimes. *Agriculture (Switzerland)*, 11(5). <https://doi.org/10.3390/agriculture11050410>
- Dlamini, L., Crespo, O., van Dam, J., & Kooistra, L. (2023). A Global Systematic Review of Improving Crop Model Estimations by Assimilating Remote Sensing Data: Implications for Small-Scale Agricultural Systems. In *Remote Sensing* (Vol. 15, Issue 16). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/rs15164066>
- Dong, Y., Zhao, C., Yang, G., Chen, L., Wang, J., & Feng, H. (2013). Integrating a very fast simulated annealing optimization algorithm for crop leaf area index variational assimilation. *Mathematical and Computer Modelling*, 58(3–4), 877–885. <https://doi.org/10.1016/j.mcm.2012.12.013>
- Doorenbos, J., Kassam, A. H., & Bentvelsen, C. I. M. (1979). *Yield response to water*. Food and Agriculture Organization of the United Nations.
- Doraiswamy, P. C., Hatfield, J. L., Jackson, T. J., Akhmedov, B., Prueger, J., & Stern, A. (2004). Crop condition and yield simulations using Landsat and

- MODIS. *Remote Sensing of Environment*, 92(4), 548–559.
<https://doi.org/10.1016/j.rse.2004.05.017>
- Dubovik, O., Schuster, G. L., Xu, F., Hu, Y., Bösch, H., Landgraf, J., & Li, Z. (2021). Grand Challenges in Satellite Remote Sensing. *Frontiers in Remote Sensing*, 2. <https://doi.org/10.3389/frsen.2021.619818>
- Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023a). Crop Prediction Model Using Machine Learning Algorithms. *Applied Sciences (Switzerland)*, 13(16).
<https://doi.org/10.3390/app13169288>
- Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023b). Crop Prediction Model Using Machine Learning Algorithms. *Applied Sciences (Switzerland)*, 13(16).
<https://doi.org/10.3390/app13169288>
- Er-Raki, S., Bouras, E., Rodriguez, J. C., Watts, C. J., Lizarraga-Celaya, C., & Chehbouni, A. (2021). Parameterization of the AquaCrop model for simulating table grapes growth and water productivity in an arid region of Mexico. *Agricultural Water Management*, 245.
<https://doi.org/10.1016/j.agwat.2020.106585>
- Feng, L., Wang, Y., Zhang, Z., & Du, Q. (2021). *Geographically and temporally weighted neural network for winter wheat yield prediction*.
- Feng, X., Tian, H., Cong, J., & Zhao, C. (2023). A method review of the climate change impact on crop yield. In *Frontiers in Forests and Global Change* (Vol. 6). Frontiers Media SA. <https://doi.org/10.3389/ffgc.2023.1198186>
- Fischer, C., Montmerle, T., Berre, L., Auger, L., & ȘTEFĂNESCU, S. E. (2005). An overview of the variational assimilation in the ALADIN/France numerical weather-prediction system. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3477–3492.
<https://doi.org/10.1256/qj.05.115>

- Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L. B., Obersteiner, M., & Van Der Velde, M. (2016). Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. *Nature Communications*, 7. <https://doi.org/10.1038/ncomms11872>
- 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. <https://doi.org/10.1016/j.agwat.2016.11.015>
- Gamon, J. A., Field, C. B., Goulden, M. L., Griffin, K. L., Hartley, A. E., Joel, G., Penuelas, J., & Valentini, R. (1995). Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecological Applications*, 5(1), 28–41. <https://doi.org/10.2307/1942049>
- Gavasso-Rita, Y. L., Papalexiou, S. M., Li, Y., Elshorbagy, A., Li, Z., & Schuster-Wallace, C. (2024). Crop models and their use in assessing crop production and food security: A review. In *Food and Energy Security* (Vol. 13, Issue 1). John Wiley and Sons Inc. <https://doi.org/10.1002/fes3.503>
- Gerard, F. F., George, C. T., Hayman, G., Chavana-Bryant, C., & Weedon, G. P. (2020). Leaf phenology amplitude derived from MODIS NDVI and EVI: Maps of leaf phenology synchrony for Meso- and South America. *Geoscience Data Journal*, 7(1), 13–26. <https://doi.org/10.1002/gdj3.87>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017a). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017b). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>

- Greaves, G. E., & Wang, Y. M. (2016). Assessment of fao aquacrop model for simulating maize growth and productivity under deficit irrigation in a tropical environment. *Water (Switzerland)*, 8(12). <https://doi.org/10.3390/w8120557>
- Guo, D., Olesen, J. E., Pullens, J. W. M., Guo, C., & Ma, X. (2021). Calibrating AquaCrop model using genetic algorithm with multi-objective functions applying different weight factors. *Agronomy Journal*, 113(2), 1420–1438. <https://doi.org/10.1002/agj2.20588>
- Han, H., Liu, Z., Li, J., & Zeng, Z. (2024). Challenges in remote sensing based climate and crop monitoring: navigating the complexities using AI. In *Journal of Cloud Computing* (Vol. 13, Issue 1). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1186/s13677-023-00583-8>
- Hariharan, G., & Kannan, K. (2014). Review of wavelet methods for the solution of reaction–diffusion problems in science and engineering. *Applied Mathematical Modelling*, 38(3), 799–813. <https://doi.org/https://doi.org/10.1016/j.apm.2013.08.003>
- He, D., Wang, E., Wang, J., & Robertson, M. J. (2017). Data requirement for effective calibration of process-based crop models. *Agricultural and Forest Meteorology*, 234–235, 136–148. <https://doi.org/10.1016/j.agrformet.2016.12.015>
- Hoogenboom, G., & White, J. W. (2003). Improving Physiological Assumptions Of Simulation Models By Using Gene-Based Approaches. *Agronomy Journal*, 95(1), 82–89. <https://doi.org/10.2134/agronj2003.8200>
- Hsiao, T. C., Heng, L., Steduto, P., Rojas-Lara, B., Raes, D., & Fereres, E. (2009). Aquacrop-The FAO crop model to simulate yield response to water: III. Parameterization and testing for maize. *Agronomy Journal*, 101(3), 448–459. <https://doi.org/10.2134/agronj2008.0218s>
- Hu, T., Zhang, X., Bohrer, G., Liu, Y., Zhou, Y., Martin, J., Li, Y., & Zhao, K. (2023). Crop yield prediction via explainable AI and interpretable machine

- learning: Dangers of black box models for evaluating climate change impacts on crop yield. *Agricultural and Forest Meteorology*, 336, 109458. <https://doi.org/https://doi.org/10.1016/j.agrformet.2023.109458>
- Huang, H., Huang, J., & Wu, Y. (2020). Markov Chain Monte Carlo and Four-Dimensional Variational Approach Based Winter Wheat Yield Estimation. *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 5290–5293. <https://doi.org/10.1109/IGARSS39084.2020.9324208>
- Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., Tian, L., Zhang, X. (Frank), Fan, J., & Wenbin, 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.
- Huemmrich, K. F., Vargas Zesati, S., Campbell, P., & Tweedie, C. (2021). Canopy reflectance models illustrate varying NDVI responses to change in high latitude ecosystems. *Ecological Applications*, 31(8). <https://doi.org/10.1002/eap.2435>
- Huete, A. R. (1988). A Soil-Adjusted Vegetation Index (SAVI) 295. In *REMOTE SENSING OF ENVIRONMENT* (Vol. 25).
- Iqbal, M. A., Shen, Y., Stricevic, R., Pei, H., Sun, H., Amiri, E., Penas, A., & del Rio, S. (2014). Evaluation of the FAO AquaCrop model for winter wheat on the North China Plain under deficit irrigation from field experiment to regional yield simulation. *Agricultural Water Management*, 135, 61–72. <https://doi.org/10.1016/j.agwat.2013.12.012>
- Ishihara, M., Inoue, Y., Ono, K., Shimizu, M., & Matsuura, S. (2015). The impact of sunlight conditions on the consistency of vegetation indices in croplands- Effective usage of vegetation indices from continuous ground-based spectral measurements. *Remote Sensing*, 7(10), 14079–14098. <https://doi.org/10.3390/rs71014079>

- Jain, M., Balwinder-Singh, Rao, P., Srivastava, A. K., Poonia, S., Blesh, J., Azzari, G., McDonald, A. J., & Lobell, D. B. (2019). The impact of agricultural interventions can be doubled by using satellite data. *Nature Sustainability*, 2(10), 931–934. <https://doi.org/10.1038/s41893-019-0396-x>
- Javan, K., & Darestani, M. (2024). Assessing environmental sustainability of a vital crop in a critical region: Investigating climate change impacts on agriculture using the SWAT model and HWA method. *Heliyon*, 10(3). <https://doi.org/10.1016/j.heliyon.2024.e25326>
- Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PLoS ONE*, 11(6). <https://doi.org/10.1371/journal.pone.0156571>
- Ji, Z., Pan, Y., Zhu, X., Wang, J., & Li, Q. (2021). Prediction of crop yield using phenological information extracted from remote sensing vegetation index. *Sensors (Switzerland)*, 21(4), 1–17. <https://doi.org/10.3390/s21041406>
- 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. In *European Journal of Agronomy* (Vol. 92, pp. 141–152). Elsevier B.V. <https://doi.org/10.1016/j.eja.2017.11.002>
- 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(12). <https://doi.org/10.3390/rs8120972>
- 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. <https://doi.org/10.1016/j.agwat.2019.105846>

- 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. <https://doi.org/10.1016/j.isprsjprs.2017.02.001>
- Jin, Z., Prasad, R., Shriver, J., & Zhuang, Q. (2017). Crop model- and satellite imagery-based recommendation tool for variable rate N fertilizer application for the US Corn system. *Precision Agriculture*, 18(5), 779–800. <https://doi.org/10.1007/s11119-016-9488-z>
- Jing, Q., McConkey, B., Qian, B., Smith, W., Grant, B., Shang, J., Liu, J., Bindraban, P., & Luce, M. S. (2021). Assessing water management effects on spring wheat yield in the Canadian Prairies using DSSAT wheat models. *Agricultural Water Management*, 244. <https://doi.org/10.1016/j.agwat.2020.106591>
- Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). *The DSSAT cropping system model*. www.elsevier.com/locate/eja
- Joshi, N., Mitchard, E. T. A., Brolly, M., Schumacher, J., Fernández-Landa, A., Johannsen, V. K., Marchamalo, M., & Fensholt, R. (2017). Understanding “saturation” of radar signals over forests. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-03469-3>
- 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. <https://doi.org/10.1016/j.rse.2019.04.005>
- Karaboga, D. (2005). *An Idea Based on Honey Bee Swarm for Numerical Optimization*.

- Kasampalis, D. A., Alexandridis, T. K., Deva, C., Challinor, A., Moshou, D., & Zalidis, G. (2018). Contribution of remote sensing on crop models: A review. In *Journal of Imaging* (Vol. 4, Issue 4). MDPI. <https://doi.org/10.3390/jimaging4040052>
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, 4, 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
- Kephe, P. N., Ayisi, K. K., & Petja, B. M. (2021a). Challenges and opportunities in crop simulation modelling under seasonal and projected climate change scenarios for crop production in South Africa. In *Agriculture and Food Security* (Vol. 10, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s40066-020-00283-5>
- Kephe, P. N., Ayisi, K. K., & Petja, B. M. (2021b). Challenges and opportunities in crop simulation modelling under seasonal and projected climate change scenarios for crop production in South Africa. In *Agriculture and Food Security* (Vol. 10, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s40066-020-00283-5>
- Khabba, S., Duchemin, B., Hadria, R., Er-Raki, S., Ezzahar, J., Chehbouni, A., Lahrouni, A., & Hanich, L. (2009). Evaluation of digital hemispherical photography and plant canopy analyzer for measuring vegetation area index of orange orchards. *Journal of Agronomy*, 8(2), 67–72. <https://doi.org/10.3923/ja.2009.67.72>
- Khabba, S., Er-Raki, S., Toumi, J., Ezzahar, J., Hssaine, B. A., Page, M. le, & Chehbouni, A. (2020). A simple light-use-efficiency model to estimate wheat yield in the semi-Arid areas. *Agronomy*, 10(10). <https://doi.org/10.3390/agronomy10101524>
- Khabbazan, S., Vermunt, P., Steele-Dunne, S., Arntz, L. R., Marinetti, C., van der Valk, D., Iannini, L., Molijn, R., Westerdijk, K., & van der Sande, C. (2019).

- Crop monitoring using Sentinel-1 data: A case study from The Netherlands. *Remote Sensing*, 11(16). <https://doi.org/10.3390/rs11161887>
- Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.00621>
- Khaki, S., Wang, L., & Archontoulis, S. V. (2020). A CNN-RNN Framework for Crop Yield Prediction. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.01750>
- Klocke, N., Stone, L., Briggeman, S., & Bolton, D. (2010). Technical Note: Scheduling for Deficit Irrigation—Crop Yield Predictor. *Applied Engineering in Agriculture*, 26, 413–418. <https://doi.org/10.13031/2013.29956>
- Koo, J., Bostick, W., Jones, J., Gijsman, A., & Naab, J. (2003). Estimating Soil Carbon in Agricultural Systems using Ensemble Kalman Filter and DSSAT-CENTURY. In *Transactions of the ASABE* (Vol. 50). <https://doi.org/10.13031/2013.13847>
- Krishnan, M., Gugercin, S., & Tarazaga, P. A. (2023). A wavelet-based dynamic mode decomposition for modeling mechanical systems from partial observations. *Mechanical Systems and Signal Processing*, 187, 109919. [https://doi.org/https://doi.org/10.1016/j.ymssp.2022.109919](https://doi.org/10.1016/j.ymssp.2022.109919)
- Kumar, U., Hansen, E. M., Thomsen, I. K., & Vogeler, I. (2023). Performance of APSIM to Simulate the Dynamics of Winter Wheat Growth, Phenology, and Nitrogen Uptake from Early Growth Stages to Maturity in Northern Europe. *Plants*, 12(5). <https://doi.org/10.3390/plants12050986>
- Lee, D.-Y., Lee, D.-S., Cha, Y., Min, J.-H., & Park, Y.-S. (2023). Data-driven models for predicting community changes in freshwater ecosystems: A review. *Ecological Informatics*, 77, 102163. [https://doi.org/https://doi.org/10.1016/j.ecoinf.2023.102163](https://doi.org/10.1016/j.ecoinf.2023.102163)
- Leukel, J., Zimpel, T., & Stumpe, C. (2023). Machine learning technology for early prediction of grain yield at the field scale: A systematic review. *Computers*

- and Electronics in Agriculture*, 207, 107721.
<https://doi.org/https://doi.org/10.1016/j.compag.2023.107721>
- Li, J., & Chen, B. (2020). Global revisit interval analysis of landsat-8-9 and sentinel-2a-2b data for terrestrial monitoring. *Sensors (Switzerland)*, 20(22), 1–15. <https://doi.org/10.3390/s20226631>
- Li, K., & Chen, Y. (2018). A Genetic Algorithm-based urban cluster automatic threshold method by combining VIIRS DNB, NDVI, and NDBI to monitor urbanization. *Remote Sensing*, 10(2). <https://doi.org/10.3390/rs10020277>
- Li, R., Li, C., Liu, F., Yang, X., & Wang, J. (2010). Methods and algorithms of data assimilation and its application in agriculture. *2010 World Automation Congress*, 29–34.
- Li, Z., He, J., Xu, X., Jin, X., Huang, W., Clark, B., Yang, G., & Li, Z. (2018). Estimating genetic parameters of DSSAT-CERES model with the GLUE method for winter wheat (*Triticum aestivum* L.) production. *Computers and Electronics in Agriculture*, 154, 213–221.
<https://doi.org/https://doi.org/10.1016/j.compag.2018.09.009>
- Li, Z., Jin, X., Zhao, C., Wang, J., Xu, X., Yang, G., Li, C., & Shen, J. (2015). Estimating wheat yield and quality by coupling the DSSAT-CERES model and proximal remote sensing. *European Journal of Agronomy*, 71, 53–62.
<https://doi.org/10.1016/j.eja.2015.08.006>
- Li, Z., Zhao, Y., Taylor, J., Gaulton, R., Jin, X., Song, X., Li, Z., Meng, Y., Chen, P., Feng, H., Wang, C., Guo, W., Xu, X., Chen, L., & Yang, G. (2022). Comparison and transferability of thermal, temporal and phenological-based in-season predictions of above-ground biomass in wheat crops from proximal crop reflectance data. *Remote Sensing of Environment*, 273, 112967. <https://doi.org/https://doi.org/10.1016/j.rse.2022.112967>
- Liao, C., Wang, J., Dong, T., Shang, J., Liu, J., & Song, Y. (2019). Using spatio-temporal fusion of Landsat-8 and MODIS data to derive phenology, biomass

- and yield estimates for corn and soybean. *Science of the Total Environment*, 650, 1707–1721. <https://doi.org/10.1016/j.scitotenv.2018.09.308>
- Lin, F., Li, X., Jia, N., Feng, F., Huang, H., Huang, J., Fan, S., Ciais, P., & Song, X. P. (2023). The impact of Russia-Ukraine conflict on global food security. *Global Food Security*, 36. <https://doi.org/10.1016/j.gfs.2022.100661>
- Linker, R. (2020). Unified framework for model-based optimal allocation of crop areas and water. *Agricultural Water Management*, 228, 105859. <https://doi.org/https://doi.org/10.1016/j.agwat.2019.105859>
- López-Urrea, R., Domínguez, A., Pardo, J. J., Montoya, F., García-Vila, M., & Martínez-Romero, A. (2020). Parameterization and comparison of the AquaCrop and MOPECO models for a high-yielding barley cultivar under different irrigation levels. *Agricultural Water Management*, 230. <https://doi.org/10.1016/j.agwat.2019.105931>
- Lorenz, A. J. (2013). Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. *G3: Genes, Genomes, Genetics*, 3(3), 481–491. <https://doi.org/10.1534/g3.112.004911>
- Lu, Y., Chibarabada, T. P., Ziliani, M. G., Onema, J. M. K., McCabe, M. F., & Sheffield, J. (2021). Assimilation of soil moisture and canopy cover data improves maize simulation using an under-calibrated crop model. *Agricultural Water Management*, 252. <https://doi.org/10.1016/j.agwat.2021.106884>
- Lyu, J., Jiang, Y., Xu, C., Liu, Y., Su, Z., Liu, J., & He, J. (2022). Multi-objective winter wheat irrigation strategies optimization based on coupling AquaCrop-OSPy and NSGA-III: A case study in Yangling, China. *Science of the Total Environment*, 843. <https://doi.org/10.1016/j.scitotenv.2022.157104>
- Ma, G., Huang, J., Wu, W., Fan, J., Zou, J., & Wu, S. (2013). Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat

- yield. *Mathematical and Computer Modelling*, 58(3), 634–643.
<https://doi.org/https://doi.org/10.1016/j.mcm.2011.10.038>
- 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(9), 4225–4236. <https://doi.org/10.1109/JSTARS.2017.2711482>
- Ma, Y., Yang, Z., Huang, Q., & Zhang, Z. (2023). Improving the Transferability of Deep Learning Models for Crop Yield Prediction: A Partial Domain Adaptation Approach. *Remote Sensing*, 15(18), 4562. <https://doi.org/10.3390/rs15184562>
- Ma, Y., Zhang, Z., Kang, Y., & Özdoğan, M. (2021). Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sensing of Environment*, 259, 112408. <https://doi.org/https://doi.org/10.1016/j.rse.2021.112408>
- Maestrini, B., Mimić, G., van Oort, P. A. J., Jindo, K., Brdar, S., van Evert, F. K., & Athanasiados, I. (2022a). Mixing process-based and data-driven approaches in yield prediction. In *European Journal of Agronomy* (Vol. 139). Elsevier B.V. <https://doi.org/10.1016/j.eja.2022.126569>
- Maestrini, B., Mimić, G., van Oort, P. A. J., Jindo, K., Brdar, S., van Evert, F. K., & Athanasiados, I. (2022b). Mixing process-based and data-driven approaches in yield prediction. In *European Journal of Agronomy* (Vol. 139). Elsevier B.V. <https://doi.org/10.1016/j.eja.2022.126569>
- Manschadi, A. M., Eitzinger, J., Breisch, M., Fuchs, W., Neubauer, T., & Soltani, A. (2021). Full Parameterisation Matters for the Best Performance of Crop Models: Inter-comparison of a Simple and a Detailed Maize Model. *International Journal of Plant Production*, 15(1), 61–78. <https://doi.org/10.1007/s42106-020-00116-2>

- Martínez-Ruiz, A., Ruiz-García, A., Prado-Hernández, J. V., López-Cruz, I. L., Valencia-Islas, J. O., & Pineda-Pineda, J. (2021). Global sensitivity analysis and calibration by differential evolution algorithm of HORTSYST crop model for fertigation management. *Water (Switzerland)*, 13(5). <https://doi.org/10.3390/w13050610>
- Martinho, V. J. P. D., Cunha, C. A. da S., Pato, M. L., Costa, P. J. L., Sánchez-Carreira, M. C., Georgantzís, N., Rodrigues, R. N., & Coronado, F. (2022). Machine Learning and Food Security: Insights for Agricultural Spatial Planning in the Context of Agriculture 4.0. *Applied Sciences (Switzerland)*, 12(22). <https://doi.org/10.3390/app122211828>
- Marzukhi, F., Elahami, A. L., & Bohari, S. N. (2016). Detecting nutrients deficiencies of oil palm trees using remotely sensed data. *IOP Conference Series: Earth and Environmental Science*, 37(1). <https://doi.org/10.1088/1755-1315/37/1/012040>
- McRae, G. J., Tilden, J. W., & Seinfeld, J. H. (1982). Global sensitivity analysis—a computational implementation of the Fourier Amplitude Sensitivity Test (FAST). *Computers & Chemical Engineering*, 6(1), 15–25. [https://doi.org/https://doi.org/10.1016/0098-1354\(82\)80003-3](https://doi.org/https://doi.org/10.1016/0098-1354(82)80003-3)
- Meddi, M., & Eslamian, S. (2021). Uncertainties in Rainfall and Water Resources in Maghreb Countries Under Climate Change. In *African Handbook of Climate Change Adaptation: With 610 Figures and 361 Tables* (pp. 1967–2003). Springer International Publishing. https://doi.org/10.1007/978-3-030-45106-6_114
- Meroni, M., Fasbender, D., Rembold, F., Atzberger, C., & Klisch, A. (2019). Near real-time vegetation anomaly detection with MODIS NDVI: Timeliness vs. accuracy and effect of anomaly computation options. *Remote Sensing of Environment*, 221, 508–521. <https://doi.org/https://doi.org/10.1016/j.rse.2018.11.041>

- Mohanty, B. P., Cosh, M. H., Lakshmi, V., & Montzka, C. (2017). Soil Moisture Remote Sensing: State-of-the-Science. *Vadose Zone Journal*, 16(1), 1–9. <https://doi.org/10.2136/vzj2016.10.0105>
- Monteiro, L. A., Ramos, R. M., Battisti, R., Soares, J. R., Oliveira, J. C., Figueiredo, G. K. D. A., Lamparelli, R. A. C., Nendel, C., & Lana, M. A. (2022). Potential Use of Data-Driven Models to Estimate and Predict Soybean Yields at National Scale in Brazil. *International Journal of Plant Production*, 16(4), 691–703. <https://doi.org/10.1007/s42106-022-00209-0>
- Moran, M. S., Inoue, Y., & Barnes, E. M. (1997). Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment*, 61(3), 319–346. [https://doi.org/https://doi.org/10.1016/S0034-4257\(97\)00045-X](https://doi.org/https://doi.org/10.1016/S0034-4257(97)00045-X)
- Morel, J., Bégué, A., Todoroff, P., Martiné, J. F., Lebourgeois, V., & Petit, M. (2014a). Coupling a sugarcane crop model with the remotely sensed time series of fIPAR to optimise the yield estimation. *European Journal of Agronomy*, 61, 60–68. <https://doi.org/10.1016/j.eja.2014.08.004>
- Morel, J., Bégué, A., Todoroff, P., Martiné, J.-F., Lebourgeois, V., & Petit, M. (2014b). Coupling a sugarcane crop model with the remotely sensed time series of fIPAR to optimise the yield estimation. *European Journal of Agronomy*, 61, 60–68. <https://doi.org/https://doi.org/10.1016/j.eja.2014.08.004>
- Morris, M. D. (1991a). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2), 161–174. <https://doi.org/10.1080/00401706.1991.10484804>
- Morris, M. D. (1991b). Factorial Sampling Plans for Preliminary Computational Experiments. In *TECHNOMETRICS* (Vol. 33, Issue 2).
- Mubvuma, M. T., Ogola, J. B. O., & Mhizha, T. (2021a). AquaCrop model calibration and validation for chickpea (*Cicer arietinum*) in Southern Africa.

- Cogent Food and Agriculture*, 7(1).
<https://doi.org/10.1080/23311932.2021.1898135>
- Mubvuma, M. T., Ogola, J. B. O., & Mhizha, T. (2021b). AquaCrop model calibration and validation for chickpea (*Cicer arietinum*) in Southern Africa. *Cogent Food and Agriculture*, 7(1).
<https://doi.org/10.1080/23311932.2021.1898135>
- Müller, C., Jägermeyr, J., Franke, J. A., Ruane, A. C., Balkovic, J., Ciais, P., Dury, M., Falloon, P., Folberth, C., Hank, T., Hoffmann, M., Izaurralde, R. C., Jacquemin, I., Khabarov, N., Liu, W., Olin, S., Pugh, T. A. M., Wang, X., Williams, K., ... Elliott, J. W. (2024). Substantial Differences in Crop Yield Sensitivities Between Models Call for Functionality-Based Model Evaluation. *Earth's Future*, 12(3). <https://doi.org/10.1029/2023EF003773>
- Nansen, C., Geremias, L. D., Xue, Y., Huang, F., & Parra, J. R. (2013). Agricultural Case Studies of Classification Accuracy, Spectral Resolution, and Model Over-Fitting. *Applied Spectroscopy*, 67(11), 1332–1338.
<https://doi.org/10.1366/12-06933>
- Nash, J. E., & Sutcliffe, J. v. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/https://doi.org/10.1016/0022-1694(70)90255-6)
- Newlands, N. K., Ghahari, A., Gel, Y. R., Lyubchich, V., Agriculture, T. M., & Canada, A.-F. (n.d.). *Deep Learning for Improved Agricultural Risk Management*. <https://oceanservice.noaa.gov/facts/>
- Niu, Z., He, H., Peng, S., Ren, X., Zhang, L., Gu, F., Zhu, G., Peng, C., Li, P., Wang, J., Ge, R., Zeng, N., Zhu, X., Lv, Y., Chang, Q., Xu, Q., Zhang, M., & Liu, W. (2021). A Process-Based Model Integrating Remote Sensing Data for Evaluating Ecosystem Services. *Journal of Advances in Modeling Earth Systems*, 13(6). <https://doi.org/10.1029/2020MS002451>

- Nossent, J., Elsen, P., & Bauwens, W. (2011). Sobol' sensitivity analysis of a complex environmental model. *Environmental Modelling & Software*, 26(12), 1515–1525. [https://doi.org/https://doi.org/10.1016/j.envsoft.2011.08.010](https://doi.org/10.1016/j.envsoft.2011.08.010)
- Oikonomidis, A., Catal, C., & Kassahun, A. (2023). Deep learning for crop yield prediction: a systematic literature review. In *New Zealand Journal of Crop and Horticultural Science* (Vol. 51, Issue 1, pp. 1–26). Taylor and Francis Ltd. <https://doi.org/10.1080/01140671.2022.2032213>
- Oulaid, B., Milne, A. E., Waine, T., El Alami, R., Rafiqi, M., & Corstanje, R. (2024). Stepwise model parametrisation using satellite imagery and hemispherical photography: Tuning AquaCrop sensitive parameters for improved winter wheat yield predictions in semi-arid regions. *Field Crops Research*, 309. <https://doi.org/10.1016/j.fcr.2024.109327>
- Paleari, L., Movedi, E., Zoli, M., Burato, A., Cecconi, I., Errahouly, J., Pecollo, E., Sorvillo, C., & Confalonieri, R. (2021). Sensitivity analysis using Morris: Just screening or an effective ranking method? *Ecological Modelling*, 455, 109648. [https://doi.org/https://doi.org/10.1016/j.ecolmodel.2021.109648](https://doi.org/10.1016/j.ecolmodel.2021.109648)
- Paredes, P., Wei, Z., Liu, Y., Xu, D., Xin, Y., Zhang, B., & Pereira, L. S. (2015). Performance assessment of the FAO aquaCrop model for soil water, soil evaporation, biomass and yield of soybeans in north china plain. *Agricultural Water Management*, 152, 57–71. <https://doi.org/10.1016/j.agwat.2014.12.007>
- Parent, B., & Tardieu, F. (2014). Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature? In *Journal of Experimental Botany* (Vol. 65, Issue 21, pp. 6179–6189). Oxford University Press. <https://doi.org/10.1093/jxb/eru223>
- Pasquel, D., Roux, S., Richetti, J., Cammarano, D., Tisseyre, B., & Taylor, J. A. (2022). A review of methods to evaluate crop model performance at multiple

- and changing spatial scales. In *Precision Agriculture* (Vol. 23, Issue 4, pp. 1489–1513). Springer. <https://doi.org/10.1007/s11119-022-09885-4>
- Patel, K., & Patel, H. B. (2021). A Comparative Analysis of Supervised Machine Learning Algorithm for Agriculture Crop Prediction. *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 1–5. <https://doi.org/10.1109/ICECCT52121.2021.9616731>
- Patel, U., Pathan, M., Kathiria, P., & Patel, V. (2023). Crop type classification with hyperspectral images using deep learning : a transfer learning approach. *Modeling Earth Systems and Environment*, 9(2), 1977–1987. <https://doi.org/10.1007/s40808-022-01608-y>
- Pelosi, A., Belfiore, O. R., D'Urso, G., & Chirico, G. B. (2022). Assessing Crop Water Requirement and Yield by Combining ERA5-Land Reanalysis Data with CM-SAF Satellite-Based Radiation Data and Sentinel-2 Satellite Imagery. *Remote Sensing*, 14(24). <https://doi.org/10.3390/rs14246233>
- Peng, Y., Zhang, M., Xu, Z., Yang, T., Su, Y., Zhou, T., Wang, H., Wang, Y., & Lin, Y. (2020). Estimation of leaf nutrition status in degraded vegetation based on field survey and hyperspectral data. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-61294-7>
- Petersen, L. K. (2018). Real-time prediction of crop yields from MODIS relative vegetation health: A continent-wide analysis of Africa. *Remote Sensing*, 10(11). <https://doi.org/10.3390/rs10111726>
- Poempool, L., Kruatrachue, B., & Siriboon, K. (2018, August 13). Combine multi particle swarm in supporting trapping in local optima. *ICEAST 2018 - 4th International Conference on Engineering, Applied Sciences and Technology: Exploring Innovative Solutions for Smart Society*. <https://doi.org/10.1109/ICEAST.2018.8434398>
- Prudnikova, E., Savin, I., Vindeker, G., Grubina, P., Shishkonakova, E., & Sharychev, D. (2019). Influence of soil background on spectral reflectance

- of winter wheat crop canopy. *Remote Sensing*, 11(16).
<https://doi.org/10.3390/rs11161932>
- Psiroukis, V., Darra, N., Kasimati, A., Trojacek, P., Hasanli, G., & Fountas, S. (2022). Development of a Multi-Scale Tomato Yield Prediction Model in Azerbaijan Using Spectral Indices from Sentinel-2 Imagery. *Remote Sensing*, 14(17). <https://doi.org/10.3390/rs14174202>
- Puntel, L. A., Sawyer, J. E., Barker, D. W., Dietzel, R., Poffenbarger, H., Castellano, M. J., Moore, K. J., Thorburn, P., & Archontoulis, S. V. (2016). Modeling long-term corn yield response to nitrogen rate and crop rotation. *Frontiers in Plant Science*, 7(November 2016).
<https://doi.org/10.3389/fpls.2016.01630>
- Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2009). Aquacrop-The FAO crop model to simulate yield response to water: II. main algorithms and software description. *Agronomy Journal*, 101(3), 438–447.
<https://doi.org/10.2134/agronj2008.0140s>
- Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2022). *Chapter 3 Calculation procedures AquaCrop Reference manual August 2022*.
- Rani, S., Mishra, A. K., Kataria, A., Mallik, S., & Qin, H. (2023). Machine learning-based optimal crop selection system in smart agriculture. *Scientific Reports*, 13(1), 15997. <https://doi.org/10.1038/s41598-023-42356-y>
- Razzaghi, F., Zhou, Z., Andersen, M. N., & Plauborg, F. (2017). Simulation of potato yield in temperate condition by the AquaCrop model. *Agricultural Water Management*, 191, 113–123.
<https://doi.org/10.1016/j.agwat.2017.06.008>
- Rezzoug, W., Gabrielle, B., Suleiman, A., & Benabdeli, K. (2008). Application and evaluation of the DSSAT-wheat in the Tiaret region of Algeria. In *African Journal of Agricultural Research* (Vol. 3, Issue 4).
<http://www.academicjournals.org/AJAR>

- Rivas-Fandiño, P., Acuña-Alonso, C., Novo, A., Pacheco, F. A. L., & Álvarez, X. (2023). Assessment of high spatial resolution satellite imagery for monitoring riparian vegetation: riverine management in the smallholding. *Environmental Monitoring and Assessment*, 195(1). <https://doi.org/10.1007/s10661-022-10667-8>
- Roberts, M. J., Braun, N. O., Sinclair, T. R., Lobell, D. B., & Schlenker, W. (2017). Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, 12(9). <https://doi.org/10.1088/1748-9326/aa7f33>
- Rostami, A., Raeini-Sarjaz, M., Chabokpour, J., Azamathulla, H. M., & Kumar, S. (2022). Determination of rainfed wheat agriculture potential through assimilation of remote sensing data with SWAT model case study: ZarrinehRoud Basin, Iran. *Water Supply*, 22(5), 5331–5334. <https://doi.org/10.2166/ws.2022.160>
- Rozenstein, O., Cohen, Y., Alchanatis, V., Behrendt, K., Bonfil, D. J., Eshel, G., Harari, A., Harris, W. E., Klapp, I., Laor, Y., Linker, R., Paz-Kagan, T., Peets, S., Rutter, S. M., Salzer, Y., & Lowenberg-DeBoer, J. (2024). Data-driven agriculture and sustainable farming: friends or foes? *Precision Agriculture*, 25(1), 520–531. <https://doi.org/10.1007/s11119-023-10061-5>
- Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. *Agricultural and Forest Meteorology*, 200, 233–248. <https://doi.org/10.1016/j.agrformet.2014.09.016>
- Sabzzadeh, I., & Shourian, M. (2020). Maximizing crops yield net benefit in a groundwater-irrigated plain constrained to aquifer stable depletion using a coupled PSO-SWAT-MODFLOW hydro-agronomic model. *Journal of Cleaner Production*, 262. <https://doi.org/10.1016/j.jclepro.2020.121349>
- Sahbeni, G., Székely, B., Musyimi, P. K., Timár, G., & Sahajpal, R. (2023). Crop Yield Estimation Using Sentinel-3 SLSTR, Soil Data, and Topographic

- Features Combined with Machine Learning Modeling: A Case Study of Nepal. *AgriEngineering*, 5(4), 1766–1788. <https://doi.org/10.3390/agriengineering5040109>
- Sajid, S. S., & Hu, G. (2022). Optimizing Crop Planting Schedule Considering Planting Window and Storage Capacity. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.762446>
- Salman Maher, García-Vila Margarita, Fereres Elias, Raes Dirk, & Steduto Steduto. (2021). The AquaCrop model – Enhancing crop water productivity. In *The AquaCrop model – Enhancing crop water productivity*. FAO. <https://doi.org/10.4060/cb7392en>
- Saltelli, A., Tarantola, S., & Chan, K. P.-S. (1999). A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*, 41(1), 39–56. <https://doi.org/10.1080/00401706.1999.10485594>
- Sarrazin, F., Pianosi, F., & Wagener, T. (2016). Global Sensitivity Analysis of environmental models: Convergence and validation. *Environmental Modelling and Software*, 79, 135–152. <https://doi.org/10.1016/j.envsoft.2016.02.005>
- Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., & Gao, F. (n.d.). *Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description*. <http://www.usgs.gov/pubprod>
- Schumacher, B. L., Burchfield, E. K., Bean, B., & Yost, M. A. (2023). Leveraging Important Covariate Groups for Corn Yield Prediction. *Agriculture (Switzerland)*, 13(3). <https://doi.org/10.3390/agriculture13030618>
- Shang, K., Yao, Y., Di, Z., Jia, K., Zhang, X., Fisher, J. B., Chen, J., Guo, X., Yang, J., Yu, R., Xie, Z., Liu, L., Ning, J., & Zhang, L. (2023). Coupling physical constraints with machine learning for satellite-derived

- evapotranspiration of the Tibetan Plateau. *Remote Sensing of Environment*, 289, 113519. <https://doi.org/https://doi.org/10.1016/j.rse.2023.113519>
- Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning. *IEEE Access*, 11, 111255–111264. <https://doi.org/10.1109/ACCESS.2023.3321861>
- Sharp, R. T., Henrys, P. A., Jarvis, S. G., Whitmore, A. P., Milne, A. E., Coleman, K., Mohankumar, S. E. P., & Metcalfe, H. (2021). Simulating cropping sequences using earth observation data. *Computers and Electronics in Agriculture*, 188, 106330. <https://doi.org/https://doi.org/10.1016/j.compag.2021.106330>
- Shrestha, A., Bheemanahalli, R., Adeli, A., Samiappan, S., Czarnecki, J. M. P., McCraine, C. D., Reddy, K. R., & Moorhead, R. (2023). Phenological stage and vegetation index for predicting corn yield under rainfed environments. *Frontiers in Plant Science*, 14. <https://doi.org/10.3389/fpls.2023.1168732>
- Shuai, G., & Basso, B. (2022). Subfield maize yield prediction improves when in-season crop water deficit is included in remote sensing imagery-based models. *Remote Sensing of Environment*, 272. <https://doi.org/10.1016/j.rse.2022.112938>
- Silva, J. V., & Giller, K. E. (2020). Grand challenges for the 21st century: What crop models can and can't (yet) do. *Journal of Agricultural Science*, 158(10), 794–805. <https://doi.org/10.1017/S0021859621000150>
- 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(11). <https://doi.org/10.1371/journal.pone.0187485>

- Simms, D. M., Hamer, A. M., Zeiler, I., Vita, L., & Waine, T. W. (2023). Mapping Agricultural Land in Afghanistan's Opium Provinces Using a Generalised Deep Learning Model and Medium Resolution Satellite Imagery. *Remote Sensing*, 15(19). <https://doi.org/10.3390/rs15194714>
- Siritanaratkul, B. (2022). Generalizability and limitations of machine learning for yield prediction of oxidative coupling of methane. *Digital Chemical Engineering*, 2. <https://doi.org/10.1016/j.dche.2022.100013>
- Sodari, L., Alam, J., & Das, K. K. (2021). *Performance Assessment of Aquacrop Model for Estimating Canopy Cover, Aboveground Biomass and Grain Yield of Winter Wheat in Saptari District: An Application to Irrigation Management*.
- Solomatine, D., See, L. M., & Abrahart, R. J. (2008). Data-Driven Modelling: Concepts, Approaches and Experiences. In R. J. Abrahart, L. M. See, & D. P. Solomatine (Eds.), *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications* (pp. 17–30). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-79881-1_2
- Soltani, A., Alimagham, S. M., Nehbandani, A., Torabi, B., Zeinali, E., Dadrasi, A., Zand, E., Ghassemi, S., Pourshirazi, S., Alasti, O., Hosseini, R. S., Zahed, M., Arabameri, R., Mohammadzadeh, Z., Rahban, S., Kamari, H., Fayazi, H., Mohammadi, S., Keramat, S., ... Sinclair, T. R. (2020). SSM-iCrop2: A simple model for diverse crop species over large areas. *Agricultural Systems*, 182. <https://doi.org/10.1016/j.agsy.2020.102855>
- Specka, X., Nendel, C., & Wieland, R. (2015). Analysing the parameter sensitivity of the agro-ecosystem model MONICA for different crops. *European Journal of Agronomy*, 71, 73–87. <https://doi.org/10.1016/j.eja.2015.08.004>
- Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009a). Aquacrop-the FAO crop model to simulate yield response to water: I. concepts and underlying principles. *Agronomy Journal*, 101(3), 426–437. <https://doi.org/10.2134/agronj2008.0139s>

- Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009b). AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. *Agronomy Journal*, 101(3), 426–437. <https://doi.org/https://doi.org/10.2134/agronj2008.0139s>
- Stricevic, R., Cosic, M., Djurovic, N., Pejic, B., & Maksimovic, L. (2011). Assessment of the FAO AquaCrop model in the simulation of rainfed and supplementally irrigated maize, sugar beet and sunflower. *Agricultural Water Management*, 98(10), 1615–1621. <https://doi.org/10.1016/j.agwat.2011.05.011>
- Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. *Sensors (Switzerland)*, 19(20). <https://doi.org/10.3390/s19204363>
- Sun, Y. (2022). Enhanced Weather-Based Index Insurance Design for Hedging Crop Yield Risk. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.895183>
- Szu, H., & Hartley, R. (1987). Fast simulated annealing. *Physics Letters A*, 122(3–4). [https://doi.org/10.1016/0375-9601\(87\)90796-1](https://doi.org/10.1016/0375-9601(87)90796-1)
- Tai, M. M. (1994). A Mathematical Model for the Determination of Total Area Under Glucose Tolerance and Other Metabolic Curves $\bullet^{n-1} = X_n \sim X_{n-l}$

$$\text{Total Area} = \sum_{l=1}^n (y_{2-l} - y_l) + \sum_{l=1}^n X_l^4 - \sum_{l=1}^n X_{l-3}^{3/4-3/3} - X_{l-3}^{3/3} \setminus (K_{yy} 1 + X_1 y_2 + X_{\#} 2^4 - X_l^4 - X_{3/3}^{3/3} + 4 - X_l^4 - X_{\#} 5^4 - \dots - X_l^4 - X_l^i).$$
In *DIABETES CARE* (Vol. 17, Issue 2). <http://diabetesjournals.org/care/article-pdf/17/2/152/341381/17-2-152.pdf>
- Takács, S., Csengeri, E., Pék, Z., Bíró, T., Szuvandzsiev, P., Palotás, G., & Helyes, L. (2021). Performance evaluation of aquacrop model in processing tomato biomass, fruit yield and water stress indicator modelling. *Water (Switzerland)*, 13(24). <https://doi.org/10.3390/w13243587>

- Tan, B., Gao, F., Wolfe, R. E., Pedelty, J. A., Nightingale, J., Morisette, J. T., & Ederer, G. A. (2011). An Enhanced TIMESAT Algorithm for Estimating Vegetation Phenology Metrics From MODIS Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(2), 361–371. <https://doi.org/10.1109/JSTARS.2010.2075916>
- Tang, B.-H., Shao, K., Li, Z.-L., Wu, H., & Tang, R. (2015). An improved NDVI-based threshold method for estimating land surface emissivity using MODIS satellite data. *International Journal of Remote Sensing*, 36(19–20), 4864–4878. <https://doi.org/10.1080/01431161.2015.1040132>
- Tao, H., Feng, H., Xu, L., Miao, M., Yang, G., Yang, X., & Fan, L. (2020). Estimation of the yield and plant height of winter wheat using UAV-based hyperspectral images. *Sensors (Switzerland)*, 20(4). <https://doi.org/10.3390/s20041231>
- Tao, R., Zhao, P., Wu, J., Martin, N., Harrison, M. T., Ferreira, C., Kalantari, Z., & Hovakimyan, N. (2023). *Optimizing Crop Management with Reinforcement Learning and Imitation Learning*.
- Tenreiro, T. R., García-Vila, M., Gómez, J. A., Jiménez-Berni, J. A., & Fereres, E. (2021). Using NDVI for the assessment of canopy cover in agricultural crops within modelling research. *Computers and Electronics in Agriculture*, 182. <https://doi.org/10.1016/j.compag.2021.106038>
- Tesfaye, K., Takele, R., Shelia, V., Hayi, E., Dabale, A., Traore, P., & Solomon, D. (2022). High Spatial Resolution Seasonal Crop Yield Forecasting for Heterogeneous Maize Environments in the Oromia Regional State, Ethiopia. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4074628>
- Toumi, J., Er-Raki, S., Ezzahar, J., Khabba, S., Jarlan, L., & Chehbouni, A. (2016a). Performance assessment of AquaCrop model for estimating evapotranspiration, soil water content and grain yield of winter wheat in Tensift Al Haouz (Morocco): Application to irrigation management.

- Agricultural Water Management*, 163, 219–235.
<https://doi.org/10.1016/j.agwat.2015.09.007>
- Toumi, J., Er-Raki, S., Ezzahar, J., Khabba, S., Jarlan, L., & Chehbouni, A. (2016b). Performance assessment of AquaCrop model for estimating evapotranspiration, soil water content and grain yield of winter wheat in Tensift Al Haouz (Morocco): Application to irrigation management. *Agricultural Water Management*, 163, 219–235.
<https://doi.org/10.1016/j.agwat.2015.09.007>
- Tsakmakis, I. D., Gikas, G. D., & Sylaios, G. K. (2021). Integration of Sentinel-derived NDVI to reduce uncertainties in the operational field monitoring of maize. *Agricultural Water Management*, 255, 106998.
<https://doi.org/https://doi.org/10.1016/j.agwat.2021.106998>
- Tsakmakis, I. D., Zidou, M., Gikas, G. D., & Sylaios, G. K. (2018). Impact of Irrigation Technologies and Strategies on Cotton Water Footprint Using AquaCrop and CROPWAT Models. *Environmental Processes*, 5, 181–199.
<https://doi.org/10.1007/s40710-018-0289-4>
- 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(16). <https://doi.org/10.3390/RS12162666>
- van der Velde, M., Biavetti, I., El-Aydam, M., Niemeyer, S., Santini, F., & van den Berg, M. (2019). Use and relevance of European Union crop monitoring and yield forecasts. *Agricultural Systems*, 168, 224–230.
<https://doi.org/10.1016/j.agry.2018.05.001>
- Vannoppen, A., & Gobin, A. (2021). Estimating farm wheat yields from NDVI and meteorological data. *Agronomy*, 11(5).
<https://doi.org/10.3390/agronomy11050946>

- Vanuytrecht, E., Raes, D., & Willems, P. (2014). Global sensitivity analysis of yield output from the water productivity model. *Environmental Modelling & Software*, 51, 323–332. <https://doi.org/https://doi.org/10.1016/j.envsoft.2013.10.017>
- Vazquez-Cruz, M. A., Guzman-Cruz, R., Lopez-Cruz, I. L., Cornejo-Perez, O., Torres-Pacheco, I., & Guevara-Gonzalez, R. G. (2014). Global sensitivity analysis by means of EFAST and Sobol' methods and calibration of reduced state-variable TOMGRO model using genetic algorithms. *Computers and Electronics in Agriculture*, 100, 1–12. <https://doi.org/https://doi.org/10.1016/j.compag.2013.10.006>
- Viswambharan, S., Kumaramkandath, I. T., & Tali, J. A. (2022). A geospatial approach in monitoring the variations on surface soil moisture and vegetation water content: a case study of Palakkad District, Kerala, India. *Environmental Earth Sciences*, 81(20). <https://doi.org/10.1007/s12665-022-10611-6>
- Wagner, M. P., Slawig, T., Taravat, A., & Oppelt, N. (2020). Remote sensing data assimilation in dynamic crop models using particle swarm optimization. *ISPRS International Journal of Geo-Information*, 9(2). <https://doi.org/10.3390/ijgi9020105>
- Wahba, G. (1975). Smoothing noisy data with spline functions. *Numerische Mathematik*, 24(5), 383–393. <https://doi.org/10.1007/BF01437407>
- Waine, T., Simms, D., Taylor, J., & Juniper, G. (2014). Towards improving the accuracy of opium yield estimates with remote sensing. *International Journal of Remote Sensing*, 35, 1–18. <https://doi.org/10.1080/01431161.2014.951743>
- Wale, A., Dessie, M., & Kendie, H. (2022). Evaluating the Performance of AquaCrop Model for Potato Production Under Deficit Irrigation. *Air, Soil and Water Research*, 15. <https://doi.org/10.1177/11786221221108216>

- Wallach, D., Palosuo, T., Thorburn, P., Hochman, Z., Gourdain, E., Andrianasolo, F., Asseng, S., Basso, B., Buis, S., Crout, N., Dibari, C., Dumont, B., Ferrise, R., Gaiser, T., Garcia, C., Gayler, S., Ghahramani, A., Hiremath, S., Hoek, S., ... Seidel, S. J. (2021). The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. *Environmental Modelling and Software*, 145. <https://doi.org/10.1016/j.envsoft.2021.105206>
- Wang, G., Mehmood, F., Zain, M., Hamani, A. K. M., Xue, J., Gao, Y., & Duan, A. (2022). AquaCrop Model Evaluation for Winter Wheat under Different Irrigation Management Strategies: A Case Study on the North China Plain. *Agronomy*, 12(12). <https://doi.org/10.3390/agronomy12123184>
- Wang, J., Li, X., Lu, L., & Fang, F. (2013). Estimating near future regional corn yields by integrating multi-source observations into a crop growth model. *European Journal of Agronomy*, 49, 126–140. <https://doi.org/10.1016/j.eja.2013.03.005>
- Wang, L. (2021). Data Driven Explanation of Temporal and Spatial Variability of Maize Yield in the United States. *Frontiers in Plant Science*, 12. <https://doi.org/10.3389/fpls.2021.701192>
- Wang, L., & Qu, J. J. (2009). Satellite remote sensing applications for surface soil moisture monitoring: A review. In *Frontiers of Earth Science in China* (Vol. 3, Issue 2, pp. 237–247). Higher Education Press. <https://doi.org/10.1007/s11707-009-0023-7>
- Weiss M, & Baret F. (2016). *Sentinel2 ToolBox Level2 Products S2ToolBox Level 2 products: LAI, FAPAR, FCOVER Version 1.1*.
- Wellens, J., Raes, D., Fereres, E., Diels, J., Coppys, C., Adiele, J. G., Ezui, K. S. G., Becerra, L. A., Selvaraj, M. G., Dercon, G., & Heng, L. K. (2022). Calibration and validation of the FAO AquaCrop water productivity model for cassava (*Manihot esculenta* Crantz). *Agricultural Water Management*, 263. <https://doi.org/10.1016/j.agwat.2022.107491>

- Wood, D., McNairn, H., Brown, R. J., & Dixon, R. (2002). The effect of dew on the use of RADARSAT-1 for crop monitoring: Choosing between ascending and descending orbits. *Remote Sensing of Environment*, 80(2), 241–247. [https://doi.org/https://doi.org/10.1016/S0034-4257\(01\)00304-2](https://doi.org/https://doi.org/10.1016/S0034-4257(01)00304-2)
- Wooster, M. J., Roberts, G. J., Giglio, L., Roy, D. P., Freeborn, P. H., Boschetti, L., Justice, C., Ichoku, C., Schroeder, W., Davies, D., Smith, A. M. S., Setzer, A., Csiszar, I., Strydom, T., Frost, P., Zhang, T., Xu, W., de Jong, M. C., Johnston, J. M., ... San-Miguel-Ayanz, J. (2021). Satellite remote sensing of active fires: History and current status, applications and future requirements. *Remote Sensing of Environment*, 267, 112694. <https://doi.org/https://doi.org/10.1016/j.rse.2021.112694>
- Wu, H., & Li, Z. L. (2009). Scale issues in remote sensing: A review on analysis, processing and modeling. In *Sensors* (Vol. 9, Issue 3, pp. 1768–1793). <https://doi.org/10.3390/s90301768>
- Wu, S., Huang, J., Liu, X., Fan, J., Ma, G., Zou, J., & Huang, J. (2011). *Assimilating MODIS-LAI into Crop Growth Model with EnKF to Predict Regional Crop Yield*. 410–418. <https://doi.org/10.1007/978-3-642-27275>
- Wu, S., Yang, P., Ren, J., Chen, Z., & Li, H. (2021). Regional winter wheat yield estimation based on the WOFOST model and a novel VW-4DEnSRF assimilation algorithm. *Remote Sensing of Environment*, 255, 112276. <https://doi.org/https://doi.org/10.1016/j.rse.2020.112276>
- Wu, Y., Xu, W., Huang, H., & Huang, J. (2022). Bayesian Posterior-Based Winter Wheat Yield Estimation at the Field Scale through Assimilation of Sentinel-2 Data into WOFOST Model. *Remote Sensing*, 14(15). <https://doi.org/10.3390/rs14153727>
- Xavier, A., Hall, B., Casteel, S., Muir, W., & Rainey, K. M. (2017). Using unsupervised learning techniques to assess interactions among complex traits in soybeans. *Euphytica*, 213(8). <https://doi.org/10.1007/s10681-017-1975-4>

- Xiao, J., Aggarwal, A. K., Duc, N. H., Arya, A., Rage, U. K., & Avtar, R. (2023). A review of remote sensing image spatiotemporal fusion: Challenges, applications and recent trends. *Remote Sensing Applications: Society and Environment*, 32, 101005. <https://doi.org/https://doi.org/10.1016/j.rsase.2023.101005>
- Xin, Q., Li, J., Li, Z., Li, Y., & Zhou, X. (2020). Evaluations and comparisons of rule-based and machine-learning-based methods to retrieve satellite-based vegetation phenology using MODIS and USA National Phenology Network data. *International Journal of Applied Earth Observation and Geoinformation*, 93. <https://doi.org/10.1016/j.jag.2020.102189>
- XING, H. min, XU, X. gang, LI, Z. hai, CHEN, Y. jin, FENG, H. kuan, YANG, G. jun, & CHEN, Z. xia. (2017a). 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(11), 2444–2458. [https://doi.org/10.1016/S2095-3119\(16\)61626-X](https://doi.org/10.1016/S2095-3119(16)61626-X)
- XING, H. min, XU, X. gang, LI, Z. hai, CHEN, Y. jin, FENG, H. kuan, YANG, G. jun, & CHEN, Z. xia. (2017b). 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(11), 2444–2458. [https://doi.org/10.1016/S2095-3119\(16\)61626-X](https://doi.org/10.1016/S2095-3119(16)61626-X)
- Xu, C., & Gertner, G. (2011). Understanding and comparisons of different sampling approaches for the Fourier Amplitudes Sensitivity Test (FAST). *Computational Statistics and Data Analysis*, 55(1), 184–198. <https://doi.org/10.1016/j.csda.2010.06.028>
- Xu, G., & Yu, G. (2018). On convergence analysis of particle swarm optimization algorithm. *Journal of Computational and Applied Mathematics*, 333, 65–73. <https://doi.org/10.1016/j.cam.2017.10.026>

- Xu, H., Twine, T. E., & Yang, X. (2014). Evaluating remotely sensed phenological metrics in a dynamic ecosystem model. *Remote Sensing*, 6(6), 4660–4686. <https://doi.org/10.3390/rs6064660>
- Xu, Y., Yang, Y., Chen, X., & Liu, Y. (2022). Bibliometric Analysis of Global NDVI Research Trends from 1985 to 2021. In *Remote Sensing* (Vol. 14, Issue 16). MDPI. <https://doi.org/10.3390/rs14163967>
- Yang, H., Yang, X., Heskell, M., Sun, S., & Tang, J. (2017). Seasonal variations of leaf and canopy properties tracked by ground-based NDVI imagery in a temperate forest. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-01260-y>
- Yang, Y., Anderson, M. C., Gao, F., Johnson, D. M., Yang, Y., Sun, L., Dulaney, W., Hain, C. R., Otkin, J. A., Prueger, J., Meyers, T. P., Bernacchi, C. J., & Moore, C. E. (2021). Phenological corrections to a field-scale, ET-based crop stress indicator: An application to yield forecasting across the U.S. Corn Belt. *Remote Sensing of Environment*, 257, 112337. <https://doi.org/10.1016/j.rse.2021.112337>
- Yang, Y., Tao, B., Liang, L., Huang, Y., Matocha, C., Lee, C. D., Sama, M., El Masri, B., & Ren, W. (2021). Detecting recent crop phenology dynamics in corn and soybean cropping systems of Kentucky. *Remote Sensing*, 13(9). <https://doi.org/10.3390/rs13091615>
- Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., ... Sequeira, A. M. M. (2018). Outstanding Challenges in the Transferability of Ecological Models. In *Trends in Ecology and Evolution* (Vol. 33, Issue 10, pp. 790–802). Elsevier Ltd. <https://doi.org/10.1016/j.tree.2018.08.001>

- Ye, J., Bao, W., Liao, C., Chen, D., & Hu, H. (2023). Corn Phenology Detection Using the Derivative Dynamic Time Warping Method and Sentinel-2 Time Series. *Remote Sensing*, 15(14). <https://doi.org/10.3390/rs15143456>
- Yli-Heikkilä, M., Wittke, S., Luotamo, M., Puttonen, E., Sulkava, M., Pellikka, P., Heiskanen, J., & Klami, A. (2022). Scalable Crop Yield Prediction with Sentinel-2 Time Series and Temporal Convolutional Network. *Remote Sensing*, 14(17). <https://doi.org/10.3390/rs14174193>
- Yu, J., Smith, A., & Sumner, D. A. (2018). Effects of crop insurance premium subsidies on crop acreage. *American Journal of Agricultural Economics*, 100(1), 91–114. <https://doi.org/10.1093/ajae/aax058>
- Zampieri, M., Weissteiner, C. J., Grizzetti, B., Toreti, A., van den Berg, M., & Dentener, F. (2020). Estimating resilience of crop production systems: From theory to practice. *Science of the Total Environment*, 735. <https://doi.org/10.1016/j.scitotenv.2020.139378>
- Zeng, J., Yuan, X., Ji, P., & Shi, C. (2021). Effects of meteorological forcings and land surface model on soil moisture simulation over China. *Journal of Hydrology*, 603. <https://doi.org/10.1016/j.jhydrol.2021.126978>
- Zeng, L., Wardlow, B. D., Xiang, D., Hu, S., & Li, D. (2020). A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment*, 237, 111511. <https://doi.org/https://doi.org/10.1016/j.rse.2019.111511>
- ZHANG, F., ZHANG, L.-W., SHI, J.-J., & HUANG, J.-F. (2014). Soil Moisture Monitoring Based on Land Surface Temperature-Vegetation Index Space Derived from MODIS Data. *Pedosphere*, 24(4), 450–460. [https://doi.org/https://doi.org/10.1016/S1002-0160\(14\)60031-X](https://doi.org/https://doi.org/10.1016/S1002-0160(14)60031-X)
- Zhang, J., Pan, B., Shi, W., & Zhang, Y. (2023). Monitoring Waterlogging Damage of Winter Wheat Based on HYDRUS-1D and WOFOST Coupled Model and

- Assimilated Soil Moisture Data of Remote Sensing. *Remote Sensing*, 15(17). <https://doi.org/10.3390/rs15174133>
- 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. <https://doi.org/https://doi.org/10.1016/j.compag.2019.105052>
- Zhang, Y., Walker, J. P., & Pauwels, V. R. N. (2022). Assimilation of wheat and soil states for improved yield prediction: The APSIM-EnKF framework. *Agricultural Systems*, 201. <https://doi.org/10.1016/j.agry.2022.103456>
- Zhang, Y., Walker, J. P., Pauwels, V. R. N., & Sadeh, Y. (2022a). Assimilation of wheat and soil states into the apsim-wheat crop model: A case study. *Remote Sensing*, 14(1). <https://doi.org/10.3390/rs14010065>
- Zhang, Y., Walker, J. P., Pauwels, V. R. N., & Sadeh, Y. (2022b). Assimilation of wheat and soil states into the apsim-wheat crop model: A case study. *Remote Sensing*, 14(1). <https://doi.org/10.3390/rs14010065>
- Zhao, H., Yang, C., Guo, W., Zhang, L., & Zhang, D. (2020). Automatic estimation of crop disease severity levels based on vegetation index normalization. *Remote Sensing*, 12(12). <https://doi.org/10.3390/rs12121930>
- Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y., & Yu, S. (2016). Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. *Remote Sensing*, 8(6). <https://doi.org/10.3390/rs8060469>
- Zomer, R. J., Xu, J., & Trabucco, A. (2022). Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. *Scientific Data*, 9(1). <https://doi.org/10.1038/s41597-022-01493-1>
- Zou, X., & Möttus, M. (2017). Sensitivity of common vegetation indices to the canopy structure of field crops. *Remote Sensing*, 9(10). <https://doi.org/10.3390/rs9100994>

2 Stepwise model parametrisation using satellite imagery and hemispherical photography: Tuning AquaCrop sensitive parameters for improved winter wheat yield predictions in semi-arid and arid regions

Bader Oulaid ^{a, b, c}, Alice E. Milne ^b, Toby Waine ^a, Rafiq El Alami ^c, Maryam Rafiqi ^c, Ron Corstanje ^a

^a Cranfield University, College Road, Cranfield, Bedford, MK43 0AL, United Kingdom

^b Net-Zero and Resilient Farming, Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom

^c Mohammed VI Polytechnic University, Lot 660, Moulay Rachid, Ben Guerir 43150, Morocco

Abstract

Crop models are complex with many parameters, which has limited their application. Here we present an approach which both removes the model complexity through reducing the parameter dimensionality through sensitivity analysis and presents a subsequent efficient approach to model parameterisation using swarm optimisation. We do this for two key model outputs, crop canopy and yield, and for two types of observational data, hemispheric photographs and Landsat7 imagery. Importantly we compare the usefulness of these two sources of data in terms of accurate yield prediction. The results showed that the dominant model parameters that predict canopy cover were generally consistent across the fields, with the exception of those related water stress. Although mid-season canopy cover extracted from Landsat7 was underestimated, good agreement was found between the simulated and observed canopy cover for both sources of data. Subsequently, less accurate yield predictions were achieved with the Landsat7 compared to the hemispherical photography-based parametrizations. Despite the small differences in the canopy predictions, the implications for yield prediction were substantial with the parametrization based on hemispherical photography providing far more accurate estimates of yield. There are, however, additional resource implications associated with hemispherical photography. We evaluate these trade-offs, providing model parametrization sets and

demonstrating the potential of satellite imagery to assist AquaCrop, particularly on large scales where ground measurements are challenging.

2.1 Introduction

Crop growth models use empirical and mathematical relationships to simulate the growth and development of crops under diverse conditions, making them important tools to inform agricultural decisions (Soltani et al., 2020). Several crop growth models have been successfully used to assist sustainable crop production including DSSAT (Rezzoug et al., 2008), CROPWAT (Tsakmakis et al., 2018) and APSIM (Y. Zhang, Walker, & Pauwels, 2022). While these models are effective in simulating yield under various management practices (Choruma et al., 2021; Jing et al., 2021; Y. Zhang, Walker, Pauwels, et al., 2022a), their broader use has been restricted due to their complexity: they need a relatively large number of input parameters, most of which are difficult to obtain in practice. Given this complexity, the FAO devised a simple and resilient model, AquaCrop (Raes et al., 2009), which requires a smaller number of intuitive input parameters. This model has been used in many studies to predict yield and irrigation requirements (Khabba et al., 2020; Wang et al., 2022) and has also been tested to be effective for a wide variety of crops, such as wheat (Toumi et al., 2016), potato (Razzaghi et al., 2017; Wale et al., 2022), sunflower (Stricevic et al., 2011), tomato (Takács et al., 2021), maize (Greaves & Wang, 2016), and table grape (Er-Raki et al., 2021). According to the findings of all these studies, the AquaCrop model is able to reliably simulate responses to a variety of field management practices.

The usefulness of processed-based models like AquaCrop is usually constrained by the fact that they are often calibrated for a specific circumstance and so may not be transferable to other case studies. For example, Paredes et al. (2015) found that uncalibrated, the AquaCrop model was unable to accurately forecast the yield and the biomass of soybeans in North China Plain since there was a noticeable tendency for under-estimation for both predictions with

deviations of more than 27% and 20%, respectively. Thus, areas where the model may be used can be expanded only by calibrating it across many sites with various sowing dates, seeding rates, fertilization programmes, and irrigation programmes. Given that AquaCrop is well suited for modelling crop production and that AquaCrop is widely used in Morocco (Salman et al., 2021), the creation of calibrated regional parameterization would be advantageous.

A number of approaches have been used to calibrate the AquaCrop for a specific area. Most of these studies estimate the parameters by matching phenological observations to certain parameters and adopting default values for others (Abi Saab et al., 2021; Dirwai et al., 2021; Lu et al., 2021). In many cases the observational field data was used to parameterize AquaCrop through a process of trial-and-error where data comes from several plots or fields and then the mean of the parameter values is used (Toumi et al., 2016). This provides a simple but pragmatic approach to calibration. More generally, process model parameterisation has been explored by a number of authors. The first step is often to use a sensitivity analysis to identify which parameters most influence the outcome of interest; these then become the focus of the parameterisation. Sensitivity analysis includes evaluating the local and global influence of individual parameters or combinations thereof. Local sensitivity analysis facilitates identifying parameters that have a direct effect on model predictions by examining the effect of minor changes in a single parameter on the model output. In contrast, global sensitivity analysis takes into consideration interactions and non-linear effects. Techniques such as Sobol indices (Nossent et al., 2011; Vazquez-Cruz et al., 2014; XING et al., 2017a), Morris method (Morris, 1991a; Paleari et al., 2021; Specka et al., 2015), and FAST (McRae et al., 1982; C. Xu & Gertner, 2011) are frequently adopted. Building on the insights gained from sensitivity analysis, the next step is model parameterization. The advancement in parameterization techniques for crop growth models covers forcing, calibration, and updating methods, each contributing to the refinement and reliability of model predictions (X. Jin et al., 2018). Forcing methods leverage remote sensing and field measured data using algorithms like linear interpolation (Dhillon et al., 2020;

Pelosi et al., 2022) and wavelet methods(Hariharan & Kannan, 2014; Krishnan et al., 2023) to substitute missing or incomplete crop model simulations with information derived from observed or remotely sensed sources. Findings from Casa et al. (2012) and Morel et al.(2014) highlight the effectiveness of forcing techniques in overcoming limitations related to the lack of detailed knowledge about management practices or soil characteristics. Calibration involves comparing the model's predictions or simulations to observed data and then adjusting the model parameters or inputs to reduce discrepancies between the model output and actual observations. Calibration is often performed using optimization techniques to find the best combination of parameter values that minimizes the difference between model predictions and observed data (X. Jin et al., 2016; Lyu et al., 2022). For example, Li et al. (2018) coupled the GLUE method and experimental field data to parametrise DSSAT-CERES model, the method showcased its accuracy in estimating wheat genotype parameters. Dong et al. (2013) introduced a VFSA optimization algorithm for LAI assimilation into the CERES-Wheat model, and demonstrated an improved accuracy, especially for LAI in winter wheat experiments. Lastly, updating methods refer to techniques that integrate observational data into the model to improve its accuracy and reliability. The goal is to adjust the model's predictions in light of real-world observations, making it more consistent with the observed behaviour of the system. In the context of crop modelling, data assimilation can be used to repeatedly update crop model simulations, for example, when remote sensing data is available. Zhang et al. (2022) developed an ensemble Kalman filter (EnKF) data assimilation framework to assimilate plant and soil observations into the APSIM-Wheat model, and the results showed an improvement in crop development and yield simulations. In this paper we are concerned with calibration using optimisation.

There are two key considerations for obtaining a parameterised model for any given circumstance or case study; i) data, in the form of observations, are needed for the key variables of interest (canopy cover and yield) across the growing season and once obtained, ii) an efficient optimisation procedure which

is capable of considering the full model parameter space conditional to the particularities of study areas of interest. One method for obtaining canopy cover measurements is through hemispherical handheld photography. The technique is based on a determination of how much sky is visible through the canopy, by categorising each pixel of the image as either belonging to the sky or to any plant obstructing object. Although this approach offers a direct way to assess canopy cover, it is labour-intensive, time-consuming, and requires specialised equipment. Remote sensing offers an alternative means of predicting canopy cover directly using dedicated software such as S2ToolBox (Weiss et al., 2016), or indirectly through vegetation indices (VIs). While remote sensing provides a cost-free method for estimating canopy cover expansion across multiple time and space scales, it is often affected by limiting factors, such as sensor failures and/or cloud cover obscuring satellite imagery.

There are many optimization algorithms available, and these can be generally characterized as either local or global optimizations. Local methods use an iterative search beginning with the nominal value of the parameter but can become trapped in a local minimum and prematurely terminate the search (Poempool et al., 2018). Global optimization searches adopt mechanism to search a larger part of the search space and are more frequently used to parameterize crop models. Here, methods include particle swarm optimization (Kennedy & Eberhart, 1995), the fast simulated annealing (Szu & Hartley, 1987), and the Artificial bee colony (Karaboga, 2005). Particle swarm optimization (PSO) has recently increased in popularity and has been used in a variety of applications, including the enhancement of crop yield estimate accuracy (Sabzzadeh & Shourian, 2020), and the assimilation of remotely sensed data into decision support systems (Wagner et al., 2020). It is appealing because of its ease of implementation and ability to quickly converge to a reasonably good solution compared to classic optimization algorithms, which have great search capability but poor convergence speed and computing efficiency (G. Xu & Yu, 2018).

This study presents a methodological framework to determine case specific crop model parameterization. We use the Morris and FAST methods as preliminary steps to analyse the impact of different parameters on AquaCrop key responses, including canopy cover aboveground biomass, and yield. A key novelty here is that we assess the utility of canopy cover estimates obtained from two distinct sources (i.e., Landsat 7 satellite imagery and hemispherical photographs) for calibrating the AquaCrop canopy simulation. Unlike traditional calibration methods reliant on trial and error, our approach is a stepwise automatic parameterisation. In support of our framework, we enhanced the Matlab implementation of AquaCrop by incorporating a soil fertility component. We demonstrated our framework for a semi-arid region in Morocco.

2.2 Material and methods

2.2.1 AquaCrop model

AquaCrop is a multi-crop water driven model that evolved from the FAO Irrigation Drainage Paper 33 (Doorenbos et al., 1979). The model avoids the confounding effect of non-productive water consumption by separating the actual evapotranspiration (ET) into crop transpiration (Tr) and soil evaporation (E). AquaCrop uses a comparatively small number of parameters to assess the effect of environment and management practices on crop growth. The final crop yield is estimated through four steps: i) simulate the foliage development represented by the green canopy cover (CC); (ii) calculate the crop transpiration (Tr) using the reference evapotranspiration (ET_o) and the crop coefficient; (iii) produce the above-ground biomass (B) as a function of the cumulative amount of water transpired; and (iv) simulate the final crop yield by multiplying biomass (B) with a Harvest Index (HI). Details can be found in Raes et al. (2009) and Steduto et al. (2009). In this study, the MATLAB version of the AquaCrop model (AquaCrop-OS) developed by Foster et al. (2017) which is available at (www.aquacropos.com) was used. The soil fertility feature is not implemented in AquaCrop AOS v6.0a. Based on AquaCrop reference manual (Raes et al., 2022), a code to account for the influence of soil fertility on canopy cover growth (KS_{CGC}),

maximum canopy cover (KS_{CCX}), and biomass water productivity (KS_{WP^*}) was integrated to AquaCrop as defined in equations 2-1, 2-2 and 2-3 respectively.

$$KS_{CGC} = \frac{e^{S_{rel}f_{CGC}} - 1}{e^{f_{CGC}} - 1} \quad (2-1)$$

$$KS_{CCX} = \frac{e^{S_{rel}f_{CCX}} - 1}{e^{f_{CCX}} - 1} \quad (2-2)$$

$$KS_{WP^*} = \frac{e^{S_{rel}f_{WP^*}} - 1}{e^{f_{WP^*}} - 1} \quad (2-3)$$

where S_{rel} is the relative stress level in the studied fields. The coefficients f_{CGC} , f_{CCX} and f_{WP^*} are the shape factors describing the effects of soil fertility on canopy growth, maximum canopy cover and biomass water productivity, respectively.

2.2.2 Study and data description

2.2.2.1 Study site

Field experiments were conducted at an irrigated zone of the Haouz plain during two successive growing seasons: 2002–2003 and 2003–2004 by Toumi et al. (2016). The area is located 40 km east of Marrakech (Figure 7). In season 2002–2003, three field were investigated (we refer to these as A1, A2 and A3) and in season 2003–2004, six fields were investigated (referred hereafter as B1, B2, B3, B4, B5 and B6). In this area, the irrigation scheduling is managed by ORMVAH (Office Regional de Mise en Valeur Agricole du Haouz), a regional agricultural advisory body for the management of hydro-agricultural equipment and water resources for agricultural use. The region's climate is characterized as semi-Mediterranean with an average annual precipitation of about 240 mm falling between November and May.

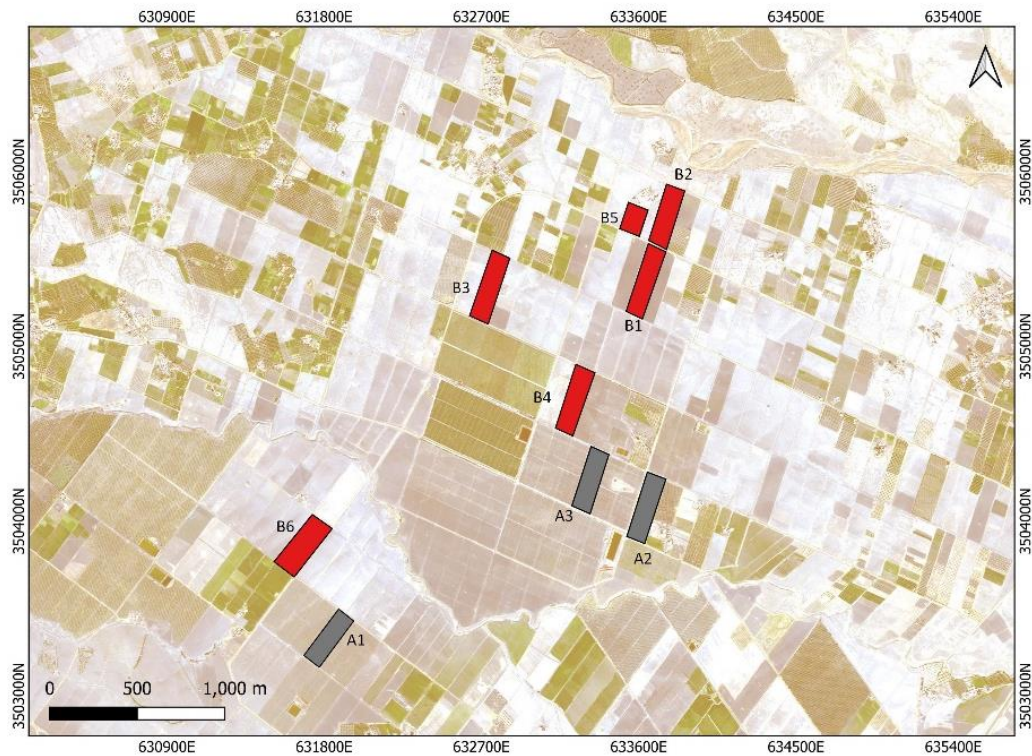


Figure 7 Location of the study area showing the study fields in which irrigated wheat was grown. The red parcels are from season 2003–2004 and the grey parcels are related to 2002–2003 growing season.

2.2.2.2 Data description

AquaCrop requires daily measurements of maximum and minimum temperatures, precipitation, and reference evapotranspiration as weather input variables. We sourced Weather data from Agafay weather stations (31.47 °N, 8.21 °W). The maximum and minimum air temperatures in 2002–2003 were 38.1 and -0.4 °C respectively while the comparative temperatures were 38.4 and -2.04 °C in 2003–2004 (for more details see Appendix B, Figure_Apx B.1-1). Field data, including canopy cover (CC), aboveground biomass and yield measurements, were collected by Toumi et al. (2016) during the 2002–2003 and 2003–2004 growing seasons. Canopy cover was measured using hemispherical canopy photographs for two seasons: Each field was differently managed in terms of sowing date and irrigation schedule. Details are given in Table 2. The geographic scope of this dataset is limited to semi-arid regions of Morocco, specifically within

the Agafay area, which may constrain the applicability of findings to regions with substantially different climatic conditions.

Table 2 Sowing and irrigation schedule for the studied fields (taken from Toumi et.al 2016). Fields A1, A2, A3 were irrigated with 30 mm per event, while fields B1, B2, B3, B4, B5, B6 received 60 mm per irrigation event.

Parcel ID	Sowing dates	Irrigation dates						Total irrigation amount (mm)
		1 st	2 nd	3 rd	4 th	5 th	6 th	
A1	17 th	28 th	22 nd	10 th				90
	December	January	February	April				
	2002	2003	2003	2003				
A2	11 th January	01 st	21 st	14 th	24 th	07 th	24 th	180
	2003	February	February	March	March	April	April	
		2003	2003	2003	2003	2003	2003	
A3	14 th January	04 th	20 th	14 th	21 st			120
	2003	February	February	April	April			
		2003	2003	2003	2003			
B1	21 st	20 th	23 rd	01 st				180
	November	January	February	April				
	2003	2004	2004	2004				
B2	21 st	16 th	17 th	28 th				180
	November	January	February	March				
	2003	2004	2004	2004				
B3	15 th	20 th	15 th	17 th				180
	December	January	February	March				
	2003	2004	2004	2004				
B4	19 th	18 th	24 th	21 st				180
	December	January	February	April				
	2003	2004	2004	2004				

B5	20 th December 2003	16 th January 2004	16 th February 2004	26 th March 2004	180
B6	24 th December 2003	26 th January 2004	21 st February 2004	27 th March 2004	180

For model calibration and validation, we required data on canopy cover and yield. We used two approaches to estimate our first state variable (CC). In the first method, the canopy cover of our calibration sites was measured over the growing season using hemispherical canopy images taken with a Nikon Coolpix 950 and FC-E8 fish-eye lens converter with a field of view of 183 degrees by Toumi et al. (2016). Following image acquisition, the CAN eye software was used for image analysis. In the pre-processing stage, all images were converted to grayscale, using 256 levels of intensity, and then transformed into binary format to distinguish between black and white pixels. Subsequently, the image with the optimal contrast between the canopy and the sky was selected. This selection process aimed to identify the image with the least number of saturated pixels, ensuring the best representation of the canopy-sky interface. The images were then classified into black and white bitmaps using the CAN EYE software and canopy cover was expressed as percentages. More details about this process are provided by Khabba et al. (2009). In the second approach, free atmospherically corrected satellite imagery was extracted from the Landsat7 ETM+ sensor and processed it using GEE Python API (Gorelick et al., 2017a) to calculate the normalized difference vegetation index (NDVI). Landsat 7 provides medium spatial resolution of 30 m images with a revisit time of 16 days. The image collections are created with the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm (Schmidt et al., 2013) and contain 4 visible bands (blue, green, red and near infrared), two short-wave infrared bands (SWIR1, SWIR2) and one thermal infrared band (TIR) as shown in Table 3.

The automated workflow for computation of the NDVI started with an extraction of an image collection covering the crop growth cycle of our calibration sites. A bitmask band was used to mask pixels that contain cirrus and high clouds, then a buffer of 30 m was applied to avoid the edges that mislead the calculation of NDVI values (Vannoppen & Gobin, 2021).

Table 3 Landsat 7 band specifications.

Band	Spatial resolution (m)	Wavelength (μm)	Description
B1	30	0.45 – 0.52	Blue
B2	30	0.52 – 0.60	Green
B3	30	0.63 – 0.69	Red
B4	30	0.77 – 0.90	Near infrared (NIR)
B5	30	1.55 – 1.75	Shortwave infrared 1 (SWIR1)
B6	60	10.40 – 12.50	Thermal infrared (TIR)
B7	30	2.08 – 2.35	Shortwave infrared 2 (SWIR2)

We used the formula by Tenreiro et al. (2021) to estimate canopy cover from Normalized Difference Vegetation Index (NDVI). Tenreiro et al. (2021) derived the formula from a meta-analysis incorporating data from 19 studies, including the significant study conducted by Er-raki et al. (2007) focusing on the Tensift Al Haouz region and aligns with our study area. The derived formula, specifically tailored for wheat and exhibits a statistically significant fit and reasonable accuracy, as evidenced by an R-squared value of 0.71 and an RMSE of 14.2. Both coefficients in the model have been verified to be statistically significant at the 5% level, providing robustness to the derived equation.

$$CC = 97.368 \text{ NDVI} - 4.942 \quad (2-4)$$

We refer to these measures of CC hereafter as CC_{L7} . To estimate grain yield, plants were sampled from five quadrates of size $0.5m \times 0.5m$ selected randomly from across the field. From each quadrate, subsamples of plants were selected at random, and the number and weight of grains were measured (Table 4).

Table 4 Average aboveground biomass and yield in the study area with the range of measured values given in brackets.

Season	2002–2003			2003–2004					
Parcel ID	A1	A2	A3	B1	B2	B3	B4	B5	B6
Biomass	6.02	4.58	1.97	6.5	4.85	5.04	6.67	2.91	5.01
(t ha⁻¹)	[5.6–6.3]	[4.2–5.2]	[1.5–2.2]	[6.2–7]	[4.5–5.3]	[4.7–5.3]	[6.4– 6.9]	[2.5–3.2]	[4.7–5.2]
Yield	2.76	2.09	0.9	2.99	2.23	2.29	3.26	1.32	2.26
(t ha⁻¹)	[2.5–3]	[1.8–2.2]	[0.7–1]	[2.5–3]	[2–2.5]	[2–2.5]	[3–3.5]	[1–1.5]	[2–2.5]

2.2.3 Sensitivity analysis methods

2.2.3.1 Morris

Morris (1991) established the elementary effects method, which was further refined by Campolongo et al. (2007), to determine the model parameters (and their interactions) that most influences the output of interest. This method is suitable for models with a large number of parameters. For a model with k independent inputs. The input space (Ω) is discretized into a number of equal segments of size Δ . The i th input factor's elementary effect is defined as:

$$EE_i = \frac{[Y(X_1, X_2, \dots, X_{i-1}, X_i + \Delta, \dots, X_k) - Y(X_1, X_2, \dots, X_k)]}{\Delta} \quad 2-5)$$

As a result, each EE_i is a measure of the local variation of the model response Y in response to a change in the related model parameter. The Elementary effects distribution F_i is determined for the i th input factor by randomly sampling different parameters (X) from Ω . Morris' approach computes two sensitivity measures: the mean (μ) and the standard deviation (σ) of the sensitivity measure. A significant standard deviation σ indicates that the input has a nonlinear effect or interacts with other inputs, while large μ indicates that the overall effect of that input on the output is greater. Certain effects may cancel out when calculating μ , resulting in a low mean value for even a significant component. To prevent such Type II errors, Campolongo et al. (2007) advocated substituting μ^* for μ , which is defined as an estimate of the mean of the distribution of the elementary effects' absolute values. Morris' measures for the i th input are defined as:

$$\mu_i = \frac{1}{r} \sum_{j=1}^r EE_i^j \quad 2-6)$$

$$\sigma_i^2 = \frac{1}{r-1} \sum_{j=1}^r (EE_i^j - \mu)^2 \quad 2-7)$$

$$\mu_i^* = \frac{1}{r} \sum_{j=1}^r |EE_i^j| \quad (2-8)$$

Where r is the number of trajectories constructed from the sampling of X from Ω and EE_i^j indicates the elementary effects relative to the i^{th} factor computed along trajectory j . A qualitative parameter ranking based on the Morris technique was established as the first step of the AquaCrop model's SA. Morris analysis samples were generated within the defined sample using the Latin Hypercube sampling design. In practice, the number of trajectories r should be as large as possible, as larger trajectories values provide a more complete coverage of the parameter space and a lower uncertainty associated with the estimated sensitivity indices (Ciric et al., 2012). $r = 50$ trajectories were applied, which is the maximum value suggested by Campolongo et al. (2007) to guarantee a high coverage of the input space. When all 36 model parameters were evaluated, a total of 50 trajectories needed a maximum of $n = 1850$ model evaluations.

2.2.3.2 Fourier Amplitude Sensitivity Test (FAST)

First-order sensitivity indices measure the variance created by a single parameter in variance-based methods. Cukier et al. (1973) developed the Classical FAST method. FAST samples the input parameter space with a transformation function. The transformation function for FAST sampling is defined as:

$$x_i = 0.5 + \frac{1}{\pi} (\arcsin (\sin(w_i s + \rho_i))) k_i \quad (2-9)$$

where x_i is the i^{th} parameter, s is the sampling range with $s \in [-\pi, \pi]$, ρ_i is a random phase shift parameter marking the starting point of the search curve, ρ_i

$\in [0, 2\pi]$, w_i ($i = 1, \dots, k$) is the individual assigned integer frequency of parameter x_i , and k_i is a scaling factor to scale the value of the transformation function that lies between 0 and 1 to the appropriate parameter range. Since the transformation function is periodic, a sampling range of 2π is sufficient for decomposition. The first-order sensitivity indices (S_i) describe the major effects of parameters by calculating how the input variance contributes to the overall output variance. Saltelli et al. (1999) improved the FAST approach to generate first order (S_i) and total sensitivity indices (TS_i) named Extended FAST that does not only account for an input's variability but also the variance caused by other parameters. While there is no broad agreement on the appropriate cut-off value for parameter significance, first-order index value of 0.01 is commonly used to differentiate sensitive from insensitive parameters, i.e., parameters with smaller indices than 0.01 contribute less than 1% of the variation in the output and are thus deemed non-influential.

2.2.3.3 Sensitivity analysis strategy

A total of 36 AquaCrop parameters (Table 5) were identified as relevant for the calibration of the canopy cover, biomass, and yield predictions. The literature was reviewed to identify plausible ranges of the model parameters, then Morris method was used to screen out parameters with marginal effects and to identify a subset of relevant model parameters. Finally, the Fourier Amplitude Sensitivity Test (FAST) method was implemented to quantify both the first and higher order parameter sensitivities on above ground biomass (AGB), canopy cover (CC) and yield for wheat crops in the nine study fields. The sensitivity analysis was implemented in MATLAB (MathWorks Inc., 2020).

Table 5 List of the parameters from AquaCrop that were investigated in the sensitivity analysis with published lower and upper bounds where available and associated references.

Name	Definition	Unit	Lower bound	Upper bound	References
Input parameters					
gEme	Time from sowing/trans planting to emergence	Days/GD D's	63	114	Toumi et al. (2016)
gZx	Time from sowing/trans planting to maximum root development	Days/GD D's	484	616	Iqbal et al. (2014)
gYld	Time from sowing/trans planting to start of yield formation	Days/GD D's	600	780	- a
gSen	Time from sowing/trans planting to start of	Days/GD D's	800	1000	-

	canopy senescence				
gMat	Time from sowing/trans planting to physiological maturity	Days/GD D's	1400	1500	-
gFlwDur	Duration of flowering	Days/GD D's	100	200	Vanuytrecht et al. (2014)
gYldDur	Duration of yield formation	Days/GD D's	300	600	-
CCX	Maximum fractional canopy cover size	-	0.73	0.99	-
CDC	Canopy decline coefficient	GDD/day	0.004	0.0067	-
CGC	Canopy growth coefficient	GDD/day	0.005	0.0093	-
Zn	Minimum effective rooting depth	Metres	0.1	0.39	Toumi et al. (2016)

Zx	Maximum effective rooting depth	Metres		0.55	2.4	Toumi et al. (2016)
sZsp	Shape factor describing the decreasing speed of root expansion over time	-		10	19	Xing et al. (2017)
WxTopZ	Maximum water extraction at the top of the root zone	m ³ day ⁻¹	m ⁻³	0.0189	0.048	Xing et al. (2017)
WxBotZ	Maximum water extraction at the bottom of the root zone	m ³ day ⁻¹	m ⁻³	0.0056	0.012	Vanuytrecht et al. (2014)
cdAge	Decline of crop coefficient due to ageing of the canopy	% day ⁻¹		0.21	0.39	Xing et al. (2017)
Kcb	Maximum crop	-		0.77	1.43	Xing et al. (2017)

	coefficient when canopy is fully developed				
Wp	Water productivity normalised for reference evapotranspi ration and atmospheric carbon dioxide	g/m ²	11	22	Xing et al. (2017)
Wpy	Adjustment of water productivity parameter in yield formation stage	% of WP	75	125	Silvestro et al. (2017)
Hi0	Reference harvest index	-	0.32	0.61	Xing et al. (2017)
exF	Excess of potential fruits that is produced by the crop	%	70	130	Xing et al. (2017)
dxHi0	Maximum possible	%	10	19	Xing et al. (2017)

	increase in harvest index above reference value				
gMnB	Minimum number of GDD's required for full biomass production	GDD's	13	15	-
Aer	Water deficit below saturation at which aeration stress begins to occur	%	3.5	6.5	Xing et al. (2017)
pExpUp	Upper soil water depletion threshold for water stress effects on canopy expansion	-	0.14	0.26	Upreti et al. (2020)
pStmUp	Upper soil water depletion	-	0.455	0.84 5	Upreti et al. (2020)

	threshold for water stress effects on stomatal control				
pSenUp	Upper soil water depletion threshold for water stress effects on canopy senescence	-	0.49	0.91	Upreti et al. (2020)
pPolUp	Upper soil water depletion threshold for water stress effects on crop pollination	-	0.455	1	Upreti et al. (2020)
pExpLo	Lower soil water depletion threshold for water stress effects on canopy expansion	-	0.445	0.84 5	Upreti et al. (2020)

pStmLo	Lower soil water depletion threshold for water stress effects on stomatal control	-	0.7	1	Upreti et al. (2020)
pSenLo	Lower soil water depletion threshold for water stress effects on canopy senescence	-	0.7	1	Upreti et al. (2020)
pPolLo	Lower soil water depletion threshold for water stress effects on crop pollination	-	0.7	1	Upreti et al. (2020)
sExp	Shape factor describing water stress effects on	-	2.1	3.9	Upreti et al. (2020)

	canopy expansion				
sStm	Shape factor describing water stress effects on stomatal control	-	1.75	3.25	Upreti et al. (2020)
sSen	Shape factor describing water stress effects on canopy senescence	-	2.1	3.9	Upreti et al. (2020)
sPol	Shape factor describing water stress effects on crop pollination	-	0.7	1.3	Upreti et al. (2020)
Output responses					
Cc	Fractional canopy cover	-			
Bio	Accumulated aboveground biomass	Kg/ha			
Yield	Crop yield	ton/ha			

^aThese bounds were adjusted from the literature values to account for the fact we used a based temperature of 5°C (which is often adopted for Wheat grown in Northern Africa) as opposed to the more standard base temperature of 0°C.

2.2.4 Stepwise model parameter optimisation

Parcels A1 – A3 were used to calibrate the model and parcels B1 – B6 served as our validation set. A two-step calibration approach was adopted by first fitting the canopy cover and then the yield. Both steps use particle swarm optimization. Each optimization phase fits only the parameters that were found to be relevant for that phase in the sensitivity analysis. The PSO optimization started by randomly allocating the initial position and velocity of the optimizer. Parameters were allowed to take values within a range not exceeding 20% of the lower and upper bounds we found in the literature (see Table 5). Then, an executable MATLAB file was run to simulate the canopy cover from input data including weather, soil, crop, and management data and taking into account the effect of soil fertility stress on both canopy cover and yield outputs. The optimization searches the parameter space aiming to minimize the objective function which we defined as the sum of squared differences between the observed and simulated measures.

For the canopy cover fitting, we used two objective functions, one using the CC_{HP_i} canopy estimates and the other using CC_{L7_i} . The constructed objective functions are given by:

$$J_{HP} = \sum_{j=1}^p \sum_{i=1}^n (CC_{S_i} - CC_{HP_i})^2 \quad (2-10)$$

$$J_{L7} = \sum_{j=1}^p \sum_{i=1}^n (CC_{S_i} - CC_{L7_i})^2 \quad (2-11)$$

where n is the number of measurements, p is the number of calibration sites, CC_{S_i} is the simulated

canopy cover, CC_{HP_i} and CC_{L7_i} are the estimated canopy cover using the hemispherical photographs and Landsat7, respectively. We then took each of these parameter sets forward to the second stage of yield fitting where our objective functions were the sum of squared differences between model and measured yield. This resulted in two sets of model parameters. The stepwise model parameterization deliberately excluded biomass fitting as our objective was to create an easily transferable and resource-efficient approach. Recognizing the challenges associated with obtaining accurate biomass measurements, the focus was placed on canopy cover (CC) data from diverse sources, such as hemispherical photographs (HP) and Landsat 7 (L7). A comprehensive conceptual framework illustrating the described methodology is presented in Figure 8 .

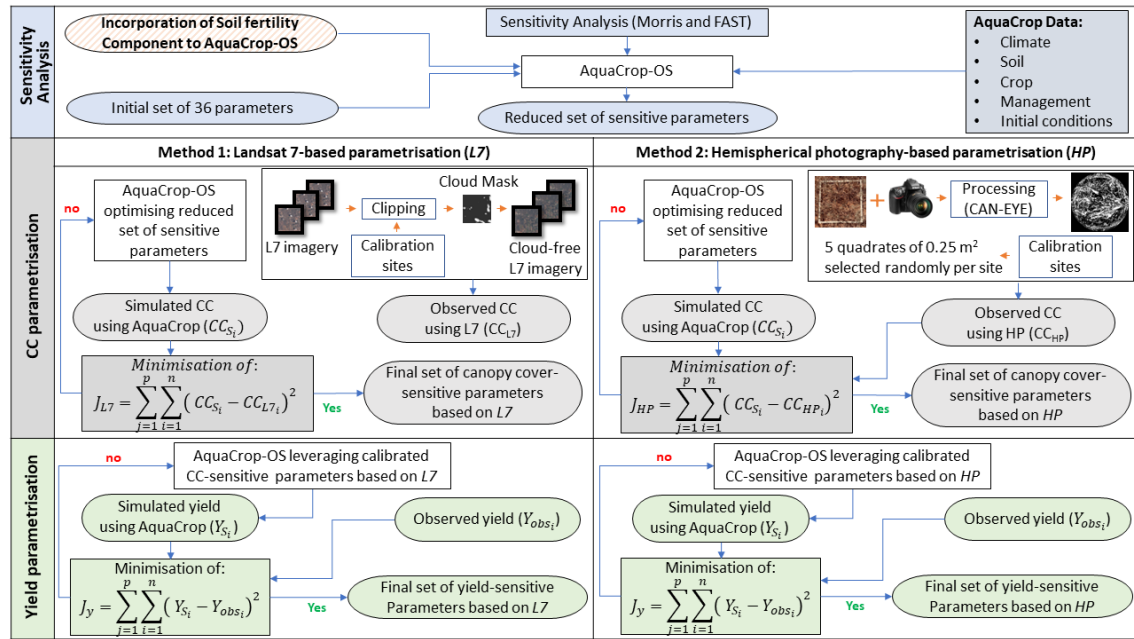


Figure 8 Conceptual framework of the sensitivity analysis and the stepwise parametrisation of AquaCrop.

2.2.5 Model evaluation

The performance of each parameter set was evaluated by predicting the responses for the six validation fields (B1, B2, B3, B4, B5 and B6) and then calculating four performance metrics. These were the coefficient of determination (R^2) which expresses the degree of collinearity between the observed and simulated data, the normalized root mean square error (NRMSE), the Nash and Sutcliffe modelling efficiency (EF) as a model performance indicator (Nash & Sutcliffe, 1970) and the Willmott's index of agreement (D-index). These are defined as follows:

$$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2 \quad 2-12)$$

$$NRMSE = \frac{100}{\bar{O}} \left[\frac{\sum_{i=1}^n (P_i - O_i)^2}{n} \right]^{0.5} \quad 2-13)$$

$$EF = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad 2-14)$$

$$D - index = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad 2-15)$$

where n is the number of measurements, P_i is the simulated value, O_i is the observed value, $\bar{O}$ is the observed mean value and $\bar{P}$ is the average of simulated data. A good agreement between measured and simulated data is indicated by NRMSE values near to zero. On the other hand, when the EF values are near to 1, the model's performance is considered satisfactory.

2.3 Results

2.3.1 Sensitivity analysis

2.3.1.1 Morris method

The five most influential parameters for the canopy cover, biomass and yield responses (as identified by the main effect μ^*) are ranked in Table 6. Only those with μ^* greater than 0.1 were ranked. The number of non-influential factors differs slightly from one field to another. Ten non-sensitive factors were identified for yield, 6 for aboveground biomass and 12 for canopy cover as shown in Table 7.

Table 6 The rank order of the mean of elementary effects (μ^*) for the Morris sensitivity analysis. Increasing ranks imply reduced sensitivity with the parameter ranked one the most sensitive (See table 5 for definitions).

Output response	Input parameters ^a	Parcels								
		A1	A2	A3	B1	B2	B3	B4	B5	B6
Yield	Wp	1	1	1	1	1	1	1	1	4
	Hi0	2	2	2	2	2	2	2	2	5
	gYld	3	5	- ^b	-	-	-	-	-	1
	CGC	4	4	4	5	4	4	4	3	-
	pExpLo	5	3	3	-	-	3	3	4	-
	Kcb	-	-	5	3	3	5	5	5	-
	Wpy	-	-	-	4	5	-	-	-	-
	gEme	-	-	-	-	-	-	-	-	2
	pStmLo	-	-	-	-	-	-	-	-	3
Biomass	Wp	1	3	3	1	1	1	2	1	1

	CGC	2	2	4	3	3	3	3	3	4
	CDC	5	-	-	4	4	-	-	4	-
	Kcb	4	4	-	2	2	5	5	5	5
	WxBotZ	-	-	-	-	-	4	4	-	3
	WxTopZ	-	5	1	-	-	-	-	-	-
	pSenUp	-	-	5	-	-	-	-	-	-
	gSen	-	-	-	5	5	-	-	-	-
	pExpLo	3	1	2	-	-	2	1	2	2
Canopy cover	CDC	2	1	4	1	1	2	2	1	-
	gSen	4	5	-	2	2	1	5	4	-
	gMat	3	4	2	4	4	-	3	3	5
	gEme	-	-	-	-	-	-	-	-	1
	CCX	-	-	-	5	-	-	-	-	-
	Zx	5	-	-	3	3	5	1	2	-
	WxTopZ	1	2	1	-	-	3	4	5	4
	WxBotZ	-	3	3	-	-	4	-	-	-
	pExpLo	-	-	5	-	-	-	-	-	-
	pStmLo	-	-	-	-	-	-	-	-	2
	pStmUp	-	-	-	-	-	-	-	-	3
	Zn	-	-	-	-	5	-	-	-	-

^a Most sensitive parameters for all parcels.

^b parameter is not ranked among the five most sensitive parameters.

The yield sensitivity analysis revealed that the most sensitive parameters for most of the fields were the normalized water productivity (Wp) and the reference harvest index ($Hi0$). Parameters related to root development (Zx , Zn , $WxTopZ$, $WxBotZ$), crop phenology ($gYld$), crop transpiration (Kcb) and water stress ($pPolLo$, $pExpLo$) exhibited high interactions between each other. Aboveground biomass was most sensitive to water productivity (Wp) and other parameters related to crop transpiration and development (Kcb , CGC). The Morris results for the canopy cover indicated that the parameters related to root development ($WxTopZ$, $WxBotZ$) were most sensitive for wheat grown in the first season. The canopy cover was most sensitive to crop development and phenology parameters (CDC , $gSen$) in the second growing season (2003–2004).

Table 7 Non-influential parameters that are common for all studied fields according to the Morris Analysis. Min and Max represent the minimum and maximum Morris mean of elementary effects (μ^*) in the studied parcels.

Output Response	Non-Influential Parameters	Min	Max
Yield	sPol	0.02	0.08
	pPolLo	0.03	0.09
	pPolUp	0.01	0.07
	Aer	0.05	0.08
	dxHi0	0.08	0.1
	sZsp	0.02	0.06
	sStm	0.04	0.09
	gZx	0.03	0.07
	Zn	0.01	0.05
	exF	0.06	0.09
Aboveground Biomass	sPol	0.02	0.05
	pPolUp	0.04	0.08
	pPolLo	0.01	0.06

	exF	0.03	0.07
	Aer	0.07	0.09
	dxHi0	0.05	0.08
Canopy Cover	dxHi0	0.04	0.06
	Wpy	0.03	0.09
	Wp	0.02	0.08
	Hi0	0.01	0.07
	exF	0.06	0.09
	sPol	0.02	0.08
	cdAge	0.03	0.09
	pPolUp	0.01	0.05
	pPolLo	0.04	0.08
	Aer	0.05	0.07
	sStm	0.03	0.06
	gYldDur	0.07	0.09

2.3.1.2 FAST method

First order sensitivity indices were calculated for the remaining input parameters from the Morris step using FAST method. Figure 9 illustrates the main effect sensitivity index time series for CC in days after sowing. The results show that *gEme*, *gMat*, *CGC* and *WxBotz* were the most sensitive parameters for the studied parcels. The canopy growth coefficient (*CGC*) and the time from sowing to emergence (*gEme*) were sensitive at the beginning of the season while the canopy decline coefficient (*CDC*), the time from sowing to maturity and senescence (i.e., *gMat*, *gSen*) were more influential at the end of the season. Finally, the most water stress sensitive parameters in mid-season were related to the stomatal control (*pStmLo*), while the canopy expansion and senescence water stress parameters (i.e., *pExpLo*, *pSenUp*) were more sensitive at the beginning of the season.

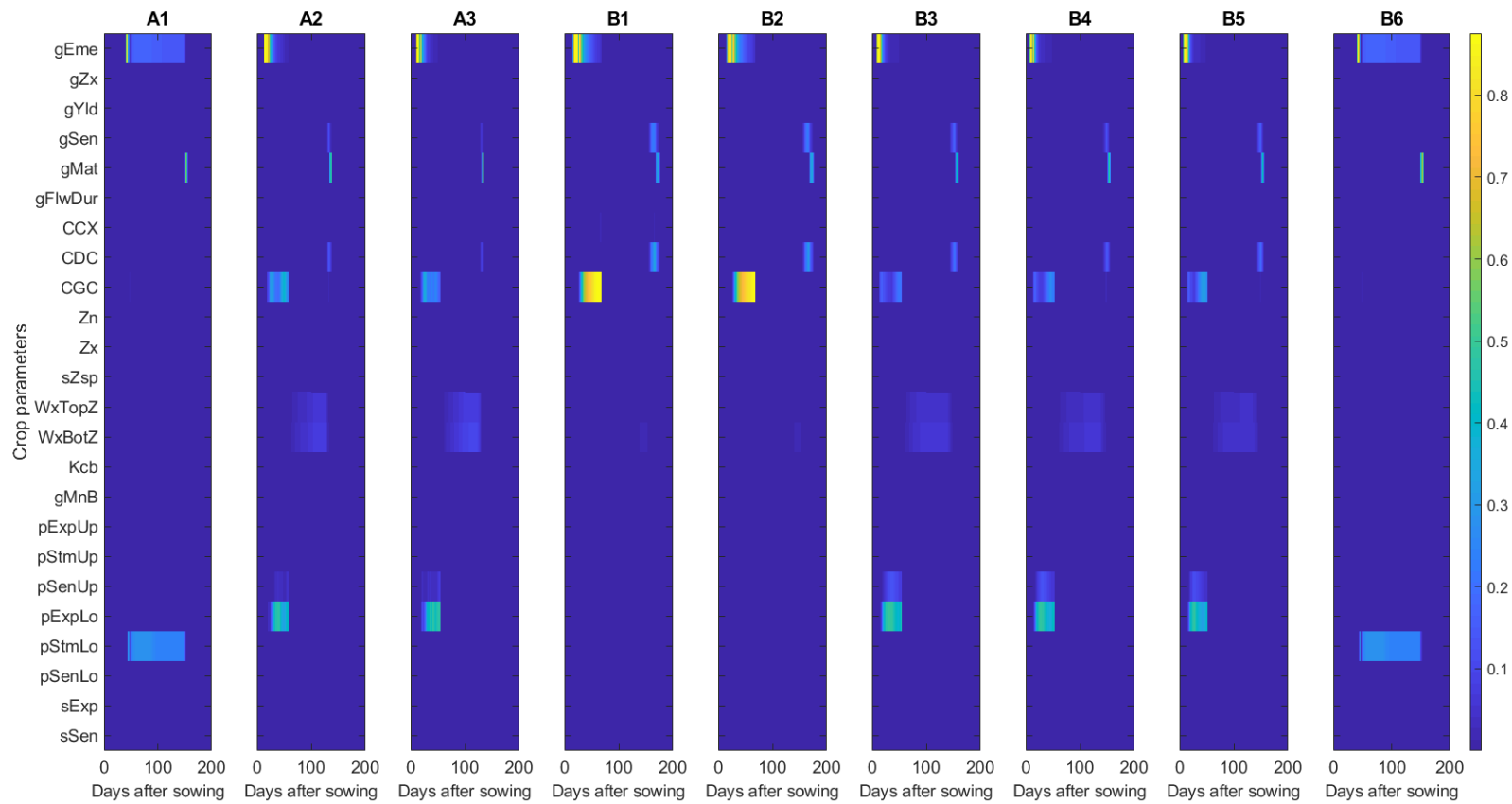


Figure 9 FAST analysis results for the canopy cover time series. Parcels B1 and B2, sown earlier, exhibited lower water stress due to additional rainfall.

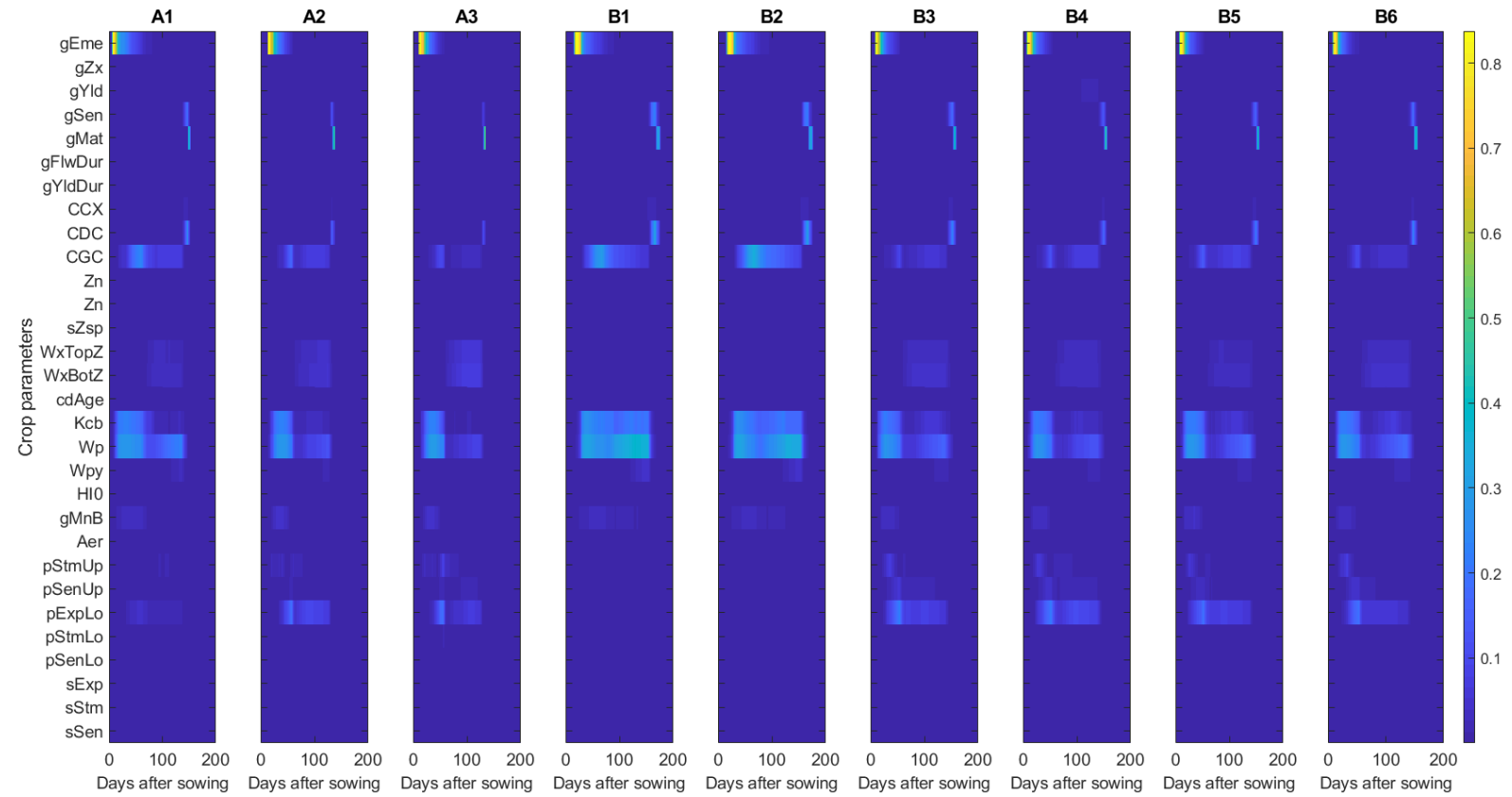


Figure 10 FAST analysis results for the aboveground biomass time series. Parcels B1 and B2, sown earlier, exhibited lower water stress due to additional rainfall.

Aboveground biomass FAST analysis findings are displayed in Figure 10. The results showed that the crop phenological parameters (i.e., *gEme*, *gMat*, *gSen*), the canopy development parameters (i.e., *CGC*, *CDC*, *CCX*, *Kcb*) and the biomass production parameters (i.e., *Wp*, *Wpy*) were the common sensitive parameters for the investigated parcels. The time from sowing to emergence (*gEme*) was sensitive at the beginning of the season and its sensitivity decreased gradually over time. Meanwhile, the time from sowing to both senescence and maturity was only sensitive at the end of the season. The parameters related to the canopy expansion and the biomass production (i.e., *CGC*, *Wp*) were influential during the whole season period. While the canopy decline coefficient (*CDC*) and the maximum water extraction at top and bottom of the root zone (*WxTopZ*, *WxBotZ*) were mostly influential at the end and the mid-season respectively.

The SA results for yield (Figure 11) showed that the most sensitive parameters were the normalized water productivity (*Wp*), the reference harvest index (*Hi0*), the maximum canopy cover coefficient (*Kcb*), the Adjustment of water productivity parameter in yield formation stage (*Wpy*), the canopy growth coefficient (*CGC*) and the Lower soil water depletion threshold for water stress effects on canopy expansion (*pExpLo*). Overall, the yield analysis showed that the reference harvest index (*Hi0*) and the normalized water productivity (*Wp*) were sensitive for all parcels.

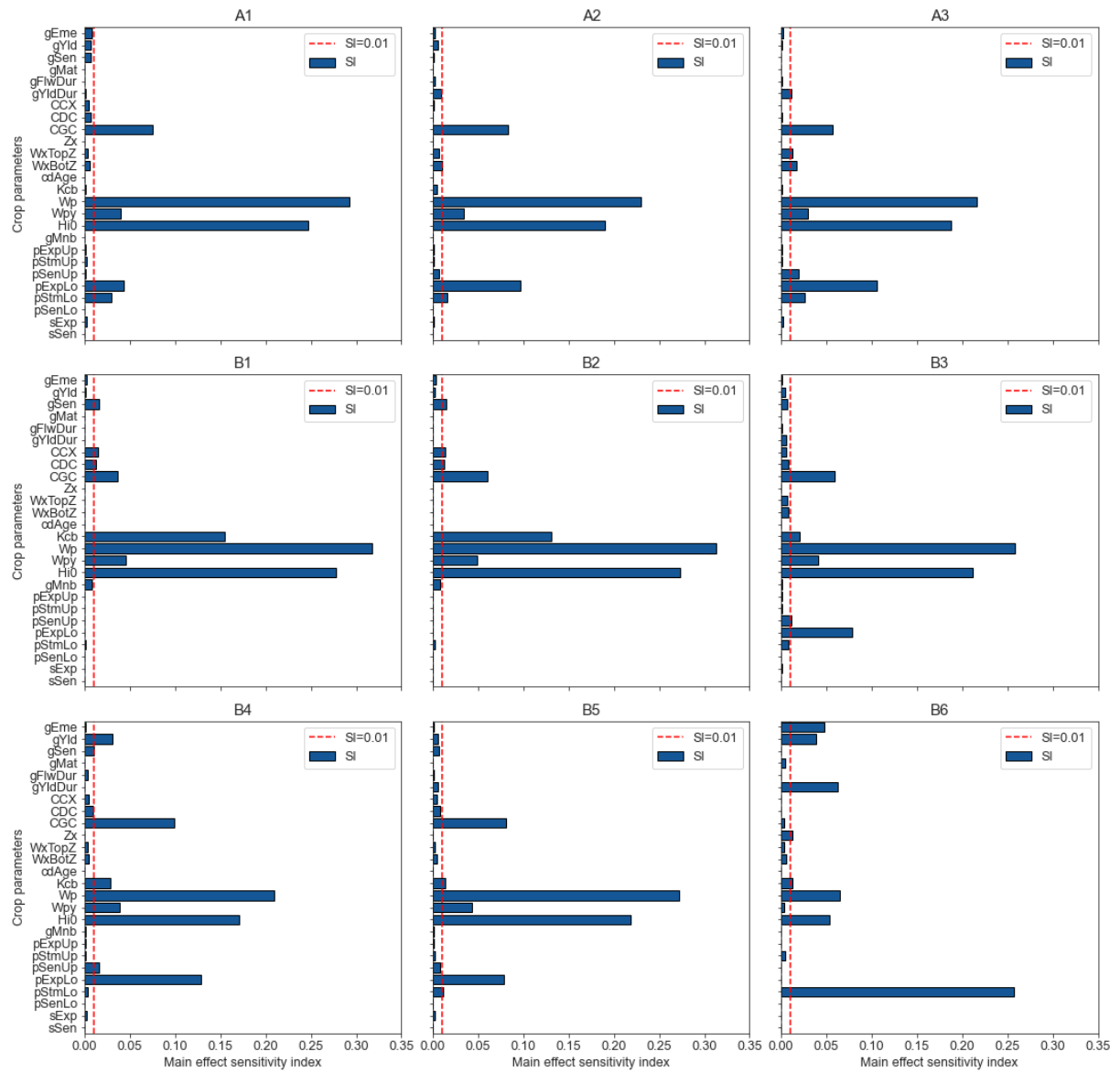


Figure 11 FAST analysis results for the yield with cut-off value (horizontal red dashed line).

2.3.2 Optimisation

2.3.2.1 Canopy cover

For the calibration set (A1 – A3), simulated and measured CC that resulted from the two fits (one based on hemispherical photography, the other on remote sensing) were in good agreement (Table 8). When CC was simulated based on hemispherical photography predictions (PSOHP), the estimated canopy cover CCHP corresponded extremely well with the observed CC with R^2 values >0.98 , low estimation errors (6.39 % NRMSE), high D-index (>0.99), and D-index (>0.98). The remote sensing-based optimization (PSOL7) in the calibration set revealed a similar but less precise match between observed and simulated canopy cover values compared to PSOHP with an R^2 , NRMSE, D-index and EF of 0.72, 28.97 %, 0.9 and 0.71 respectively.

The observed canopy cover from Landsat7 seems to be higher than that derived from hemispherical imagery at the beginning of the season as shown in Figure 12. This induced PSOL7 to have an earlier crop emergence and a higher canopy growth coefficient value ($gEme= 47.25$, $CGC=0.011$) than PSOHP ($gEme= 98.56$, $CGC= 0.01$) as indicated in Appendix B, Table_Apx B.2-1. In addition, a strong canopy decline coefficient ($CDC= 0.0049$) was identified in PSOHP, resulted in a rapid drop of the canopy and lesser growing degree days to reach the maturity ($gMat = 1472.10$), while a lesser CDC ($CDC = 0.004$) was observed in PSOL7, resulted in a delayed maturity of the crop ($gMat = 1500$). For complete model parametrization sets, see Table_Apx B.2-1 in Appendix B.

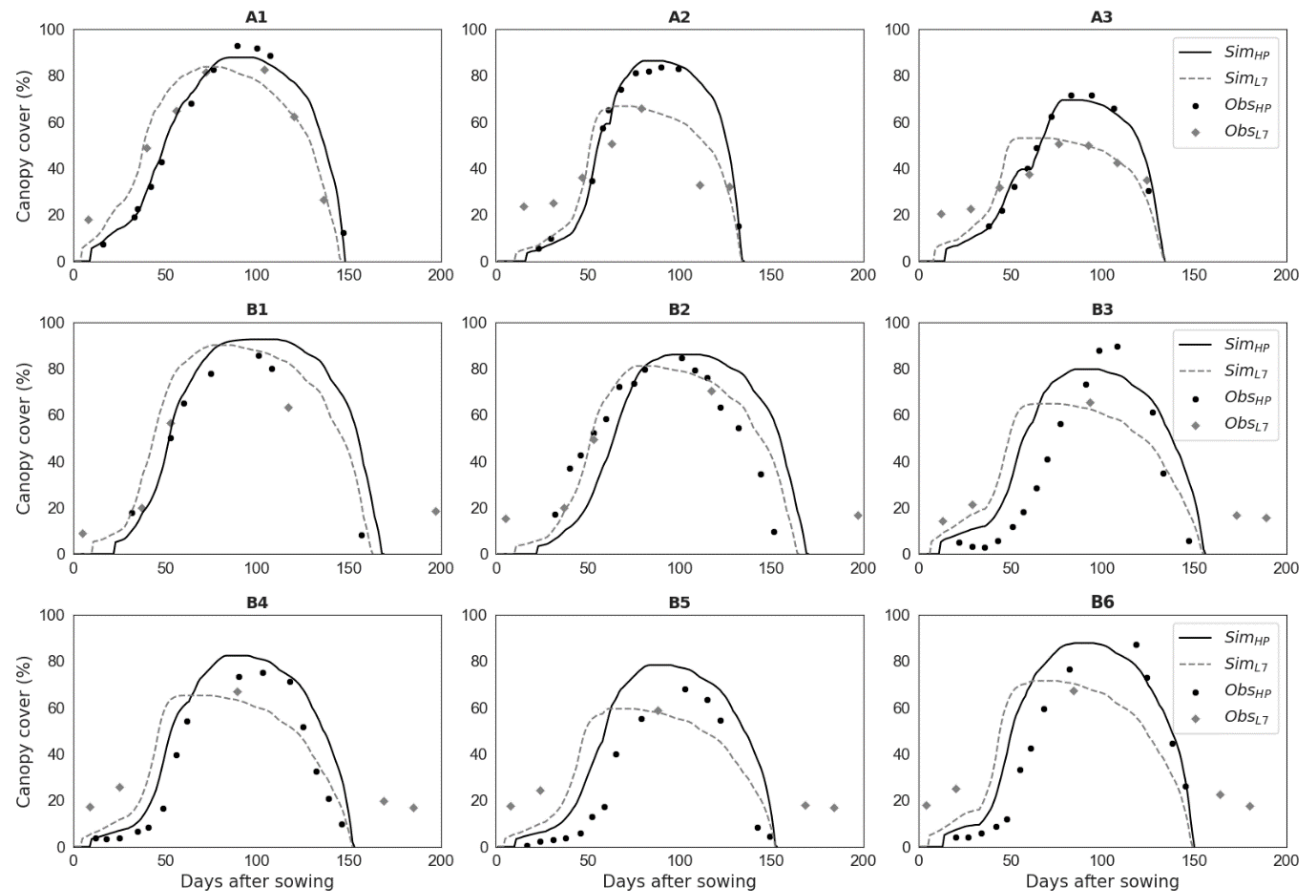


Figure 12 Comparison of model simulated canopy cover using hemispherical photography, Landsat7 and field observed canopy cover using hemispherical photography and Landsat7.

The validation set revealed that canopy cover predicted using hemispherical photographs led to better model parametrization compared to Landsat7. Overall, the observed and simulated canopy cover were in good agreement, as shown by low NRMSE, and high R², D-index and EF values in Table 8.

Table 8 Model performance indicators comparing observed and simulated canopy cover.

Performance indicator	Hemispherical photography		Landsat7	
	Calibration	Validation	Calibration	Validation
R ²	0.98	0.76	0.72	0.62
NRMSE (%)	6.39	45.87	28.97	51.37
D-index	0.99	0.90	0.90	0.86
EF	0.98	0.63	0.71	0.54

2.3.2.2 Yield

Winter wheat yields measured in the field experiments ranged from 0.9 t/ha to 3.26 t/ha, while simulated values using hemispherical photographs and remote sensing parametrizations ranged from 1.14 t/ha and 1.29 t/ha to 3.9 t/ha and 3.8 t/ha, respectively as shown in Figure 13. The R² between measured and simulated yield using PSO_{HP} and PSO_{L7} in the calibration dataset was equal to 0.89 and 0.99 (Table 9). Overall, the application of PSO with hemispherical and Landsat7 extracted canopy cover simulated the yield with low estimation errors and high efficiency in the calibration set. In the validation set, hemispherical photography-based parametrization produced better yield estimates (R²=0.44, D-index= 0.75, EF=0.04) and lower estimation errors (NRMSE= 25.4%) compared to the other approach (i.e., PSO_{L7}). It is important to note that while simulating

yield using PSO_{L7} , a negative EF was found, indicating that the mean of the observations gives a better prediction than the model.

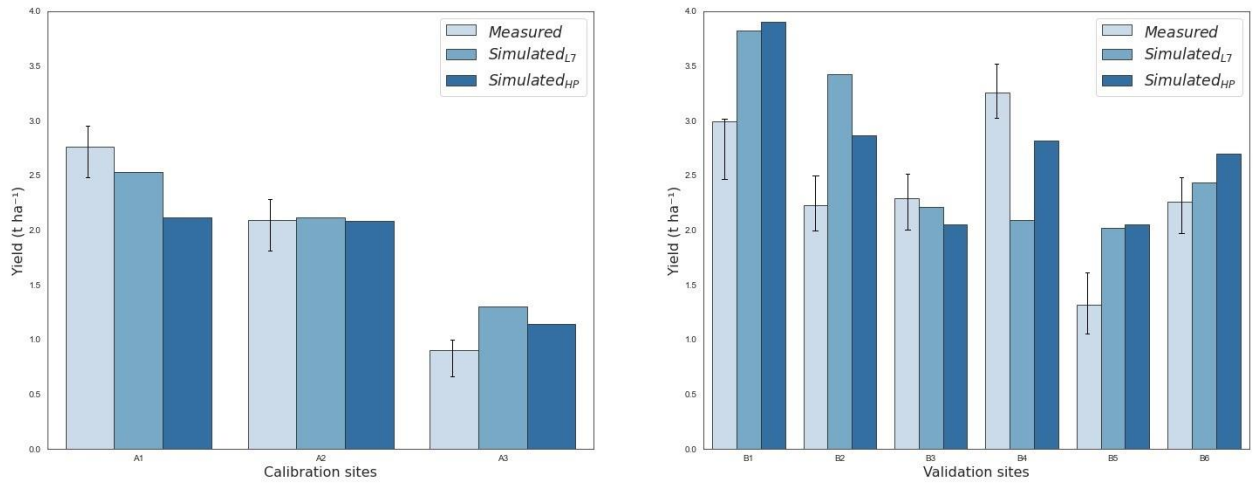


Figure 13 Comparison of model simulated yield using hemispherical photography, Landsat7 and field observed yield in the calibration sites (left) and the validation sites (right).

Table 9 Model performance indicators comparing observed and simulated yield.

Performance indicator	Hemispherical photography		Landsat7	
	Calibration	Validation	Calibration	Validation
R²	0.89	0.44	0.99	0.10
NRMSE (%)	20.63	25.4	13.86	34.23
D-index	0.89	0.75	0.95	0.55
EF	0.73	0.04	0.88	-0.73

2.4 Discussion

2.4.1 Insights from Morris and FAST sensitivity analysis of AquaCrop parameters

Many research studies have calibrated AquaCrop using field data (López-Urrea et al., 2020; Sodari et al., 2021). However, most of them have fine-tuned the parameters using trial and error. We present a practical method for parameterizing AquaCrop using a two-step optimization; first by selecting the set of the influential parameters then fitting the canopy cover and then the yield.

To inform our optimisation, we used Morris and FAST sensitivity analyses to determine the most influential parameters of AquaCrop. Many studies have demonstrated that just a few parameters out of the model's large number have a substantial influence on the model's output (DeJonge et al., 2012; Sarrazin et al., 2016). DeJonge et al. (2012) performed both Morris and Sobol variance-based methods on CERES-Maize input parameters affecting the crop growth and found that while considering the anthesis day after planting (ADAY) output, only two out of the model parameters were sensitive for both full and limited irrigation treatments. Our results agree with this observation. We discovered that only 24 out of 36 model parameters influenced the canopy cover significantly and only 26 for the yield. Moreover, our results showed that Morris and FAST do not always agree on the priority ranking of a parameter. As an example, Morris prioritized the time from sowing to start of yield formation (*gYld*) for yield prediction, but the lower soil water depletion threshold for water stress effects on stomatal control (*pStmlo*) was found to be the most influential parameter when using the FAST method for the parcel B6. This finding aligns with Confalonieri et al. (2010) who analysed different sensitivity analysis outcomes of the rice crop model WARM and concluded that in nonlinear models or when the number of model trajectories is small, Morris versus variance-based approaches may order parameters differently. They concluded that the simplest approach (i.e., Morris) achieved equivalent outcomes to those reached by computationally intensive methods. We also found that the sensitive factors were generally consistent across the fields

considered in our study. There were a few exceptions. Water stress factors ($pStmUp$, $pSenUp$, $pExpLo$) had no effect on simulating canopy cover in fields B1 and B2. These fields were sown early and so it is likely that they benefited from early rainfall mitigating effects of water stress. This has been reported in other studies including Xing et al. (2017) who reported that $pSenUp$ was more sensitive in plants under water stress scenarios. Data limitations must also be considered interpreting these results, The dataset used in this study was limited to semi-arid and arid regions of Morocco and focused on specific field conditions during the 2002–2003 and 2003–2004 growing seasons. These geographic and climatic constraints mean that while the parameter prioritisation observed here may hold for similar regions, they may not apply to areas with markedly different agro-environmental conditions, such as tropical or temperate zones. Furthermore, variations in management practices, such as irrigation or sowing timings, across region can significantly alter the sensitivity of certain parameters, as observed in fields B1 and B2. Future research should focus incorporate datasets from a wider range of climatic zones to validate the consistency of these findings across diverse environmental settings.

2.4.2 Impact of the source of canopy cover on the predictive performance of AquaCrop

Investigating the impact of various sources of canopy measures on prediction accuracy and considering the practical implications of sourcing these measures are critical aspects in understanding the influence of different parametrization methods on AquaCrop's predictions. Notably, good agreement was found between the simulated and observed canopy cover when using Landsat7 parametrization (PSO_{L7}), although it was less accurate when compared to the hemispherical photography-based parametrizations (PSO_{HP}). This could be explained by the effect of soil brightness on NDVI at low vegetative cover, which leads to an overestimation of the canopy cover at the beginning of the growing season, hence predicting an accelerated crop emergence (47.25 GDDs when simulating with CC_{L7} against 98.56 GDDs when simulating with CC_{HP}). Another aspect in favour of hemispherical photography parametrization is the saturation

effect of satellite-based products and particularly NDVI at high vegetative cover in our scenario. Consequently, the simulated canopy cover tended to be underestimated at high vegetation density at mid-season, resulting in a shorter time period for the crop to reach senescence compared to the other type of parametrization ($gSen_{L7}=610.99$, $gSen_{HP}=817.6$). Another key limitation with the Landsat 7 was the numbers of observations available across the season. At the time of the study passes occurred every 16 days. In the past five years there has been a rapid increase in the number of EO satellites, now medium resolution imagery (10 to 30 m pixels) is freely available every few days (J. Li & Chen, 2020). Other studies have explored the effect of the source of canopy cover measures on the accuracy of predictions. For example, Ma et al. (2017) considered the reliability of canopy cover retrieval from three different data types (LIDAR, aerial photography, and spaceborne imaging) and found that spaceborne WorldView2 (1.84 m pixel) data produced the highest accuracy ($R^2 = 0.58$), followed by aerial imagery ($R^2= 0.50$) and LIDAR data ($R^2= 0.33$). Jin et al. (2020) conducted a similar method of parameter optimization with PSO to determine the yield of maize in China using hyperspectral data from a spectrometer. They demonstrated that the enhanced vegetation index (EVI) could be used to estimate CC more accurately than other vegetation indices such as NDVI ($R^2 = 0.78$ and $RMSE = 9.84\%$). In a separate study, Jin et al. (2017) compared the results of the assimilation of optical and radar imaging data in China using the PSO algorithm. They found that the estimated CC obtained from RADARSAT-2 data outperformed HJ-1A/B optical data, indicating that the source of canopy cover data is a crucial factor affecting the model parametrizations.

The empirical relationship between CC and NDVI is widely used (Gamon et al., 1995; Tsakmakis et al., 2021; Yang et al., 2017). However, its accuracy is subject to uncertainties related to factors, covering atmospheric conditions (Agapiou et al., 2011), sunlight angle (Ishihara et al., 2015), soil properties (Prudnikova et al., 2019), and vegetation types (Huemmrich et al., 2021; Zou & Möttus, 2017). Despite the goodness of fit of the NDVI-CC relationship ($R^2 =$

0.71), these uncertainties may in part explain the poor performance of the L7 parametrisation as compared to the hemispherical-based parametrisation.

The accuracy of AquaCrop's yield predictions is reported to be strongly linked with the quality of the canopy cover simulations (Hsiao et al., 2009). This is consistent with our findings; our parametrization based on hemispherical photography data produced improved yield prediction results (Table 9). The obvious improvements in prediction accuracy from using the hemispherical photography-based compared with using Landsat 7 come at a cost. Hemispherical photographs required human intervention in the field (in our case eleven times per season). For each field, many photographs were taken to get an estimate of the mean canopy cover. As a result, the time and effort required grows proportionally with the number of taken photographs. Finally, the ability to use hemispherical photography is partially limited by the weather, since measurements may be missed due to rain, wind, or inaccessibility to the research site. Thus, coupling canopy cover observations extracted from both satellite imagery and hemispherical photographs is highly recommended to alleviate some of the weather-related constraints and subsequently improve yield simulations.

The parameterization approaches evaluated in this study are bounded not only by the data sources and environmental conditions considered but also by the inherent assumptions of the AquaCrop model. The use of NDVI from Landsat 7, while widely accessible and historically valuable, is constrained by different limitations. Spectral saturation at high vegetation density reduces the reliability of NDVI as a proxy for canopy cover, as it underestimates canopy cover at mid to late growth stages. This saturation effect is particularly present in dense cropping systems, where high biomass accumulation occurs, making NDVI less effective in regions with similar conditions. Soil brightness effects, especially in early growth stages, can lead to overestimations of canopy cover in sparsely vegetated areas. These inaccuracies are related to the sensitivity of NDVI to background reflectance, which is influenced by soil type, moisture content, and surface conditions. Furthermore, the limited temporal coverage of Landsat 7, with a 16-

day revisit period, often results in data gaps during critical phenological stages, especially in areas with frequent cloud cover or short growing seasons. These temporal gaps limit the ability to capture rapid changes in canopy development, reducing the accuracy of dynamic model simulations. Similarly, the reliance on hemispherical photography, while offering high accuracy, introduces significant logistical and operational challenges. The manual nature of this method necessitates frequent field visits, which may be infeasible in large-scale studies or regions with limited accessibility. Weather conditions, such as rain or wind, can further restrict data collection opportunities, leading to potential inconsistencies in measurement timing and quality. In addition to these data-related limitations, AquaCrop assumes uniformity within a field, ignoring heterogeneity in soil properties, microclimates, or pest pressures, which can introduce errors when applying the model to larger or more diverse landscapes. These simplifications, while making the model computationally efficient, limit its applicability in complex or highly variable environments. Addressing these boundary conditions requires the adoption of more advanced and flexible data acquisition and modeling strategies. For instance, multi-sensor data fusion that combines radar and hyperspectral imagery could mitigate the limitations of NDVI. Radar imagery is less sensitive to atmospheric conditions and can provide reliable canopy structure information even under dense vegetation or cloudy conditions. Hyperspectral data, with its finer spectral resolution, can capture subtle variations in vegetation traits, offering improved accuracy in estimating canopy properties and addressing saturation issues. Similarly, incorporating higher-frequency satellite datasets, such as Sentinel-2 with a 5-day revisit cycle, can reduce temporal gaps, enabling more consistent monitoring of phenological changes. Finally, integrating AquaCrop with machine learning approaches or data assimilation techniques could enhance its capacity to handle nonlinear dynamics and account for spatial heterogeneity. Such advancements would not only improve the robustness and accuracy of AquaCrop parameterization but also extend its applicability to a broader range of agro-ecological zones and operational contexts.

2.5 Conclusions

In this study, we carried out a two-stage sensitivity analysis (Morris and FAST) of crop canopy cover and yield for the FAO-AquaCrop model, in which we incorporated algorithms to account for the impact of soil fertility on canopy cover growth, maximum canopy cover, and biomass water productivity. The sensitivity analysis results revealed that canopy cover predictions were most sensitive to parameters related to emergence and maturity timing as well as the canopy growth rate. For our study area, the influential parameters were generally consistent across seasons and fields with the exception of those related to water stress. This shows the importance of selecting sensitive parameters for a site based on data from several seasons or variation in sowing time. Our findings showed that the hemispherical photography-based parameterization outperformed the satellite-based parameterization for canopy cover and yield simulations using AquaCrop. Although simulations of the Landsat 7 observed canopy were reasonable in the validation set, they were not as accurate as those from hemispherical photographs and the associated yield predictions were particularly poor. This highlights the importance of accurate canopy cover predictions in estimating final yield. Overall, our research demonstrates the usefulness and efficiency of the proposed optimization procedure in capturing the full range of model parameters using different source of data while accommodating the unique characterises of the study area. Further research may be needed to examine the potential Improvements in prediction that could be achieved by integrating different types of canopy cover data (Hybrid models, portable instruments, UAV, Radar, and optical data) or by updating the model predictions in real-time as new observations become available.

CRedit authorship contribution statement

Bader Oulaid: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alice E. Milne:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Toby Waine:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Rafiq El Alami:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Maryam Rafiqi:** Resources, Project administration, Methodology, Conceptualization. **Ron Corstanje:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

2.6 References

- A, A., E, H., A, A., & S, D. P. (2021). E-Farm Management and Yield Prediction. *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 1–6. <https://doi.org/10.1109/ICIRCA51532.2021.9544759>
- Abdelmajid, S., Mukhtar, A., Baig, M. B., & Reed, M. R. (2021). Climate change, agricultural policy and food security in Morocco. In *Emerging Challenges to Food Production and Security in Asia, Middle East, and Africa: Climate Risks and Resource Scarcity* (pp. 171–196). Springer International Publishing. https://doi.org/10.1007/978-3-030-72987-5_7
- Abi Saab, M. T., el Alam, R., Jomaa, I., Skaf, S., Fahed, S., Albrizio, R., & Todorovic, M. (2021). Coupling remote sensing data and aquacrop model for simulation of winter wheat growth under rainfed and irrigated conditions in a mediterranean environment. *Agronomy*, 11(11). <https://doi.org/10.3390/agronomy11112265>
- Adnan, A. A., Jibrin, J. M., Kamara, A. Y., Abdulrahman, B. L., Shaibu, A. S., & Garba, I. I. (2017). CERES–Maize model for determining the optimum planting dates of early maturing maize varieties in Northern Nigeria. *Frontiers in Plant Science*, 8. <https://doi.org/10.3389/fpls.2017.01118>
- Agapiou, A., Hadjimitsis, D. G., Papoutsas, C., Alexakis, D. D., & Papadavid, G. (2011). The Importance of accounting for atmospheric effects in the application of NDVI and interpretation of satellite imagery supporting archaeological research: The case studies of Palaepaphos and Nea Paphos sites in Cyprus. *Remote Sensing*, 3(12), 2605–2629. <https://doi.org/10.3390/rs3122605>
- Ait El Haj, F., Ouadif, L., & Akhssas, A. (2023). Monitoring Land Use and Land Cover Changes Using Remote Sensing Techniques and the Precipitation-Vegetation Indexes in Morocco. *Ecological Engineering and Environmental Technology*, 24(1), 272–286. <https://doi.org/10.12912/27197050/154937>

- Ali, A. M., Abouelghar, M., Belal, A. A., Saleh, N., Yones, M., Selim, A. I., Amin, M. E. S., Elwesemy, A., Kucher, D. E., Maginan, S., & Savin, I. (2022a). Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). *The Egyptian Journal of Remote Sensing and Space Science*, 25(3), 711–716. <https://doi.org/https://doi.org/10.1016/j.ejrs.2022.04.006>
- Ali, A. M., Abouelghar, M., Belal, A. A., Saleh, N., Yones, M., Selim, A. I., Amin, M. E. S., Elwesemy, A., Kucher, D. E., Maginan, S., & Savin, I. (2022b). Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). *The Egyptian Journal of Remote Sensing and Space Science*, 25(3), 711–716. <https://doi.org/https://doi.org/10.1016/j.ejrs.2022.04.006>
- Al-Shammari, D., Whelan, B. M., Wang, C., Bramley, R. G. V., Fajardo, M., & Bishop, T. F. A. (2021). Impact of spatial resolution on the quality of crop yield predictions for site-specific crop management. *Agricultural and Forest Meteorology*, 310. <https://doi.org/10.1016/j.agrformet.2021.108622>
- Amiri, E., Irmak, S., & Tari, D. B. (2024). Comparison of the AquaCrop and CERES-Maize models for simulating maize phenology, grain yield, evapotranspiration and water productivity under different irrigation and nitrogen levels. *Irrigation and Drainage*. <https://doi.org/10.1002/ird.2917>
- Amiri, N., Lahlali, R., Amiri, S., Jarroudi, M. E. L., Khebiza, M. Y., & Messouli, M. (2021). Development of an integrated model to assess the impact of agricultural practices and land use on agricultural production in morocco under climate stress over the next twenty years. *Sustainability (Switzerland)*, 13(21). <https://doi.org/10.3390/su132111943>
- Ashourloo, D., Nematollahi, H., Huete, A., Aghighi, H., Azadbakht, M., Shahrabi, H. S., & Goodarzashti, S. (2022). A new phenology-based method for mapping wheat and barley using time-series of Sentinel-2 images. *Remote Sensing of Environment*, 280, 113206. <https://doi.org/https://doi.org/10.1016/j.rse.2022.113206>

- Bai, T., Zhang, N., Mercatoris, B., & Chen, Y. (2019). Improving jujube fruit tree yield estimation at the field scale by assimilating a single Landsat remotely-sensed LAI into the WOFOST model. *Remote Sensing*, 11(9). <https://doi.org/10.3390/rs11091119>
- Bandara, W. B. M. A. C., Sakai, K., Nakandakari, T., Kapetch, P., Anan, M., Nakamura, S., Setouchi, H., & Rathnappriya, R. H. K. (2021). Global optimization of cultivar trait parameters in the simulation of sugarcane phenology using gaussian process emulation. *Agronomy*, 11(7). <https://doi.org/10.3390/agronomy11071379>
- Bandaru, V., Yaramasu, R., PNVR, K., He, J., Fernando, S., Sahajpal, R., Wardlow, B. D., Suyker, A., & Justice, C. (2020). PhenoCrop: An integrated satellite-based framework to estimate physiological growth stages of corn and soybeans. *International Journal of Applied Earth Observation and Geoinformation*, 92. <https://doi.org/10.1016/j.jag.2020.102188>
- Barbedo, J. G. A. (2022). Data Fusion in Agriculture: Resolving Ambiguities and Closing Data Gaps. In *Sensors* (Vol. 22, Issue 6). MDPI. <https://doi.org/10.3390/s22062285>
- Bassu, S., Asseng, S., Motzo, R., & Giunta, F. (2009). Optimising sowing date of durum wheat in a variable Mediterranean environment. *Field Crops Research*, 111(1–2), 109–118. <https://doi.org/10.1016/j.fcr.2008.11.002>
- Belmahi, M., Hanchane, M., Mahjoub, A., Najjari, F., Khayati, A., & Kessabi, R. (2023). Sustainability assessment of the main cereals market in Morocco: Evaluating production and import. *European Journal of Sustainable Development*. <https://api.semanticscholar.org/CorpusID:259255834>
- Bertassello, L., Winters, P., & Müller, M. F. (2023). Access to global wheat reserves determines country-level vulnerability to conflict-induced Ukrainian wheat supply disruption. *Nature Food*, 4(8), 673–676. <https://doi.org/10.1038/s43016-023-00806-w>

- Bertino, L., Evensen, G., & Wackernagel, H. (2003). Sequential Data Assimilation Techniques in Oceanography. *International Statistical Review*, 71(2), 223–241. <https://doi.org/https://doi.org/10.1111/j.1751-5823.2003.tb00194.x>
- Beveridge, L., Whitfield, S., & Challinor, A. (2018). Crop modelling: towards locally relevant and climate-informed adaptation. *Climatic Change*, 147(3–4), 475–489. <https://doi.org/10.1007/s10584-018-2160-z>
- Bhullar, A., Nadeem, K., & Ali, R. A. (2023). Simultaneous multi-crop land suitability prediction from remote sensing data using semi-supervised learning. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-33840-6>
- Bischiniotis, K., van den Hurk, B., Coughlan de Perez, E., Veldkamp, T., Nobre, G. G., & Aerts, J. (2019). Assessing time, cost and quality trade-offs in forecast-based action for floods. *International Journal of Disaster Risk Reduction*, 40. <https://doi.org/10.1016/j.ijdr.2019.101252>
- Boote, K. J., Jones, J. W., & Pickering, N. B. (1996). Potential Uses and Limitations of Crop Models. *Agronomy Journal*, 88(5), 704–716. <https://doi.org/https://doi.org/10.2134/agronj1996.00021962008800050005x>
- Bouras, E., Jarlan, L., Khabba, S., Er-Raki, S., Dezetter, A., Sghir, F., & Tramblay, Y. (2019). Assessing the impact of global climate changes on irrigated wheat yields and water requirements in a semi-arid environment of Morocco. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-55251-2>
- Breunig, M., Kuper, P., Reitze, F., Landgraf, S., Al-Doori, M., Stefanakis, E., Abdulmuttalib, H., & Kugler, Z. (2023). IMPROVING DATA QUALITY AND MANAGEMENT FOR REMOTE SENSING ANALYSIS: USE-CASES AND EMERGING RESEARCH QUESTIONS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10(1-

W1-2023), 41–49. <https://doi.org/10.5194/isprs-annals-X-1-W1-2023-41-2023>

Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, 22(10), 1509–1518. <https://doi.org/10.1016/j.envsoft.2006.10.004>

Casa, R., Varella, H., Buis, S., Guérif, M., De Solan, B., & Baret, F. (2012a). 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), 1–10. <https://doi.org/10.1016/j.eja.2011.09.004>

Casa, R., Varella, H., Buis, S., Guérif, M., De Solan, B., & Baret, F. (2012b). 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), 1–10. <https://doi.org/10.1016/j.eja.2011.09.004>

Castañeda-Vera, A., Leffelaar, P. A., Álvaro-Fuentes, J., Cantero-Martínez, C., & Mínguez, M. I. (2015). Selecting crop models for decision making in wheat insurance. *European Journal of Agronomy*, 68, 97–116. <https://doi.org/https://doi.org/10.1016/j.eja.2015.04.008>

Chang, Y., Latham, J., Licht, M., & Wang, L. (2023a). A data-driven crop model for maize yield prediction. *Communications Biology*, 6(1). <https://doi.org/10.1038/s42003-023-04833-y>

Chang, Y., Latham, J., Licht, M., & Wang, L. (2023b). A data-driven crop model for maize yield prediction. *Communications Biology*, 6(1). <https://doi.org/10.1038/s42003-023-04833-y>

Chapagain, R., Remenyi, T. A., Harris, R. M. B., Mohammed, C. L., Huth, N., Wallach, D., Rezaei, E. E., & Ojeda, J. J. (2022). Decomposing crop model

- uncertainty: A systematic review. In *Field Crops Research* (Vol. 279). Elsevier B.V. <https://doi.org/10.1016/j.fcr.2022.108448>
- Chen, Y., & Tao, F. (2022). Potential of remote sensing data-crop model assimilation and seasonal weather forecasts for early-season crop yield forecasting over a large area. *Field Crops Research*, 276, 108398. <https://doi.org/https://doi.org/10.1016/j.fcr.2021.108398>
- Cheng, E., Zhang, B., Peng, D., Zhong, L., Yu, L., Liu, Y., Xiao, C., Li, C., Li, X., Chen, Y., Ye, H., Wang, H., Yu, R., Hu, J., & Yang, S. (2022). Wheat yield estimation using remote sensing data based on machine learning approaches. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.1090970>
- Cheng, M., Wang, H., Fan, J., Xiang, Y., Liu, X., Liao, Z., Abdelghany, A. E., Zhang, F., & Li, Z. (2022). Evaluation of AquaCrop model for greenhouse cherry tomato with plastic film mulch under various water and nitrogen supplies. *Agricultural Water Management*, 274. <https://doi.org/10.1016/j.agwat.2022.107949>
- Choruma, D. J., Balkovic, J., Pietsch, S. A., & Odume, O. N. (2021). Using EPIC to simulate the effects of different irrigation and fertilizer levels on maize yield in the Eastern Cape, South Africa. *Agricultural Water Management*, 254. <https://doi.org/10.1016/j.agwat.2021.106974>
- Choudhury, A. K., Molla, M. S. H., Zahan, T., Sen, R., Biswas, J. C., Akhter, S., Ishtiaque, S., Ahmed, F., Maniruzaman, M., Hossain, M. B., Sarker, P. C., Dessoky, E. S., Hassan, M. M., & Hossain, A. (2021). Optimum sowing window and yield forecasting for maize in northern and western bangladesh using ceres maize model. *Agronomy*, 11(4). <https://doi.org/10.3390/AGRONOMY11040635>
- Ciric, C., Ciffroy, P., & Charles, S. (2012). Use of sensitivity analysis to identify influential and non-influential parameters within an aquatic ecosystem

- model. *Ecological Modelling*, 246(C), 119–130.
<https://doi.org/10.1016/j.ecolmodel.2012.06.024>
- Coleman, K., Muhammed, S. E., Milne, A. E., Todman, L. C., Dailey, A. G., Glendining, M. J., & Whitmore, A. P. (2017a). The landscape model: A model for exploring trade-offs between agricultural production and the environment. *Science of the Total Environment*, 609, 1483–1499.
<https://doi.org/10.1016/j.scitotenv.2017.07.193>
- Coleman, K., Muhammed, S. E., Milne, A. E., Todman, L. C., Dailey, A. G., Glendining, M. J., & Whitmore, A. P. (2017b). The landscape model: A model for exploring trade-offs between agricultural production and the environment. *Science of the Total Environment*, 609, 1483–1499.
<https://doi.org/10.1016/j.scitotenv.2017.07.193>
- Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., & Acutis, M. (2010). Comparison of sensitivity analysis techniques: A case study with the rice model WARM. *Ecological Modelling*, 221(16), 1897–1906.
<https://doi.org/10.1016/j.ecolmodel.2010.04.021>
- Constantin, J., Raynal, H., Casellas, E., Hoffmann, H., Bindi, M., Doro, L., Eckersten, H., Gaiser, T., Grosz, B., Haas, E., Kersebaum, K., Klatt, S., Kuhnert, M., Lewan, E., Maharjan, G., Moriondo, M., Nendel, C., Roggero, P. P., Specka, X., & Bergez, J.-E. (2019). Management and spatial resolution effects on yield and water balance at regional scale in crop models. *Agricultural and Forest Meteorology*, 275, 184–195.
<https://doi.org/10.1016/j.agrformet.2019.05.013>
- Corstanje, R., Kirk, G. J. D., & Lark, R. M. (2008). The behaviour of soil process models of ammonia volatilization at contrasting spatial scales. *European Journal of Soil Science*, 59(6), 1271–1283. <https://doi.org/10.1111/j.1365-2389.2008.01086.x>
- Corstanje, R., Mercer, T., Rickson, R., Lynda, K., Newell Price, P., Holman, I., Kechavarsi, C., & Waine, T. (2017). Physical soil quality indicators for

- monitoring British soils. *Solid Earth*, 8, 1003–1016.
<https://doi.org/10.5194/se-8-1003-2017>
- Courault, D., Hossard, L., Demarez, V., Dechatre, H., Irfan, K., Baghdadi, N., Flamain, F., & Ruget, F. (n.d.). *STICS crop model and Sentinel-2 images for monitoring rice growth and yield in the Camargue region*.
<https://doi.org/10.1007/s13593-021-00697-w/Published>
- Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., & Schaibly, J. H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics*, 59(8), 3873–3878. <https://doi.org/10.1063/1.1680571>
- Darra, N., Anastasiou, E., Kriezi, O., Lazarou, E., Kalivas, D., & Fountas, S. (2023). Can Yield Prediction Be Fully Digitized? A Systematic Review. In *Agronomy* (Vol. 13, Issue 9). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/agronomy13092441>
- De Wit, C. T. (1958). *Transpiration and Crop Yields*. Institute of Biological and Chemical Research in Field Crops and Herbage.
<https://books.google.co.uk/books?id=3YolHQAACAAJ>
- DeJonge, K. C., Ascough, J. C., Ahmadi, M., Andales, A. A., & Arabi, M. (2012). Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. *Ecological Modelling*, 231, 113–125. <https://doi.org/10.1016/j.ecolmodel.2012.01.024>
- Dhanapal, R., Ajanraj, A., Balavinayagapragathish, S., & Balaji, J. (2021). Retraction: Crop price prediction using supervised machine learning algorithms. In *Journal of Physics: Conference Series* (Vol. 1916, Issue 1). IOP Publishing Ltd. <https://doi.org/10.1088/1742-6596/1916/1/012042>
- Dhillon, M. S., Dahms, T., Kuebert-Flock, C., Borg, E., Conrad, C., & Ullmann, T. (2020). Modelling crop biomass from synthetic remote sensing time series:

- Example for the DEMMIN test site, Germany. *Remote Sensing*, 12(11).
<https://doi.org/10.3390/rs12111819>
- Di Paola, A., Valentini, R., & Santini, M. (2016). An overview of available crop growth and yield models for studies and assessments in agriculture. *Journal of the Science of Food and Agriculture*, 96(3), 709–714.
<https://doi.org/https://doi.org/10.1002/jsfa.7359>
- Diao, C., Augspurger, C. K., Zhao, Y., & Salk, C. F. (2024). A satellite-field phenological bridging framework for characterizing community-level spring forest phenology using multi-scale satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211, 83–103.
<https://doi.org/10.1016/j.isprsjprs.2024.03.018>
- Dirwai, T. L., Senzanje, A., & Mabhaudhi, T. (2021). Calibration and evaluation of the fao aquacrop model for canola (*Brassica napus*) under varied moistube irrigation regimes. *Agriculture (Switzerland)*, 11(5).
<https://doi.org/10.3390/agriculture11050410>
- Dlamini, L., Crespo, O., van Dam, J., & Kooistra, L. (2023). A Global Systematic Review of Improving Crop Model Estimations by Assimilating Remote Sensing Data: Implications for Small-Scale Agricultural Systems. In *Remote Sensing* (Vol. 15, Issue 16). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/rs15164066>
- Dong, Y., Zhao, C., Yang, G., Chen, L., Wang, J., & Feng, H. (2013). Integrating a very fast simulated annealing optimization algorithm for crop leaf area index variational assimilation. *Mathematical and Computer Modelling*, 58(3–4), 877–885. <https://doi.org/10.1016/j.mcm.2012.12.013>
- Doorenbos, J., Kassam, A. H., & Bentvelsen, C. I. M. (1979). *Yield response to water*. Food and Agriculture Organization of the United Nations.
- Doraiswamy, P. C., Hatfield, J. L., Jackson, T. J., Akhmedov, B., Prueger, J., & Stern, A. (2004). Crop condition and yield simulations using Landsat and

- MODIS. *Remote Sensing of Environment*, 92(4), 548–559.
<https://doi.org/10.1016/j.rse.2004.05.017>
- Dubovik, O., Schuster, G. L., Xu, F., Hu, Y., Bösch, H., Landgraf, J., & Li, Z. (2021). Grand Challenges in Satellite Remote Sensing. *Frontiers in Remote Sensing*, 2. <https://doi.org/10.3389/frsen.2021.619818>
- Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023a). Crop Prediction Model Using Machine Learning Algorithms. *Applied Sciences (Switzerland)*, 13(16).
<https://doi.org/10.3390/app13169288>
- Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023b). Crop Prediction Model Using Machine Learning Algorithms. *Applied Sciences (Switzerland)*, 13(16).
<https://doi.org/10.3390/app13169288>
- Er-Raki, S., Bouras, E., Rodriguez, J. C., Watts, C. J., Lizarraga-Celaya, C., & Chehbouni, A. (2021). Parameterization of the AquaCrop model for simulating table grapes growth and water productivity in an arid region of Mexico. *Agricultural Water Management*, 245.
<https://doi.org/10.1016/j.agwat.2020.106585>
- Feng, L., Wang, Y., Zhang, Z., & Du, Q. (2021). *Geographically and temporally weighted neural network for winter wheat yield prediction*.
- Feng, X., Tian, H., Cong, J., & Zhao, C. (2023). A method review of the climate change impact on crop yield. In *Frontiers in Forests and Global Change* (Vol. 6). Frontiers Media SA. <https://doi.org/10.3389/ffgc.2023.1198186>
- Fischer, C., Montmerle, T., Berre, L., Auger, L., & ȘTEFĂNESCU, S. E. (2005). An overview of the variational assimilation in the ALADIN/France numerical weather-prediction system. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3477–3492.
<https://doi.org/10.1256/qj.05.115>

- Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L. B., Obersteiner, M., & Van Der Velde, M. (2016). Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. *Nature Communications*, 7. <https://doi.org/10.1038/ncomms11872>
- 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. <https://doi.org/10.1016/j.agwat.2016.11.015>
- Gamon, J. A., Field, C. B., Goulden, M. L., Griffin, K. L., Hartley, A. E., Joel, G., Penuelas, J., & Valentini, R. (1995). Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecological Applications*, 5(1), 28–41. <https://doi.org/10.2307/1942049>
- Gavasso-Rita, Y. L., Papalexiou, S. M., Li, Y., Elshorbagy, A., Li, Z., & Schuster-Wallace, C. (2024). Crop models and their use in assessing crop production and food security: A review. In *Food and Energy Security* (Vol. 13, Issue 1). John Wiley and Sons Inc. <https://doi.org/10.1002/fes3.503>
- Gerard, F. F., George, C. T., Hayman, G., Chavana-Bryant, C., & Weedon, G. P. (2020). Leaf phenology amplitude derived from MODIS NDVI and EVI: Maps of leaf phenology synchrony for Meso- and South America. *Geoscience Data Journal*, 7(1), 13–26. <https://doi.org/10.1002/gdj3.87>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017a). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017b). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>

- Greaves, G. E., & Wang, Y. M. (2016). Assessment of fao aquacrop model for simulating maize growth and productivity under deficit irrigation in a tropical environment. *Water (Switzerland)*, 8(12). <https://doi.org/10.3390/w8120557>
- Guo, D., Olesen, J. E., Pullens, J. W. M., Guo, C., & Ma, X. (2021). Calibrating AquaCrop model using genetic algorithm with multi-objective functions applying different weight factors. *Agronomy Journal*, 113(2), 1420–1438. <https://doi.org/10.1002/agj2.20588>
- Han, H., Liu, Z., Li, J., & Zeng, Z. (2024). Challenges in remote sensing based climate and crop monitoring: navigating the complexities using AI. In *Journal of Cloud Computing* (Vol. 13, Issue 1). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1186/s13677-023-00583-8>
- Hariharan, G., & Kannan, K. (2014). Review of wavelet methods for the solution of reaction–diffusion problems in science and engineering. *Applied Mathematical Modelling*, 38(3), 799–813. <https://doi.org/https://doi.org/10.1016/j.apm.2013.08.003>
- He, D., Wang, E., Wang, J., & Robertson, M. J. (2017). Data requirement for effective calibration of process-based crop models. *Agricultural and Forest Meteorology*, 234–235, 136–148. <https://doi.org/10.1016/j.agrformet.2016.12.015>
- Hoogenboom, G., & White, J. W. (2003). Improving Physiological Assumptions Of Simulation Models By Using Gene-Based Approaches. *Agronomy Journal*, 95(1), 82–89. <https://doi.org/10.2134/agronj2003.8200>
- Hsiao, T. C., Heng, L., Steduto, P., Rojas-Lara, B., Raes, D., & Fereres, E. (2009). Aquacrop-The FAO crop model to simulate yield response to water: III. Parameterization and testing for maize. *Agronomy Journal*, 101(3), 448–459. <https://doi.org/10.2134/agronj2008.0218s>
- Hu, T., Zhang, X., Bohrer, G., Liu, Y., Zhou, Y., Martin, J., Li, Y., & Zhao, K. (2023). Crop yield prediction via explainable AI and interpretable machine

- learning: Dangers of black box models for evaluating climate change impacts on crop yield. *Agricultural and Forest Meteorology*, 336, 109458. <https://doi.org/https://doi.org/10.1016/j.agrformet.2023.109458>
- Huang, H., Huang, J., & Wu, Y. (2020). Markov Chain Monte Carlo and Four-Dimensional Variational Approach Based Winter Wheat Yield Estimation. *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 5290–5293. <https://doi.org/10.1109/IGARSS39084.2020.9324208>
- Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., Tian, L., Zhang, X. (Frank), Fan, J., & Wenbin, 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.
- Huemmrich, K. F., Vargas Zesati, S., Campbell, P., & Tweedie, C. (2021). Canopy reflectance models illustrate varying NDVI responses to change in high latitude ecosystems. *Ecological Applications*, 31(8). <https://doi.org/10.1002/eap.2435>
- Huete, A. R. (1988). A Soil-Adjusted Vegetation Index (SAVI) 295. In *REMOTE SENSING OF ENVIRONMENT* (Vol. 25).
- Iqbal, M. A., Shen, Y., Stricevic, R., Pei, H., Sun, H., Amiri, E., Penas, A., & del Rio, S. (2014). Evaluation of the FAO AquaCrop model for winter wheat on the North China Plain under deficit irrigation from field experiment to regional yield simulation. *Agricultural Water Management*, 135, 61–72. <https://doi.org/10.1016/j.agwat.2013.12.012>
- Ishihara, M., Inoue, Y., Ono, K., Shimizu, M., & Matsuura, S. (2015). The impact of sunlight conditions on the consistency of vegetation indices in croplands- Effective usage of vegetation indices from continuous ground-based spectral measurements. *Remote Sensing*, 7(10), 14079–14098. <https://doi.org/10.3390/rs71014079>

- Jain, M., Balwinder-Singh, Rao, P., Srivastava, A. K., Poonia, S., Blesh, J., Azzari, G., McDonald, A. J., & Lobell, D. B. (2019). The impact of agricultural interventions can be doubled by using satellite data. *Nature Sustainability*, 2(10), 931–934. <https://doi.org/10.1038/s41893-019-0396-x>
- Javan, K., & Darestani, M. (2024). Assessing environmental sustainability of a vital crop in a critical region: Investigating climate change impacts on agriculture using the SWAT model and HWA method. *Heliyon*, 10(3). <https://doi.org/10.1016/j.heliyon.2024.e25326>
- Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PLoS ONE*, 11(6). <https://doi.org/10.1371/journal.pone.0156571>
- Ji, Z., Pan, Y., Zhu, X., Wang, J., & Li, Q. (2021). Prediction of crop yield using phenological information extracted from remote sensing vegetation index. *Sensors (Switzerland)*, 21(4), 1–17. <https://doi.org/10.3390/s21041406>
- 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. In *European Journal of Agronomy* (Vol. 92, pp. 141–152). Elsevier B.V. <https://doi.org/10.1016/j.eja.2017.11.002>
- 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(12). <https://doi.org/10.3390/rs8120972>
- 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. <https://doi.org/10.1016/j.agwat.2019.105846>

- 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. <https://doi.org/10.1016/j.isprsjprs.2017.02.001>
- Jin, Z., Prasad, R., Shriver, J., & Zhuang, Q. (2017). Crop model- and satellite imagery-based recommendation tool for variable rate N fertilizer application for the US Corn system. *Precision Agriculture*, 18(5), 779–800. <https://doi.org/10.1007/s11119-016-9488-z>
- Jing, Q., McConkey, B., Qian, B., Smith, W., Grant, B., Shang, J., Liu, J., Bindraban, P., & Luce, M. S. (2021). Assessing water management effects on spring wheat yield in the Canadian Prairies using DSSAT wheat models. *Agricultural Water Management*, 244. <https://doi.org/10.1016/j.agwat.2020.106591>
- Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). *The DSSAT cropping system model*. www.elsevier.com/locate/eja
- Joshi, N., Mitchard, E. T. A., Brolly, M., Schumacher, J., Fernández-Landa, A., Johannsen, V. K., Marchamalo, M., & Fensholt, R. (2017). Understanding “saturation” of radar signals over forests. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-03469-3>
- 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. <https://doi.org/10.1016/j.rse.2019.04.005>
- Karaboga, D. (2005). *An Idea Based on Honey Bee Swarm for Numerical Optimization*.

- Kasampalis, D. A., Alexandridis, T. K., Deva, C., Challinor, A., Moshou, D., & Zalidis, G. (2018). Contribution of remote sensing on crop models: A review. In *Journal of Imaging* (Vol. 4, Issue 4). MDPI. <https://doi.org/10.3390/jimaging4040052>
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, 4, 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
- Kephe, P. N., Ayisi, K. K., & Petja, B. M. (2021a). Challenges and opportunities in crop simulation modelling under seasonal and projected climate change scenarios for crop production in South Africa. In *Agriculture and Food Security* (Vol. 10, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s40066-020-00283-5>
- Kephe, P. N., Ayisi, K. K., & Petja, B. M. (2021b). Challenges and opportunities in crop simulation modelling under seasonal and projected climate change scenarios for crop production in South Africa. In *Agriculture and Food Security* (Vol. 10, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s40066-020-00283-5>
- Khabba, S., Duchemin, B., Hadria, R., Er-Raki, S., Ezzahar, J., Chehbouni, A., Lahrouni, A., & Hanich, L. (2009). Evaluation of digital hemispherical photography and plant canopy analyzer for measuring vegetation area index of orange orchards. *Journal of Agronomy*, 8(2), 67–72. <https://doi.org/10.3923/ja.2009.67.72>
- Khabba, S., Er-Raki, S., Toumi, J., Ezzahar, J., Hssaine, B. A., Page, M. le, & Chehbouni, A. (2020). A simple light-use-efficiency model to estimate wheat yield in the semi-Arid areas. *Agronomy*, 10(10). <https://doi.org/10.3390/agronomy10101524>
- Khabbazan, S., Vermunt, P., Steele-Dunne, S., Arntz, L. R., Marinetti, C., van der Valk, D., Iannini, L., Molijn, R., Westerdijk, K., & van der Sande, C. (2019).

- Crop monitoring using Sentinel-1 data: A case study from The Netherlands. *Remote Sensing*, 11(16). <https://doi.org/10.3390/rs11161887>
- Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.00621>
- Khaki, S., Wang, L., & Archontoulis, S. V. (2020). A CNN-RNN Framework for Crop Yield Prediction. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.01750>
- Klocke, N., Stone, L., Briggeman, S., & Bolton, D. (2010). Technical Note: Scheduling for Deficit Irrigation—Crop Yield Predictor. *Applied Engineering in Agriculture*, 26, 413–418. <https://doi.org/10.13031/2013.29956>
- Koo, J., Bostick, W., Jones, J., Gijsman, A., & Naab, J. (2003). Estimating Soil Carbon in Agricultural Systems using Ensemble Kalman Filter and DSSAT-CENTURY. In *Transactions of the ASABE* (Vol. 50). <https://doi.org/10.13031/2013.13847>
- Krishnan, M., Gugercin, S., & Tarazaga, P. A. (2023). A wavelet-based dynamic mode decomposition for modeling mechanical systems from partial observations. *Mechanical Systems and Signal Processing*, 187, 109919. <https://doi.org/10.1016/j.ymssp.2022.109919>
- Kumar, U., Hansen, E. M., Thomsen, I. K., & Vogeler, I. (2023). Performance of APSIM to Simulate the Dynamics of Winter Wheat Growth, Phenology, and Nitrogen Uptake from Early Growth Stages to Maturity in Northern Europe. *Plants*, 12(5). <https://doi.org/10.3390/plants12050986>
- Lee, D.-Y., Lee, D.-S., Cha, Y., Min, J.-H., & Park, Y.-S. (2023). Data-driven models for predicting community changes in freshwater ecosystems: A review. *Ecological Informatics*, 77, 102163. <https://doi.org/10.1016/j.ecoinf.2023.102163>
- Leukel, J., Zimpel, T., & Stumpe, C. (2023). Machine learning technology for early prediction of grain yield at the field scale: A systematic review. *Computers*

- and Electronics in Agriculture*, 207, 107721.
<https://doi.org/https://doi.org/10.1016/j.compag.2023.107721>
- Li, J., & Chen, B. (2020). Global revisit interval analysis of landsat-8-9 and sentinel-2a-2b data for terrestrial monitoring. *Sensors (Switzerland)*, 20(22), 1–15. <https://doi.org/10.3390/s20226631>
- Li, K., & Chen, Y. (2018). A Genetic Algorithm-based urban cluster automatic threshold method by combining VIIRS DNB, NDVI, and NDBI to monitor urbanization. *Remote Sensing*, 10(2). <https://doi.org/10.3390/rs10020277>
- Li, R., Li, C., Liu, F., Yang, X., & Wang, J. (2010). Methods and algorithms of data assimilation and its application in agriculture. *2010 World Automation Congress*, 29–34.
- Li, Z., He, J., Xu, X., Jin, X., Huang, W., Clark, B., Yang, G., & Li, Z. (2018). Estimating genetic parameters of DSSAT-CERES model with the GLUE method for winter wheat (*Triticum aestivum* L.) production. *Computers and Electronics in Agriculture*, 154, 213–221.
<https://doi.org/https://doi.org/10.1016/j.compag.2018.09.009>
- Li, Z., Jin, X., Zhao, C., Wang, J., Xu, X., Yang, G., Li, C., & Shen, J. (2015). Estimating wheat yield and quality by coupling the DSSAT-CERES model and proximal remote sensing. *European Journal of Agronomy*, 71, 53–62.
<https://doi.org/10.1016/j.eja.2015.08.006>
- Li, Z., Zhao, Y., Taylor, J., Gaulton, R., Jin, X., Song, X., Li, Z., Meng, Y., Chen, P., Feng, H., Wang, C., Guo, W., Xu, X., Chen, L., & Yang, G. (2022). Comparison and transferability of thermal, temporal and phenological-based in-season predictions of above-ground biomass in wheat crops from proximal crop reflectance data. *Remote Sensing of Environment*, 273, 112967. <https://doi.org/https://doi.org/10.1016/j.rse.2022.112967>
- Liao, C., Wang, J., Dong, T., Shang, J., Liu, J., & Song, Y. (2019). Using spatio-temporal fusion of Landsat-8 and MODIS data to derive phenology, biomass

- and yield estimates for corn and soybean. *Science of the Total Environment*, 650, 1707–1721. <https://doi.org/10.1016/j.scitotenv.2018.09.308>
- Lin, F., Li, X., Jia, N., Feng, F., Huang, H., Huang, J., Fan, S., Ciais, P., & Song, X. P. (2023). The impact of Russia-Ukraine conflict on global food security. *Global Food Security*, 36. <https://doi.org/10.1016/j.gfs.2022.100661>
- Linker, R. (2020). Unified framework for model-based optimal allocation of crop areas and water. *Agricultural Water Management*, 228, 105859. <https://doi.org/https://doi.org/10.1016/j.agwat.2019.105859>
- López-Urrea, R., Domínguez, A., Pardo, J. J., Montoya, F., García-Vila, M., & Martínez-Romero, A. (2020). Parameterization and comparison of the AquaCrop and MOPECO models for a high-yielding barley cultivar under different irrigation levels. *Agricultural Water Management*, 230. <https://doi.org/10.1016/j.agwat.2019.105931>
- Lorenz, A. J. (2013). Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. *G3: Genes, Genomes, Genetics*, 3(3), 481–491. <https://doi.org/10.1534/g3.112.004911>
- Lu, Y., Chibarabada, T. P., Ziliani, M. G., Onema, J. M. K., McCabe, M. F., & Sheffield, J. (2021). Assimilation of soil moisture and canopy cover data improves maize simulation using an under-calibrated crop model. *Agricultural Water Management*, 252. <https://doi.org/10.1016/j.agwat.2021.106884>
- Lyu, J., Jiang, Y., Xu, C., Liu, Y., Su, Z., Liu, J., & He, J. (2022). Multi-objective winter wheat irrigation strategies optimization based on coupling AquaCrop-OSPy and NSGA-III: A case study in Yangling, China. *Science of the Total Environment*, 843. <https://doi.org/10.1016/j.scitotenv.2022.157104>
- Ma, G., Huang, J., Wu, W., Fan, J., Zou, J., & Wu, S. (2013). Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat

- yield. *Mathematical and Computer Modelling*, 58(3), 634–643.
<https://doi.org/https://doi.org/10.1016/j.mcm.2011.10.038>
- 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(9), 4225–4236. <https://doi.org/10.1109/JSTARS.2017.2711482>
- Ma, Y., Yang, Z., Huang, Q., & Zhang, Z. (2023). Improving the Transferability of Deep Learning Models for Crop Yield Prediction: A Partial Domain Adaptation Approach. *Remote Sensing*, 15(18), 4562. <https://doi.org/10.3390/rs15184562>
- Ma, Y., Zhang, Z., Kang, Y., & Özdoğan, M. (2021). Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sensing of Environment*, 259, 112408. <https://doi.org/https://doi.org/10.1016/j.rse.2021.112408>
- Maestrini, B., Mimić, G., van Oort, P. A. J., Jindo, K., Brdar, S., van Evert, F. K., & Athanasiados, I. (2022a). Mixing process-based and data-driven approaches in yield prediction. In *European Journal of Agronomy* (Vol. 139). Elsevier B.V. <https://doi.org/10.1016/j.eja.2022.126569>
- Maestrini, B., Mimić, G., van Oort, P. A. J., Jindo, K., Brdar, S., van Evert, F. K., & Athanasiados, I. (2022b). Mixing process-based and data-driven approaches in yield prediction. In *European Journal of Agronomy* (Vol. 139). Elsevier B.V. <https://doi.org/10.1016/j.eja.2022.126569>
- Manschadi, A. M., Eitzinger, J., Breisch, M., Fuchs, W., Neubauer, T., & Soltani, A. (2021). Full Parameterisation Matters for the Best Performance of Crop Models: Inter-comparison of a Simple and a Detailed Maize Model. *International Journal of Plant Production*, 15(1), 61–78. <https://doi.org/10.1007/s42106-020-00116-2>

- Martínez-Ruiz, A., Ruiz-García, A., Prado-Hernández, J. V., López-Cruz, I. L., Valencia-Islas, J. O., & Pineda-Pineda, J. (2021). Global sensitivity analysis and calibration by differential evolution algorithm of HORTSYST crop model for fertigation management. *Water (Switzerland)*, 13(5). <https://doi.org/10.3390/w13050610>
- Martinho, V. J. P. D., Cunha, C. A. da S., Pato, M. L., Costa, P. J. L., Sánchez-Carreira, M. C., Georgantzís, N., Rodrigues, R. N., & Coronado, F. (2022). Machine Learning and Food Security: Insights for Agricultural Spatial Planning in the Context of Agriculture 4.0. *Applied Sciences (Switzerland)*, 12(22). <https://doi.org/10.3390/app122211828>
- Marzukhi, F., Elahami, A. L., & Bohari, S. N. (2016). Detecting nutrients deficiencies of oil palm trees using remotely sensed data. *IOP Conference Series: Earth and Environmental Science*, 37(1). <https://doi.org/10.1088/1755-1315/37/1/012040>
- McRae, G. J., Tilden, J. W., & Seinfeld, J. H. (1982). Global sensitivity analysis—a computational implementation of the Fourier Amplitude Sensitivity Test (FAST). *Computers & Chemical Engineering*, 6(1), 15–25. [https://doi.org/https://doi.org/10.1016/0098-1354\(82\)80003-3](https://doi.org/https://doi.org/10.1016/0098-1354(82)80003-3)
- Meddi, M., & Eslamian, S. (2021). Uncertainties in Rainfall and Water Resources in Maghreb Countries Under Climate Change. In *African Handbook of Climate Change Adaptation: With 610 Figures and 361 Tables* (pp. 1967–2003). Springer International Publishing. https://doi.org/10.1007/978-3-030-45106-6_114
- Meroni, M., Fasbender, D., Rembold, F., Atzberger, C., & Klisch, A. (2019). Near real-time vegetation anomaly detection with MODIS NDVI: Timeliness vs. accuracy and effect of anomaly computation options. *Remote Sensing of Environment*, 221, 508–521. <https://doi.org/https://doi.org/10.1016/j.rse.2018.11.041>

- Mohanty, B. P., Cosh, M. H., Lakshmi, V., & Montzka, C. (2017). Soil Moisture Remote Sensing: State-of-the-Science. *Vadose Zone Journal*, 16(1), 1–9. <https://doi.org/10.2136/vzj2016.10.0105>
- Monteiro, L. A., Ramos, R. M., Battisti, R., Soares, J. R., Oliveira, J. C., Figueiredo, G. K. D. A., Lamparelli, R. A. C., Nendel, C., & Lana, M. A. (2022). Potential Use of Data-Driven Models to Estimate and Predict Soybean Yields at National Scale in Brazil. *International Journal of Plant Production*, 16(4), 691–703. <https://doi.org/10.1007/s42106-022-00209-0>
- Moran, M. S., Inoue, Y., & Barnes, E. M. (1997). Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment*, 61(3), 319–346. [https://doi.org/https://doi.org/10.1016/S0034-4257\(97\)00045-X](https://doi.org/https://doi.org/10.1016/S0034-4257(97)00045-X)
- Morel, J., Bégué, A., Todoroff, P., Martiné, J. F., Lebourgeois, V., & Petit, M. (2014a). Coupling a sugarcane crop model with the remotely sensed time series of fIPAR to optimise the yield estimation. *European Journal of Agronomy*, 61, 60–68. <https://doi.org/10.1016/j.eja.2014.08.004>
- Morel, J., Bégué, A., Todoroff, P., Martiné, J.-F., Lebourgeois, V., & Petit, M. (2014b). Coupling a sugarcane crop model with the remotely sensed time series of fIPAR to optimise the yield estimation. *European Journal of Agronomy*, 61, 60–68. <https://doi.org/https://doi.org/10.1016/j.eja.2014.08.004>
- Morris, M. D. (1991a). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2), 161–174. <https://doi.org/10.1080/00401706.1991.10484804>
- Morris, M. D. (1991b). Factorial Sampling Plans for Preliminary Computational Experiments. In *TECHNOMETRICS* (Vol. 33, Issue 2).
- Mubvuma, M. T., Ogola, J. B. O., & Mhizha, T. (2021a). AquaCrop model calibration and validation for chickpea (*Cicer arietinum*) in Southern Africa.

- Cogent Food and Agriculture*, 7(1).
<https://doi.org/10.1080/23311932.2021.1898135>
- Mubvuma, M. T., Ogola, J. B. O., & Mhizha, T. (2021b). AquaCrop model calibration and validation for chickpea (*Cicer arietinum*) in Southern Africa. *Cogent Food and Agriculture*, 7(1).
<https://doi.org/10.1080/23311932.2021.1898135>
- Müller, C., Jägermeyr, J., Franke, J. A., Ruane, A. C., Balkovic, J., Ciais, P., Dury, M., Falloon, P., Folberth, C., Hank, T., Hoffmann, M., Izaurralde, R. C., Jacquemin, I., Khabarov, N., Liu, W., Olin, S., Pugh, T. A. M., Wang, X., Williams, K., ... Elliott, J. W. (2024). Substantial Differences in Crop Yield Sensitivities Between Models Call for Functionality-Based Model Evaluation. *Earth's Future*, 12(3). <https://doi.org/10.1029/2023EF003773>
- Nansen, C., Geremias, L. D., Xue, Y., Huang, F., & Parra, J. R. (2013). Agricultural Case Studies of Classification Accuracy, Spectral Resolution, and Model Over-Fitting. *Applied Spectroscopy*, 67(11), 1332–1338.
<https://doi.org/10.1366/12-06933>
- Nash, J. E., & Sutcliffe, J. v. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/https://doi.org/10.1016/0022-1694(70)90255-6)
- Newlands, N. K., Ghahari, A., Gel, Y. R., Lyubchich, V., Agriculture, T. M., & Canada, A.-F. (n.d.). *Deep Learning for Improved Agricultural Risk Management*. <https://oceanservice.noaa.gov/facts/>
- Niu, Z., He, H., Peng, S., Ren, X., Zhang, L., Gu, F., Zhu, G., Peng, C., Li, P., Wang, J., Ge, R., Zeng, N., Zhu, X., Lv, Y., Chang, Q., Xu, Q., Zhang, M., & Liu, W. (2021). A Process-Based Model Integrating Remote Sensing Data for Evaluating Ecosystem Services. *Journal of Advances in Modeling Earth Systems*, 13(6). <https://doi.org/10.1029/2020MS002451>

- Nossent, J., Elsen, P., & Bauwens, W. (2011). Sobol' sensitivity analysis of a complex environmental model. *Environmental Modelling & Software*, 26(12), 1515–1525. [https://doi.org/https://doi.org/10.1016/j.envsoft.2011.08.010](https://doi.org/10.1016/j.envsoft.2011.08.010)
- Oikonomidis, A., Catal, C., & Kassahun, A. (2023). Deep learning for crop yield prediction: a systematic literature review. In *New Zealand Journal of Crop and Horticultural Science* (Vol. 51, Issue 1, pp. 1–26). Taylor and Francis Ltd. <https://doi.org/10.1080/01140671.2022.2032213>
- Oulaid, B., Milne, A. E., Waine, T., El Alami, R., Rafiqi, M., & Corstanje, R. (2024). Stepwise model parametrisation using satellite imagery and hemispherical photography: Tuning AquaCrop sensitive parameters for improved winter wheat yield predictions in semi-arid regions. *Field Crops Research*, 309. <https://doi.org/10.1016/j.fcr.2024.109327>
- Paleari, L., Movedi, E., Zoli, M., Burato, A., Cecconi, I., Errahouly, J., Pecollo, E., Sorvillo, C., & Confalonieri, R. (2021). Sensitivity analysis using Morris: Just screening or an effective ranking method? *Ecological Modelling*, 455, 109648. [https://doi.org/https://doi.org/10.1016/j.ecolmodel.2021.109648](https://doi.org/10.1016/j.ecolmodel.2021.109648)
- Paredes, P., Wei, Z., Liu, Y., Xu, D., Xin, Y., Zhang, B., & Pereira, L. S. (2015). Performance assessment of the FAO aquaCrop model for soil water, soil evaporation, biomass and yield of soybeans in north china plain. *Agricultural Water Management*, 152, 57–71. <https://doi.org/10.1016/j.agwat.2014.12.007>
- Parent, B., & Tardieu, F. (2014). Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature? In *Journal of Experimental Botany* (Vol. 65, Issue 21, pp. 6179–6189). Oxford University Press. <https://doi.org/10.1093/jxb/eru223>
- Pasquel, D., Roux, S., Richetti, J., Cammarano, D., Tisseyre, B., & Taylor, J. A. (2022). A review of methods to evaluate crop model performance at multiple

- and changing spatial scales. In *Precision Agriculture* (Vol. 23, Issue 4, pp. 1489–1513). Springer. <https://doi.org/10.1007/s11119-022-09885-4>
- Patel, K., & Patel, H. B. (2021). A Comparative Analysis of Supervised Machine Learning Algorithm for Agriculture Crop Prediction. *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 1–5. <https://doi.org/10.1109/ICECCT52121.2021.9616731>
- Patel, U., Pathan, M., Kathiria, P., & Patel, V. (2023). Crop type classification with hyperspectral images using deep learning : a transfer learning approach. *Modeling Earth Systems and Environment*, 9(2), 1977–1987. <https://doi.org/10.1007/s40808-022-01608-y>
- Pelosi, A., Belfiore, O. R., D'Urso, G., & Chirico, G. B. (2022). Assessing Crop Water Requirement and Yield by Combining ERA5-Land Reanalysis Data with CM-SAF Satellite-Based Radiation Data and Sentinel-2 Satellite Imagery. *Remote Sensing*, 14(24). <https://doi.org/10.3390/rs14246233>
- Peng, Y., Zhang, M., Xu, Z., Yang, T., Su, Y., Zhou, T., Wang, H., Wang, Y., & Lin, Y. (2020). Estimation of leaf nutrition status in degraded vegetation based on field survey and hyperspectral data. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-61294-7>
- Petersen, L. K. (2018). Real-time prediction of crop yields from MODIS relative vegetation health: A continent-wide analysis of Africa. *Remote Sensing*, 10(11). <https://doi.org/10.3390/rs10111726>
- Poempool, L., Kruatrachue, B., & Siriboon, K. (2018, August 13). Combine multi particle swarm in supporting trapping in local optima. *ICEAST 2018 - 4th International Conference on Engineering, Applied Sciences and Technology: Exploring Innovative Solutions for Smart Society*. <https://doi.org/10.1109/ICEAST.2018.8434398>
- Prudnikova, E., Savin, I., Vindeker, G., Grubina, P., Shishkonakova, E., & Sharychev, D. (2019). Influence of soil background on spectral reflectance

- of winter wheat crop canopy. *Remote Sensing*, 11(16).
<https://doi.org/10.3390/rs11161932>
- Psiroukis, V., Darra, N., Kasimati, A., Trojacek, P., Hasanli, G., & Fountas, S. (2022). Development of a Multi-Scale Tomato Yield Prediction Model in Azerbaijan Using Spectral Indices from Sentinel-2 Imagery. *Remote Sensing*, 14(17). <https://doi.org/10.3390/rs14174202>
- Puntel, L. A., Sawyer, J. E., Barker, D. W., Dietzel, R., Poffenbarger, H., Castellano, M. J., Moore, K. J., Thorburn, P., & Archontoulis, S. V. (2016). Modeling long-term corn yield response to nitrogen rate and crop rotation. *Frontiers in Plant Science*, 7(November 2016).
<https://doi.org/10.3389/fpls.2016.01630>
- Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2009). Aquacrop-The FAO crop model to simulate yield response to water: II. main algorithms and software description. *Agronomy Journal*, 101(3), 438–447.
<https://doi.org/10.2134/agronj2008.0140s>
- Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2022). *Chapter 3 Calculation procedures AquaCrop Reference manual August 2022*.
- Rani, S., Mishra, A. K., Kataria, A., Mallik, S., & Qin, H. (2023). Machine learning-based optimal crop selection system in smart agriculture. *Scientific Reports*, 13(1), 15997. <https://doi.org/10.1038/s41598-023-42356-y>
- Razzaghi, F., Zhou, Z., Andersen, M. N., & Plauborg, F. (2017). Simulation of potato yield in temperate condition by the AquaCrop model. *Agricultural Water Management*, 191, 113–123.
<https://doi.org/10.1016/j.agwat.2017.06.008>
- Rezzoug, W., Gabrielle, B., Suleiman, A., & Benabdeli, K. (2008). Application and evaluation of the DSSAT-wheat in the Tiaret region of Algeria. In *African Journal of Agricultural Research* (Vol. 3, Issue 4).
<http://www.academicjournals.org/AJAR>

- Rivas-Fandiño, P., Acuña-Alonso, C., Novo, A., Pacheco, F. A. L., & Álvarez, X. (2023). Assessment of high spatial resolution satellite imagery for monitoring riparian vegetation: riverine management in the smallholding. *Environmental Monitoring and Assessment*, 195(1). <https://doi.org/10.1007/s10661-022-10667-8>
- Roberts, M. J., Braun, N. O., Sinclair, T. R., Lobell, D. B., & Schlenker, W. (2017). Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, 12(9). <https://doi.org/10.1088/1748-9326/aa7f33>
- Rostami, A., Raeini-Sarjaz, M., Chabokpour, J., Azamathulla, H. M., & Kumar, S. (2022). Determination of rainfed wheat agriculture potential through assimilation of remote sensing data with SWAT model case study: ZarrinehRoud Basin, Iran. *Water Supply*, 22(5), 5331–5334. <https://doi.org/10.2166/ws.2022.160>
- Rozenstein, O., Cohen, Y., Alchanatis, V., Behrendt, K., Bonfil, D. J., Eshel, G., Harari, A., Harris, W. E., Klapp, I., Laor, Y., Linker, R., Paz-Kagan, T., Peets, S., Rutter, S. M., Salzer, Y., & Lowenberg-DeBoer, J. (2024). Data-driven agriculture and sustainable farming: friends or foes? *Precision Agriculture*, 25(1), 520–531. <https://doi.org/10.1007/s11119-023-10061-5>
- Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. *Agricultural and Forest Meteorology*, 200, 233–248. <https://doi.org/10.1016/j.agrformet.2014.09.016>
- Sabzzadeh, I., & Shourian, M. (2020). Maximizing crops yield net benefit in a groundwater-irrigated plain constrained to aquifer stable depletion using a coupled PSO-SWAT-MODFLOW hydro-agronomic model. *Journal of Cleaner Production*, 262. <https://doi.org/10.1016/j.jclepro.2020.121349>
- Sahbeni, G., Székely, B., Musyimi, P. K., Timár, G., & Sahajpal, R. (2023). Crop Yield Estimation Using Sentinel-3 SLSTR, Soil Data, and Topographic

- Features Combined with Machine Learning Modeling: A Case Study of Nepal. *AgriEngineering*, 5(4), 1766–1788. <https://doi.org/10.3390/agriengineering5040109>
- Sajid, S. S., & Hu, G. (2022). Optimizing Crop Planting Schedule Considering Planting Window and Storage Capacity. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.762446>
- Salman Maher, García-Vila Margarita, Fereres Elias, Raes Dirk, & Steduto Steduto. (2021). The AquaCrop model – Enhancing crop water productivity. In *The AquaCrop model – Enhancing crop water productivity*. FAO. <https://doi.org/10.4060/cb7392en>
- Saltelli, A., Tarantola, S., & Chan, K. P.-S. (1999). A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*, 41(1), 39–56. <https://doi.org/10.1080/00401706.1999.10485594>
- Sarrazin, F., Pianosi, F., & Wagener, T. (2016). Global Sensitivity Analysis of environmental models: Convergence and validation. *Environmental Modelling and Software*, 79, 135–152. <https://doi.org/10.1016/j.envsoft.2016.02.005>
- Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., & Gao, F. (n.d.). *Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description*. <http://www.usgs.gov/pubprod>
- Schumacher, B. L., Burchfield, E. K., Bean, B., & Yost, M. A. (2023). Leveraging Important Covariate Groups for Corn Yield Prediction. *Agriculture (Switzerland)*, 13(3). <https://doi.org/10.3390/agriculture13030618>
- Shang, K., Yao, Y., Di, Z., Jia, K., Zhang, X., Fisher, J. B., Chen, J., Guo, X., Yang, J., Yu, R., Xie, Z., Liu, L., Ning, J., & Zhang, L. (2023). Coupling physical constraints with machine learning for satellite-derived

- evapotranspiration of the Tibetan Plateau. *Remote Sensing of Environment*, 289, 113519. <https://doi.org/https://doi.org/10.1016/j.rse.2023.113519>
- Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning. *IEEE Access*, 11, 111255–111264. <https://doi.org/10.1109/ACCESS.2023.3321861>
- Sharp, R. T., Henrys, P. A., Jarvis, S. G., Whitmore, A. P., Milne, A. E., Coleman, K., Mohankumar, S. E. P., & Metcalfe, H. (2021). Simulating cropping sequences using earth observation data. *Computers and Electronics in Agriculture*, 188, 106330. <https://doi.org/https://doi.org/10.1016/j.compag.2021.106330>
- Shrestha, A., Bheemanahalli, R., Adeli, A., Samiappan, S., Czarnecki, J. M. P., McCraine, C. D., Reddy, K. R., & Moorhead, R. (2023). Phenological stage and vegetation index for predicting corn yield under rainfed environments. *Frontiers in Plant Science*, 14. <https://doi.org/10.3389/fpls.2023.1168732>
- Shuai, G., & Basso, B. (2022). Subfield maize yield prediction improves when in-season crop water deficit is included in remote sensing imagery-based models. *Remote Sensing of Environment*, 272. <https://doi.org/10.1016/j.rse.2022.112938>
- Silva, J. V., & Giller, K. E. (2020). Grand challenges for the 21st century: What crop models can and can't (yet) do. *Journal of Agricultural Science*, 158(10), 794–805. <https://doi.org/10.1017/S0021859621000150>
- 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(11). <https://doi.org/10.1371/journal.pone.0187485>

- Simms, D. M., Hamer, A. M., Zeiler, I., Vita, L., & Waine, T. W. (2023). Mapping Agricultural Land in Afghanistan's Opium Provinces Using a Generalised Deep Learning Model and Medium Resolution Satellite Imagery. *Remote Sensing*, 15(19). <https://doi.org/10.3390/rs15194714>
- Siritanaratkul, B. (2022). Generalizability and limitations of machine learning for yield prediction of oxidative coupling of methane. *Digital Chemical Engineering*, 2. <https://doi.org/10.1016/j.dche.2022.100013>
- Sodari, L., Alam, J., & Das, K. K. (2021). *Performance Assessment of Aquacrop Model for Estimating Canopy Cover, Aboveground Biomass and Grain Yield of Winter Wheat in Saptari District: An Application to Irrigation Management*.
- Solomatine, D., See, L. M., & Abrahart, R. J. (2008). Data-Driven Modelling: Concepts, Approaches and Experiences. In R. J. Abrahart, L. M. See, & D. P. Solomatine (Eds.), *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications* (pp. 17–30). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-79881-1_2
- Soltani, A., Alimagham, S. M., Nehbandani, A., Torabi, B., Zeinali, E., Dadrasi, A., Zand, E., Ghassemi, S., Pourshirazi, S., Alasti, O., Hosseini, R. S., Zahed, M., Arabameri, R., Mohammadzadeh, Z., Rahban, S., Kamari, H., Fayazi, H., Mohammadi, S., Keramat, S., ... Sinclair, T. R. (2020). SSM-iCrop2: A simple model for diverse crop species over large areas. *Agricultural Systems*, 182. <https://doi.org/10.1016/j.agsy.2020.102855>
- Specka, X., Nendel, C., & Wieland, R. (2015). Analysing the parameter sensitivity of the agro-ecosystem model MONICA for different crops. *European Journal of Agronomy*, 71, 73–87. <https://doi.org/10.1016/j.eja.2015.08.004>
- Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009a). Aquacrop-the FAO crop model to simulate yield response to water: I. concepts and underlying principles. *Agronomy Journal*, 101(3), 426–437. <https://doi.org/10.2134/agronj2008.0139s>

- Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009b). AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. *Agronomy Journal*, 101(3), 426–437. <https://doi.org/10.2134/agronj2008.0139s>
- Stricevic, R., Cosic, M., Djurovic, N., Pejic, B., & Maksimovic, L. (2011). Assessment of the FAO AquaCrop model in the simulation of rainfed and supplementally irrigated maize, sugar beet and sunflower. *Agricultural Water Management*, 98(10), 1615–1621. <https://doi.org/10.1016/j.agwat.2011.05.011>
- Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. *Sensors (Switzerland)*, 19(20). <https://doi.org/10.3390/s19204363>
- Sun, Y. (2022). Enhanced Weather-Based Index Insurance Design for Hedging Crop Yield Risk. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.895183>
- Szu, H., & Hartley, R. (1987). Fast simulated annealing. *Physics Letters A*, 122(3–4). [https://doi.org/10.1016/0375-9601\(87\)90796-1](https://doi.org/10.1016/0375-9601(87)90796-1)
- Tai, M. M. (1994). A Mathematical Model for the Determination of Total Area Under Glucose Tolerance and Other Metabolic Curves •ⁿ⁻¹ = X_n ~ X_{n-l}
Total Area = $\sum (y_2 - y_1) + X^4 - x^3 (3/4 - 3/3) - X^{3/3} \setminus (K y_1 + X_1 y_2 + X^2 - X^4 - X^{3/3} + 4 - X^4 - X^5 - \dots - X^i)$. In *DIABETES CARE* (Vol. 17, Issue 2). <http://diabetesjournals.org/care/article-pdf/17/2/152/341381/17-2-152.pdf>
- Takács, S., Csengeri, E., Pék, Z., Bíró, T., Szuvandzsiev, P., Palotás, G., & Helyes, L. (2021). Performance evaluation of aquacrop model in processing tomato biomass, fruit yield and water stress indicator modelling. *Water (Switzerland)*, 13(24). <https://doi.org/10.3390/w13243587>

- Tan, B., Gao, F., Wolfe, R. E., Pedelty, J. A., Nightingale, J., Morisette, J. T., & Ederer, G. A. (2011). An Enhanced TIMESAT Algorithm for Estimating Vegetation Phenology Metrics From MODIS Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(2), 361–371. <https://doi.org/10.1109/JSTARS.2010.2075916>
- Tang, B.-H., Shao, K., Li, Z.-L., Wu, H., & Tang, R. (2015). An improved NDVI-based threshold method for estimating land surface emissivity using MODIS satellite data. *International Journal of Remote Sensing*, 36(19–20), 4864–4878. <https://doi.org/10.1080/01431161.2015.1040132>
- Tao, H., Feng, H., Xu, L., Miao, M., Yang, G., Yang, X., & Fan, L. (2020). Estimation of the yield and plant height of winter wheat using UAV-based hyperspectral images. *Sensors (Switzerland)*, 20(4). <https://doi.org/10.3390/s20041231>
- Tao, R., Zhao, P., Wu, J., Martin, N., Harrison, M. T., Ferreira, C., Kalantari, Z., & Hovakimyan, N. (2023). *Optimizing Crop Management with Reinforcement Learning and Imitation Learning*.
- Tenreiro, T. R., García-Vila, M., Gómez, J. A., Jiménez-Berni, J. A., & Fereres, E. (2021). Using NDVI for the assessment of canopy cover in agricultural crops within modelling research. *Computers and Electronics in Agriculture*, 182. <https://doi.org/10.1016/j.compag.2021.106038>
- Tesfaye, K., Takele, R., Shelia, V., Hayi, E., Dabale, A., Traore, P., & Solomon, D. (2022). High Spatial Resolution Seasonal Crop Yield Forecasting for Heterogeneous Maize Environments in the Oromia Regional State, Ethiopia. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4074628>
- Toumi, J., Er-Raki, S., Ezzahar, J., Khabba, S., Jarlan, L., & Chehbouni, A. (2016a). Performance assessment of AquaCrop model for estimating evapotranspiration, soil water content and grain yield of winter wheat in Tensift Al Haouz (Morocco): Application to irrigation management.

- Agricultural Water Management*, 163, 219–235.
<https://doi.org/10.1016/j.agwat.2015.09.007>
- Toumi, J., Er-Raki, S., Ezzahar, J., Khabba, S., Jarlan, L., & Chehbouni, A. (2016b). Performance assessment of AquaCrop model for estimating evapotranspiration, soil water content and grain yield of winter wheat in Tensift Al Haouz (Morocco): Application to irrigation management. *Agricultural Water Management*, 163, 219–235.
<https://doi.org/10.1016/j.agwat.2015.09.007>
- Tsakmakis, I. D., Gikas, G. D., & Sylaios, G. K. (2021). Integration of Sentinel-derived NDVI to reduce uncertainties in the operational field monitoring of maize. *Agricultural Water Management*, 255, 106998.
<https://doi.org/https://doi.org/10.1016/j.agwat.2021.106998>
- Tsakmakis, I. D., Zidou, M., Gikas, G. D., & Sylaios, G. K. (2018). Impact of Irrigation Technologies and Strategies on Cotton Water Footprint Using AquaCrop and CROPWAT Models. *Environmental Processes*, 5, 181–199.
<https://doi.org/10.1007/s40710-018-0289-4>
- 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(16). <https://doi.org/10.3390/RS12162666>
- van der Velde, M., Biavetti, I., El-Aydam, M., Niemeyer, S., Santini, F., & van den Berg, M. (2019). Use and relevance of European Union crop monitoring and yield forecasts. *Agricultural Systems*, 168, 224–230.
<https://doi.org/10.1016/j.agry.2018.05.001>
- Vannoppen, A., & Gobin, A. (2021). Estimating farm wheat yields from NDVI and meteorological data. *Agronomy*, 11(5).
<https://doi.org/10.3390/agronomy11050946>

- Vanuytrecht, E., Raes, D., & Willems, P. (2014). Global sensitivity analysis of yield output from the water productivity model. *Environmental Modelling & Software*, 51, 323–332. <https://doi.org/https://doi.org/10.1016/j.envsoft.2013.10.017>
- Vazquez-Cruz, M. A., Guzman-Cruz, R., Lopez-Cruz, I. L., Cornejo-Perez, O., Torres-Pacheco, I., & Guevara-Gonzalez, R. G. (2014). Global sensitivity analysis by means of EFAST and Sobol' methods and calibration of reduced state-variable TOMGRO model using genetic algorithms. *Computers and Electronics in Agriculture*, 100, 1–12. <https://doi.org/https://doi.org/10.1016/j.compag.2013.10.006>
- Viswambharan, S., Kumaramkandath, I. T., & Tali, J. A. (2022). A geospatial approach in monitoring the variations on surface soil moisture and vegetation water content: a case study of Palakkad District, Kerala, India. *Environmental Earth Sciences*, 81(20). <https://doi.org/10.1007/s12665-022-10611-6>
- Wagner, M. P., Slawig, T., Taravat, A., & Oppelt, N. (2020). Remote sensing data assimilation in dynamic crop models using particle swarm optimization. *ISPRS International Journal of Geo-Information*, 9(2). <https://doi.org/10.3390/ijgi9020105>
- Wahba, G. (1975). Smoothing noisy data with spline functions. *Numerische Mathematik*, 24(5), 383–393. <https://doi.org/10.1007/BF01437407>
- Waine, T., Simms, D., Taylor, J., & Juniper, G. (2014). Towards improving the accuracy of opium yield estimates with remote sensing. *International Journal of Remote Sensing*, 35, 1–18. <https://doi.org/10.1080/01431161.2014.951743>
- Wale, A., Dessie, M., & Kendie, H. (2022). Evaluating the Performance of AquaCrop Model for Potato Production Under Deficit Irrigation. *Air, Soil and Water Research*, 15. <https://doi.org/10.1177/11786221221108216>

- Wallach, D., Palosuo, T., Thorburn, P., Hochman, Z., Gourdain, E., Andrianasolo, F., Asseng, S., Basso, B., Buis, S., Crout, N., Dibari, C., Dumont, B., Ferrise, R., Gaiser, T., Garcia, C., Gayler, S., Ghahramani, A., Hiremath, S., Hoek, S., ... Seidel, S. J. (2021). The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. *Environmental Modelling and Software*, 145. <https://doi.org/10.1016/j.envsoft.2021.105206>
- Wang, G., Mehmood, F., Zain, M., Hamani, A. K. M., Xue, J., Gao, Y., & Duan, A. (2022). AquaCrop Model Evaluation for Winter Wheat under Different Irrigation Management Strategies: A Case Study on the North China Plain. *Agronomy*, 12(12). <https://doi.org/10.3390/agronomy12123184>
- Wang, J., Li, X., Lu, L., & Fang, F. (2013). Estimating near future regional corn yields by integrating multi-source observations into a crop growth model. *European Journal of Agronomy*, 49, 126–140. <https://doi.org/10.1016/j.eja.2013.03.005>
- Wang, L. (2021). Data Driven Explanation of Temporal and Spatial Variability of Maize Yield in the United States. *Frontiers in Plant Science*, 12. <https://doi.org/10.3389/fpls.2021.701192>
- Wang, L., & Qu, J. J. (2009). Satellite remote sensing applications for surface soil moisture monitoring: A review. In *Frontiers of Earth Science in China* (Vol. 3, Issue 2, pp. 237–247). Higher Education Press. <https://doi.org/10.1007/s11707-009-0023-7>
- Weiss M, & Baret F. (2016). *Sentinel2 ToolBox Level2 Products S2ToolBox Level 2 products: LAI, FAPAR, FCOVER Version 1.1*.
- Wellens, J., Raes, D., Fereres, E., Diels, J., Coppys, C., Adiele, J. G., Ezui, K. S. G., Becerra, L. A., Selvaraj, M. G., Dercon, G., & Heng, L. K. (2022). Calibration and validation of the FAO AquaCrop water productivity model for cassava (*Manihot esculenta* Crantz). *Agricultural Water Management*, 263. <https://doi.org/10.1016/j.agwat.2022.107491>

- Wood, D., McNairn, H., Brown, R. J., & Dixon, R. (2002). The effect of dew on the use of RADARSAT-1 for crop monitoring: Choosing between ascending and descending orbits. *Remote Sensing of Environment*, 80(2), 241–247. [https://doi.org/https://doi.org/10.1016/S0034-4257\(01\)00304-2](https://doi.org/https://doi.org/10.1016/S0034-4257(01)00304-2)
- Wooster, M. J., Roberts, G. J., Giglio, L., Roy, D. P., Freeborn, P. H., Boschetti, L., Justice, C., Ichoku, C., Schroeder, W., Davies, D., Smith, A. M. S., Setzer, A., Csiszar, I., Strydom, T., Frost, P., Zhang, T., Xu, W., de Jong, M. C., Johnston, J. M., ... San-Miguel-Ayanz, J. (2021). Satellite remote sensing of active fires: History and current status, applications and future requirements. *Remote Sensing of Environment*, 267, 112694. <https://doi.org/https://doi.org/10.1016/j.rse.2021.112694>
- Wu, H., & Li, Z. L. (2009). Scale issues in remote sensing: A review on analysis, processing and modeling. In *Sensors* (Vol. 9, Issue 3, pp. 1768–1793). <https://doi.org/10.3390/s90301768>
- Wu, S., Huang, J., Liu, X., Fan, J., Ma, G., Zou, J., & Huang, J. (2011). *Assimilating MODIS-LAI into Crop Growth Model with EnKF to Predict Regional Crop Yield*. 410–418. <https://doi.org/10.1007/978-3-642-27275>
- Wu, S., Yang, P., Ren, J., Chen, Z., & Li, H. (2021). Regional winter wheat yield estimation based on the WOFOST model and a novel VW-4DEnSRF assimilation algorithm. *Remote Sensing of Environment*, 255, 112276. <https://doi.org/https://doi.org/10.1016/j.rse.2020.112276>
- Wu, Y., Xu, W., Huang, H., & Huang, J. (2022). Bayesian Posterior-Based Winter Wheat Yield Estimation at the Field Scale through Assimilation of Sentinel-2 Data into WOFOST Model. *Remote Sensing*, 14(15). <https://doi.org/10.3390/rs14153727>
- Xavier, A., Hall, B., Casteel, S., Muir, W., & Rainey, K. M. (2017). Using unsupervised learning techniques to assess interactions among complex traits in soybeans. *Euphytica*, 213(8). <https://doi.org/10.1007/s10681-017-1975-4>

- Xiao, J., Aggarwal, A. K., Duc, N. H., Arya, A., Rage, U. K., & Avtar, R. (2023). A review of remote sensing image spatiotemporal fusion: Challenges, applications and recent trends. *Remote Sensing Applications: Society and Environment*, 32, 101005. <https://doi.org/https://doi.org/10.1016/j.rsase.2023.101005>
- Xin, Q., Li, J., Li, Z., Li, Y., & Zhou, X. (2020). Evaluations and comparisons of rule-based and machine-learning-based methods to retrieve satellite-based vegetation phenology using MODIS and USA National Phenology Network data. *International Journal of Applied Earth Observation and Geoinformation*, 93. <https://doi.org/10.1016/j.jag.2020.102189>
- XING, H. min, XU, X. gang, LI, Z. hai, CHEN, Y. jin, FENG, H. kuan, YANG, G. jun, & CHEN, Z. xia. (2017a). 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(11), 2444–2458. [https://doi.org/10.1016/S2095-3119\(16\)61626-X](https://doi.org/10.1016/S2095-3119(16)61626-X)
- XING, H. min, XU, X. gang, LI, Z. hai, CHEN, Y. jin, FENG, H. kuan, YANG, G. jun, & CHEN, Z. xia. (2017b). 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(11), 2444–2458. [https://doi.org/10.1016/S2095-3119\(16\)61626-X](https://doi.org/10.1016/S2095-3119(16)61626-X)
- Xu, C., & Gertner, G. (2011). Understanding and comparisons of different sampling approaches for the Fourier Amplitudes Sensitivity Test (FAST). *Computational Statistics and Data Analysis*, 55(1), 184–198. <https://doi.org/10.1016/j.csda.2010.06.028>
- Xu, G., & Yu, G. (2018). On convergence analysis of particle swarm optimization algorithm. *Journal of Computational and Applied Mathematics*, 333, 65–73. <https://doi.org/10.1016/j.cam.2017.10.026>

- Xu, H., Twine, T. E., & Yang, X. (2014). Evaluating remotely sensed phenological metrics in a dynamic ecosystem model. *Remote Sensing*, 6(6), 4660–4686. <https://doi.org/10.3390/rs6064660>
- Xu, Y., Yang, Y., Chen, X., & Liu, Y. (2022). Bibliometric Analysis of Global NDVI Research Trends from 1985 to 2021. In *Remote Sensing* (Vol. 14, Issue 16). MDPI. <https://doi.org/10.3390/rs14163967>
- Yang, H., Yang, X., Heskell, M., Sun, S., & Tang, J. (2017). Seasonal variations of leaf and canopy properties tracked by ground-based NDVI imagery in a temperate forest. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-01260-y>
- Yang, Y., Anderson, M. C., Gao, F., Johnson, D. M., Yang, Y., Sun, L., Dulaney, W., Hain, C. R., Otkin, J. A., Prueger, J., Meyers, T. P., Bernacchi, C. J., & Moore, C. E. (2021). Phenological corrections to a field-scale, ET-based crop stress indicator: An application to yield forecasting across the U.S. Corn Belt. *Remote Sensing of Environment*, 257, 112337. <https://doi.org/10.1016/j.rse.2021.112337>
- Yang, Y., Tao, B., Liang, L., Huang, Y., Matocha, C., Lee, C. D., Sama, M., El Masri, B., & Ren, W. (2021). Detecting recent crop phenology dynamics in corn and soybean cropping systems of Kentucky. *Remote Sensing*, 13(9). <https://doi.org/10.3390/rs13091615>
- Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., ... Sequeira, A. M. M. (2018). Outstanding Challenges in the Transferability of Ecological Models. In *Trends in Ecology and Evolution* (Vol. 33, Issue 10, pp. 790–802). Elsevier Ltd. <https://doi.org/10.1016/j.tree.2018.08.001>

- Ye, J., Bao, W., Liao, C., Chen, D., & Hu, H. (2023). Corn Phenology Detection Using the Derivative Dynamic Time Warping Method and Sentinel-2 Time Series. *Remote Sensing*, 15(14). <https://doi.org/10.3390/rs15143456>
- Yli-Heikkilä, M., Wittke, S., Luotamo, M., Puttonen, E., Sulkava, M., Pellikka, P., Heiskanen, J., & Klami, A. (2022). Scalable Crop Yield Prediction with Sentinel-2 Time Series and Temporal Convolutional Network. *Remote Sensing*, 14(17). <https://doi.org/10.3390/rs14174193>
- Yu, J., Smith, A., & Sumner, D. A. (2018). Effects of crop insurance premium subsidies on crop acreage. *American Journal of Agricultural Economics*, 100(1), 91–114. <https://doi.org/10.1093/ajae/aax058>
- Zampieri, M., Weissteiner, C. J., Grizzetti, B., Toreti, A., van den Berg, M., & Dentener, F. (2020). Estimating resilience of crop production systems: From theory to practice. *Science of the Total Environment*, 735. <https://doi.org/10.1016/j.scitotenv.2020.139378>
- Zeng, J., Yuan, X., Ji, P., & Shi, C. (2021). Effects of meteorological forcings and land surface model on soil moisture simulation over China. *Journal of Hydrology*, 603. <https://doi.org/10.1016/j.jhydrol.2021.126978>
- Zeng, L., Wardlow, B. D., Xiang, D., Hu, S., & Li, D. (2020). A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment*, 237, 111511. <https://doi.org/https://doi.org/10.1016/j.rse.2019.111511>
- ZHANG, F., ZHANG, L.-W., SHI, J.-J., & HUANG, J.-F. (2014). Soil Moisture Monitoring Based on Land Surface Temperature-Vegetation Index Space Derived from MODIS Data. *Pedosphere*, 24(4), 450–460. [https://doi.org/https://doi.org/10.1016/S1002-0160\(14\)60031-X](https://doi.org/https://doi.org/10.1016/S1002-0160(14)60031-X)
- Zhang, J., Pan, B., Shi, W., & Zhang, Y. (2023). Monitoring Waterlogging Damage of Winter Wheat Based on HYDRUS-1D and WOFOST Coupled Model and

- Assimilated Soil Moisture Data of Remote Sensing. *Remote Sensing*, 15(17). <https://doi.org/10.3390/rs15174133>
- 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. <https://doi.org/https://doi.org/10.1016/j.compag.2019.105052>
- Zhang, Y., Walker, J. P., & Pauwels, V. R. N. (2022). Assimilation of wheat and soil states for improved yield prediction: The APSIM-EnKF framework. *Agricultural Systems*, 201. <https://doi.org/10.1016/j.agry.2022.103456>
- Zhang, Y., Walker, J. P., Pauwels, V. R. N., & Sadeh, Y. (2022a). Assimilation of wheat and soil states into the apsim-wheat crop model: A case study. *Remote Sensing*, 14(1). <https://doi.org/10.3390/rs14010065>
- Zhang, Y., Walker, J. P., Pauwels, V. R. N., & Sadeh, Y. (2022b). Assimilation of wheat and soil states into the apsim-wheat crop model: A case study. *Remote Sensing*, 14(1). <https://doi.org/10.3390/rs14010065>
- Zhao, H., Yang, C., Guo, W., Zhang, L., & Zhang, D. (2020). Automatic estimation of crop disease severity levels based on vegetation index normalization. *Remote Sensing*, 12(12). <https://doi.org/10.3390/rs12121930>
- Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y., & Yu, S. (2016). Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. *Remote Sensing*, 8(6). <https://doi.org/10.3390/rs8060469>
- Zomer, R. J., Xu, J., & Trabucco, A. (2022). Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. *Scientific Data*, 9(1). <https://doi.org/10.1038/s41597-022-01493-1>
- Zou, X., & Mörtus, M. (2017). Sensitivity of common vegetation indices to the canopy structure of field crops. *Remote Sensing*, 9(10). <https://doi.org/10.3390/rs9100994>

3 Towards a better understanding of rainfed wheat yield predictions in arid and semi-arid regions: How early, precise, and transferable can these models be?

Bader Oulaid ^{a,b,c}, Tadesse Gashaw Asrat ^b, Alice E. Milne ^b, Toby Waine ^a, Rafiq El Alami ^c, Stephan M. Haefele ^b, Abdellah Hamma ^c, Nabil Farah ^c, Ron Corstanje ^a

^a*Cranfield University, College Road, Cranfield, Bedford, MK43 0AL, United Kingdom*

^b*Net-Zero and Resilient Farming, Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom*

^c*Mohammed VI Polytechnic University, Lot 660, Moulay Rachid, Ben Guerir 43150, Morocco*

Abstract

Precisely predicting wheat yields is essential for food security and agricultural planning, but implementing a useful prediction model is a complex task as their utility for management interventions rapidly diminishes as the growing season progresses. It involves achieving sufficient precision for models to reliably inform practices early in the season, which can be readily applied across agricultural landscapes. This study explores yield prediction in semi-arid and arid regions of Morocco. It has three primary objectives: i) Compare the in-season predictive performance of three machine learning algorithms (i.e., Multiple Linear Regression, Support Vector Regression, and Random Forests) across four prediction stages. ii) Quantify the relative feature contribution to yield prediction, ii) Assess model generalisation and transferability when moving between climatic zones. Our analysis was conducted using an extensive dataset comprising wheat yield, climate, management, and remotely sensed phenology metrics. The results showed that excessive heat during reproductive phases negatively impacted wheat yield in arid climates, while early-season heat had a positive association with yield in semi-arid regions. Rainfall, especially before the peak of canopy cover had a positive influence on yield in both climates. Model performance improved as the season progressed. In arid regions, SVR displayed the lowest prediction errors (0.23 t ha^{-1} RMSE at the canopy senescence). For both semiarid and global models, RF consistently outperformed other models, with an accuracy

of 0.50 t ha⁻¹ and 0.46 t ha⁻¹ *RMSE* for predictions made at the time of canopy senescence, respectively. Feature contribution analysis revealed distinct trends across climate types. Notably, model for arid climate showed a more evenly distributed contribution of various factors to yield predictions. However, rainfall from the peak of the season onward was a dominant factor, contributing to over than 40% of predictions in semi-arid regions. Model transfer between different climates resulted in significantly higher *RMSE* values (up to 72.8% and 79.4% higher *RMSE* compared to baseline arid and semiarid scenarios). Finally, localised models consistently outperformed generalized models in the arid climate throughout all stages but were outperformed by generalized models in semiarid climate at peak of season's prediction stage.

Keywords: Wheat, Machine learning, Remote sensing, Early yield prediction, Transferability analysis, Feature contribution analysis

3.1 Introduction

Precise in-season yield prediction has become essential for the sustainability of rainfed farming in arid and semi-arid countries, given the growing water scarcity and unpredictable climate patterns (L. Feng et al., 2021; Shuai & Basso, 2022). Prior studies underscored the significance of early yield prediction models to facilitate crop management. For farmers, these models assisted in determining the optimal timing for planting and harvesting (Choudhury et al., 2021), optimizing irrigation amount and schedule (Klocke et al., 2010), selecting suitable crops based on regional conditions (Rani et al., 2023), thus increasing yields, reducing costs, and ensuring profitability even in adverse conditions. Policy makers also benefit from yield prediction models to make informed decisions about resource allocations and food security (Martinho et al., 2022). In addition, Newlands et al. (2019) indicate that yield prediction models had the potential to reduce crop insurance underwriting inefficiencies and coverage costs associated with more imprecise yield-based metrics (i.e., average yield).

Yield prediction models can be divided into two main categories: process-based and data-driven models. Process-based models, as seen in tools such as

AquaCrop (Steduto et al., 2009b), DSSAT (Jones et al., 2003) , and the Rothamsted Landscape model (Coleman et al., 2017b) are based on an understanding of the physiological processes behind crop growth. These models use mathematical equations to describe how environmental factors, soil properties, and plant physiology interact. Process-based models, while insightful, have limitations (Corstanje et al., 2008). Their data requirements can be challenging, necessitating abundant information on environmental factors, soil properties, and crop specifics. Calibration needs are also demanding, requiring historical data and expertise, which can limit accessibility (He et al., 2017). The complexity of these models, with their detailed representation of physiological and environmental processes, presents difficulties as well, particularly for users unfamiliar with the science underlying them (Kephe et al., 2021b). Data-driven models, such as those based on machine learning, offer an alternative method for yield prediction. These models use large datasets to identify predictive patterns and relationships (Y. Ma et al., 2021; Solomatine et al., 2008). In contrast to process-based models, data-driven models do not necessitate any understanding of the underlying biological processes. They derive information from the data patterns to develop models used forecast crop yields. The data dependency makes them less effective in regions with limited historical data (Chang et al., 2023b; Monteiro et al., 2022; L. Wang, 2021). Both process-based and data-driven models are challenged when it comes to data acquisition (Shang et al., 2023), a process that often demands significant resources in terms of both cost and labour. This raises the question of whether remote sensing methods can offer an efficient solution for acquiring data, potentially alleviating these constraints.

Remote sensing (RS) has been used to enhance the capabilities of different models across a wide range of research areas (Moran et al., 1997; Wooster et al., 2021; Xiao et al., 2023). For example, based on RS data and policy-related soil functions, Corstanje et al. (2017) assessed whether physical soil properties can be used an indicator for the quality of British soils. Similarly, Sharp et al. (2021) used satellite-based crop maps to provide critical model

parameters for models simulating cropping sequences. Simms et al. (2023) leveraged Sentinel, Landsat, and DMC imagery to stratify Afghanistan's agricultural lands with an accuracy comparable to that of manual image interpretation and providing a rapid information on land use change with minimal manual human input. Remote sensing (RS) has also emerged as a valuable resource for generating essential features related to soil, climate, crop health, and overall climate conditions that contribute to the development of effective yield prediction models. For instance, Sahbeni et al. (2023) used SRTM-based topographic features in combination with machine learning techniques to forecast wheat yield in Nepal. Sun et al. (2019) leveraged MODIS Land Surface Temperature (LST) data to predict soybean yield at the county level, based on CNN-LSTM deep learning models. Schumacher et al. (2023) explored the spatiotemporal efficacy of random forests (RF) in predicting county-level corn yields using bioclimatic variables. Moreover, different studies have extracted phenological metrics from a range of sources, including Unmanned Aerial Vehicles (Waine et al., 2014), Sentinel2 (Ashourloo et al., 2022; Shrestha et al., 2023), Landsat (Kang & Özdoğan, 2019; Liao et al., 2019) and used them as proxies for yield prediction. A key advantage of many remotely sensed data products is that they are available in-season and as the growing season progresses, more observation (i.e., more features) are available allowing to inform yield predictions. As we progress with advanced models that use the high-resolution remote sensing data at our disposal, a fundamental question comes to the forefront: should the primary objective of our models be the accuracy or their ability to provide predictions earlier in the season?

A crucial consideration for all agricultural stakeholders is the trade-off between model prediction precision and the time required for deliver those predictions (Bischiniotis et al., 2019; Meroni et al., 2019). In one sense, precise models can provide detailed insights into crop growth (Ji et al., 2021; Y. Yang et al., 2021), allowing for better informed management (Sajid & Hu, 2022; B. Li et al., 2018; Paredes et al., 2014). Moreover, they are valuable for policymakers enabling them to allocate resources efficiently and anticipate market dynamics with greater

accuracy (Jain et al., 2019; van der Velde et al., 2019). Models which can supply early yield predictions provide crucial information that aids in advance planning by stakeholders for major agricultural events, including market strategy, disease control, and weather alertness. Delivering these advantages requires simplified and quickly deployable models, which use readily accessible data. For example, Tesfaye et al. (2022) found that CCAFS Regional Agricultural Toolbox (CRAFT) was able to produce accurate yield forecasts and has the potential to be used as an early warning system for smallholder farming systems in Ethiopia. Moreover, Y. Sun. (2022) used a machine learning framework incorporating approaches to develop weather-yield models that are parsimonious (i.e., do not include too many features). These models resulted in reducing farmers revenue risk without government subsidies.

Precision and earliness in prediction are crucial considerations when developing yield prediction models, but they are not the sole factors that demand attention. Generalisation and transferability are equally essential aspects of these models. First, achieving a high degree of precision and providing early yield forecasts are valuable, but these attributes must also be paired with the model's ability to generalize its predictions and to extrapolate them to new conditions within its original context (Siritanaratkul, 2022). Transferability concerns the model's ability to go beyond its initial context and provide accurate predictions across different geographic areas, and time periods (Y. Ma et al., 2023). This is a more complex challenge, as it requires the model to adapt its knowledge to new scenarios effectively (Z. Li et al., 2022). Due to their reliance on data without prior knowledge of crop physiology (Maestrini et al., 2022b), machine learning models used in yield prediction often struggle when used outside their original implementation areas. Factors such as disparate scales between reference and target systems, unexpected stochastic events in evaluation data, sampling biases, spatial and temporal autocorrelation contribute to this limitation, as highlighted by Yates et al. (2018). Thus, exploring the combination of machine learning and remote sensing could offer a promising approach to address these challenges and enhance models' generalisation and transferability.

In this paper, we evaluate and compare the predictive performances of different machine learning algorithms for in-season yield predictions within and across arid and semiarid regions of Morocco at four distinct prediction stages. Recognizing the challenge of interpretability in machine learning models, we quantify the contribution of each model feature to the models' predictions. This approach will enable future model users to make informed decisions based on the identified influential features. Considering the interdependency of machine learning model effectiveness and the underlying training data, we assess the performance of generalized models as compared with climate-specific models. We explore the extent to which our models demonstrate transferability when tested beyond the boundaries of their initial implementation area. This comprehensive examination provides valuable perspectives regarding the resilience, interpretability, and practicality of our models across different domains, thereby enhancing our understanding of their capabilities and performances.

3.2 Study area and data

3.2.1 Study area

We investigated a dataset on wheat yields from 390 distinct sites from across four regions in Morocco (Figure 14). The regions of Fes-Meknes (34.0339° N, 4.9994° W) and Rabat-Sale-Kenitra (34.0209° N, 6.8417° W) fall within the northern semi-arid zone of Morocco and collectively cover a cultivated land area of 2178 501 ha. These regions are characterized by wet winters, warm summers and typically receive a total of 400 mm to 800 mm precipitation per year. The other two regions of interest, Marrakesh-Safi (31.6295° N, 8.0083° W) and Casablanca-Settat (33.5731° N, 7.5898° W), are situated in the centre part of Morocco and exhibit dry arid and semi-arid climatic conditions with hot dry summers and cooler winters. The yearly rainfall in these areas ranges from 200 mm to 600 mm. These regions together cover a cultivated land area of around 3319 020 ha.

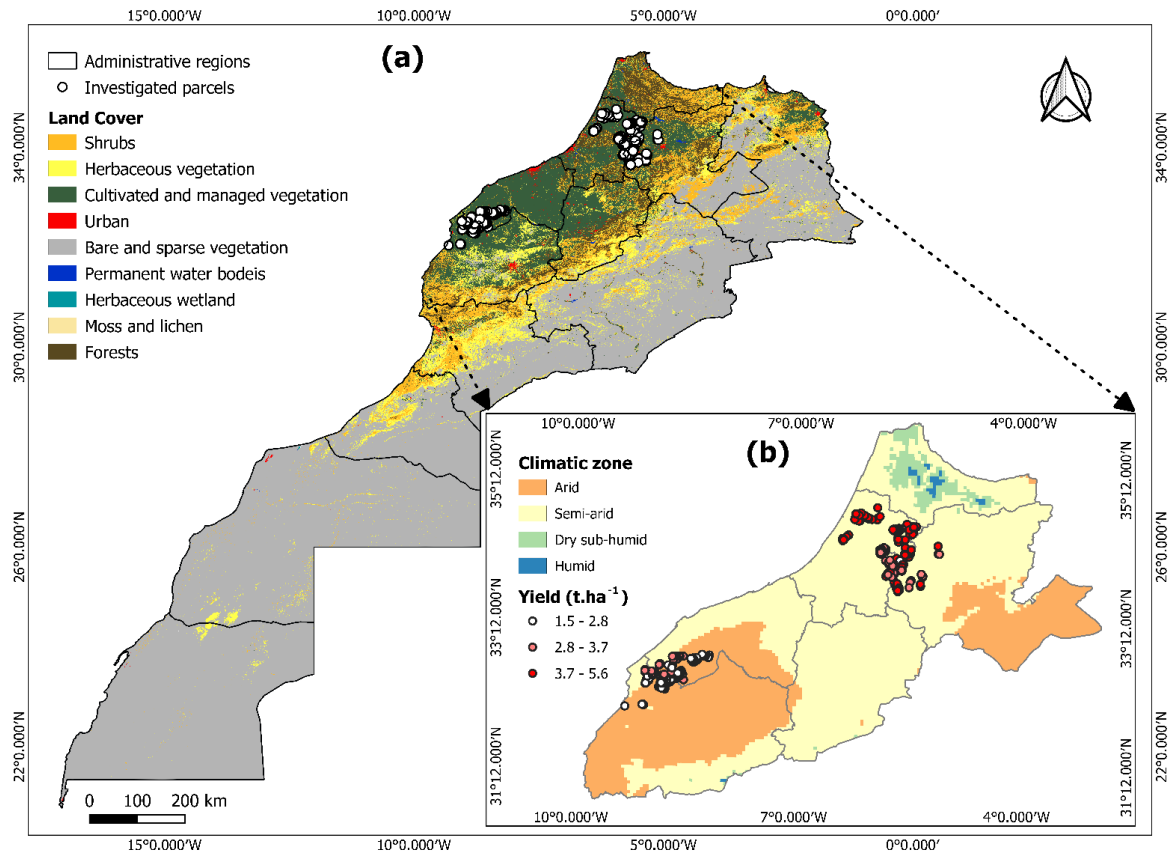


Figure 14 Morocco study area with (a) the land-cover distribution and (b) the climatic zones with wheat yield distribution for investigated sites for the 2020–2021 cropping season.

3.2.2 Yield and associated management data

We had access to data on the yield from rainfed wheat crops grown in the 2020–2021 cropping season by commercial farmers (Figure 14). The data is provided by Al Moutmir, Morocco. In all cases the cultivar was Bandera. The data set comprised location of the field, yield ($t \cdot ha^{-1}$), sowing dates, and applied fertilizer *NPK* ($kg \cdot ha^{-1}$). The yield data were collected directly from farmers by agronomists working with Al Moutmir initiative, ensuring accurate records of management practices and site-specific conditions. In line with our objective of developing early yield prediction models customized for regions located in arid and semi-arid climatic zones; the sites were categorized in to semi-arid ($n = 220$) and arid ($n = 170$) samples following the methodology outlined by Zomer et al. (2022) in which

the aridity was expressed as a ratio of precipitation over reference evapotranspiration (ET_0).

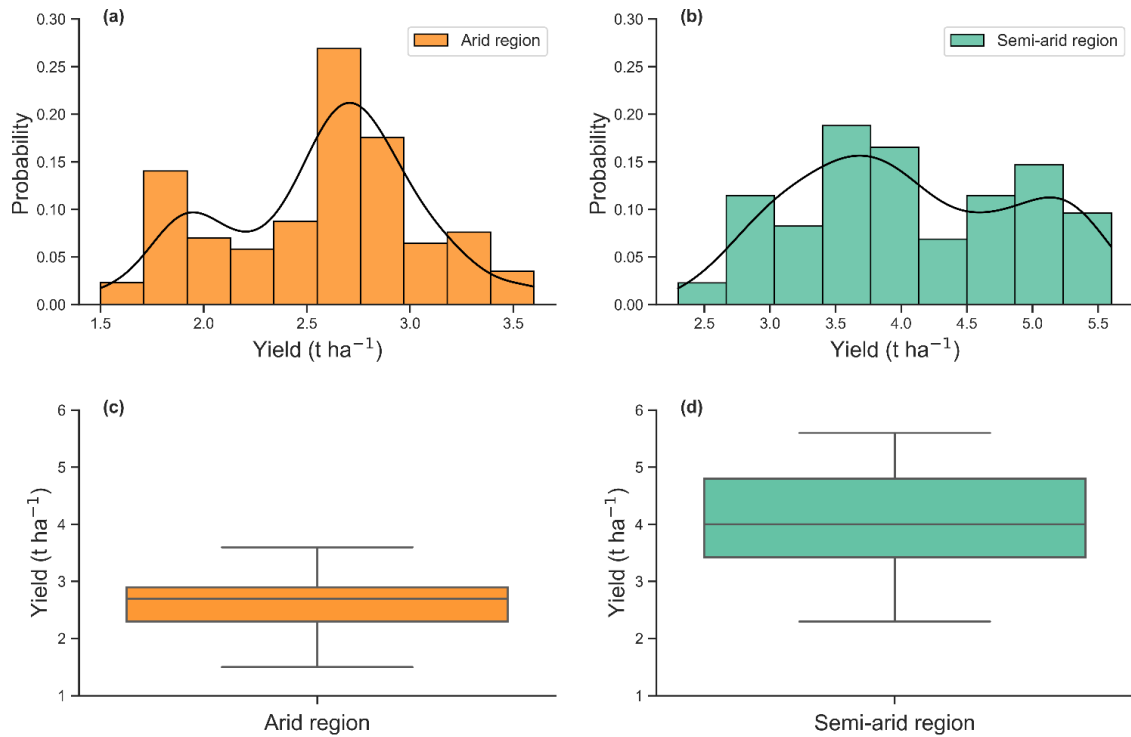


Figure 15 Wheat yield distribution and box plots for arid and semi-arid regions (2020–2021 season)

Yields in the semi-arid climate are generally larger than those in the arid climate (Table 10) and in both cases the distributions appear bimodal (Figure 15). Similarly, in both semi-arid and arid regions, nitrogen applications exhibit a bimodal distribution and similar variations, but with a slightly higher mean of in the semi-arid climate than arid (Figure 16 and Table 10). In contrast, phosphorus and potassium fertilizer rates, while generally higher in the semi-arid region, exhibit narrower ranges in the arid climate. In the semi-arid climate, sowing ranged from late September to late December with a preference for early sowing, as indicated by its negatively-skewed distribution. Conversely, in arid climate there were more concentrated sowing times (i.e., early November to early December). These sowing patterns highlight the adaptive strategies employed by

farmers to align with the diverse climatic conditions in each region (Bassu et al., 2009).

Table 10 Summary statistics of wheat yield and applied fertilizer and sowing dates in 2020–2021 cropping season .

Data	Dataset	Min	Max	Mean	Std	C.V (%)	Skewness	Kurtosis
Yield (t ha⁻¹)	Semi-arid (n=220)	2.3	5.6	4.07	0.84	20.79	0.13	-1.09
	Arid (n=170)	1.5	3.6	2.57	0.45	17.87	-0.22	-0.37
Nitrogen (Kg ha⁻¹)	Semi-arid (n=220)	0	138	58.72	41.87	71.29	0.01	-1.61
	Arid (n=170)	5	130	44.4	48.74	109.78	0.67	-1.46
Phosphorus (Kg ha⁻¹)	Semi-arid (n=220)	0	99	43.97	20.60	46.850	-0.15	-0.73
	Arid (n=170)	10	60	26.63	11.73	44.04	0.67	0.63
Potassium (Kg ha⁻¹)	Semi-arid (n=220)	0	93	22.46	13.59	60.51	3.06	12.16
	Arid (n=170)	5	40	20.49	11.19	54.61	0.54	-0.98
Sowing dates (DOY)	Semi-arid (n=220)	283	355	322.34	10.40	3.22	-1.20	3.71
	Arid (n=170)	307	330	316.44	6.84	2.16	0.62	-0.84

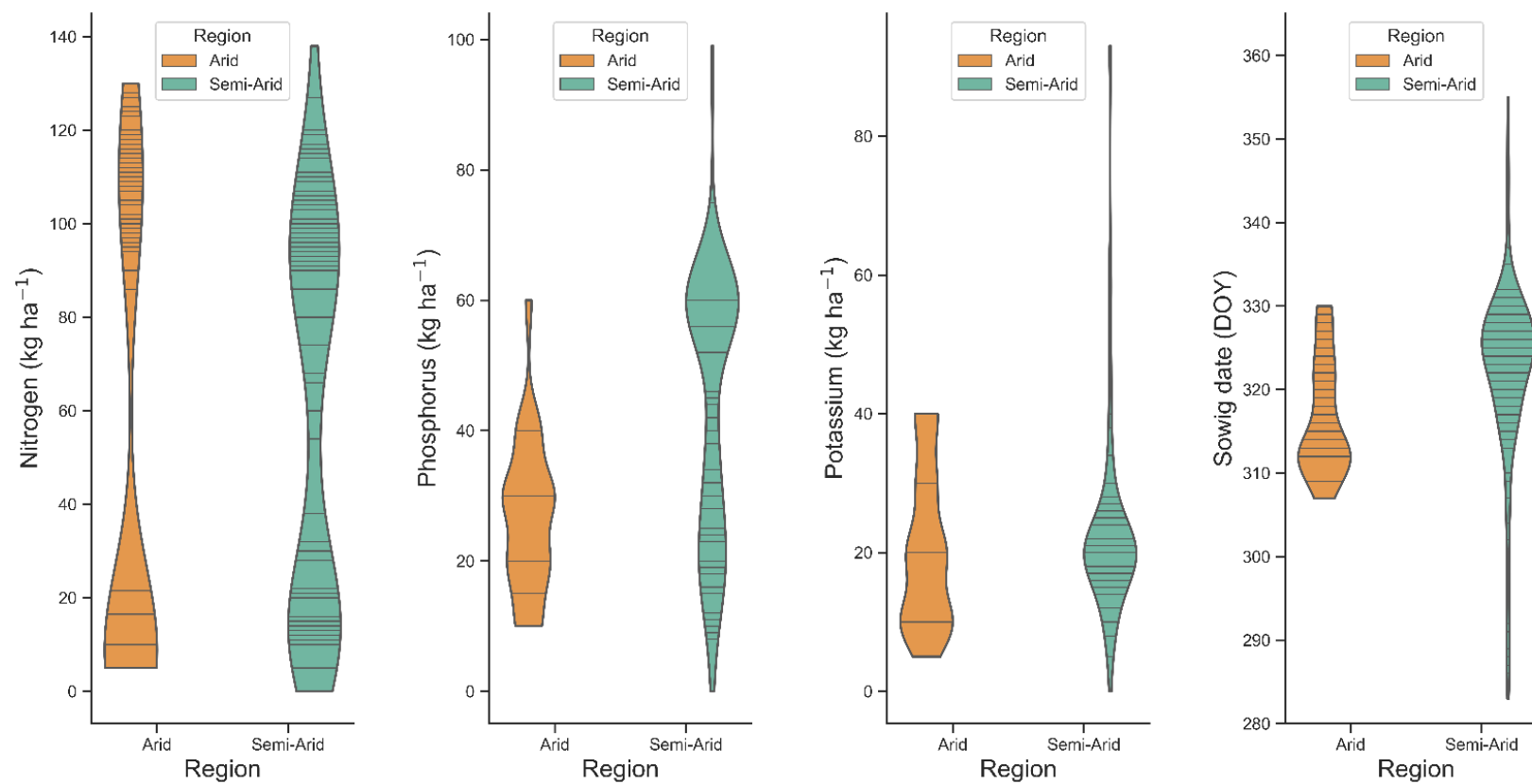


Figure 16 Distribution of applied fertilizer (NPK) and sowing dates in arid and semi-arid sites (2020–2021 season).

3.2.3 Satellite data

3.2.3.1 Satellite imagery extraction

Level-2A orthorectified atmospherically corrected surface reflectance satellite imagery was extracted from Sentinel-2 sensor and processed using GEE Python API (Gorelick et al., 2017b). Sentinel-2 is a wide-swath, high-resolution, multi-spectral imaging mission supporting Copernicus Land Monitoring studies. The multi-spectral instrument (MSI) has four 10-m bands, six 20-m bands, and three 60-m bands (Table 11). The automated workflow started with an extraction of an image collection covering the crop growth cycle of our study sites (September 2020 to June 2021). A bitmask band was used to mask pixels that contains cirrus and high clouds, then a buffer of 10 m was applied to avoid the edges that mislead the calculation of vegetation indices (Vannoppen & Gobin, 2021). The bands used in this study are shown in Table 11.

Table 11 Sentinel-2 MSI L2A band characteristics for both S2A and S2B sensors
The spectral bands used in this study are in bold.

Band	Spatial resolution (m)	Wavelength (nm)		Description
		S2A	S2B	
B1	60	443.9	442.3	Aerosols
B2	10	496.6	492.1	Blue
B3	10	560.0	559.0	Green
B4	10	664.5	665.0	Red
B5	20	703.9	703.8	Red Edge 1
B6	20	740.2	739.1	Red Edge 2
B7	20	782.5	779.7	Red Edge 3

B8	10	835.1	833.0	Near infrared
B8A	20	864.8	864.0	Red Edge 4
B9	60	945.0	943.2	Water vapor
B11	20	1613.7	1610.4	Short-wave infrared 1
B12	20	2202.4	2185.7	Short-wave infrared 2

3.2.3.2 Smoothing and phenological measures extraction

Five vegetation indices were calculated from the satellite imagery. A piecewise cubic spline smoother (Wahba, 1975) was used to smooth the computed temporal profiles. Hence, for n observations during the growing season, the vegetation indices were interpolated in $n - 1$ periods to fill the gaps and their corresponding values at time t in the i th period ($VI_i(t)$) is generated by the function:

$$VI_i(t) = a_i(t - t_i)^3 + b_i(t - t_i)^2 + c_i(t - t_i) + d_i \quad (3-1)$$

where a_i , b_i , c_i , d_i are the piecewise cubic spline parameters at the i^{th} period.

Key phenological phases in wheat were identified using first and second derivatives of the smoothed temporal profiles (see Figure 17). The peak of the canopy (where the canopy is assumed to be densest) was attained when the first derivative was found to be equal to zero, the maximum rates of increase and decrease during green-up and senescence phases were determined by identifying the global maximum and minimum of the first derivative of the temporal profiles (Bandaru et al., 2020), while maximum second derivative values during the vegetative and the reproductive phases were used as indicators of the start of the green-up and end of crop senescence respectively (Tan et al., 2011; H. Xu et al., 2014; Y. Yang, Tao, et al., 2021). Four distinct early yield prediction stages were established. *Stage 1* encompasses the period between the start of green-up and the maximum growth rate. *Stage 2* extends from the maximum increase

rate to the peak of the canopy. *Stage 3* spans from the day of the peak of the canopy to the maximum rate of decrease. Finally, *Stage 4* covers the interval from the day when the crop reaches its maximum rate of decline to the end of senescence as shown in Figure 17. Then, Trapezoidal method (Tai, 1994) was used for each site to estimate the area under the smoothed temporal profiles between the key crop growth stages.

Table 12 Vegetation indices used in this study, where NIR, RED, GREEN and SWIR1 correspond to Sentinel's B8, B4, B3 and B11, respectively.

Index	Abbreviation	Definition	Reference
Normalized Difference Vegetation Index	NDVI	$\frac{NIR - RED}{NIR + RED}$	Y. Xu et al. (2022)
Green Normalized Difference Vegetation Index	GNDVI	$\frac{NIR - GREEN}{NIR + GREEN}$	Psiroukis et al. (2022)
Normalized Difference Moisture Index	NDMI	$\frac{NIR - SWIR1}{NIR + SWIR1}$	Viswambharan et al. (2022)
Pigment Specific Simple Ratio	PSSR	$\frac{NIR}{RED}$	Tao et al., (2020)
Soil-Adjusted Vegetation Index	SAVI	$\frac{(1 + L) * (NIR - RED)}{(NIR + RED + L)}$	(Huete, 1988)

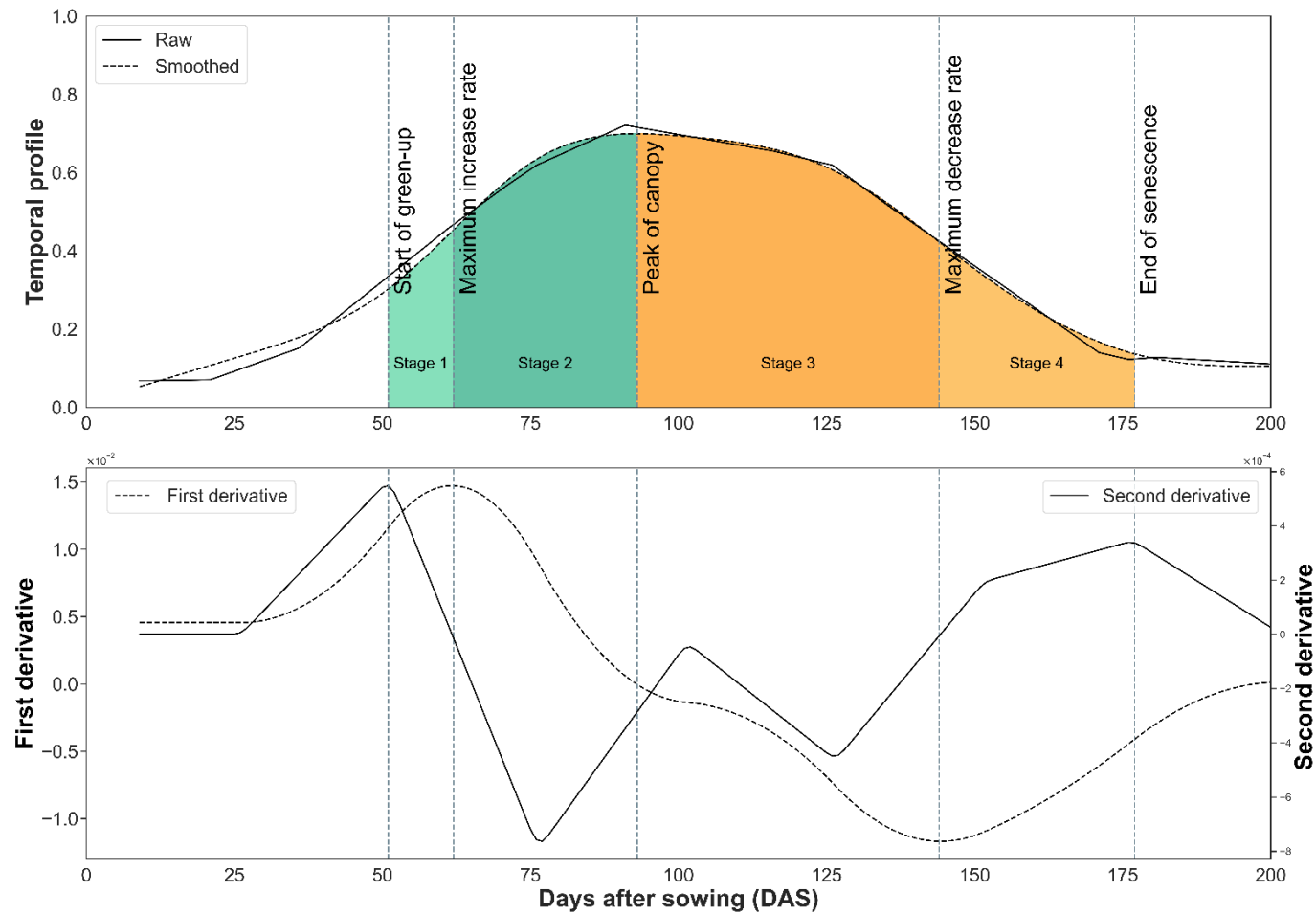


Figure 17 Vegetation index smoothing and phenology extraction methodology

3.2.3.3 Weather data

To assess the effect of weather on the yield predictions, weather variables from nearby stations were used to calculate the accumulated rainfall and accumulated growing degree days for each site at the four stages previously described. Growing degree days (GDD) were calculated by:

$$GDD = T_{\text{avg}} - T_{\text{base}} \quad \text{where} \quad T_{\text{avg}} = \frac{T_x + T_n}{2} \quad (3-2)$$

where T_{base} is the temperature below which crop development is assumed to be suspended, T_x and T_n are the daily maximum and minimum air temperatures. We bounded T_{avg} between 0 and T_{upper} as advised by Raes et al. (2022). For T_{base} and T_{upper} , we selected the values 5°C and 33°C, respectively, which are widely used for Moroccan wheat crops in our study area (Toumi et al., 2016a). The set of variables used in the predictive models are reported in Table 13. The table covers features specific to different growth stages, such as the five vegetation index values (VI) and cumulative data at key phenological stages, as well as stage-independent attributes, including information about the site's topography and the management practices related to fertiliser application in each site.

Table 13 Detailed list of features used for model development.

Feature Class	Features Type	Feature	Description
Stage specific	VI values	VI _{Greenup}	Vegetation index value at the start of green-up
		VI _{MIR}	Vegetation index value at maximum increase rate
		VI _{Peak}	Vegetation index value at peak of canopy
		VI _{MDR}	Vegetation index value at maximum decrease rate
		VI _{Senescence}	Vegetation index value at end of senescence
	Cumulative values at Stage 1	VI _{Stage1}	Cumulative vegetation index from the start of green-up to maximum increase rate
		GDD _{Stage1}	Cumulative GDD from the start of green-up to maximum increase rate (°C)
		Rainfall _{Stage1}	Cumulative rainfall from the start of green-up to maximum increase rate (mm)

Cumulative values at Stage 2	VI _{Stage2}	Cumulative vegetation index from the maximum increase rate to the peak of canopy
	GDD _{Stage2}	Cumulative GDD from the maximum increase rate to the peak of canopy (°C)
	Rainfall _{Stage2}	Cumulative rainfall from the maximum increase rate to the peak of canopy (mm)
Cumulative values at Stage 3	VI _{Stage3}	Cumulative vegetation index from the peak of the canopy to the maximum decrease rate
	GDD _{Stage3}	Cumulative GDD from the peak of the canopy to the maximum decrease rate (°C)
	Rainfall _{Stage3}	Cumulative rainfall from the peak of the canopy to the maximum decrease rate (mm)
Cumulative values at Stage 4	VI _{Stage4}	Cumulative vegetation index from the maximum decrease rate to the end of senescence
	GDD _{Stage4}	Cumulative GDD from the maximum decrease rate to the end of senescence (°C)

		Rainfall _{Stage4}	Cumulative rainfall from the maximum decrease rate to the end of senescence (mm)
Stage independent	Topography	Slope	Slope of the site (degree)
		Elevation	Elevation of the site (m)
	Fertilizer management	N	Applied Nitrogen (kg ha ⁻¹)
		P	Applied Phosphorus (kg ha ⁻¹)
		K	Applied Potassium (kg ha ⁻¹)

3.3 Methodology

3.3.1 Feature selection

Machine learning models require distinct training and testing datasets (Shobha & Rangaswamy, 2018). We used a stratified split on yield values to ensure a representative train-test distribution (80%:20%). This maintained a balanced representation of yield in both training and testing sets, ensuring more reliable model evaluations (Rendleman et al., 2022; Singh et al., 2021).

For feature selection, we first assessed collinearity between of features. By setting a collinearity threshold of 0.85, we identified pairs of features with correlation coefficients meeting or exceeding this value, indicating a strong linear relationship between them. In such cases, we retained only the feature with the higher absolute correlation with the yield, as this mitigates potential overfitting and instability in our predictive models.

To remove those features with marginal effect, we used Forward Feature Elimination (FFE). This selection algorithm enabled us to systematically identify the features that most influence crop yield. By iteratively adding and removing features based on their impact on model performance, we aimed to maximize the predictive accuracy of the models while avoiding overfitting. The process was conducted for each yield prediction stage. The performance of the models was evaluated using the Mean Absolute Percentage Error (MAPE) on the corresponding test sets.

3.3.2 Yield modelling techniques

Three modelling schemes were used: multiple linear regression, support vector regression, and a random forest model. All of them shared the same set of features at each specific prediction stage. Multiple Linear Regression (MLR) was used to model the relationships between the different independent and dependent variable stated in Table 13. *MLR* estimates coefficients for each independent variable, revealing their individual contributions and directions of

impact. We fitted the *MLR* model using the Scikit-learn package in Python (Pedregosa et al., 2011) without considering interactions between features. *MLR* served as a baseline model for our analysis due to its simplicity and interpretability. Support Vector regression (*SVR*) is a robust machine learning algorithm. It is designed specifically for regression problems and was introduced as an extension of Support Vector Machines (*SVMs*). Initially, it transforms the features into a higher dimensional feature space using a kernel function that could be linear (Bukowiecki et al., 2020), polynomial (Padierna et al., 2018), radial-basis function (Izonin et al., 2021) or sigmoid (Jebadurai & Peter, 2017). This allows the model to learn non-linear relationships between the dependent and independent variables, in contrast to *MLR*. *SVR*'s robustness against outliers and noise further enhances its suitability for noisy datasets. It defines an epsilon-tube around the predicted values and minimises the errors within this tube. This ensures that the model is more resilient to outliers and noise. The solution to the constrained optimization problem generates the coefficients of the hyperplane in the transformed space. Three hyperparameters were examined in the *SVR* tuning. The first consideration was the kernel type (*K*). We investigated four unique kernels to describe data relationships. In addition to kernel selection, we established ranges for two critical hyperparameters. For the regularisation parameter (*C*), which balances margin maximisation and error reduction, the range was from 1×10^{-4} to 1. For the hyperparameter (ϵ), which represented the margin of error tolerated in the *SVR* model, the range was set to vary from very tight margins (e.g., 1×10^{-10}) to more relaxed margins (e.g., 1), providing flexibility in adjusting the model's margin width to suit specific data characteristics and prediction stage. The equation for the hyperplane in *SVR* is given by:

$$f(x) = \sum_{i=1}^n \alpha_i K(x, x_i) + b \quad (3-3)$$

where $f(x)$ is the predicted yield, α_i are the Lagrange coefficients associated with each support vector, $K(x, x_i)$ is the kernel function and b is the bias term.

The third modelling approach was Random Forests (*RF*), which is an ensemble learning method formulated by Breiman (2001) for regression and classification tasks. The method builds a large number of decision trees during a training phase and their predictions are then combined. The approach involves a sequence of four steps. First is the bootstrap resampling phase, which generates multiple bootstrapped training sets by arbitrarily selecting samples from the initial dataset. Next, the phase of random feature selection begins. At each node of the decision tree, a random subset of the features (F_{subset}) is selected to prevent excessive reliance on a single feature. In the third stage, unique decision trees (T_i) are generated for each bootstrapped training set. The optimal feature splits (S_i) for these trees are determined by metrics such as Gini coefficients and mean squared error for classification or regression tasks, respectively. The final stage is where the process of ensemble aggregation begins. Here, individual decision tree predictions combine to form the final prediction using the equation:

$$P_i = \frac{1}{T} \sum_{t=1}^T P_i^t \quad (3-4)$$

where for a given sample i , P_i is the final aggregated prediction, T is the number of trees and P_i^t is the prediction of the decision tree t . *RF* operates with a distinct set of hyperparameters customized to its ensemble of decision trees. These settings, or hyperparameters, included the number of decision trees, the complexity of individual trees, and the minimum sample requirements for node splitting and leaf designation. This adjustment process ensures that the *RF* model is able to capture complex data relationships between our features and the final yield, while preventing overfitting, and delivering accurate predictions across different time leads.

3.3.3 Hyperparameter tuning

Hyperparameter tuning is an important step to avoid model overfitting, improve model performance and generalisability (L. Yang & Shami, 2020). In our case, our objective was to identify the hyperparameter combination H_i that optimizes the performance metric P_i (i.e., *MAPE* at each prediction stage). We began by defining a range of variation of our hyperparameters (Table_Apx C.2-1), then conducted a systematic grid search to explore all possible combinations. For each combination, we constructed Support Vector Regressor (SVR) or Random Forest (RF) and evaluated their performance using 10-fold cross validation procedure, in which the training set is divided into ten subsets, each serving as a test set while the rest are used for the training (Figure_Apx C.3-1).

3.3.4 Model Evaluation

Model evaluations were carried out to assess performances and generalization abilities using an independent test dataset. The coefficient of determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (3-5)$$

served as a measure to quantify the extend of linear relationships between the observed and predicted yield, where n is the number of observations, O_i , $\bar{O}$ and P_i are the observed, mean of observed and predicted rainfed wheat yield values, respectively. The Mean Absolute Error (MAE) was used to measure the average deviation between the predicted and the actual yield values according to

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - P_i| \quad (3-6)$$

The Root Mean Square Error (RMSE) was used to assess the overall discrepancy between the actual and the predicted yield according to

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (3-7)$$

and the Mean Absolute Percentage Error (MAPE) was calculated to express errors as a percentage of the actual yield values

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{O_i - P_i}{O_i} \right| \quad (3-8)$$

3.3.5 Percentage feature contribution

We used SHAP (Lundberg & Lee, 2017) to determine the proportional contributions of the features to yield predictions. This analysis was conducted after the feature selection and model tuning phases were completed. For each feature j at each site i , we calculated the SHAP value $S_i(j)$, representing the magnitude and direction of its influence on our model's predictions. Subsequently, we calculated the Mean absolute SHAP value (MASV) for each feature j , which provided an average measure of the feature's influence on the yield predictions.

$$MASV_j = \frac{1}{n} \sum_{i=1}^n |S_i(j)| \quad (3-9)$$

To further assess the relative importance of each feature, we determined their percentage contributions concerning the total MASV of all features. The formula for the percentage contribution C_j of a feature j is given by:

$$C_j = \frac{MASV_j}{\sum_{i=1}^m MASV_i} \quad (3-10)$$

where m is the total of number of selected features in the arid, semiarid, and global models. This calculation allowed us to rank the features by their relative significance.

3.3.6 Transferability and generalisation

To provide a comprehensive analysis of model generalization and transferability in varying climatic conditions, we investigated six distinct scenarios. First, we tested the models developed for arid and semiarid conditions with their respective test data sets. We refer to these as *A2A* and *SA2SA*. These act as baseline metrics for the model performance. To test model transferability, we tested the performance of the model developed for arid conditions on the semi-arid test data set (*A2SA*), and conversely the semi-arid models were tested on the arid test data (*SA2A*). Lastly, to assess model generalisation, we tested the performance of the global models (trained on both arid and semi-arid data) on the semi-arid and arid test sets separately. We refer to these as *G2SA* and *G2A*, respectively. The percentage of change in *RMSE* (*PC*) was used to quantify the extent of deterioration in the performance of our transferred models as shown in equation:

$$PC = 100 * \left(\frac{RMSE_{Target} - RMSE_{Base}}{RMSE_{Target}} \right) \quad (3-11)$$

where $RMSE_{Base}$ is the root mean squared error of the models trained and tested in the same climate (i.e., *A2A* and *SA2SA*) while $RMSE_{Target}$ is the root mean square error of the models trained in one climate and then tested in a different climate (i.e., *A2SA*, *SA2A*, *G2SA* and *G2A*).

3.4 Results

3.4.1 Feature selection

Figure 18 shows the correlations between selected features and yield in the global, semiarid, and arid data sets. Only non-correlated features, determined by a threshold of 0.85, were chosen for the model development. In arid climates (Figure 18.a) there was a negative correlation between the growing degree days (specifically GDD_{stage3} and GDD_{stage4}), and wheat yield. There was a positive correlation between seasonal rainfall effect (i.e., $Rainfall_{Season}$) and yield. This was stronger in semi-arid climates (Figure 18.b) where rainfall was more abundant. There were negative correlations between yield and vegetation indices

related to early green-up (i.e., $GNDVI_{Greenup}$ and $NDVI_{Greenup}$). For semi-arid climates (Figure 18.b), temperature at the beginning of the season (GDD_{Stage1}) showed a relatively weak positive correlation with yield. As the season progresses towards its peak, the impact of the temperature on the yield became more pronounced, indicating that warmer conditions at this stage are more favourable for wheat production. Rainfall during the season ($Rainfall_{Season}$) had a substantial influence on the yield, with a particularly strong correlation before the peak of the canopy ($Rainfall_{Stage2}$), but the rainfall effect became weaker in the later stages.

Correlation analysis for the global dataset revealed a strong positive correlation between the final yield and the seasonal cumulative rainfall and temperature ($Rainfall_{Season}$ and GDD_{Season}). Somewhat predictably, fertiliser application was found to have a positive correlation with yield. High cumulative $GNDVI$ values between the maximum decrease rate and senescence were negatively linked with yield production. However, high cumulative Green Normalized Difference Vegetation Index values between the maximum decrease rate and senescence stages were negatively linked with yield production. This suggests that in these specific growth stages, excessive green vegetation may not necessarily translate to higher wheat yield. In-depth details on the relationships between the features are provided in Appendix C.1.

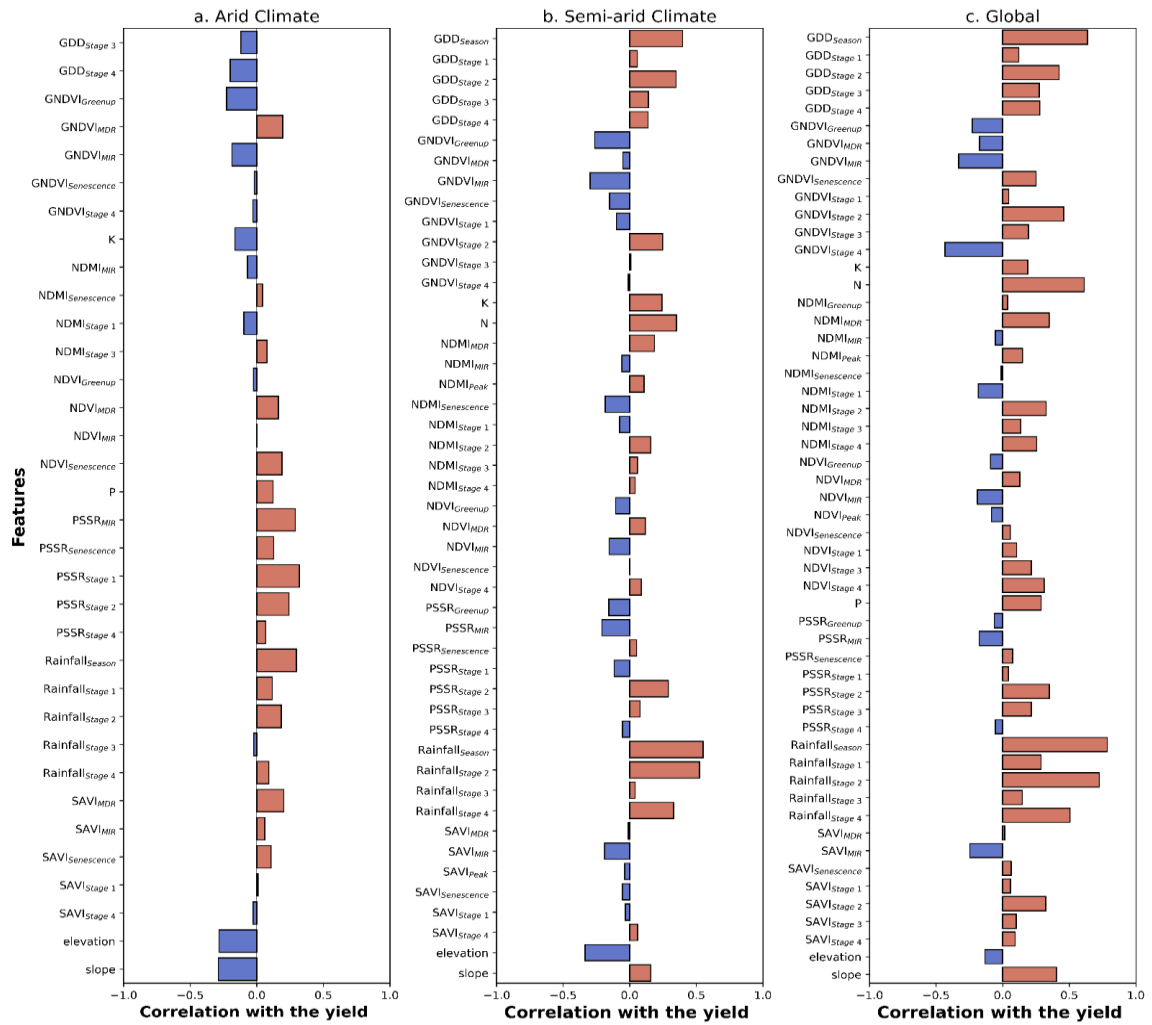


Figure 18 Correlation with the yield and correlation based selected features in arid, semiarid, and global datasets.

The recursive feature elimination with cross-validation (RFECV) analysis revealed insights into main factors to be used for yield prediction (Figure 19). In arid regions, the number of selected features gradually increases from 8 to 12 across the four stages, signifying the growing relevance of diverse factors as wheat crops progress through their growth stages. In semiarid regions, the number of selected features remains relatively stable from Stage 1 to Stage 2 but increased to 20 features in Stage 4, suggesting the growing complexity and sensitivity of crop-environment interactions in later stages. The same trend was seen in the global dataset, but with a gradual increase in selected features from 12 to 19.

The Venn diagrams (Figure 20) show the numbers of features that overlap between arid and semi-arid climates across four growth stages. At all prediction stages, key shared factors include Potassium (K), Slope, and vegetation indices values at the maximum increase rate (i.e., NDMI at stages 1 to 3 and PSSR at stage 4). Unique features specific to semi-arid climates included Nitrogen (N) and GDD_{Stage2} . In contrast, specific features to arid climates cover Phosphorus (P), Elevation, $Rainfall_{Stage1}$, $Rainfall_{Stage2}$, and phenology features extracted from SAVI ($SAVI_{Stage1}$, $SAVI_{Stage4}$, $SAVI_{MIR}$, $SAVI_{MDR}$).

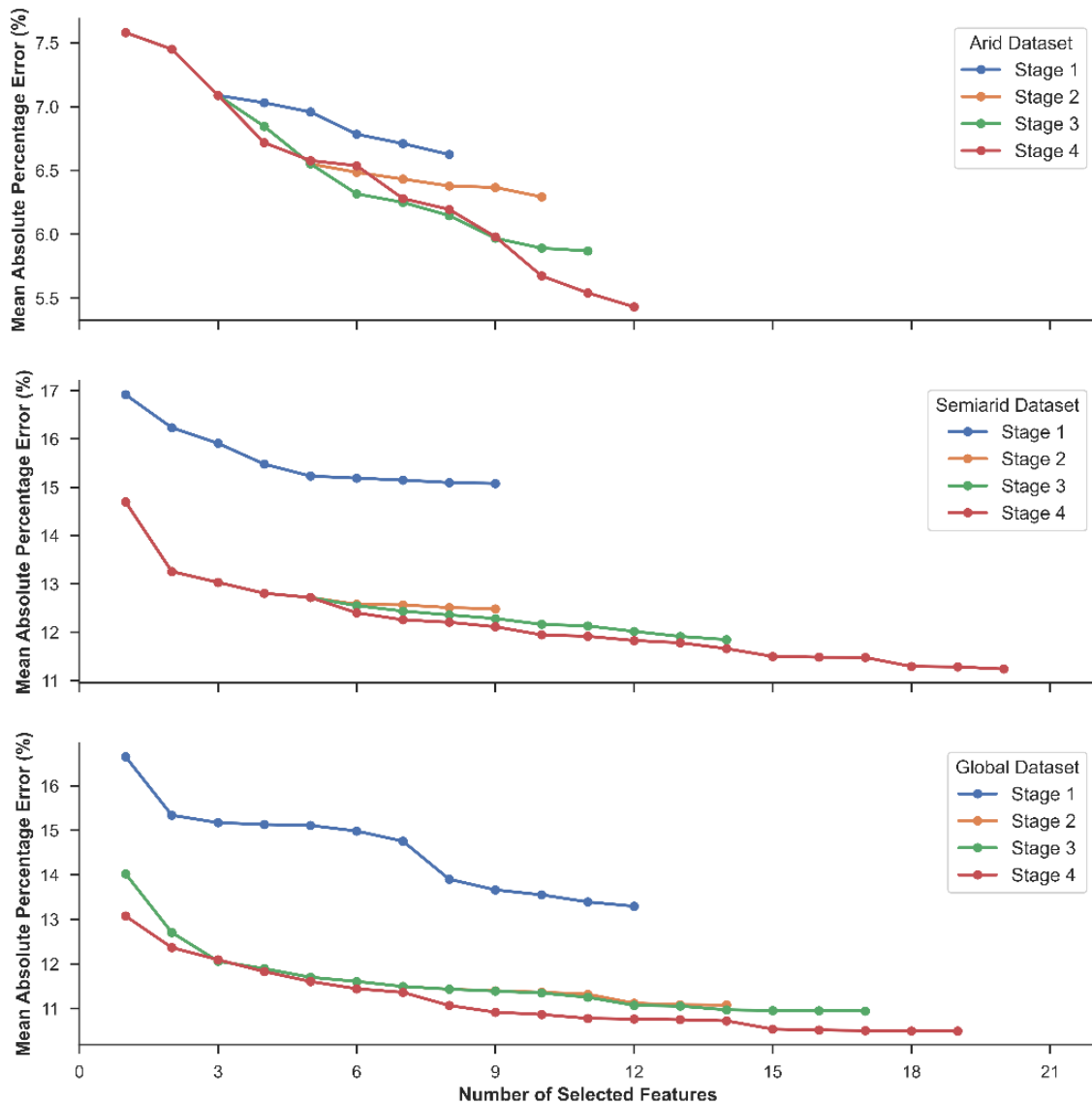


Figure 19 Stage-specific recursive feature elimination for yield prediction in the arid, semi-arid and global datasets.

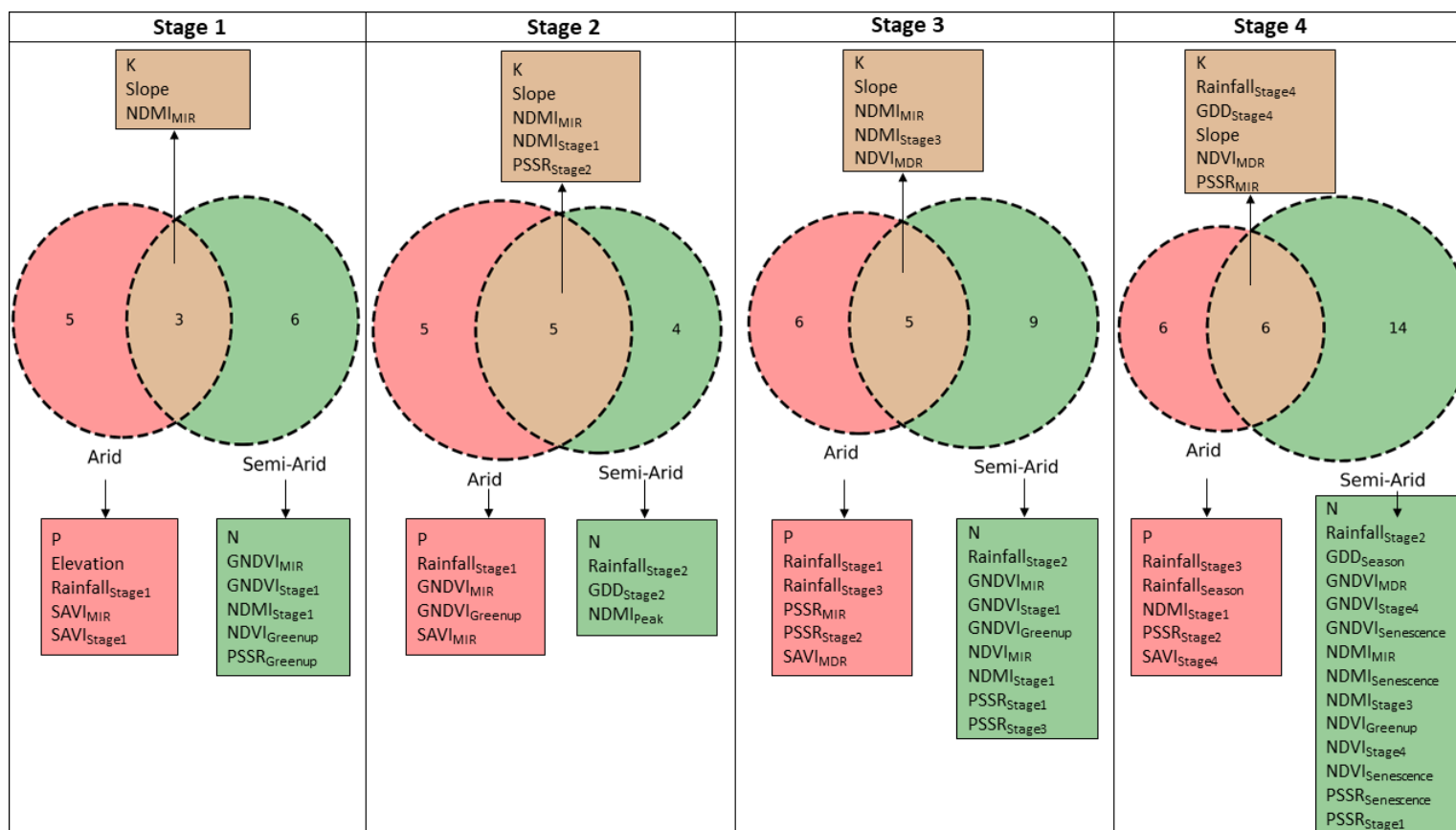


Figure 20 Venn diagrams showing the unique and shared features between arid and semi-arid climate at the four prediction stages.

3.4.2 Model evaluation

At the first prediction stage in the arid climate (Figure 21.c), Multiple Linear Regression (*MLR*) demonstrated the lowest RMSE of 0.30 t ha⁻¹. However, as we progressed through the stages and more features were available, Support Vector Regression (*SVR*) consistently exhibited an enhanced performance. Predictions at the senescence (i.e., stage 4) using *SVR* achieved the lowest RMSE of 0.23 t ha⁻¹. In contrast, Random Forest (*RF*) had slightly higher RMSE values at all stages within this climate category.

A substantial shift in performance patterns becomes evident in the semiarid region (Figure 21.b). Throughout all prediction stages, Random Forest (*RF*) emerged as the most accurate model. This model demonstrated an improvement in predictive accuracy, transitioning from an initial RMSE of 0.62 t ha⁻¹ at Stage 1 to 0.5 t ha⁻¹ at Stage 4.

The global models (Figure 21.a) shared a similar pattern to the localised semiarid models. Random Forest (*RF*) once again emerged as the most accurate model across all prediction stages. Starting at an RMSE of 0.68 t ha⁻¹ at the earliest prediction stage, *RF*'s accuracy steadily improved to reach a lower RMSE of 0.46 t ha⁻¹ when predictions were made at the senescence (i.e., Stage 4).

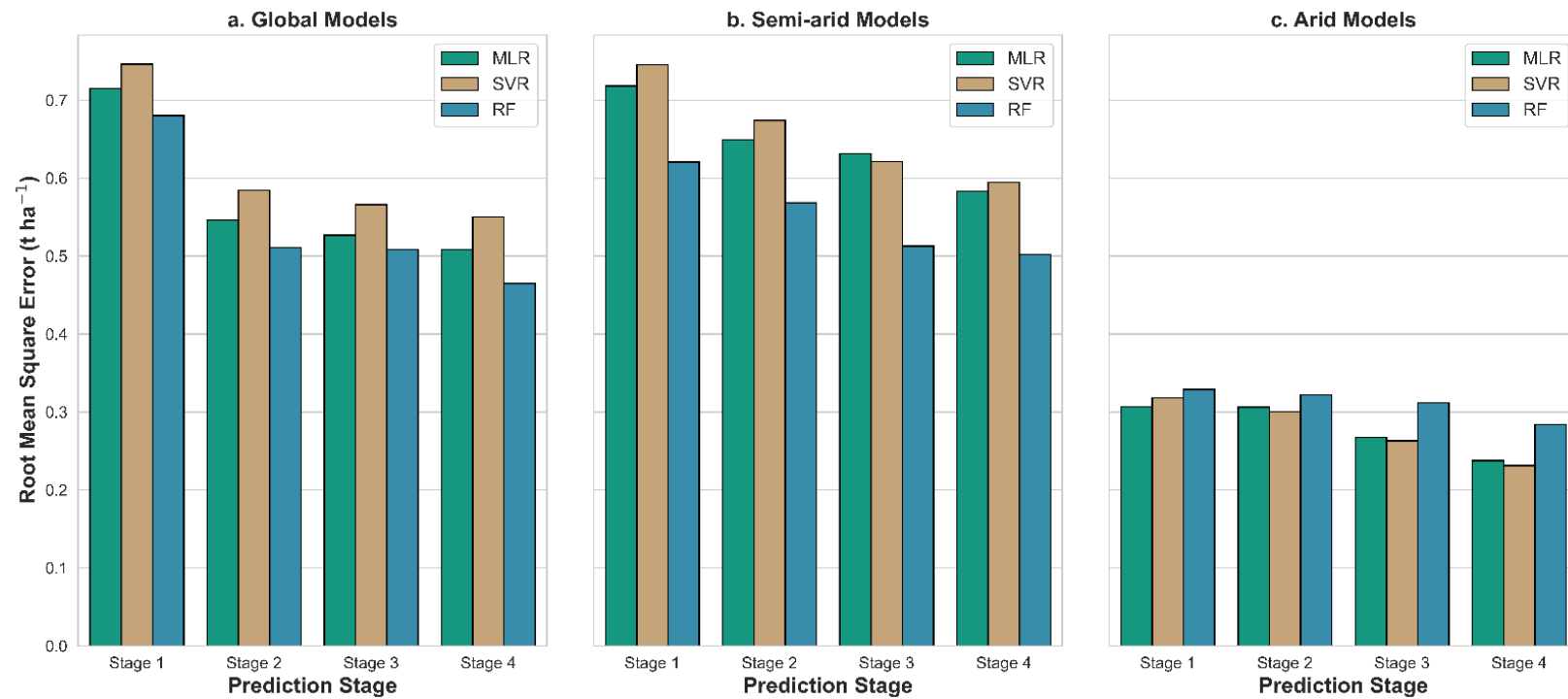


Figure 21 Stage-specific evaluation of the global (a) and localised models (i.e., b and c).

3.4.3 Feature contribution

There were substantial differences in the contributions of features to the predictions across the four stages (Figure 22). At the first prediction stage of the semi-arid models, $GNDVI_{MIR}$ had the greatest contribution (31.1%), closely followed by applied nitrogen (25.4%). In contrast to the arid models, topographical, soil related vegetation indices and fertilizer attributes came to the fore, with $slope$, P , $SAVI_{MIR}$ and K emerging as the most influential features (contributions of 24.30%, 17.04%, 15.94 % and 13.07%, respectively). Contribution analysis of *Stage 2* arid models consistently underscored the importance of $slope$, P , and K , with respective contributions of 20.83%, 13.56%, and 12.26%. respectively. In the other *Stage 2* models (i.e., semiarid, and global), $Rainfall_{Stage2}$ emerged as a dominant feature, with a substantial contribution of 45.07 and 51.81% of semiarid and global model predictions. Transitioning to *Stage 3*, the contribution of key important features to the models' predictions was almost similar to *Stage 2* for semi-arid and global models. In the arid climate, $slope$, P , and K remained important, with contributions of 17.95%, 12.36%, and 12.17%, respectively. In contrast, the semi-arid climate models were influenced by $Rainfall_{Stage2}$ (46.98%). Globally, $Rainfall_{Stage2}$ was a dominant factor, contributing 50.45%, along with $slope$ and N at 19% and 11.55%, respectively, highlighting the impact of climate patterns, topography, and nitrogen applications on global yield predictions. Finally, in *Stage4*, $slope$, P , and K contributions of 17.58%, 11.47%, and 10.08%, respectively, were observed in arid climates alongside with $NDVI_{MDR}$. In semi-arid regions, $Rainfall_{Stage2}$ contributed 39.42% to predictions, while global models highlighted $Rainfall_{Season}$ (50.45%), $Rainfall_{Stage2}$ (14.36%) and Nitrogen (N) at *Stage4* (11.55%), respectively, as key contributors to yield predictions.

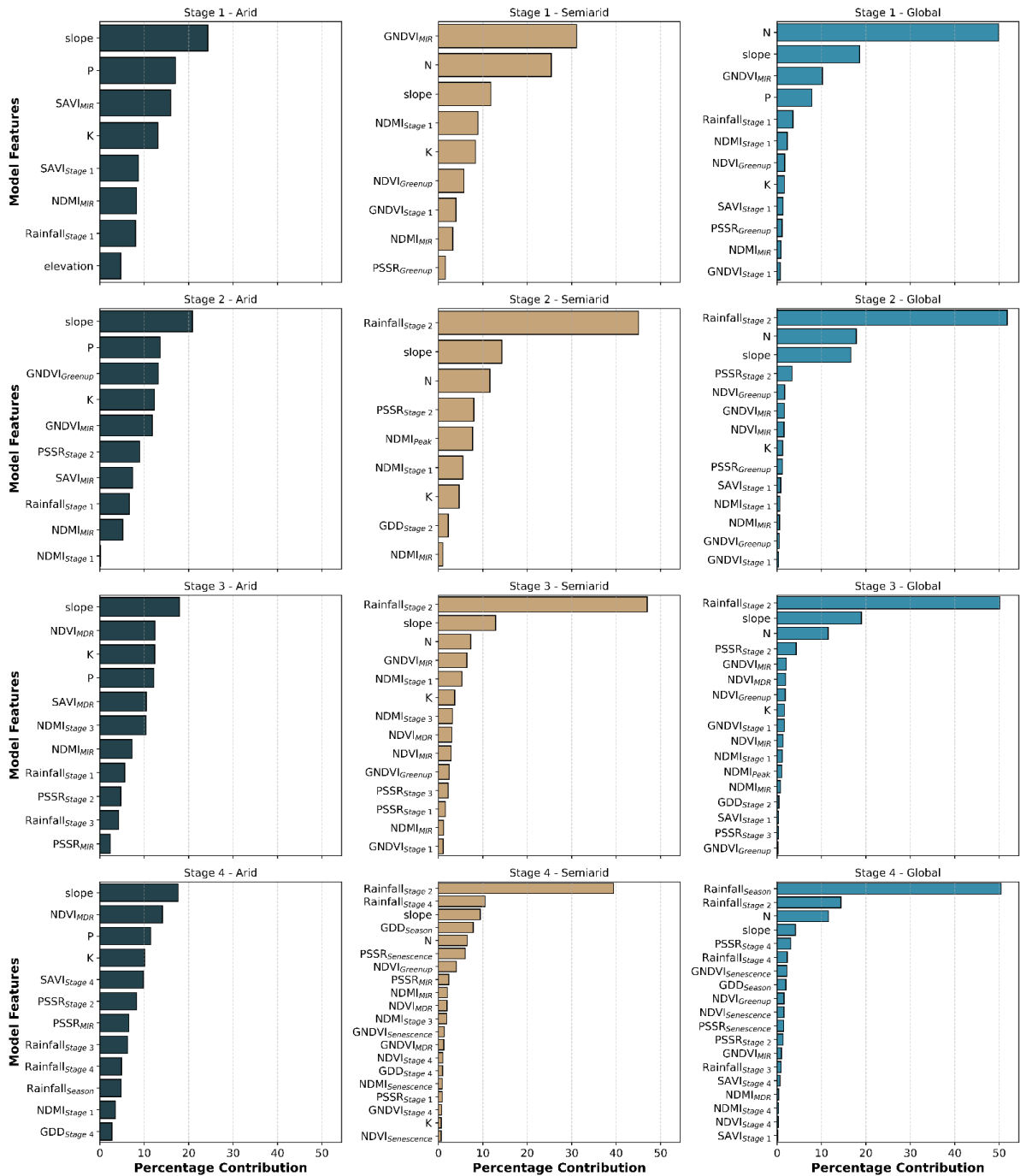


Figure 22 Percentage contribution of each feature to the yield prediction at 4 stages in arid (dark blue), semi-arid (brown) and global (light blue) models based on SHAP values. Features are ranked from highest to lowest contributors.

3.4.4 Transferability and generalisation analysis

In the baseline scenarios (A2A and SA2SA), the models exhibited strong performance across all growth stages, with *RMSE* values ranging from 0.23 to 0.30 t ha⁻¹ in the arid climate and from 0.50 to 0.62 t ha⁻¹ in the semiarid climate. However, when examining the transferability of the models, a substantial decrease in accuracy was evident compared to the baseline scenarios, with *RMSE* values ranging from 0.64 to 1.12 t ha⁻¹ in the SA2A scenario and from 1.53 to 2.48 t ha⁻¹ in the A2SA scenario (Figure 23). While the accuracy of baseline scenarios increased with advancing prediction stages, this trend was not observed in the transferred models. Regarding the generalised scenarios, some G2SA cases achieved lower *RMSE* values than SA2SA baselines, especially with *MLR* predictions from the peak of the season onward, along with *SVR* and *RF* predictions at the canopy's peak (i.e., Stage 2). However, baseline models consistently outperformed generalised models in the arid climate across all prediction stages. Generally, although generalised models demonstrated lower performance compared to baseline models, they still outperformed the transferred models across all prediction stages. Further information on the percentage change in *RMSE* of the transferred and generalised models can be found in appendix C.5 (Figure_Apx C.4-2).

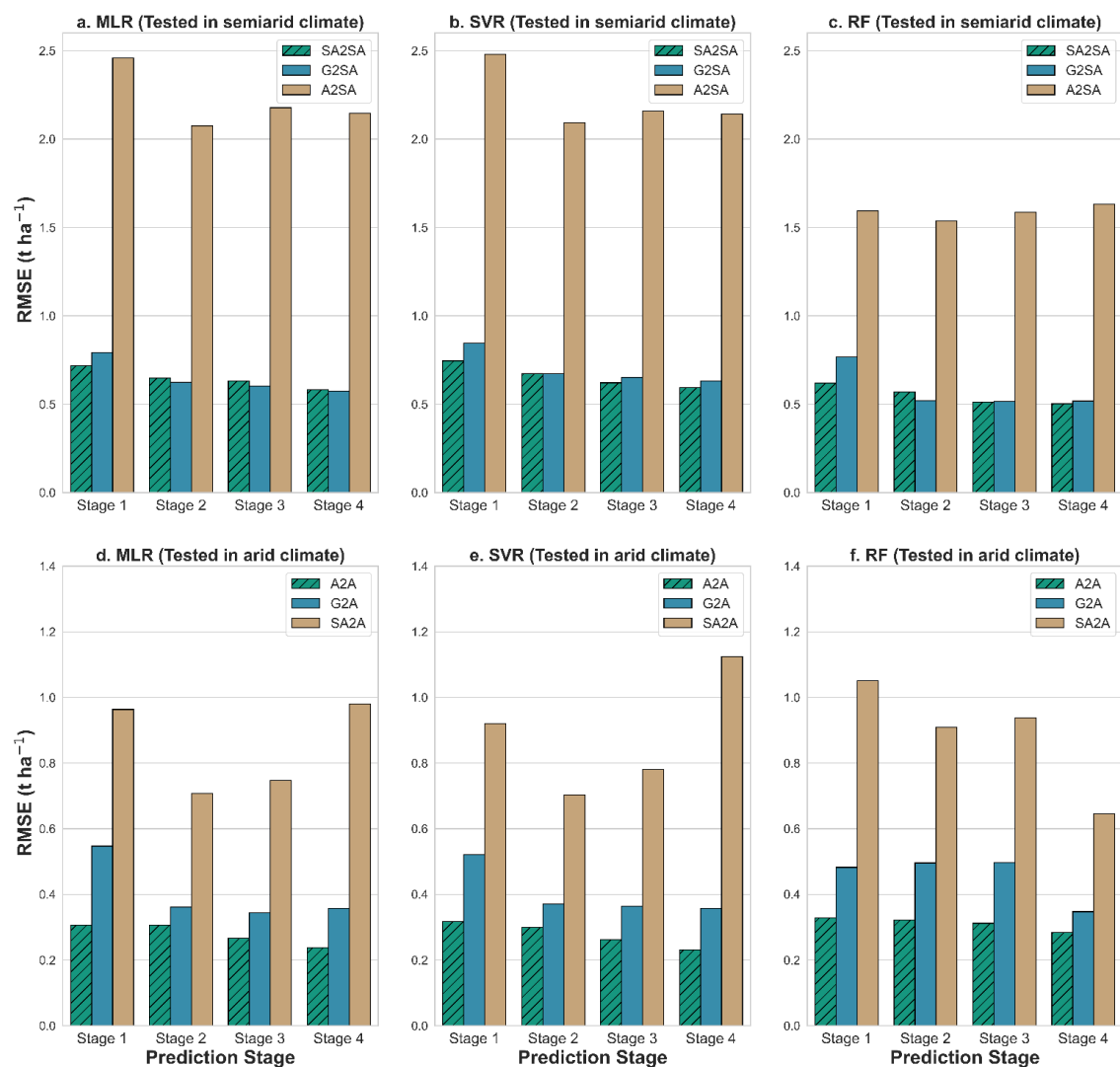


Figure 23 Transferability and generalisation evaluation metrics for MLR, SVR and RF trained to predict the yield at four different prediction stages. SA2SA and SA2A represent models trained in semi-arid and tested in semi-arid and arid regions, respectively.

3.5 Discussion

3.5.1 Key factors contributing to yield prediction across varying climates

For the semi-arid regions, where water availability is relatively higher compared to arid region, we observed a positive correlation between GDD for all four of our prediction stages (Figure 18.b), suggesting that in this region wheat production benefits from higher temperatures. This aligns with research highlighting the positive effect of increased temperature on crop yield in semi-arid regions (Vashisth et al., 2020). In contrast, we found that the arid regions exhibited a negative correlation between cumulative GDD during reproductive phases (Stages 3 and 4) and yield, indicating that excessive heat during this critical period reduced yield in arid climates. Previous studies have highlighted the sensitivity of wheat to high temperatures, for instance, Al-Karaki (2012) reported a field study conducted under rainfed conditions and determined that grain yield was positively correlated with GDD accumulated until to the heading but negatively correlated with cumulative GDD to maturity. Here, the influence of precipitation on crop yield varied throughout growth stages, showing the most substantial positive impact during Stages 1 and Stage 2. This highlights the critical role of water in the early phases of crop development, emphasizing the need for adequate water supply during this initial period. These findings align with a study by Yu et al. (2014), who examined historical wheat data from 1889 to 2004 in Australia, which indicated that precipitation during the vegetative stage played a dominant role in enhancing wheat yields. Our results also revealed that elevated values of specific vegetation indices, including GNDVI, NDVI, and PSSR, during wheat green-up (Figure 18.c) were linked to decreased wheat yields. This association could indicate an excessive vegetative growth, resulting in an increased transpiration of the crop which was also underlined by a study by Panek & Gozdowski. (2020) in Germany. In areas with limited water resources, this increased water demand exerts stress on the crop, which eventually results in diminished yields (Lake & Sadras, 2021).

Our examination of feature contributions using SHAP values reveals critical insights into the complex relationships between various features, going beyond simple correlation analysis. Feature contributions to yield in arid climates appear to be more evenly distributed, with several factors, such as slope, applied P and K, having similar percentage contribution in yield predictions. This suggests that in arid regions, various attributes collectively influence crop yields without a single feature dominating the predictions. In contrast, the semi-arid climate exhibits a striking reliance on Rainfall from Stage2 onward, contributing over 40% to yield predictions. This emphasizes the significant impact of rainfall during specific growth stages in semi-arid regions. Finally, our findings demonstrated that the number of selected features increases as the crop grows. This suggests that with the progression of crop development, more satellite imagery and additional features will become valuable for accurate yield predictions (Vavlas et al., 2020). Across all sites, it should be noted that the study focused exclusively on one wheat variety, i.e., Bandera. Given that the selected features are significantly affected by the temporal profiles of vegetation indices, it is important to emphasize that the feature set used in this study may not be suitable for other wheat varieties. Various wheat varieties display distinct temporal profiles as indicated by Khan et al (2018). Consequently, for precise yield predictions, it may be necessary to customize the features to the specific characteristics and growth patterns of wheat varieties with different durations.

The yield prediction framework developed in this study is influenced by several boundary conditions that constrain its applicability and generalizability. First, the analysis focused exclusively on a single wheat variety, Bandera, which limits the transferability of the results to other varieties with distinct temporal profiles and growth patterns. Variations in traits such as phenology, root structure, or canopy development among wheat varieties can alter the relevance of selected features, such as vegetation indices, and necessitate recalibration of the model. Second, the study sites were confined to semi-arid and arid regions, where precipitation, temperature, and soil moisture conditions differ from other agro-climatic zones. Consequently, the feature importance rankings and SHAP contributions identified

here may not hold in areas with different climate dynamics, such as cooler regions where heat stress during reproductive stages is not a dominant factor. Lastly, the framework does not explicitly account for nonlinear interactions or extreme events, such as heatwaves or droughts, which can cause abrupt shifts in crop growth and yield. Future research should address these limitations by expanding the dataset to include diverse wheat varieties, testing the model across multiple agro-climatic regions, and incorporating higher-frequency remote sensing data to capture rapid phenological changes and improve feature selection robustness.

3.5.2 Model performance across growth stages and climate zone suitability

Our research offers novel perspectives on how well prediction models function in different climatic zones. Generally, our models for arid regions performed better than other models with respect to RMSE but it should be noted that yields from the arid regions were generally lower. Prediction accuracies increased as the season progressed, which can be justified by the fact that more phenology metrics data became available. In semiarid climates, where more samples and features were available, complex models like Random Forests (*RF*) tended to outperform the other models (i.e., *MLR* and *SVR*). This is because *RF* can effectively leverage the abundance of data to capture the relationships between the features and the yield. Similar results were found by Cao et al. (2022) where *RF* outperformed linear models in predicting the yield in China because most relationships between yield and different variables were non-linear. Conversely, in arid climates, the limited availability of features made it challenging for *RF* to perform optimally. Multiple Linear Regression (*MLR*) excelled at early prediction stages outperforming the rest of the models. This can be attributed to the linearity between the features and the yield, which aligns with the nature of the *MLR* models we used. A similar finding was highlighted by Feng et al. (2022) where simple *MLR* models outperformed *RF* in estimating the growth index (*CGI*) based on UAV-*RGB* data. However, the performance of these models must be contextualised within the limitations of the training datasets. The data used in this

study were specific to semi-arid and arid regions of Morocco and focused on a single growing season, which may limit the generalisability of the findings to other regions or climatic zones. Further studies incorporating datasets from diverse geographic areas could validate the broader applicability of these models. Additionally, the reliance on data collected during a single season may not capture the interannual variability in weather and management practices, which may impact the model robustness. Future research should integrate multi-season and multi-region datasets to validate and enhance the broader applicability of these models. Several approaches can be explored to enhance the model performance and applicability. First, incorporating higher resolution satellite imagery could refine data quality and subsequently contribute to more precise models. Additionally, future research should focus on establishing links between RS-derived phenology metrics and actual wheat phenology, enhancing the models' biological relevance.

3.5.3 Model transferability

Our work has shown that transferring predictive models between climatic zones can lead to a significant decline in their performances. In the A2SA and SA2A scenarios (Figure 23), *RMSE* values exhibited an increase ranging from approximately 56.02% to 79.44% and 61.02% to 72.8% across all prediction stages when compared to their baseline semiarid and arid scenarios, respectively. The core issue lies in the distinct selected features and their contributions within these climatic zones. In semiarid climates, crop yield predictions relied heavily on rainfall-related features and nitrogen application, while in arid models, topographic features, *P* and *K* were more prominent. Additionally, the modelling methodology could restrict the extent to which the models can be transferred. Linear models, such as MLR, assume linear relationships between predictors and target variables, which may not accurately capture the underlying dynamics in domains with non-linear relationships (Matsumura et al., 2015). Moreover, fine-tuning RF hyperparameters to one specific area may also result in overfitting to training data, making them less effective on unseen climatic conditions. Findings from Jin et al. (2018) revealed

a 40% increase in *RMSE* when yield prediction models were applied to areas with different vegetation types. Another limitation in transferring predictive models across climatic zones is the risk of bias developed during training, particularly when the model is tuned to predict a specific range of yields. For instance, arid climate exhibited consistently lower yields (mean yield of 2.57 t ha⁻¹) compared to the semiarid climate with relatively higher yields (mean yield of 4.07 t ha⁻¹). If a model is exclusively trained on arid climate data, it will be biased towards this lower yield range, and may struggle to adapt when applied to semiarid climates with medium yields, resulting in inaccuracies in predictions.

Our research has shown that generalised models provide a viable solution to the problem of reduced accuracy in transferred models. Despite demonstrating lower performance compared to baseline models, generalised models still outperformed transferred models in all prediction stages. This is because generalised models incorporate a robust feature representation that is more adaptable to different climatic conditions (Figure 23). Integrating crop and machine learning models is another promising approach to address the challenges associated with models' transferability. By combining the strengths of crop models—which provide a mechanical understanding of crop growth—with those of data-driven machine learning methods, this method improves the model's ability to understand complex interactions between environmental factors and crop development in different regions. For instance, Shahhosseini et al. (2021) used *APSIM* to generate new features that were taken into account by six machine learning models, resulting in a decrease of the *RMSE* from 7 to 20%. Domain adaptation, as demonstrated by Ma et al. (2021) through their Adaptive Domain Adversarial Neural network (*ADANN*), effectively reduces the distribution discrepancy between base and target domains. By projecting data from two different domains into a common space, *ADANN* achieved higher performance compared to the random forest model. For instance, when transferring from the Eastern Temperate Forests (*ETF*) to the Great Plains (*GP*) in the U.S. corn belt, *ADANN* resulted in a *RMSE* of 1.23 t ha⁻¹, as compared to random forest model (*RMSE* of 1.78 t ha⁻¹).

3.5.4 Model generalisation

In arid climates, generalised models consistently demonstrated lower prediction accuracy across all prediction stages compared to baseline A2A models, showing an increase in prediction error ranging between 15% and 43% with a standard deviation of 9.2%. These findings suggest that while generalised models remain a practical choice in resource-intensive scenarios, where developing and maintaining climate-specific models is challenging, they did not effectively leverage the broad training domain and were consistently outperformed by climate-specific models. This performance shift can be attributed to two main factors. First, the spatial heterogeneity of features, particularly the remotely sensed phenology metrics extracted from distinct arid and semiarid regions. These two regions exhibit different rainfall patterns, management practices, and topographic characteristics, leading to heterogeneous VI responses. This aligns with the findings of Gao et al. (2023), who observed a spatial distribution of precipitation leading to *NDVI* heterogeneity in the yellow river basin of China. Second, the diminished performance of generalised models could be linked to the presence of spatial non-stationarity in yield and nutrient variables. The statistical properties of these variables vary across regions, introducing complexities that challenge the assumptions of spatial stationarity in generalised models. This observation aligns with a study by Feng et al. (2021) indicating that spatial non-stationarity is a limitation to the generalization ability of yield prediction models.

The analysis of generalised semiarid models (G2SA) revealed compelling insights into predictive performances, especially during critical prediction stages like the peak of the canopy. Interestingly, all G2SA cases exhibited lower Root Mean Squared Error (*RMSE*) values, signifying enhanced accuracy compared to baseline scenarios. This trend is notably pronounced in the G2SA-*RF* at Stage 2, surpassing SA2SA models with a substantial 9% reduction in *RMSE*. These findings align with the research by Pang et al. (2022) which demonstrated that regional-scale models outperformed local-scale Random Forest (*RF*) models in yield prediction, reporting R^2 values of 0.45 and 0.86 for regional and local scales,

respectively. The improved performance of the generalised models can be attributed to their integration of features crucial to baseline predictions, augmented by additional training and features. Notably, the features constituting to 90.1% of SA2SA predictions, significantly contributed to 92% of the global model predictions (Figure 22). This underscores the effectiveness of leveraging key features from specific models to enhance the accuracy of generalised models in predicting agricultural yields.

3.6 Conclusion

This study provides valuable perspectives regarding the performance, interpretability, and practicality of different in-season yield predictive models across different prediction stages. Arid environments revealed a detrimental effect of excessive heat during reproductive stages, while in semi-arid regions, early-season temperature was beneficial for yield production. Notably, rainfall, particularly preceding the canopy's peak, emerged as a pivotal determinant in both climatic contexts. In terms of model performance, *SVR* demonstrated the lowest prediction errors in arid climates, especially in later prediction stages. In contrast, *RF* consistently outperformed other models in semi-arid and global models. The analysis of feature contributions demonstrated distinct climate-specific trends. Arid regions exhibited an even distribution of contributions from different factors, emphasizing the collective influence of multiple features on yield predictions. In contrast, semi-arid regions displayed a strong dependence on rainfall from the canopy's peak onward. Our results revealed also that model transfer between arid and semi-arid climates comes with significant increases in prediction errors (increases of 72.8% to 79.44%). Specific *G2SA* cases exhibited enhanced accuracy at particular stages, contrasting with *G2A* models that performed worse than baseline models in arid climates. Despite the generalized models' overall lower performance, they consistently outperformed transferred models.

Credit authorship contribution statement

Bader Oulaid: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Alice E. Milne:** Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Toby Waine:** Writing – review & editing. **Rafiq El Alami:** Resources, Writing –review & editing, Supervision, Project administration, Funding acquisition. **Tadesse Gashaw Asrat:** Resources Data curation, Writing – review & editing. **Stephan Haefele:** Writing – review & editing. **Abdellah Hamma:** Resources, review & editing. **Ron Corstanje:** Resources, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

3.7 References

- A, A., E, H., A, A., & S, D. P. (2021). E-Farm Management and Yield Prediction. *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*, 1–6. <https://doi.org/10.1109/ICIRCA51532.2021.9544759>
- Abdelmajid, S., Mukhtar, A., Baig, M. B., & Reed, M. R. (2021). Climate change, agricultural policy and food security in Morocco. In *Emerging Challenges to Food Production and Security in Asia, Middle East, and Africa: Climate Risks and Resource Scarcity* (pp. 171–196). Springer International Publishing. https://doi.org/10.1007/978-3-030-72987-5_7
- Abi Saab, M. T., el Alam, R., Jomaa, I., Skaf, S., Fahed, S., Albrizio, R., & Todorovic, M. (2021). Coupling remote sensing data and aquacrop model for simulation of winter wheat growth under rainfed and irrigated conditions in a mediterranean environment. *Agronomy*, 11(11). <https://doi.org/10.3390/agronomy11112265>
- Adnan, A. A., Jibrin, J. M., Kamara, A. Y., Abdulrahman, B. L., Shaibu, A. S., & Garba, I. I. (2017). CERES–Maize model for determining the optimum planting dates of early maturing maize varieties in Northern Nigeria. *Frontiers in Plant Science*, 8. <https://doi.org/10.3389/fpls.2017.01118>
- Agapiou, A., Hadjimitsis, D. G., Papoutsas, C., Alexakis, D. D., & Papadavid, G. (2011). The Importance of accounting for atmospheric effects in the application of NDVI and interpretation of satellite imagery supporting archaeological research: The case studies of Palaepaphos and Nea Paphos sites in Cyprus. *Remote Sensing*, 3(12), 2605–2629. <https://doi.org/10.3390/rs3122605>
- Ait El Haj, F., Ouadif, L., & Akhssas, A. (2023). Monitoring Land Use and Land Cover Changes Using Remote Sensing Techniques and the Precipitation-Vegetation Indexes in Morocco. *Ecological Engineering and Environmental Technology*, 24(1), 272–286. <https://doi.org/10.12912/27197050/154937>

- Ali, A. M., Abouelghar, M., Belal, A. A., Saleh, N., Yones, M., Selim, A. I., Amin, M. E. S., Elwesemy, A., Kucher, D. E., Maginan, S., & Savin, I. (2022a). Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). *The Egyptian Journal of Remote Sensing and Space Science*, 25(3), 711–716. <https://doi.org/https://doi.org/10.1016/j.ejrs.2022.04.006>
- Ali, A. M., Abouelghar, M., Belal, A. A., Saleh, N., Yones, M., Selim, A. I., Amin, M. E. S., Elwesemy, A., Kucher, D. E., Maginan, S., & Savin, I. (2022b). Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). *The Egyptian Journal of Remote Sensing and Space Science*, 25(3), 711–716. <https://doi.org/https://doi.org/10.1016/j.ejrs.2022.04.006>
- Al-Shammari, D., Whelan, B. M., Wang, C., Bramley, R. G. V., Fajardo, M., & Bishop, T. F. A. (2021). Impact of spatial resolution on the quality of crop yield predictions for site-specific crop management. *Agricultural and Forest Meteorology*, 310. <https://doi.org/10.1016/j.agrformet.2021.108622>
- Amiri, E., Irmak, S., & Tari, D. B. (2024). Comparison of the AquaCrop and CERES-Maize models for simulating maize phenology, grain yield, evapotranspiration and water productivity under different irrigation and nitrogen levels. *Irrigation and Drainage*. <https://doi.org/10.1002/ird.2917>
- Amiri, N., Lahlali, R., Amiri, S., Jarroudi, M. E. L., Khebiza, M. Y., & Messouli, M. (2021). Development of an integrated model to assess the impact of agricultural practices and land use on agricultural production in morocco under climate stress over the next twenty years. *Sustainability (Switzerland)*, 13(21). <https://doi.org/10.3390/su132111943>
- Ashourloo, D., Nematollahi, H., Huete, A., Aghighi, H., Azadbakht, M., Shahrabi, H. S., & Goodarzashti, S. (2022). A new phenology-based method for mapping wheat and barley using time-series of Sentinel-2 images. *Remote Sensing of Environment*, 280, 113206. <https://doi.org/https://doi.org/10.1016/j.rse.2022.113206>

- Bai, T., Zhang, N., Mercatoris, B., & Chen, Y. (2019). Improving jujube fruit tree yield estimation at the field scale by assimilating a single Landsat remotely-sensed LAI into the WOFOST model. *Remote Sensing*, 11(9). <https://doi.org/10.3390/rs11091119>
- Bandara, W. B. M. A. C., Sakai, K., Nakandakari, T., Kapetch, P., Anan, M., Nakamura, S., Setouchi, H., & Rathnappriya, R. H. K. (2021). Global optimization of cultivar trait parameters in the simulation of sugarcane phenology using gaussian process emulation. *Agronomy*, 11(7). <https://doi.org/10.3390/agronomy11071379>
- Bandaru, V., Yaramasu, R., PNVR, K., He, J., Fernando, S., Sahajpal, R., Wardlow, B. D., Suyker, A., & Justice, C. (2020). PhenoCrop: An integrated satellite-based framework to estimate physiological growth stages of corn and soybeans. *International Journal of Applied Earth Observation and Geoinformation*, 92. <https://doi.org/10.1016/j.jag.2020.102188>
- Barbedo, J. G. A. (2022). Data Fusion in Agriculture: Resolving Ambiguities and Closing Data Gaps. In *Sensors* (Vol. 22, Issue 6). MDPI. <https://doi.org/10.3390/s22062285>
- Bassu, S., Asseng, S., Motzo, R., & Giunta, F. (2009). Optimising sowing date of durum wheat in a variable Mediterranean environment. *Field Crops Research*, 111(1–2), 109–118. <https://doi.org/10.1016/j.fcr.2008.11.002>
- Belmahi, M., Hanchane, M., Mahjoub, A., Najjari, F., Khayati, A., & Kessabi, R. (2023). Sustainability assessment of the main cereals market in Morocco: Evaluating production and import. *European Journal of Sustainable Development*. <https://api.semanticscholar.org/CorpusID:259255834>
- Bertassello, L., Winters, P., & Müller, M. F. (2023). Access to global wheat reserves determines country-level vulnerability to conflict-induced Ukrainian wheat supply disruption. *Nature Food*, 4(8), 673–676. <https://doi.org/10.1038/s43016-023-00806-w>

- Bertino, L., Evensen, G., & Wackernagel, H. (2003). Sequential Data Assimilation Techniques in Oceanography. *International Statistical Review*, 71(2), 223–241. <https://doi.org/https://doi.org/10.1111/j.1751-5823.2003.tb00194.x>
- Beveridge, L., Whitfield, S., & Challinor, A. (2018). Crop modelling: towards locally relevant and climate-informed adaptation. *Climatic Change*, 147(3–4), 475–489. <https://doi.org/10.1007/s10584-018-2160-z>
- Bhullar, A., Nadeem, K., & Ali, R. A. (2023). Simultaneous multi-crop land suitability prediction from remote sensing data using semi-supervised learning. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-33840-6>
- Bischiniotis, K., van den Hurk, B., Coughlan de Perez, E., Veldkamp, T., Nobre, G. G., & Aerts, J. (2019). Assessing time, cost and quality trade-offs in forecast-based action for floods. *International Journal of Disaster Risk Reduction*, 40. <https://doi.org/10.1016/j.ijdr.2019.101252>
- Boote, K. J., Jones, J. W., & Pickering, N. B. (1996). Potential Uses and Limitations of Crop Models. *Agronomy Journal*, 88(5), 704–716. <https://doi.org/https://doi.org/10.2134/agronj1996.00021962008800050005x>
- Bouras, E., Jarlan, L., Khabba, S., Er-Raki, S., Dezetter, A., Sghir, F., & Tramblay, Y. (2019). Assessing the impact of global climate changes on irrigated wheat yields and water requirements in a semi-arid environment of Morocco. *Scientific Reports*, 9(1). <https://doi.org/10.1038/s41598-019-55251-2>
- Breunig, M., Kuper, P., Reitze, F., Landgraf, S., Al-Doori, M., Stefanakis, E., Abdulmuttalib, H., & Kugler, Z. (2023). IMPROVING DATA QUALITY AND MANAGEMENT FOR REMOTE SENSING ANALYSIS: USE-CASES AND EMERGING RESEARCH QUESTIONS. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10(1-

W1-2023), 41–49. <https://doi.org/10.5194/isprs-annals-X-1-W1-2023-41-2023>

Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, 22(10), 1509–1518. <https://doi.org/10.1016/j.envsoft.2006.10.004>

Casa, R., Varella, H., Buis, S., Guérif, M., De Solan, B., & Baret, F. (2012a). 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), 1–10. <https://doi.org/10.1016/j.eja.2011.09.004>

Casa, R., Varella, H., Buis, S., Guérif, M., De Solan, B., & Baret, F. (2012b). 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), 1–10. <https://doi.org/10.1016/j.eja.2011.09.004>

Castañeda-Vera, A., Leffelaar, P. A., Álvaro-Fuentes, J., Cantero-Martínez, C., & Mínguez, M. I. (2015). Selecting crop models for decision making in wheat insurance. *European Journal of Agronomy*, 68, 97–116. <https://doi.org/https://doi.org/10.1016/j.eja.2015.04.008>

Chang, Y., Latham, J., Licht, M., & Wang, L. (2023a). A data-driven crop model for maize yield prediction. *Communications Biology*, 6(1). <https://doi.org/10.1038/s42003-023-04833-y>

Chang, Y., Latham, J., Licht, M., & Wang, L. (2023b). A data-driven crop model for maize yield prediction. *Communications Biology*, 6(1). <https://doi.org/10.1038/s42003-023-04833-y>

Chapagain, R., Remenyi, T. A., Harris, R. M. B., Mohammed, C. L., Huth, N., Wallach, D., Rezaei, E. E., & Ojeda, J. J. (2022). Decomposing crop model

- uncertainty: A systematic review. In *Field Crops Research* (Vol. 279). Elsevier B.V. <https://doi.org/10.1016/j.fcr.2022.108448>
- Chen, Y., & Tao, F. (2022). Potential of remote sensing data-crop model assimilation and seasonal weather forecasts for early-season crop yield forecasting over a large area. *Field Crops Research*, 276, 108398. <https://doi.org/https://doi.org/10.1016/j.fcr.2021.108398>
- Cheng, E., Zhang, B., Peng, D., Zhong, L., Yu, L., Liu, Y., Xiao, C., Li, C., Li, X., Chen, Y., Ye, H., Wang, H., Yu, R., Hu, J., & Yang, S. (2022). Wheat yield estimation using remote sensing data based on machine learning approaches. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.1090970>
- Cheng, M., Wang, H., Fan, J., Xiang, Y., Liu, X., Liao, Z., Abdelghany, A. E., Zhang, F., & Li, Z. (2022). Evaluation of AquaCrop model for greenhouse cherry tomato with plastic film mulch under various water and nitrogen supplies. *Agricultural Water Management*, 274. <https://doi.org/10.1016/j.agwat.2022.107949>
- Choruma, D. J., Balkovic, J., Pietsch, S. A., & Odume, O. N. (2021). Using EPIC to simulate the effects of different irrigation and fertilizer levels on maize yield in the Eastern Cape, South Africa. *Agricultural Water Management*, 254. <https://doi.org/10.1016/j.agwat.2021.106974>
- Choudhury, A. K., Molla, M. S. H., Zahan, T., Sen, R., Biswas, J. C., Akhter, S., Ishtiaque, S., Ahmed, F., Maniruzaman, M., Hossain, M. B., Sarker, P. C., Dessoky, E. S., Hassan, M. M., & Hossain, A. (2021). Optimum sowing window and yield forecasting for maize in northern and western bangladesh using ceres maize model. *Agronomy*, 11(4). <https://doi.org/10.3390/AGRONOMY11040635>
- Ciric, C., Ciffroy, P., & Charles, S. (2012). Use of sensitivity analysis to identify influential and non-influential parameters within an aquatic ecosystem

- model. *Ecological Modelling*, 246(C), 119–130.
<https://doi.org/10.1016/j.ecolmodel.2012.06.024>
- Coleman, K., Muhammed, S. E., Milne, A. E., Todman, L. C., Dailey, A. G., Glendining, M. J., & Whitmore, A. P. (2017a). The landscape model: A model for exploring trade-offs between agricultural production and the environment. *Science of the Total Environment*, 609, 1483–1499.
<https://doi.org/10.1016/j.scitotenv.2017.07.193>
- Coleman, K., Muhammed, S. E., Milne, A. E., Todman, L. C., Dailey, A. G., Glendining, M. J., & Whitmore, A. P. (2017b). The landscape model: A model for exploring trade-offs between agricultural production and the environment. *Science of the Total Environment*, 609, 1483–1499.
<https://doi.org/10.1016/j.scitotenv.2017.07.193>
- Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., & Acutis, M. (2010). Comparison of sensitivity analysis techniques: A case study with the rice model WARM. *Ecological Modelling*, 221(16), 1897–1906.
<https://doi.org/10.1016/j.ecolmodel.2010.04.021>
- Constantin, J., Raynal, H., Casellas, E., Hoffmann, H., Bindi, M., Doro, L., Eckersten, H., Gaiser, T., Grosz, B., Haas, E., Kersebaum, K., Klatt, S., Kuhnert, M., Lewan, E., Maharjan, G., Moriondo, M., Nendel, C., Roggero, P. P., Specka, X., & Bergez, J.-E. (2019). Management and spatial resolution effects on yield and water balance at regional scale in crop models. *Agricultural and Forest Meteorology*, 275, 184–195.
<https://doi.org/10.1016/j.agrformet.2019.05.013>
- Corstanje, R., Kirk, G. J. D., & Lark, R. M. (2008). The behaviour of soil process models of ammonia volatilization at contrasting spatial scales. *European Journal of Soil Science*, 59(6), 1271–1283. <https://doi.org/10.1111/j.1365-2389.2008.01086.x>
- Corstanje, R., Mercer, T., Rickson, R., Lynda, K., Newell Price, P., Holman, I., Kechavarsi, C., & Waive, T. (2017). Physical soil quality indicators for

- monitoring British soils. *Solid Earth*, 8, 1003–1016.
<https://doi.org/10.5194/se-8-1003-2017>
- Courault, D., Hossard, L., Demarez, V., Dechatre, H., Irfan, K., Baghdadi, N., Flamain, F., & Ruget, F. (n.d.). *STICS crop model and Sentinel-2 images for monitoring rice growth and yield in the Camargue region*.
<https://doi.org/10.1007/s13593-021-00697-w/Published>
- Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., & Schaibly, J. H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics*, 59(8), 3873–3878. <https://doi.org/10.1063/1.1680571>
- Darra, N., Anastasiou, E., Kriezi, O., Lazarou, E., Kalivas, D., & Fountas, S. (2023). Can Yield Prediction Be Fully Digitized? A Systematic Review. In *Agronomy* (Vol. 13, Issue 9). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/agronomy13092441>
- De Wit, C. T. (1958). *Transpiration and Crop Yields*. Institute of Biological and Chemical Research in Field Crops and Herbage.
<https://books.google.co.uk/books?id=3YolHQAACAAJ>
- DeJonge, K. C., Ascough, J. C., Ahmadi, M., Andales, A. A., & Arabi, M. (2012). Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. *Ecological Modelling*, 231, 113–125. <https://doi.org/10.1016/j.ecolmodel.2012.01.024>
- Dhanapal, R., Ajanraj, A., Balavinayagapragathish, S., & Balaji, J. (2021). Retraction: Crop price prediction using supervised machine learning algorithms. In *Journal of Physics: Conference Series* (Vol. 1916, Issue 1). IOP Publishing Ltd. <https://doi.org/10.1088/1742-6596/1916/1/012042>
- Dhillon, M. S., Dahms, T., Kuebert-Flock, C., Borg, E., Conrad, C., & Ullmann, T. (2020). Modelling crop biomass from synthetic remote sensing time series:

- Example for the DEMMIN test site, Germany. *Remote Sensing*, 12(11). <https://doi.org/10.3390/rs12111819>
- Di Paola, A., Valentini, R., & Santini, M. (2016). An overview of available crop growth and yield models for studies and assessments in agriculture. *Journal of the Science of Food and Agriculture*, 96(3), 709–714. <https://doi.org/https://doi.org/10.1002/jsfa.7359>
- Diao, C., Augspurger, C. K., Zhao, Y., & Salk, C. F. (2024). A satellite-field phenological bridging framework for characterizing community-level spring forest phenology using multi-scale satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211, 83–103. <https://doi.org/10.1016/j.isprsjprs.2024.03.018>
- Dirwai, T. L., Senzanje, A., & Mabhaudhi, T. (2021). Calibration and evaluation of the fao aquacrop model for canola (*Brassica napus*) under varied moistube irrigation regimes. *Agriculture (Switzerland)*, 11(5). <https://doi.org/10.3390/agriculture11050410>
- Dlamini, L., Crespo, O., van Dam, J., & Kooistra, L. (2023). A Global Systematic Review of Improving Crop Model Estimations by Assimilating Remote Sensing Data: Implications for Small-Scale Agricultural Systems. In *Remote Sensing* (Vol. 15, Issue 16). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/rs15164066>
- Dong, Y., Zhao, C., Yang, G., Chen, L., Wang, J., & Feng, H. (2013). Integrating a very fast simulated annealing optimization algorithm for crop leaf area index variational assimilation. *Mathematical and Computer Modelling*, 58(3–4), 877–885. <https://doi.org/10.1016/j.mcm.2012.12.013>
- Doorenbos, J., Kassam, A. H., & Bentvelsen, C. I. M. (1979). *Yield response to water*. Food and Agriculture Organization of the United Nations.
- Doraiswamy, P. C., Hatfield, J. L., Jackson, T. J., Akhmedov, B., Prueger, J., & Stern, A. (2004). Crop condition and yield simulations using Landsat and

- MODIS. *Remote Sensing of Environment*, 92(4), 548–559.
<https://doi.org/10.1016/j.rse.2004.05.017>
- Dubovik, O., Schuster, G. L., Xu, F., Hu, Y., Bösch, H., Landgraf, J., & Li, Z. (2021). Grand Challenges in Satellite Remote Sensing. *Frontiers in Remote Sensing*, 2. <https://doi.org/10.3389/frsen.2021.619818>
- Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023a). Crop Prediction Model Using Machine Learning Algorithms. *Applied Sciences (Switzerland)*, 13(16).
<https://doi.org/10.3390/app13169288>
- Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023b). Crop Prediction Model Using Machine Learning Algorithms. *Applied Sciences (Switzerland)*, 13(16).
<https://doi.org/10.3390/app13169288>
- Er-Raki, S., Bouras, E., Rodriguez, J. C., Watts, C. J., Lizarraga-Celaya, C., & Chehbouni, A. (2021). Parameterization of the AquaCrop model for simulating table grapes growth and water productivity in an arid region of Mexico. *Agricultural Water Management*, 245.
<https://doi.org/10.1016/j.agwat.2020.106585>
- Feng, L., Wang, Y., Zhang, Z., & Du, Q. (2021). *Geographically and temporally weighted neural network for winter wheat yield prediction*.
- Feng, X., Tian, H., Cong, J., & Zhao, C. (2023). A method review of the climate change impact on crop yield. In *Frontiers in Forests and Global Change* (Vol. 6). Frontiers Media SA. <https://doi.org/10.3389/ffgc.2023.1198186>
- Fischer, C., Montmerle, T., Berre, L., Auger, L., & ȘTEFĂNESCU, S. E. (2005). An overview of the variational assimilation in the ALADIN/France numerical weather-prediction system. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3477–3492.
<https://doi.org/10.1256/qj.05.115>

- Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L. B., Obersteiner, M., & Van Der Velde, M. (2016). Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. *Nature Communications*, 7. <https://doi.org/10.1038/ncomms11872>
- 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. <https://doi.org/10.1016/j.agwat.2016.11.015>
- Gamon, J. A., Field, C. B., Goulden, M. L., Griffin, K. L., Hartley, A. E., Joel, G., Penuelas, J., & Valentini, R. (1995). Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecological Applications*, 5(1), 28–41. <https://doi.org/10.2307/1942049>
- Gavasso-Rita, Y. L., Papalexiou, S. M., Li, Y., Elshorbagy, A., Li, Z., & Schuster-Wallace, C. (2024). Crop models and their use in assessing crop production and food security: A review. In *Food and Energy Security* (Vol. 13, Issue 1). John Wiley and Sons Inc. <https://doi.org/10.1002/fes3.503>
- Gerard, F. F., George, C. T., Hayman, G., Chavana-Bryant, C., & Weedon, G. P. (2020). Leaf phenology amplitude derived from MODIS NDVI and EVI: Maps of leaf phenology synchrony for Meso- and South America. *Geoscience Data Journal*, 7(1), 13–26. <https://doi.org/10.1002/gdj3.87>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017a). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017b). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>

- Greaves, G. E., & Wang, Y. M. (2016). Assessment of fao aquacrop model for simulating maize growth and productivity under deficit irrigation in a tropical environment. *Water (Switzerland)*, 8(12). <https://doi.org/10.3390/w8120557>
- Guo, D., Olesen, J. E., Pullens, J. W. M., Guo, C., & Ma, X. (2021). Calibrating AquaCrop model using genetic algorithm with multi-objective functions applying different weight factors. *Agronomy Journal*, 113(2), 1420–1438. <https://doi.org/10.1002/agj2.20588>
- Han, H., Liu, Z., Li, J., & Zeng, Z. (2024). Challenges in remote sensing based climate and crop monitoring: navigating the complexities using AI. In *Journal of Cloud Computing* (Vol. 13, Issue 1). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1186/s13677-023-00583-8>
- Hariharan, G., & Kannan, K. (2014). Review of wavelet methods for the solution of reaction–diffusion problems in science and engineering. *Applied Mathematical Modelling*, 38(3), 799–813. <https://doi.org/https://doi.org/10.1016/j.apm.2013.08.003>
- He, D., Wang, E., Wang, J., & Robertson, M. J. (2017). Data requirement for effective calibration of process-based crop models. *Agricultural and Forest Meteorology*, 234–235, 136–148. <https://doi.org/10.1016/j.agrformet.2016.12.015>
- Hoogenboom, G., & White, J. W. (2003). Improving Physiological Assumptions Of Simulation Models By Using Gene-Based Approaches. *Agronomy Journal*, 95(1), 82–89. <https://doi.org/10.2134/agronj2003.8200>
- Hsiao, T. C., Heng, L., Steduto, P., Rojas-Lara, B., Raes, D., & Fereres, E. (2009). Aquacrop-The FAO crop model to simulate yield response to water: III. Parameterization and testing for maize. *Agronomy Journal*, 101(3), 448–459. <https://doi.org/10.2134/agronj2008.0218s>
- Hu, T., Zhang, X., Bohrer, G., Liu, Y., Zhou, Y., Martin, J., Li, Y., & Zhao, K. (2023). Crop yield prediction via explainable AI and interpretable machine

- learning: Dangers of black box models for evaluating climate change impacts on crop yield. *Agricultural and Forest Meteorology*, 336, 109458. <https://doi.org/https://doi.org/10.1016/j.agrformet.2023.109458>
- Huang, H., Huang, J., & Wu, Y. (2020). Markov Chain Monte Carlo and Four-Dimensional Variational Approach Based Winter Wheat Yield Estimation. *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 5290–5293. <https://doi.org/10.1109/IGARSS39084.2020.9324208>
- Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., Tian, L., Zhang, X. (Frank), Fan, J., & Wenbin, 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.
- Huemmrich, K. F., Vargas Zesati, S., Campbell, P., & Tweedie, C. (2021). Canopy reflectance models illustrate varying NDVI responses to change in high latitude ecosystems. *Ecological Applications*, 31(8). <https://doi.org/10.1002/eap.2435>
- Huete, A. R. (1988). A Soil-Adjusted Vegetation Index (SAVI) 295. In *REMOTE SENSING OF ENVIRONMENT* (Vol. 25).
- Iqbal, M. A., Shen, Y., Stricevic, R., Pei, H., Sun, H., Amiri, E., Penas, A., & del Rio, S. (2014). Evaluation of the FAO AquaCrop model for winter wheat on the North China Plain under deficit irrigation from field experiment to regional yield simulation. *Agricultural Water Management*, 135, 61–72. <https://doi.org/10.1016/j.agwat.2013.12.012>
- Ishihara, M., Inoue, Y., Ono, K., Shimizu, M., & Matsuura, S. (2015). The impact of sunlight conditions on the consistency of vegetation indices in croplands- Effective usage of vegetation indices from continuous ground-based spectral measurements. *Remote Sensing*, 7(10), 14079–14098. <https://doi.org/10.3390/rs71014079>

- Jain, M., Balwinder-Singh, Rao, P., Srivastava, A. K., Poonia, S., Blesh, J., Azzari, G., McDonald, A. J., & Lobell, D. B. (2019). The impact of agricultural interventions can be doubled by using satellite data. *Nature Sustainability*, 2(10), 931–934. <https://doi.org/10.1038/s41893-019-0396-x>
- Javan, K., & Darestani, M. (2024). Assessing environmental sustainability of a vital crop in a critical region: Investigating climate change impacts on agriculture using the SWAT model and HWA method. *Heliyon*, 10(3). <https://doi.org/10.1016/j.heliyon.2024.e25326>
- Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PLoS ONE*, 11(6). <https://doi.org/10.1371/journal.pone.0156571>
- Ji, Z., Pan, Y., Zhu, X., Wang, J., & Li, Q. (2021). Prediction of crop yield using phenological information extracted from remote sensing vegetation index. *Sensors (Switzerland)*, 21(4), 1–17. <https://doi.org/10.3390/s21041406>
- 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. In *European Journal of Agronomy* (Vol. 92, pp. 141–152). Elsevier B.V. <https://doi.org/10.1016/j.eja.2017.11.002>
- 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(12). <https://doi.org/10.3390/rs8120972>
- 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. <https://doi.org/10.1016/j.agwat.2019.105846>

- 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. <https://doi.org/10.1016/j.isprsjprs.2017.02.001>
- Jin, Z., Prasad, R., Shriver, J., & Zhuang, Q. (2017). Crop model- and satellite imagery-based recommendation tool for variable rate N fertilizer application for the US Corn system. *Precision Agriculture*, 18(5), 779–800. <https://doi.org/10.1007/s11119-016-9488-z>
- Jing, Q., McConkey, B., Qian, B., Smith, W., Grant, B., Shang, J., Liu, J., Bindraban, P., & Luce, M. S. (2021). Assessing water management effects on spring wheat yield in the Canadian Prairies using DSSAT wheat models. *Agricultural Water Management*, 244. <https://doi.org/10.1016/j.agwat.2020.106591>
- Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). *The DSSAT cropping system model*. www.elsevier.com/locate/eja
- Joshi, N., Mitchard, E. T. A., Brolly, M., Schumacher, J., Fernández-Landa, A., Johannsen, V. K., Marchamalo, M., & Fensholt, R. (2017). Understanding “saturation” of radar signals over forests. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-03469-3>
- 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. <https://doi.org/10.1016/j.rse.2019.04.005>
- Karaboga, D. (2005). *An Idea Based on Honey Bee Swarm for Numerical Optimization*.

- Kasampalis, D. A., Alexandridis, T. K., Deva, C., Challinor, A., Moshou, D., & Zalidis, G. (2018). Contribution of remote sensing on crop models: A review. In *Journal of Imaging* (Vol. 4, Issue 4). MDPI. <https://doi.org/10.3390/jimaging4040052>
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, 4, 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
- Kephe, P. N., Ayisi, K. K., & Petja, B. M. (2021a). Challenges and opportunities in crop simulation modelling under seasonal and projected climate change scenarios for crop production in South Africa. In *Agriculture and Food Security* (Vol. 10, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s40066-020-00283-5>
- Kephe, P. N., Ayisi, K. K., & Petja, B. M. (2021b). Challenges and opportunities in crop simulation modelling under seasonal and projected climate change scenarios for crop production in South Africa. In *Agriculture and Food Security* (Vol. 10, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s40066-020-00283-5>
- Khabba, S., Duchemin, B., Hadria, R., Er-Raki, S., Ezzahar, J., Chehbouni, A., Lahrouni, A., & Hanich, L. (2009). Evaluation of digital hemispherical photography and plant canopy analyzer for measuring vegetation area index of orange orchards. *Journal of Agronomy*, 8(2), 67–72. <https://doi.org/10.3923/ja.2009.67.72>
- Khabba, S., Er-Raki, S., Toumi, J., Ezzahar, J., Hssaine, B. A., Page, M. le, & Chehbouni, A. (2020). A simple light-use-efficiency model to estimate wheat yield in the semi-Arid areas. *Agronomy*, 10(10). <https://doi.org/10.3390/agronomy10101524>
- Khabbazan, S., Vermunt, P., Steele-Dunne, S., Arntz, L. R., Marinetti, C., van der Valk, D., Iannini, L., Molijn, R., Westerdijk, K., & van der Sande, C. (2019).

- Crop monitoring using Sentinel-1 data: A case study from The Netherlands. *Remote Sensing*, 11(16). <https://doi.org/10.3390/rs11161887>
- Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.00621>
- Khaki, S., Wang, L., & Archontoulis, S. V. (2020). A CNN-RNN Framework for Crop Yield Prediction. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.01750>
- Klocke, N., Stone, L., Briggeman, S., & Bolton, D. (2010). Technical Note: Scheduling for Deficit Irrigation—Crop Yield Predictor. *Applied Engineering in Agriculture*, 26, 413–418. <https://doi.org/10.13031/2013.29956>
- Koo, J., Bostick, W., Jones, J., Gijsman, A., & Naab, J. (2003). Estimating Soil Carbon in Agricultural Systems using Ensemble Kalman Filter and DSSAT-CENTURY. In *Transactions of the ASABE* (Vol. 50). <https://doi.org/10.13031/2013.13847>
- Krishnan, M., Gugercin, S., & Tarazaga, P. A. (2023). A wavelet-based dynamic mode decomposition for modeling mechanical systems from partial observations. *Mechanical Systems and Signal Processing*, 187, 109919. <https://doi.org/10.1016/j.ymssp.2022.109919>
- Kumar, U., Hansen, E. M., Thomsen, I. K., & Vogeler, I. (2023). Performance of APSIM to Simulate the Dynamics of Winter Wheat Growth, Phenology, and Nitrogen Uptake from Early Growth Stages to Maturity in Northern Europe. *Plants*, 12(5). <https://doi.org/10.3390/plants12050986>
- Lee, D.-Y., Lee, D.-S., Cha, Y., Min, J.-H., & Park, Y.-S. (2023). Data-driven models for predicting community changes in freshwater ecosystems: A review. *Ecological Informatics*, 77, 102163. <https://doi.org/10.1016/j.ecoinf.2023.102163>
- Leukel, J., Zimpel, T., & Stumpe, C. (2023). Machine learning technology for early prediction of grain yield at the field scale: A systematic review. *Computers*

- and Electronics in Agriculture*, 207, 107721.
<https://doi.org/https://doi.org/10.1016/j.compag.2023.107721>
- Li, J., & Chen, B. (2020). Global revisit interval analysis of landsat-8-9 and sentinel-2a-2b data for terrestrial monitoring. *Sensors (Switzerland)*, 20(22), 1–15. <https://doi.org/10.3390/s20226631>
- Li, K., & Chen, Y. (2018). A Genetic Algorithm-based urban cluster automatic threshold method by combining VIIRS DNB, NDVI, and NDBI to monitor urbanization. *Remote Sensing*, 10(2). <https://doi.org/10.3390/rs10020277>
- Li, R., Li, C., Liu, F., Yang, X., & Wang, J. (2010). Methods and algorithms of data assimilation and its application in agriculture. *2010 World Automation Congress*, 29–34.
- Li, Z., He, J., Xu, X., Jin, X., Huang, W., Clark, B., Yang, G., & Li, Z. (2018). Estimating genetic parameters of DSSAT-CERES model with the GLUE method for winter wheat (*Triticum aestivum* L.) production. *Computers and Electronics in Agriculture*, 154, 213–221.
<https://doi.org/https://doi.org/10.1016/j.compag.2018.09.009>
- Li, Z., Jin, X., Zhao, C., Wang, J., Xu, X., Yang, G., Li, C., & Shen, J. (2015). Estimating wheat yield and quality by coupling the DSSAT-CERES model and proximal remote sensing. *European Journal of Agronomy*, 71, 53–62.
<https://doi.org/10.1016/j.eja.2015.08.006>
- Li, Z., Zhao, Y., Taylor, J., Gaulton, R., Jin, X., Song, X., Li, Z., Meng, Y., Chen, P., Feng, H., Wang, C., Guo, W., Xu, X., Chen, L., & Yang, G. (2022). Comparison and transferability of thermal, temporal and phenological-based in-season predictions of above-ground biomass in wheat crops from proximal crop reflectance data. *Remote Sensing of Environment*, 273, 112967. <https://doi.org/https://doi.org/10.1016/j.rse.2022.112967>
- Liao, C., Wang, J., Dong, T., Shang, J., Liu, J., & Song, Y. (2019). Using spatio-temporal fusion of Landsat-8 and MODIS data to derive phenology, biomass

- and yield estimates for corn and soybean. *Science of the Total Environment*, 650, 1707–1721. <https://doi.org/10.1016/j.scitotenv.2018.09.308>
- Lin, F., Li, X., Jia, N., Feng, F., Huang, H., Huang, J., Fan, S., Ciais, P., & Song, X. P. (2023). The impact of Russia-Ukraine conflict on global food security. *Global Food Security*, 36. <https://doi.org/10.1016/j.gfs.2022.100661>
- Linker, R. (2020). Unified framework for model-based optimal allocation of crop areas and water. *Agricultural Water Management*, 228, 105859. <https://doi.org/https://doi.org/10.1016/j.agwat.2019.105859>
- López-Urrea, R., Domínguez, A., Pardo, J. J., Montoya, F., García-Vila, M., & Martínez-Romero, A. (2020). Parameterization and comparison of the AquaCrop and MOPECO models for a high-yielding barley cultivar under different irrigation levels. *Agricultural Water Management*, 230. <https://doi.org/10.1016/j.agwat.2019.105931>
- Lorenz, A. J. (2013). Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. *G3: Genes, Genomes, Genetics*, 3(3), 481–491. <https://doi.org/10.1534/g3.112.004911>
- Lu, Y., Chibarabada, T. P., Ziliani, M. G., Onema, J. M. K., McCabe, M. F., & Sheffield, J. (2021). Assimilation of soil moisture and canopy cover data improves maize simulation using an under-calibrated crop model. *Agricultural Water Management*, 252. <https://doi.org/10.1016/j.agwat.2021.106884>
- Lyu, J., Jiang, Y., Xu, C., Liu, Y., Su, Z., Liu, J., & He, J. (2022). Multi-objective winter wheat irrigation strategies optimization based on coupling AquaCrop-OSPy and NSGA-III: A case study in Yangling, China. *Science of the Total Environment*, 843. <https://doi.org/10.1016/j.scitotenv.2022.157104>
- Ma, G., Huang, J., Wu, W., Fan, J., Zou, J., & Wu, S. (2013). Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat

- yield. *Mathematical and Computer Modelling*, 58(3), 634–643.
<https://doi.org/https://doi.org/10.1016/j.mcm.2011.10.038>
- 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(9), 4225–4236. <https://doi.org/10.1109/JSTARS.2017.2711482>
- Ma, Y., Yang, Z., Huang, Q., & Zhang, Z. (2023). Improving the Transferability of Deep Learning Models for Crop Yield Prediction: A Partial Domain Adaptation Approach. *Remote Sensing*, 15(18), 4562. <https://doi.org/10.3390/rs15184562>
- Ma, Y., Zhang, Z., Kang, Y., & Özdoğan, M. (2021). Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sensing of Environment*, 259, 112408. <https://doi.org/https://doi.org/10.1016/j.rse.2021.112408>
- Maestrini, B., Mimić, G., van Oort, P. A. J., Jindo, K., Brdar, S., van Evert, F. K., & Athanasiados, I. (2022a). Mixing process-based and data-driven approaches in yield prediction. In *European Journal of Agronomy* (Vol. 139). Elsevier B.V. <https://doi.org/10.1016/j.eja.2022.126569>
- Maestrini, B., Mimić, G., van Oort, P. A. J., Jindo, K., Brdar, S., van Evert, F. K., & Athanasiados, I. (2022b). Mixing process-based and data-driven approaches in yield prediction. In *European Journal of Agronomy* (Vol. 139). Elsevier B.V. <https://doi.org/10.1016/j.eja.2022.126569>
- Manschadi, A. M., Eitzinger, J., Breisch, M., Fuchs, W., Neubauer, T., & Soltani, A. (2021). Full Parameterisation Matters for the Best Performance of Crop Models: Inter-comparison of a Simple and a Detailed Maize Model. *International Journal of Plant Production*, 15(1), 61–78. <https://doi.org/10.1007/s42106-020-00116-2>

- Martínez-Ruiz, A., Ruiz-García, A., Prado-Hernández, J. V., López-Cruz, I. L., Valencia-Islas, J. O., & Pineda-Pineda, J. (2021). Global sensitivity analysis and calibration by differential evolution algorithm of HORTSYST crop model for fertigation management. *Water (Switzerland)*, 13(5). <https://doi.org/10.3390/w13050610>
- Martinho, V. J. P. D., Cunha, C. A. da S., Pato, M. L., Costa, P. J. L., Sánchez-Carreira, M. C., Georgantzís, N., Rodrigues, R. N., & Coronado, F. (2022). Machine Learning and Food Security: Insights for Agricultural Spatial Planning in the Context of Agriculture 4.0. *Applied Sciences (Switzerland)*, 12(22). <https://doi.org/10.3390/app122211828>
- Marzukhi, F., Elahami, A. L., & Bohari, S. N. (2016). Detecting nutrients deficiencies of oil palm trees using remotely sensed data. *IOP Conference Series: Earth and Environmental Science*, 37(1). <https://doi.org/10.1088/1755-1315/37/1/012040>
- McRae, G. J., Tilden, J. W., & Seinfeld, J. H. (1982). Global sensitivity analysis—a computational implementation of the Fourier Amplitude Sensitivity Test (FAST). *Computers & Chemical Engineering*, 6(1), 15–25. [https://doi.org/https://doi.org/10.1016/0098-1354\(82\)80003-3](https://doi.org/https://doi.org/10.1016/0098-1354(82)80003-3)
- Meddi, M., & Eslamian, S. (2021). Uncertainties in Rainfall and Water Resources in Maghreb Countries Under Climate Change. In *African Handbook of Climate Change Adaptation: With 610 Figures and 361 Tables* (pp. 1967–2003). Springer International Publishing. https://doi.org/10.1007/978-3-030-45106-6_114
- Meroni, M., Fasbender, D., Rembold, F., Atzberger, C., & Klisch, A. (2019). Near real-time vegetation anomaly detection with MODIS NDVI: Timeliness vs. accuracy and effect of anomaly computation options. *Remote Sensing of Environment*, 221, 508–521. <https://doi.org/https://doi.org/10.1016/j.rse.2018.11.041>

- Mohanty, B. P., Cosh, M. H., Lakshmi, V., & Montzka, C. (2017). Soil Moisture Remote Sensing: State-of-the-Science. *Vadose Zone Journal*, 16(1), 1–9. <https://doi.org/10.2136/vzj2016.10.0105>
- Monteiro, L. A., Ramos, R. M., Battisti, R., Soares, J. R., Oliveira, J. C., Figueiredo, G. K. D. A., Lamparelli, R. A. C., Nendel, C., & Lana, M. A. (2022). Potential Use of Data-Driven Models to Estimate and Predict Soybean Yields at National Scale in Brazil. *International Journal of Plant Production*, 16(4), 691–703. <https://doi.org/10.1007/s42106-022-00209-0>
- Moran, M. S., Inoue, Y., & Barnes, E. M. (1997). Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment*, 61(3), 319–346. [https://doi.org/https://doi.org/10.1016/S0034-4257\(97\)00045-X](https://doi.org/https://doi.org/10.1016/S0034-4257(97)00045-X)
- Morel, J., Bégué, A., Todoroff, P., Martiné, J. F., Lebourgeois, V., & Petit, M. (2014a). Coupling a sugarcane crop model with the remotely sensed time series of fIPAR to optimise the yield estimation. *European Journal of Agronomy*, 61, 60–68. <https://doi.org/10.1016/j.eja.2014.08.004>
- Morel, J., Bégué, A., Todoroff, P., Martiné, J.-F., Lebourgeois, V., & Petit, M. (2014b). Coupling a sugarcane crop model with the remotely sensed time series of fIPAR to optimise the yield estimation. *European Journal of Agronomy*, 61, 60–68. <https://doi.org/https://doi.org/10.1016/j.eja.2014.08.004>
- Morris, M. D. (1991a). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2), 161–174. <https://doi.org/10.1080/00401706.1991.10484804>
- Morris, M. D. (1991b). Factorial Sampling Plans for Preliminary Computational Experiments. In *TECHNOMETRICS* (Vol. 33, Issue 2).
- Mubvuma, M. T., Ogola, J. B. O., & Mhizha, T. (2021a). AquaCrop model calibration and validation for chickpea (*Cicer arietinum*) in Southern Africa.

- Cogent Food and Agriculture*, 7(1).
<https://doi.org/10.1080/23311932.2021.1898135>
- Mubvuma, M. T., Ogola, J. B. O., & Mhizha, T. (2021b). AquaCrop model calibration and validation for chickpea (*Cicer arietinum*) in Southern Africa. *Cogent Food and Agriculture*, 7(1).
<https://doi.org/10.1080/23311932.2021.1898135>
- Müller, C., Jägermeyr, J., Franke, J. A., Ruane, A. C., Balkovic, J., Ciais, P., Dury, M., Falloon, P., Folberth, C., Hank, T., Hoffmann, M., Izaurralde, R. C., Jacquemin, I., Khabarov, N., Liu, W., Olin, S., Pugh, T. A. M., Wang, X., Williams, K., ... Elliott, J. W. (2024). Substantial Differences in Crop Yield Sensitivities Between Models Call for Functionality-Based Model Evaluation. *Earth's Future*, 12(3). <https://doi.org/10.1029/2023EF003773>
- Nansen, C., Geremias, L. D., Xue, Y., Huang, F., & Parra, J. R. (2013). Agricultural Case Studies of Classification Accuracy, Spectral Resolution, and Model Over-Fitting. *Applied Spectroscopy*, 67(11), 1332–1338.
<https://doi.org/10.1366/12-06933>
- Nash, J. E., & Sutcliffe, J. v. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/https://doi.org/10.1016/0022-1694(70)90255-6)
- Newlands, N. K., Ghahari, A., Gel, Y. R., Lyubchich, V., Agriculture, T. M., & Canada, A.-F. (n.d.). *Deep Learning for Improved Agricultural Risk Management*. <https://oceanservice.noaa.gov/facts/>
- Niu, Z., He, H., Peng, S., Ren, X., Zhang, L., Gu, F., Zhu, G., Peng, C., Li, P., Wang, J., Ge, R., Zeng, N., Zhu, X., Lv, Y., Chang, Q., Xu, Q., Zhang, M., & Liu, W. (2021). A Process-Based Model Integrating Remote Sensing Data for Evaluating Ecosystem Services. *Journal of Advances in Modeling Earth Systems*, 13(6). <https://doi.org/10.1029/2020MS002451>

- Nossent, J., Elsen, P., & Bauwens, W. (2011). Sobol' sensitivity analysis of a complex environmental model. *Environmental Modelling & Software*, 26(12), 1515–1525. [https://doi.org/https://doi.org/10.1016/j.envsoft.2011.08.010](https://doi.org/10.1016/j.envsoft.2011.08.010)
- Oikonomidis, A., Catal, C., & Kassahun, A. (2023). Deep learning for crop yield prediction: a systematic literature review. In *New Zealand Journal of Crop and Horticultural Science* (Vol. 51, Issue 1, pp. 1–26). Taylor and Francis Ltd. <https://doi.org/10.1080/01140671.2022.2032213>
- Oulaid, B., Milne, A. E., Waine, T., El Alami, R., Rafiqi, M., & Corstanje, R. (2024). Stepwise model parametrisation using satellite imagery and hemispherical photography: Tuning AquaCrop sensitive parameters for improved winter wheat yield predictions in semi-arid regions. *Field Crops Research*, 309. <https://doi.org/10.1016/j.fcr.2024.109327>
- Paleari, L., Movedi, E., Zoli, M., Burato, A., Cecconi, I., Errahouly, J., Pecollo, E., Sorvillo, C., & Confalonieri, R. (2021). Sensitivity analysis using Morris: Just screening or an effective ranking method? *Ecological Modelling*, 455, 109648. [https://doi.org/https://doi.org/10.1016/j.ecolmodel.2021.109648](https://doi.org/10.1016/j.ecolmodel.2021.109648)
- Paredes, P., Wei, Z., Liu, Y., Xu, D., Xin, Y., Zhang, B., & Pereira, L. S. (2015). Performance assessment of the FAO aquaCrop model for soil water, soil evaporation, biomass and yield of soybeans in north china plain. *Agricultural Water Management*, 152, 57–71. <https://doi.org/10.1016/j.agwat.2014.12.007>
- Parent, B., & Tardieu, F. (2014). Can current crop models be used in the phenotyping era for predicting the genetic variability of yield of plants subjected to drought or high temperature? In *Journal of Experimental Botany* (Vol. 65, Issue 21, pp. 6179–6189). Oxford University Press. <https://doi.org/10.1093/jxb/eru223>
- Pasquel, D., Roux, S., Richetti, J., Cammarano, D., Tisseyre, B., & Taylor, J. A. (2022). A review of methods to evaluate crop model performance at multiple

- and changing spatial scales. In *Precision Agriculture* (Vol. 23, Issue 4, pp. 1489–1513). Springer. <https://doi.org/10.1007/s11119-022-09885-4>
- Patel, K., & Patel, H. B. (2021). A Comparative Analysis of Supervised Machine Learning Algorithm for Agriculture Crop Prediction. *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 1–5. <https://doi.org/10.1109/ICECCT52121.2021.9616731>
- Patel, U., Pathan, M., Kathiria, P., & Patel, V. (2023). Crop type classification with hyperspectral images using deep learning : a transfer learning approach. *Modeling Earth Systems and Environment*, 9(2), 1977–1987. <https://doi.org/10.1007/s40808-022-01608-y>
- Pelosi, A., Belfiore, O. R., D'Urso, G., & Chirico, G. B. (2022). Assessing Crop Water Requirement and Yield by Combining ERA5-Land Reanalysis Data with CM-SAF Satellite-Based Radiation Data and Sentinel-2 Satellite Imagery. *Remote Sensing*, 14(24). <https://doi.org/10.3390/rs14246233>
- Peng, Y., Zhang, M., Xu, Z., Yang, T., Su, Y., Zhou, T., Wang, H., Wang, Y., & Lin, Y. (2020). Estimation of leaf nutrition status in degraded vegetation based on field survey and hyperspectral data. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-61294-7>
- Petersen, L. K. (2018). Real-time prediction of crop yields from MODIS relative vegetation health: A continent-wide analysis of Africa. *Remote Sensing*, 10(11). <https://doi.org/10.3390/rs10111726>
- Poempool, L., Kruatrachue, B., & Siriboon, K. (2018, August 13). Combine multi particle swarm in supporting trapping in local optima. *ICEAST 2018 - 4th International Conference on Engineering, Applied Sciences and Technology: Exploring Innovative Solutions for Smart Society*. <https://doi.org/10.1109/ICEAST.2018.8434398>
- Prudnikova, E., Savin, I., Vindeker, G., Grubina, P., Shishkonakova, E., & Sharychev, D. (2019). Influence of soil background on spectral reflectance

- of winter wheat crop canopy. *Remote Sensing*, 11(16).
<https://doi.org/10.3390/rs11161932>
- Psiroukis, V., Darra, N., Kasimati, A., Trojacek, P., Hasanli, G., & Fountas, S. (2022). Development of a Multi-Scale Tomato Yield Prediction Model in Azerbaijan Using Spectral Indices from Sentinel-2 Imagery. *Remote Sensing*, 14(17). <https://doi.org/10.3390/rs14174202>
- Puntel, L. A., Sawyer, J. E., Barker, D. W., Dietzel, R., Poffenbarger, H., Castellano, M. J., Moore, K. J., Thorburn, P., & Archontoulis, S. V. (2016). Modeling long-term corn yield response to nitrogen rate and crop rotation. *Frontiers in Plant Science*, 7(November 2016).
<https://doi.org/10.3389/fpls.2016.01630>
- Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2009). Aquacrop-The FAO crop model to simulate yield response to water: II. main algorithms and software description. *Agronomy Journal*, 101(3), 438–447.
<https://doi.org/10.2134/agronj2008.0140s>
- Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. (2022). *Chapter 3 Calculation procedures AquaCrop Reference manual August 2022*.
- Rani, S., Mishra, A. K., Kataria, A., Mallik, S., & Qin, H. (2023). Machine learning-based optimal crop selection system in smart agriculture. *Scientific Reports*, 13(1), 15997. <https://doi.org/10.1038/s41598-023-42356-y>
- Razzaghi, F., Zhou, Z., Andersen, M. N., & Plauborg, F. (2017). Simulation of potato yield in temperate condition by the AquaCrop model. *Agricultural Water Management*, 191, 113–123.
<https://doi.org/10.1016/j.agwat.2017.06.008>
- Rezzoug, W., Gabrielle, B., Suleiman, A., & Benabdeli, K. (2008). Application and evaluation of the DSSAT-wheat in the Tiaret region of Algeria. In *African Journal of Agricultural Research* (Vol. 3, Issue 4).
<http://www.academicjournals.org/AJAR>

- Rivas-Fandiño, P., Acuña-Alonso, C., Novo, A., Pacheco, F. A. L., & Álvarez, X. (2023). Assessment of high spatial resolution satellite imagery for monitoring riparian vegetation: riverine management in the smallholding. *Environmental Monitoring and Assessment*, 195(1). <https://doi.org/10.1007/s10661-022-10667-8>
- Roberts, M. J., Braun, N. O., Sinclair, T. R., Lobell, D. B., & Schlenker, W. (2017). Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, 12(9). <https://doi.org/10.1088/1748-9326/aa7f33>
- Rostami, A., Raeini-Sarjaz, M., Chabokpour, J., Azamathulla, H. M., & Kumar, S. (2022). Determination of rainfed wheat agriculture potential through assimilation of remote sensing data with SWAT model case study: ZarrinehRoud Basin, Iran. *Water Supply*, 22(5), 5331–5334. <https://doi.org/10.2166/ws.2022.160>
- Rozenstein, O., Cohen, Y., Alchanatis, V., Behrendt, K., Bonfil, D. J., Eshel, G., Harari, A., Harris, W. E., Klapp, I., Laor, Y., Linker, R., Paz-Kagan, T., Peets, S., Rutter, S. M., Salzer, Y., & Lowenberg-DeBoer, J. (2024). Data-driven agriculture and sustainable farming: friends or foes? *Precision Agriculture*, 25(1), 520–531. <https://doi.org/10.1007/s11119-023-10061-5>
- Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. *Agricultural and Forest Meteorology*, 200, 233–248. <https://doi.org/10.1016/j.agrformet.2014.09.016>
- Sabzzadeh, I., & Shourian, M. (2020). Maximizing crops yield net benefit in a groundwater-irrigated plain constrained to aquifer stable depletion using a coupled PSO-SWAT-MODFLOW hydro-agronomic model. *Journal of Cleaner Production*, 262. <https://doi.org/10.1016/j.jclepro.2020.121349>
- Sahbeni, G., Székely, B., Musyimi, P. K., Timár, G., & Sahajpal, R. (2023). Crop Yield Estimation Using Sentinel-3 SLSTR, Soil Data, and Topographic

- Features Combined with Machine Learning Modeling: A Case Study of Nepal. *AgriEngineering*, 5(4), 1766–1788. <https://doi.org/10.3390/agriengineering5040109>
- Sajid, S. S., & Hu, G. (2022). Optimizing Crop Planting Schedule Considering Planting Window and Storage Capacity. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.762446>
- Salman Maher, García-Vila Margarita, Fereres Elias, Raes Dirk, & Steduto Steduto. (2021). The AquaCrop model – Enhancing crop water productivity. In *The AquaCrop model – Enhancing crop water productivity*. FAO. <https://doi.org/10.4060/cb7392en>
- Saltelli, A., Tarantola, S., & Chan, K. P.-S. (1999). A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*, 41(1), 39–56. <https://doi.org/10.1080/00401706.1999.10485594>
- Sarrazin, F., Pianosi, F., & Wagener, T. (2016). Global Sensitivity Analysis of environmental models: Convergence and validation. *Environmental Modelling and Software*, 79, 135–152. <https://doi.org/10.1016/j.envsoft.2016.02.005>
- Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., & Gao, F. (n.d.). *Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description*. <http://www.usgs.gov/pubprod>
- Schumacher, B. L., Burchfield, E. K., Bean, B., & Yost, M. A. (2023). Leveraging Important Covariate Groups for Corn Yield Prediction. *Agriculture (Switzerland)*, 13(3). <https://doi.org/10.3390/agriculture13030618>
- Shang, K., Yao, Y., Di, Z., Jia, K., Zhang, X., Fisher, J. B., Chen, J., Guo, X., Yang, J., Yu, R., Xie, Z., Liu, L., Ning, J., & Zhang, L. (2023). Coupling physical constraints with machine learning for satellite-derived

- evapotranspiration of the Tibetan Plateau. *Remote Sensing of Environment*, 289, 113519. <https://doi.org/https://doi.org/10.1016/j.rse.2023.113519>
- Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning. *IEEE Access*, 11, 111255–111264. <https://doi.org/10.1109/ACCESS.2023.3321861>
- Sharp, R. T., Henrys, P. A., Jarvis, S. G., Whitmore, A. P., Milne, A. E., Coleman, K., Mohankumar, S. E. P., & Metcalfe, H. (2021). Simulating cropping sequences using earth observation data. *Computers and Electronics in Agriculture*, 188, 106330. <https://doi.org/https://doi.org/10.1016/j.compag.2021.106330>
- Shrestha, A., Bheemanahalli, R., Adeli, A., Samiappan, S., Czarnecki, J. M. P., McCraine, C. D., Reddy, K. R., & Moorhead, R. (2023). Phenological stage and vegetation index for predicting corn yield under rainfed environments. *Frontiers in Plant Science*, 14. <https://doi.org/10.3389/fpls.2023.1168732>
- Shuai, G., & Basso, B. (2022). Subfield maize yield prediction improves when in-season crop water deficit is included in remote sensing imagery-based models. *Remote Sensing of Environment*, 272. <https://doi.org/10.1016/j.rse.2022.112938>
- Silva, J. V., & Giller, K. E. (2020). Grand challenges for the 21st century: What crop models can and can't (yet) do. *Journal of Agricultural Science*, 158(10), 794–805. <https://doi.org/10.1017/S0021859621000150>
- 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(11). <https://doi.org/10.1371/journal.pone.0187485>

- Simms, D. M., Hamer, A. M., Zeiler, I., Vita, L., & Waine, T. W. (2023). Mapping Agricultural Land in Afghanistan's Opium Provinces Using a Generalised Deep Learning Model and Medium Resolution Satellite Imagery. *Remote Sensing*, 15(19). <https://doi.org/10.3390/rs15194714>
- Siritanaratkul, B. (2022). Generalizability and limitations of machine learning for yield prediction of oxidative coupling of methane. *Digital Chemical Engineering*, 2. <https://doi.org/10.1016/j.dche.2022.100013>
- Sodari, L., Alam, J., & Das, K. K. (2021). *Performance Assessment of Aquacrop Model for Estimating Canopy Cover, Aboveground Biomass and Grain Yield of Winter Wheat in Saptari District: An Application to Irrigation Management*.
- Solomatine, D., See, L. M., & Abrahart, R. J. (2008). Data-Driven Modelling: Concepts, Approaches and Experiences. In R. J. Abrahart, L. M. See, & D. P. Solomatine (Eds.), *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications* (pp. 17–30). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-79881-1_2
- Soltani, A., Alimagham, S. M., Nehbandani, A., Torabi, B., Zeinali, E., Dadrasi, A., Zand, E., Ghassemi, S., Pourshirazi, S., Alasti, O., Hosseini, R. S., Zahed, M., Arabameri, R., Mohammadzadeh, Z., Rahban, S., Kamari, H., Fayazi, H., Mohammadi, S., Keramat, S., ... Sinclair, T. R. (2020). SSM-iCrop2: A simple model for diverse crop species over large areas. *Agricultural Systems*, 182. <https://doi.org/10.1016/j.agsy.2020.102855>
- Specka, X., Nendel, C., & Wieland, R. (2015). Analysing the parameter sensitivity of the agro-ecosystem model MONICA for different crops. *European Journal of Agronomy*, 71, 73–87. <https://doi.org/10.1016/j.eja.2015.08.004>
- Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009a). Aquacrop-the FAO crop model to simulate yield response to water: I. concepts and underlying principles. *Agronomy Journal*, 101(3), 426–437. <https://doi.org/10.2134/agronj2008.0139s>

- Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009b). AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. *Agronomy Journal*, 101(3), 426–437. <https://doi.org/10.2134/agronj2008.0139s>
- Stricevic, R., Cosic, M., Djurovic, N., Pejic, B., & Maksimovic, L. (2011). Assessment of the FAO AquaCrop model in the simulation of rainfed and supplementally irrigated maize, sugar beet and sunflower. *Agricultural Water Management*, 98(10), 1615–1621. <https://doi.org/10.1016/j.agwat.2011.05.011>
- Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. *Sensors (Switzerland)*, 19(20). <https://doi.org/10.3390/s19204363>
- Sun, Y. (2022). Enhanced Weather-Based Index Insurance Design for Hedging Crop Yield Risk. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.895183>
- Szu, H., & Hartley, R. (1987). Fast simulated annealing. *Physics Letters A*, 122(3–4). [https://doi.org/10.1016/0375-9601\(87\)90796-1](https://doi.org/10.1016/0375-9601(87)90796-1)
- Tai, M. M. (1994). A Mathematical Model for the Determination of Total Area Under Glucose Tolerance and Other Metabolic Curves $\sum_{i=1}^{n-1} X_i \sim X_n$
Total Area = $\sum_{i=1}^{n-1} (y_i - y_{i+1}) + X_n - y_n$
 $\sum_{i=1}^{n-1} (y_i - y_{i+1}) + X_n - y_n = \sum_{i=1}^{n-1} (y_i - y_{i+1}) + X_n - y_n$
 $\sum_{i=1}^{n-1} (y_i - y_{i+1}) + X_n - y_n = \sum_{i=1}^{n-1} (y_i - y_{i+1}) + X_n - y_n$
In *DIABETES CARE* (Vol. 17, Issue 2). <http://diabetesjournals.org/care/article-pdf/17/2/152/341381/17-2-152.pdf>
- Takács, S., Csengeri, E., Pék, Z., Bíró, T., Szuvandzsiev, P., Palotás, G., & Helyes, L. (2021). Performance evaluation of aquacrop model in processing tomato biomass, fruit yield and water stress indicator modelling. *Water (Switzerland)*, 13(24). <https://doi.org/10.3390/w13243587>

- Tan, B., Gao, F., Wolfe, R. E., Pedelty, J. A., Nightingale, J., Morissette, J. T., & Ederer, G. A. (2011). An Enhanced TIMESAT Algorithm for Estimating Vegetation Phenology Metrics From MODIS Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(2), 361–371. <https://doi.org/10.1109/JSTARS.2010.2075916>
- Tang, B.-H., Shao, K., Li, Z.-L., Wu, H., & Tang, R. (2015). An improved NDVI-based threshold method for estimating land surface emissivity using MODIS satellite data. *International Journal of Remote Sensing*, 36(19–20), 4864–4878. <https://doi.org/10.1080/01431161.2015.1040132>
- Tao, H., Feng, H., Xu, L., Miao, M., Yang, G., Yang, X., & Fan, L. (2020). Estimation of the yield and plant height of winter wheat using UAV-based hyperspectral images. *Sensors (Switzerland)*, 20(4). <https://doi.org/10.3390/s20041231>
- Tao, R., Zhao, P., Wu, J., Martin, N., Harrison, M. T., Ferreira, C., Kalantari, Z., & Hovakimyan, N. (2023). *Optimizing Crop Management with Reinforcement Learning and Imitation Learning*.
- Tenreiro, T. R., García-Vila, M., Gómez, J. A., Jiménez-Berni, J. A., & Fereres, E. (2021). Using NDVI for the assessment of canopy cover in agricultural crops within modelling research. *Computers and Electronics in Agriculture*, 182. <https://doi.org/10.1016/j.compag.2021.106038>
- Tesfaye, K., Takele, R., Shelia, V., Hayi, E., Dabale, A., Traore, P., & Solomon, D. (2022). High Spatial Resolution Seasonal Crop Yield Forecasting for Heterogeneous Maize Environments in the Oromia Regional State, Ethiopia. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4074628>
- Toumi, J., Er-Raki, S., Ezzahar, J., Khabba, S., Jarlan, L., & Chehbouni, A. (2016a). Performance assessment of AquaCrop model for estimating evapotranspiration, soil water content and grain yield of winter wheat in Tensift Al Haouz (Morocco): Application to irrigation management.

- Agricultural Water Management*, 163, 219–235.
<https://doi.org/10.1016/j.agwat.2015.09.007>
- Toumi, J., Er-Raki, S., Ezzahar, J., Khabba, S., Jarlan, L., & Chehbouni, A. (2016b). Performance assessment of AquaCrop model for estimating evapotranspiration, soil water content and grain yield of winter wheat in Tensift Al Haouz (Morocco): Application to irrigation management. *Agricultural Water Management*, 163, 219–235.
<https://doi.org/10.1016/j.agwat.2015.09.007>
- Tsakmakis, I. D., Gikas, G. D., & Sylaios, G. K. (2021). Integration of Sentinel-derived NDVI to reduce uncertainties in the operational field monitoring of maize. *Agricultural Water Management*, 255, 106998.
<https://doi.org/https://doi.org/10.1016/j.agwat.2021.106998>
- Tsakmakis, I. D., Zidou, M., Gikas, G. D., & Sylaios, G. K. (2018). Impact of Irrigation Technologies and Strategies on Cotton Water Footprint Using AquaCrop and CROPWAT Models. *Environmental Processes*, 5, 181–199.
<https://doi.org/10.1007/s40710-018-0289-4>
- 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(16). <https://doi.org/10.3390/RS12162666>
- van der Velde, M., Biavetti, I., El-Aydam, M., Niemeyer, S., Santini, F., & van den Berg, M. (2019). Use and relevance of European Union crop monitoring and yield forecasts. *Agricultural Systems*, 168, 224–230.
<https://doi.org/10.1016/j.agry.2018.05.001>
- Vannoppen, A., & Gobin, A. (2021). Estimating farm wheat yields from NDVI and meteorological data. *Agronomy*, 11(5).
<https://doi.org/10.3390/agronomy11050946>

- Vanuytrecht, E., Raes, D., & Willems, P. (2014). Global sensitivity analysis of yield output from the water productivity model. *Environmental Modelling & Software*, 51, 323–332. <https://doi.org/https://doi.org/10.1016/j.envsoft.2013.10.017>
- Vazquez-Cruz, M. A., Guzman-Cruz, R., Lopez-Cruz, I. L., Cornejo-Perez, O., Torres-Pacheco, I., & Guevara-Gonzalez, R. G. (2014). Global sensitivity analysis by means of EFAST and Sobol' methods and calibration of reduced state-variable TOMGRO model using genetic algorithms. *Computers and Electronics in Agriculture*, 100, 1–12. <https://doi.org/https://doi.org/10.1016/j.compag.2013.10.006>
- Viswambharan, S., Kumaramkandath, I. T., & Tali, J. A. (2022). A geospatial approach in monitoring the variations on surface soil moisture and vegetation water content: a case study of Palakkad District, Kerala, India. *Environmental Earth Sciences*, 81(20). <https://doi.org/10.1007/s12665-022-10611-6>
- Wagner, M. P., Slawig, T., Taravat, A., & Oppelt, N. (2020). Remote sensing data assimilation in dynamic crop models using particle swarm optimization. *ISPRS International Journal of Geo-Information*, 9(2). <https://doi.org/10.3390/ijgi9020105>
- Wahba, G. (1975). Smoothing noisy data with spline functions. *Numerische Mathematik*, 24(5), 383–393. <https://doi.org/10.1007/BF01437407>
- Waine, T., Simms, D., Taylor, J., & Juniper, G. (2014). Towards improving the accuracy of opium yield estimates with remote sensing. *International Journal of Remote Sensing*, 35, 1–18. <https://doi.org/10.1080/01431161.2014.951743>
- Wale, A., Dessie, M., & Kendie, H. (2022). Evaluating the Performance of AquaCrop Model for Potato Production Under Deficit Irrigation. *Air, Soil and Water Research*, 15. <https://doi.org/10.1177/11786221221108216>

- Wallach, D., Palosuo, T., Thorburn, P., Hochman, Z., Gourdain, E., Andrianasolo, F., Asseng, S., Basso, B., Buis, S., Crout, N., Dibari, C., Dumont, B., Ferrise, R., Gaiser, T., Garcia, C., Gayler, S., Ghahramani, A., Hiremath, S., Hoek, S., ... Seidel, S. J. (2021). The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. *Environmental Modelling and Software*, 145. <https://doi.org/10.1016/j.envsoft.2021.105206>
- Wang, G., Mehmood, F., Zain, M., Hamani, A. K. M., Xue, J., Gao, Y., & Duan, A. (2022). AquaCrop Model Evaluation for Winter Wheat under Different Irrigation Management Strategies: A Case Study on the North China Plain. *Agronomy*, 12(12). <https://doi.org/10.3390/agronomy12123184>
- Wang, J., Li, X., Lu, L., & Fang, F. (2013). Estimating near future regional corn yields by integrating multi-source observations into a crop growth model. *European Journal of Agronomy*, 49, 126–140. <https://doi.org/10.1016/j.eja.2013.03.005>
- Wang, L. (2021). Data Driven Explanation of Temporal and Spatial Variability of Maize Yield in the United States. *Frontiers in Plant Science*, 12. <https://doi.org/10.3389/fpls.2021.701192>
- Wang, L., & Qu, J. J. (2009). Satellite remote sensing applications for surface soil moisture monitoring: A review. In *Frontiers of Earth Science in China* (Vol. 3, Issue 2, pp. 237–247). Higher Education Press. <https://doi.org/10.1007/s11707-009-0023-7>
- Weiss M, & Baret F. (2016). *Sentinel2 ToolBox Level2 Products S2ToolBox Level 2 products: LAI, FAPAR, FCOVER Version 1.1*.
- Wellens, J., Raes, D., Fereres, E., Diels, J., Coppys, C., Adiele, J. G., Ezui, K. S. G., Becerra, L. A., Selvaraj, M. G., Dercon, G., & Heng, L. K. (2022). Calibration and validation of the FAO AquaCrop water productivity model for cassava (*Manihot esculenta* Crantz). *Agricultural Water Management*, 263. <https://doi.org/10.1016/j.agwat.2022.107491>

- Wood, D., McNairn, H., Brown, R. J., & Dixon, R. (2002). The effect of dew on the use of RADARSAT-1 for crop monitoring: Choosing between ascending and descending orbits. *Remote Sensing of Environment*, 80(2), 241–247. [https://doi.org/https://doi.org/10.1016/S0034-4257\(01\)00304-2](https://doi.org/https://doi.org/10.1016/S0034-4257(01)00304-2)
- Wooster, M. J., Roberts, G. J., Giglio, L., Roy, D. P., Freeborn, P. H., Boschetti, L., Justice, C., Ichoku, C., Schroeder, W., Davies, D., Smith, A. M. S., Setzer, A., Csiszar, I., Strydom, T., Frost, P., Zhang, T., Xu, W., de Jong, M. C., Johnston, J. M., ... San-Miguel-Ayanz, J. (2021). Satellite remote sensing of active fires: History and current status, applications and future requirements. *Remote Sensing of Environment*, 267, 112694. <https://doi.org/https://doi.org/10.1016/j.rse.2021.112694>
- Wu, H., & Li, Z. L. (2009). Scale issues in remote sensing: A review on analysis, processing and modeling. In *Sensors* (Vol. 9, Issue 3, pp. 1768–1793). <https://doi.org/10.3390/s90301768>
- Wu, S., Huang, J., Liu, X., Fan, J., Ma, G., Zou, J., & Huang, J. (2011). *Assimilating MODIS-LAI into Crop Growth Model with EnKF to Predict Regional Crop Yield*. 410–418. <https://doi.org/10.1007/978-3-642-27275>
- Wu, S., Yang, P., Ren, J., Chen, Z., & Li, H. (2021). Regional winter wheat yield estimation based on the WOFOST model and a novel VW-4DEnSRF assimilation algorithm. *Remote Sensing of Environment*, 255, 112276. <https://doi.org/https://doi.org/10.1016/j.rse.2020.112276>
- Wu, Y., Xu, W., Huang, H., & Huang, J. (2022). Bayesian Posterior-Based Winter Wheat Yield Estimation at the Field Scale through Assimilation of Sentinel-2 Data into WOFOST Model. *Remote Sensing*, 14(15). <https://doi.org/10.3390/rs14153727>
- Xavier, A., Hall, B., Casteel, S., Muir, W., & Rainey, K. M. (2017). Using unsupervised learning techniques to assess interactions among complex traits in soybeans. *Euphytica*, 213(8). <https://doi.org/10.1007/s10681-017-1975-4>

- Xiao, J., Aggarwal, A. K., Duc, N. H., Arya, A., Rage, U. K., & Avtar, R. (2023). A review of remote sensing image spatiotemporal fusion: Challenges, applications and recent trends. *Remote Sensing Applications: Society and Environment*, 32, 101005. <https://doi.org/https://doi.org/10.1016/j.rsase.2023.101005>
- Xin, Q., Li, J., Li, Z., Li, Y., & Zhou, X. (2020). Evaluations and comparisons of rule-based and machine-learning-based methods to retrieve satellite-based vegetation phenology using MODIS and USA National Phenology Network data. *International Journal of Applied Earth Observation and Geoinformation*, 93. <https://doi.org/10.1016/j.jag.2020.102189>
- XING, H. min, XU, X. gang, LI, Z. hai, CHEN, Y. jin, FENG, H. kuan, YANG, G. jun, & CHEN, Z. xia. (2017a). 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(11), 2444–2458. [https://doi.org/10.1016/S2095-3119\(16\)61626-X](https://doi.org/10.1016/S2095-3119(16)61626-X)
- XING, H. min, XU, X. gang, LI, Z. hai, CHEN, Y. jin, FENG, H. kuan, YANG, G. jun, & CHEN, Z. xia. (2017b). 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(11), 2444–2458. [https://doi.org/10.1016/S2095-3119\(16\)61626-X](https://doi.org/10.1016/S2095-3119(16)61626-X)
- Xu, C., & Gertner, G. (2011). Understanding and comparisons of different sampling approaches for the Fourier Amplitudes Sensitivity Test (FAST). *Computational Statistics and Data Analysis*, 55(1), 184–198. <https://doi.org/10.1016/j.csda.2010.06.028>
- Xu, G., & Yu, G. (2018). On convergence analysis of particle swarm optimization algorithm. *Journal of Computational and Applied Mathematics*, 333, 65–73. <https://doi.org/10.1016/j.cam.2017.10.026>

- Xu, H., Twine, T. E., & Yang, X. (2014). Evaluating remotely sensed phenological metrics in a dynamic ecosystem model. *Remote Sensing*, 6(6), 4660–4686. <https://doi.org/10.3390/rs6064660>
- Xu, Y., Yang, Y., Chen, X., & Liu, Y. (2022). Bibliometric Analysis of Global NDVI Research Trends from 1985 to 2021. In *Remote Sensing* (Vol. 14, Issue 16). MDPI. <https://doi.org/10.3390/rs14163967>
- Yang, H., Yang, X., Heskell, M., Sun, S., & Tang, J. (2017). Seasonal variations of leaf and canopy properties tracked by ground-based NDVI imagery in a temperate forest. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-01260-y>
- Yang, Y., Anderson, M. C., Gao, F., Johnson, D. M., Yang, Y., Sun, L., Dulaney, W., Hain, C. R., Otkin, J. A., Prueger, J., Meyers, T. P., Bernacchi, C. J., & Moore, C. E. (2021). Phenological corrections to a field-scale, ET-based crop stress indicator: An application to yield forecasting across the U.S. Corn Belt. *Remote Sensing of Environment*, 257, 112337. <https://doi.org/10.1016/j.rse.2021.112337>
- Yang, Y., Tao, B., Liang, L., Huang, Y., Matocha, C., Lee, C. D., Sama, M., El Masri, B., & Ren, W. (2021). Detecting recent crop phenology dynamics in corn and soybean cropping systems of Kentucky. *Remote Sensing*, 13(9). <https://doi.org/10.3390/rs13091615>
- Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., Parnell, S., Fielding, A. H., Bamford, A. J., Ban, S., Barbosa, A. M., Dormann, C. F., Elith, J., Embling, C. B., Ervin, G. N., Fisher, R., Gould, S., Graf, R. F., Gregr, E. J., Halpin, P. N., ... Sequeira, A. M. M. (2018). Outstanding Challenges in the Transferability of Ecological Models. In *Trends in Ecology and Evolution* (Vol. 33, Issue 10, pp. 790–802). Elsevier Ltd. <https://doi.org/10.1016/j.tree.2018.08.001>

- Ye, J., Bao, W., Liao, C., Chen, D., & Hu, H. (2023). Corn Phenology Detection Using the Derivative Dynamic Time Warping Method and Sentinel-2 Time Series. *Remote Sensing*, 15(14). <https://doi.org/10.3390/rs15143456>
- Yli-Heikkilä, M., Wittke, S., Luotamo, M., Puttonen, E., Sulkava, M., Pellikka, P., Heiskanen, J., & Klami, A. (2022). Scalable Crop Yield Prediction with Sentinel-2 Time Series and Temporal Convolutional Network. *Remote Sensing*, 14(17). <https://doi.org/10.3390/rs14174193>
- Yu, J., Smith, A., & Sumner, D. A. (2018). Effects of crop insurance premium subsidies on crop acreage. *American Journal of Agricultural Economics*, 100(1), 91–114. <https://doi.org/10.1093/ajae/aax058>
- Zampieri, M., Weissteiner, C. J., Grizzetti, B., Toreti, A., van den Berg, M., & Dentener, F. (2020). Estimating resilience of crop production systems: From theory to practice. *Science of the Total Environment*, 735. <https://doi.org/10.1016/j.scitotenv.2020.139378>
- Zeng, J., Yuan, X., Ji, P., & Shi, C. (2021). Effects of meteorological forcings and land surface model on soil moisture simulation over China. *Journal of Hydrology*, 603. <https://doi.org/10.1016/j.jhydrol.2021.126978>
- Zeng, L., Wardlow, B. D., Xiang, D., Hu, S., & Li, D. (2020). A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment*, 237, 111511. <https://doi.org/https://doi.org/10.1016/j.rse.2019.111511>
- ZHANG, F., ZHANG, L.-W., SHI, J.-J., & HUANG, J.-F. (2014). Soil Moisture Monitoring Based on Land Surface Temperature-Vegetation Index Space Derived from MODIS Data. *Pedosphere*, 24(4), 450–460. [https://doi.org/https://doi.org/10.1016/S1002-0160\(14\)60031-X](https://doi.org/https://doi.org/10.1016/S1002-0160(14)60031-X)
- Zhang, J., Pan, B., Shi, W., & Zhang, Y. (2023). Monitoring Waterlogging Damage of Winter Wheat Based on HYDRUS-1D and WOFOST Coupled Model and

- Assimilated Soil Moisture Data of Remote Sensing. *Remote Sensing*, 15(17). <https://doi.org/10.3390/rs15174133>
- 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. <https://doi.org/https://doi.org/10.1016/j.compag.2019.105052>
- Zhang, Y., Walker, J. P., & Pauwels, V. R. N. (2022). Assimilation of wheat and soil states for improved yield prediction: The APSIM-EnKF framework. *Agricultural Systems*, 201. <https://doi.org/10.1016/j.agsy.2022.103456>
- Zhang, Y., Walker, J. P., Pauwels, V. R. N., & Sadeh, Y. (2022a). Assimilation of wheat and soil states into the apsim-wheat crop model: A case study. *Remote Sensing*, 14(1). <https://doi.org/10.3390/rs14010065>
- Zhang, Y., Walker, J. P., Pauwels, V. R. N., & Sadeh, Y. (2022b). Assimilation of wheat and soil states into the apsim-wheat crop model: A case study. *Remote Sensing*, 14(1). <https://doi.org/10.3390/rs14010065>
- Zhao, H., Yang, C., Guo, W., Zhang, L., & Zhang, D. (2020). Automatic estimation of crop disease severity levels based on vegetation index normalization. *Remote Sensing*, 12(12). <https://doi.org/10.3390/rs12121930>
- Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y., & Yu, S. (2016). Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. *Remote Sensing*, 8(6). <https://doi.org/10.3390/rs8060469>
- Zomer, R. J., Xu, J., & Trabucco, A. (2022). Version 3 of the Global Aridity Index and Potential Evapotranspiration Database. *Scientific Data*, 9(1). <https://doi.org/10.1038/s41597-022-01493-1>
- Zou, X., & Mörtus, M. (2017). Sensitivity of common vegetation indices to the canopy structure of field crops. *Remote Sensing*, 9(10). <https://doi.org/10.3390/rs9100994>

4 Enhancing wheat yield prediction through data assimilation of satellite-derived canopy cover into AquaCrop: A comparative study of assimilation techniques and timings

Bader Oulaid ^{a,b,c}, Tadesse Gashaw Asrat ^b, Alice E. Milne ^b, Toby Waine ^a, Rafiq El Alami ^c, Stephan M. Haefele ^b, Abdellah Hamma ^c, Nabil Farah ^c, Ron Corstanje ^a

^aCranfield University, College Road, Cranfield, Bedford, MK43 0AL, United Kingdom

^bNet-Zero and Resilient Farming, Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom

^cMohammed VI Polytechnic University, Lot 660, Moulay Rachid, Ben Guerir 43150, Morocco

Abstract

Accurate yield prediction is crucial for food security, resource optimisation, and agricultural policymaking. Traditional calibration methods do not account for the inherent errors in the models, often leading to significant prediction errors. This study investigates the enhancement of wheat yield prediction by integrating satellite-derived canopy cover data into the AquaCrop model using three data assimilation techniques: Ensemble Kalman Filter (ENKF), Ensemble Transform Kalman Filter (ETKF), and Bayesian assimilation. The research focused on 185 sites spread across four semi-arid and arid regions in Morocco. We considered canopy cover as our state variable in the assimilation and used remote sensing data from Landsat 7 and Sentinel-2, trained through a Lasso regression model to provide observations of this. We compared the results from data assimilation using remotely-sensed data acquired across the whole season, from sowing to anthesis, and from anthesis to maturity. This analysis aimed to determine the relative importance of data assimilation at different stages of the crop cycle. We investigated whether coupling data from both sensors improve the assimilation accuracy. Our findings demonstrate that Bayesian assimilation consistently provided the highest accuracy in yield predictions, reducing RMSE by 25% and MAPE by 76.5% compared to no assimilation. The assimilation period was found to be important, with Bayesian methods performing best when integrating canopy

cover data throughout the entire growing season or from anthesis to wheat maturity. ENKF and ETKF were slightly effective during the emergence-to-anthesis stage. The combined use of Sentinel 2 and Landsat 7 in the data assimilation did not consistently outperform the individual sensors, highlighting that increasing temporal resolution through sensor fusion does not necessarily enhance prediction accuracy.

Keywords: Data assimilation, Machine learning, wheat, Sensor fusion.

4.1 Introduction

Crop yield prediction is fundamental to ensuring food security, optimising resource use, and informing agricultural policy (Ziliani et al., 2022a). Accurate predictions enable farmers to make informed decisions regarding crop management (e.g., irrigation and harvesting), which can maximise profitability and reduce losses (Chang et al., 2023). For policy makers, reliable yield forecasts aid in strategic planning and food distribution, trade policies, and emergency preparedness (Li et al., 2021). Crop insurance companies also rely on these predictions to assess risks and set premiums accurately (Zhu, 2024). Despite its significance, crop yield prediction faces notable limitations. The primary challenge is the reliance on data aggregated at broad administrative levels, such as countries, states, or provinces (Engen et al., 2021). This aggregated data often fails to capture the micro-level variability inherent in agricultural fields and could lead to high prediction errors reaching up to 60% (Maharjan et al., 2019). Furthermore, traditional methods of crop yield prediction, such as simple crop yield averaging and basic linear statistical techniques, often fall short in capturing the complex interaction between environment variables and crop responses (Ansarifar et al., 2021). These methods typically do not account for dynamic factors such as climate change (Lobell & Burke, 2010), pest outbreaks (Domingues et al., 2022), or shifts in farming practices (Nyéki & Neményi, 2022), leading to potential inaccuracies in yield forecasts. Consequently, these limitations highlight the need for more detailed and precise methods for yield estimation.

The evolution of crop yield estimation methods reflects the continuous improvement in agricultural modelling technologies. Early methods relied on empirical approaches, using historical data averaging to predict yield outcomes. While straightforward, these methods were limited by their inability to account for changing environmental conditions and agricultural practices (Florence et al., 2021). In response, crop models were developed to simulate the growth of crops based on inputs such as soil properties, weather conditions and crop management practices (Singels, 2013). Prominent examples include the Rothamsted Landscape Model (RLM), the decision Support System for Agrotechnology Transfer (DSSAT) and AquaCrop (Coleman et al., 2017; Jones et al., 2003; Vanuytrecht et al., 2014). The RLM has a detailed representation of soil-plant-atmosphere interactions, making it useful for long-term simulations and landscape-scale studies (Coleman et al., 2021). DSSAT is mainly used because of its versatility and the extensive database of crop varieties it supports (Thorp et al., 2008). AquaCrop, developed by the Food and Agriculture Organization (FAO), is particularly suited for arid and semi-arid conditions. It is designed to simulate the yield response of crops to water, making it an essential tool for water-limited environments (Oulaid et al., 2024a). These models offered a significant improvement over empirical methods by integrating biophysical processes, but they still faced challenges in parametrization and calibration, particularly when applied to diverse farming systems (Kephe et al., 2021). Data-driven models, leveraging machine learning and statistical techniques, have also been used for crop yield prediction (Oikonomidis et al., 2023; Panigrahi et al., 2023). These models can handle large datasets and infer complex patterns, making them suitable for yield prediction. However, their performance is heavily dependent on the quality and quantity of available data (Meroni et al., 2021). Moreover, they often operate as “black boxes”, providing limited details into the underlying processes driving crop growth and subsequently yield (Khaki & Wang, 2019).

Crop yield prediction using process-based models is subject to uncertainties due to model structure, parameterisation, and input data (Dlamini

et al., 2023). Accurate parametrization is essential to reflect the biophysical responses crop models aim to simulate (canopy growth, yield accumulation), and any discrepancies in these parameters may lead to substantial prediction errors. For instance, Tao et al. (2018) demonstrated that varying parameters in WOFOST model could lead yield predictions varying by 66% for wheat crops. Additionally, uncertainties in soil data can compromise predictions. These inaccuracies are often due to the variability in measurement techniques and spatial heterogeneity (Hu & Mo, 2011; Liman Harou et al., 2021). Data-driven models, which often rely on sensor and remote sensing data, are also subject to uncertainties related to data quality, integration (Dong et al., 2009), smoothing and interpolation (Enting et al., 2006). These uncertainties can lead to significant errors that adversely affect yield predictions.

Data assimilation has been proposed as a method to account for uncertainties in both crop models and observations (Jin et al., 2018). Crop models generate predictions with inherent errors, which can be quantified through ensemble generation and parameter sampling. Similarly, observations are prone to errors due to measurement inaccuracies. Data assimilation accounts for these quantified errors to continuously update the state variable of the model, which represent the true condition of the system at a given time. Data assimilation is often done through techniques like the Kalman Filters (KF) and Bayesian methods (de Wit & van Diepen, 2007a; Knowling et al., 2023). Kalman Filters computes a gain which determine the weight given to model predictions versus observations, based on their error covariances. This ensures an optimal integration of information, improving the state estimate (Liu et al., 2010) . However, traditional Kalman Filters has limitations with non-linear systems and large-scale models, since they assume linearity of model dynamics (Sun et al., 2016; Wang et al., 2023). The Ensemble Kalam Filter (EnKF) was developed to address these limitations by using an ensemble of model states to represent prediction uncertainty. Despite requiring significant computational resources, EnKF enhances the robustness and accuracy of crop forecast by continuously integrating new observational data into the models (Evensen, 2003). Another

variation of KFs is the Ensemble Transform Kalman Filters (ETKF). The ETKF improves upon the EnKF by transforming the ensemble members in a way that maintains the ensemble spread and improves the computational efficiency (Bowler & Mylne, 2009). The ENKF and ETKF are sequential data assimilation methods used to directly update the state of the model (in our case, canopy cover) based on crop observations. These observations are predictions of canopy cover derived from remote sensing data. First, during the initialisation step an ensemble of model trajectories is created by perturbing the initial state variables to account for uncertainties. Each member of the ensemble is run forward in time until the first satellite observation becomes available. The predicted canopy state is then updated based on the observation and its associated uncertainties.

Bayesian methods are another approach used in data assimilation to integrate uncertainties in crop models. These methods incorporate prior knowledge from the model predictions and observational data to update the probability distributions of the model states. Unlike the KF approaches which directly update the state variable, the Bayesian assimilation (Bayes) process involves updating the distribution of model parameters based on the likelihood of the ensemble members given the observations. This approach is particularly useful in complex systems where uncertainties are significant, and data is limited.

Data assimilation has been used to enhance the accuracy of crop yield simulations across various models and geographic locations. For instance, Ziliani et al., (2022b) demonstrated that integrating high-resolution satellite imagery from CubeSat into the APSIM model significantly reduced yield prediction errors for maize in the USA, decreasing from 1 to 0.2 Mg/ha. Similarly, Li et al. (2014) utilized Landsat 7 data to retrieve maize leaf area index, which was assimilated into the WOFOST model to produce yield maps in northwest China. Ren et al. (2024) applied the Ensemble Kalman Filter (EnKF) to assimilate remote sensing observations with the WOFOST model, resulting in improved yield simulation accuracy for maize in China. Another study conducted by Curnel et al. (2011) investigated the effect of EnKF assimilation on WOFOST yield predictions for wheat crops and found that, rather than reducing errors, the EnKF increased the

average errors in final grain yield estimation by more than 50%. Ines et al. (2013) integrated both MODIS-LAI and AMSR-E soil moisture maps into the DSSAT-Maize model using an EnKF, highlighting that while the independent assimilation of LAI or soil moisture only slightly improved yield correlations, the simultaneous assimilation of both significantly enhanced the correlation between observed and simulated yields. Lu et al. (2021) conducted a study that involved inverting field-based leaf area measurements of maize into canopy cover, which was then assimilated into AquaCrop using an EnKF for rainfed fields in Nebraska over six growing seasons, resulting in a reduction of errors from 2.48 to 2.03 t/ha RMSE. However, these studies and others in literature present some limitations. Many rely on data from a limited number of sites, which may not capture the full spatial variability of fields and the impact of different management practices. Some studies rely on ground-based measurements for state variables which introduce logistical challenges and can be destructive (Lu et al., 2021). Others used high resolution products which are costly and may limit the scalability of these method (Ziliani et al., 2022b). Lastly, most studies focus on a single assimilation approach and sensor, which may not fully determine the best assimilation configuration minimizing the yield prediction errors.

In this study, we evaluate the effectiveness of the integrating of satellite-derived canopy cover data into the AquaCrop model to improve wheat yield predictions for 185 wheat fields across four semi-arid and arid regions in Morocco. We apply three data assimilation techniques: the Ensemble Kalman Filter (EnKF), the Ensemble Transform Kalman Filter (ETKF), and Bayesian assimilation to determine which of the methods was most likely to give the best results. We explored the use of remote sensing data from two freely available projects, Landsat 7 and Sentinel 2 to understand any relative benefits each might bring. We also assessed the relative importance of data assimilation at different crop growth stages as compared to assimilations at the whole season. Finally, we assessed whether combining data from both satellite sensors provides any additional benefits.

4.2 Material and methods

4.2.1 Study site

This study focuses on four regions in Morocco: Fes-Meknes (34.0339° N, 4.9994° W), Rabat-Sale-Kenitra (34.0209° N, 6.8417° W), Marrakesh-Safi (31.6295° N, 8.0083° W), and Casablanca-Settat (33.5731° N, 7.5898° W). Fes-Meknes and Rabat-Sale-Kenitra, located in the northern semi-arid zone, collectively cover a cultivated land area of 2178 501 hectares. These regions experience wet winters, warm summers, and an annual precipitation ranging from 400 mm to 800 mm. In contrast, Marrakesh-Safi and Casablanca-Settat, situated in central Morocco, cover approximately 3319 020 hectares of cultivated land, and are characterized by dry arid and semi-arid conditions, with hot summers and cooler winters.

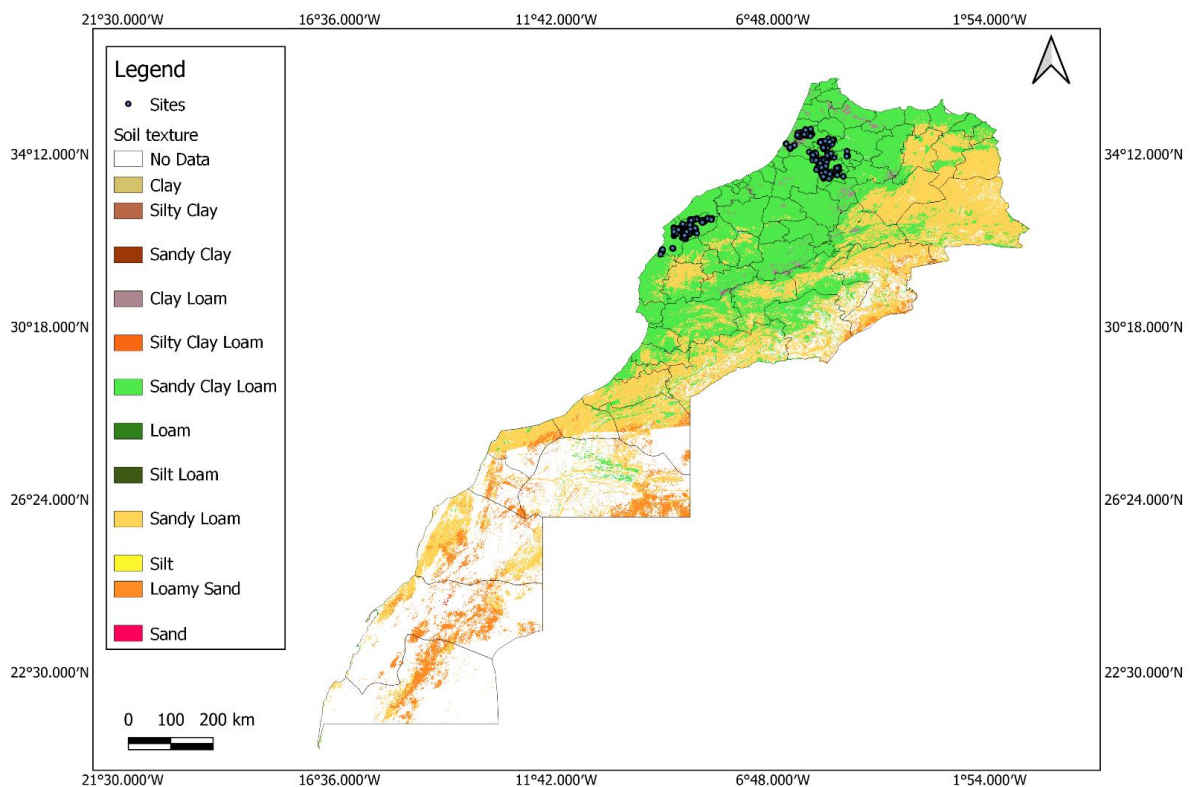


Figure 24 Location of the study area in Morocco showing the soil texture distribution of the studies sites.

4.2.2 Crop and management data

Data were collected from a network of 185 commercial farmers distributed across four regions. The data collection was facilitated by agronomists as part of the Al Moutmir initiative, supported by OCP, Morocco. The dataset included the coordinates of each field, yield, sowing dates, and key wheat phenological dates. The yield reported ranged from a minimum of 1.5 t ha⁻¹ in arid southern regions (i.e., Marrakesh-Safi) to a maximum of 5.6 t ha⁻¹ in the north of Morocco, where precipitations are more frequent. The mean yield across all sites was approximately 3.18 t/ha, with a coefficient of variation of 29%, indicating substantial variation in yield across sites (Figure 25).

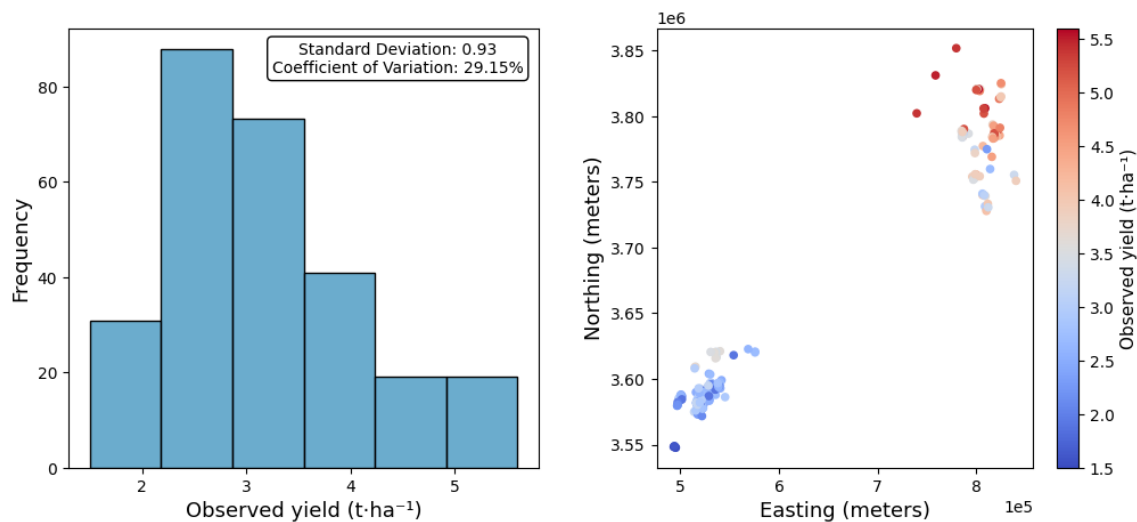


Figure 25 Histogram and spatial distribution of the observed yield (t ha⁻¹) across the studied sites.

The planting dates ranged from late October to mid-November (DOY 295 to 325), emergence dates from mid-November to late December (DOY 320 to 355), anthesis dates from early January to early February (DOY 5 to 40), and maturity dates from early April to early May (DOY 90 to 130). Spatial analysis (Figure 26) revealed distinct clustering patterns. In the northern region, earlier planting and emergence was associated with earlier anthesis and maturity. These areas benefit from more favourable conditions, such as higher precipitation and milder

temperatures. Conversely, the southern region had later planting and emergence, were associated with later anthesis and maturity stages.

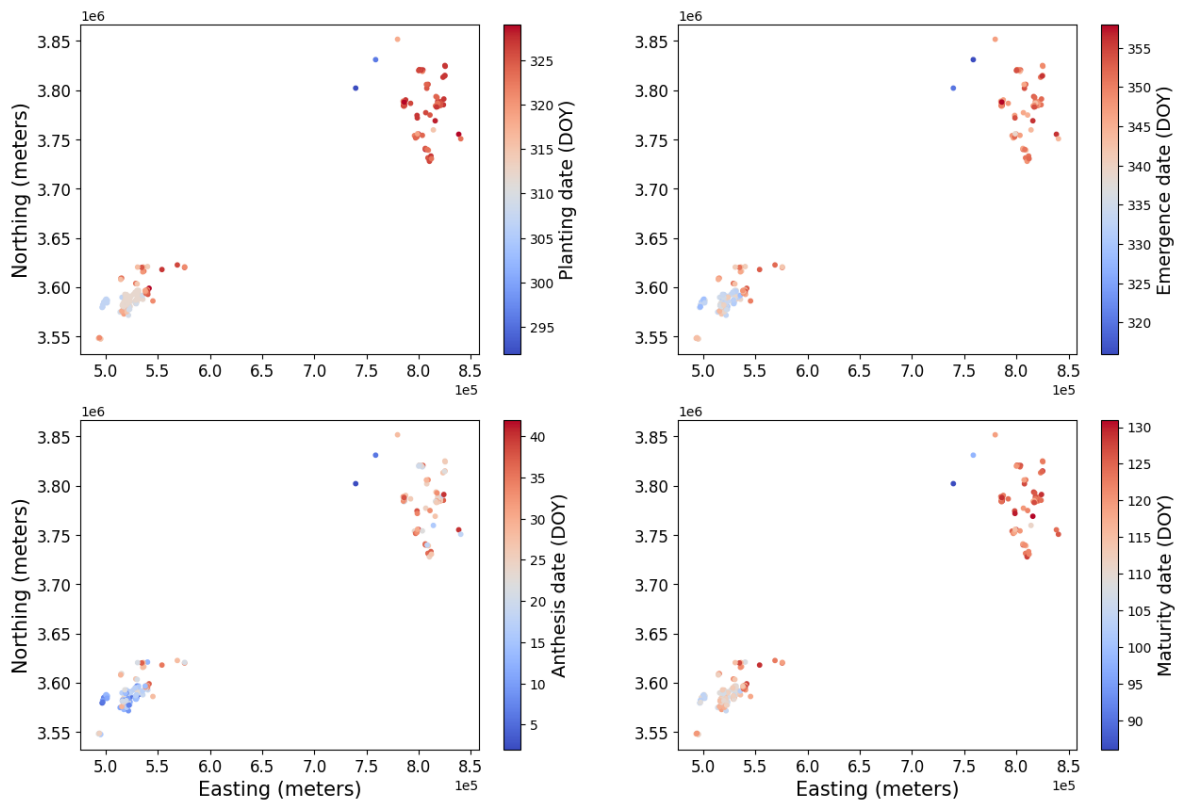


Figure 26 Spatial distribution of planting and key phenological dates for the studies sites.

4.2.3 Crop growth model

AquaCrop is a crop growth model that simulates the yield response to management and environment with a focus on water productivity. The model integrates weather data, soil texture, water table, and management data, to simulate daily crop responses including canopy cover (our state variable of interest), biomass, and crop root development. The final crop yield is simulated by multiplying the biomass with a Harvest Index (HI).

The weather data were sourced from Copernicus ERA5. These data indicated moderate variability in temperatures with an average minimum temperature of

12.82°C and an average maximum temperature of 24.11°C. Precipitation was infrequent but variable, with a mean of 1.16 mm day⁻¹, suggesting dry conditions, while reference daily evapotranspiration showed clear variability with a mean value of 4.54 mm. The soil data for each field were sourced from the iSDA soil database (Hengl et al., 2021). The data revealed 82.7% of sites had sandy clay loam texture, 10.2% of sites were clay loam, and 7.1% of sites were sandy clay.

4.2.4 Remote sensing data

Level-2A orthorectified atmospherically corrected surface reflectance satellite imagery was extracted from both the Landsat 7 ETM+ and Sentinel-2 MSI sensors and processed using the *GEE* Python API. Sentinel-2 provides a 5-day revisit imagery with 10-meter resolution, while Landsat 7 provided a 30-meter spatial resolution across visible, near-infrared, and shortwave infrared bands, with a 16-day revisit time. The bands used in this study for both sensors are shown in Table 14. We extracted images covering the crop growth cycle from September 2020 to June 2021, masking pixels with cirrus and high clouds, and applying a 10-meter buffer to avoid edge effects. To ensure compatibility between datasets from different sensors, images were resampled to a common spatial resolution of 30 meters. Using both Landsat 7 and Sentinel-2 increased the number of satellite images available, enhancing the temporal resolution and enabling the assessment of sensor type impact on the assimilation accuracy.

Table 14 Sentinel-2 MSI and Landsat 7 ETM+ bands used in this study.

Band	Name		Spatial resolution (m)		Wavelength (µm)	
	Sentinel2	Landsat7	Sentinel2	Landsat7	Sentinel2	Landsat7
Red	B4	B3	10	30	0.66-0.68	0.63-0.69
Near Infrared (NIR)	B8	B4	10	30	0.78-0.90	0.77-0.90

Shortwave Infrared (SWIR)	B11	B5	10	30	1.56-1.65	1.55-1.75
--	-----	----	----	----	-----------	-----------

Eight vegetation indices were computed for each of the studied sites. These were the Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), Normalized Difference Water Index (NDWI), Modified Soil Adjusted Vegetation Index (MSAVI), Normalized Difference Infrared Index (NDII), Infrared Percentage Vegetation Index (IPVI), Global Environmental Monitoring Index (GEMI), and the Normalized Burn Ratio (NBR). The list of equations of all the indices is given in Appendix D.1 (Table_Apx D.1-1). Based on the Lasso model, three indices (Equations 4-1, 4-2, and 4-3) were subsequently used to predict the canopy cover (see Section 4.2.5).

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (4-1)$$

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} \quad (4-2)$$

$$SAVI = 1.5 \frac{NIR - RED}{NIR + SWIR + 0.5} \quad (4-3)$$

4.2.5 Data assimilation framework

4.2.5.1 canopy cover estimation from remote sensed data

In our study, canopy cover served as the state variable. The RS data does not provide a direct measure of this but rather the spectral indices listed above from which it can be inferred. We used Toumi et al. (2016) to derive an empirical relationship between these predictor variables and canopy cover. Toumi et al. (2016) provided canopy cover data for nine sites in Morocco during the 2002–2003 and 2003–2004 growing seasons. We computed the vegetation indices listed above for these sites using only Landsat 7 imagery only because Sentinel

2 was not orbiting during that period. These indices were used as predictors in the Lasso (Least Absolute Shrinkage and Selection Operator) regression model to estimate canopy cover.

Lasso regression is a type of linear regression that includes a regularisation term in its cost function, which helps prevent overfitting by penalizing large coefficients. It is particularly suitable for datasets with collinearity among predictors and is effective for small datasets. The regularisation in Lasso allows it to handle multicollinearity by shrinking coefficients of correlated predictors, which improves the model's stability and predictive power. The Lasso regression cost is given by:

$$\frac{1}{2n} \sum_{i=1}^n (y_i - X_i b_i)^2 + a \sum_{j=1}^p |b_j| \quad (4-4)$$

where y_i is the observed canopy cover, X_i are the predictor variables (Vegetation indices), b are the regression coefficients, a is the regularisation parameter and, n and p are the number of observations and predictors, respectively. The model was fitted using leave-one-out cross-validation (LOOCV) to ensure its robustness. LOOCV involves using a single observation from the dataset as the validation data and the remaining observations as the training data, repeating this process for each observation. The optimal value of the regularisation parameter was selected based on the minimum mean squared error across the cross-validation iterations. The model fitting retained just three predictors and the equation derived from the Lasso regression using Landsat 7 was:

$$CC_{L7} = \beta_{L0} + \beta_{L1} NDVI_{L7} + \beta_{L2} NDMI_{L7} + \beta_{L3} SAVI_{L7} \quad (4-5)$$

Uncertainty in the remote sensing data was set to the standard deviation of the residuals of the fitted Lasso model. The use of Sentinel2 for CC estimation required ensuring data compatibility with the existing model. Sentinel 2 imagery was acquired for each study site within a strict +/- 2-day window of the corresponding Landsat 7 acquisitions to minimise the impact of potential growth stage variations between the two datasets. Subsequently, transformation

equations were established to convert vegetation indices from Landsat7 to Sentinel 2, ensuring the model, trained on Landsat-7 data, could be applied for CC prediction using Sentinel 2 imagery. Using collocated estimates of each of the vegetation indices we derived a set of linear equations to relate indices derived from Sentinel 2 data to those derived from Landsat7 data.

$$NDVI_{L7} = \alpha_1 NDVI_{S2} + \alpha_{S1} \quad (4-6)$$

$$NDMI_{L7} = \alpha_2 NDMI_{S2} + \alpha_{S2} \quad (4-7)$$

$$SAVI_{L7} = \alpha_3 SAVI_{S2} + \alpha_{S3} \quad (4-8)$$

The lasso regression model was then applied to Sentinel2 data, resulting in the following equation:

$$CC_{S2} = \alpha_0 + \beta_{L1} \alpha_1 NDVI_{S2} + \beta_{L2} \alpha_2 NDMI_{S2} + \beta_{L3} \alpha_3 SAVI_{S2} \quad (4-9)$$

where $\alpha_0 = \beta_{L0} + \beta_{L1} \alpha_{S1} + \beta_{L2} \alpha_{S2} + \beta_{L3} \alpha_{S3}$

4.2.5.2 Ensemble generation

Oulaid et al. (2024) found that the canopy cover prediction by Aquacrop was sensitive to six key parameters: days to emergence (Eme), days to senescence (Sen), days to maturity (Mat), maximum canopy cover (CCx), canopy decline coefficient (CDC), and canopy growth coefficient (CGC). In our assimilation these were treated as Gaussian random variables. Taking a conservative approach and based on Oulaid et al., (2024), we assume the bounds of these parameters to be $\pm 30\%$ of the AquaCrop wheat default values (Raes et al., (2009). For each parameter, 250 samples were drawn from a Gaussian distribution with the mean set as the default value and the standard deviation calculated to ensure 95% coverage within these bounds. The sampled parameters were clipped to ensure they remained within their respective bounds. Figure 27 provides a visual representation of the distribution and bounds of the sampled parameters.

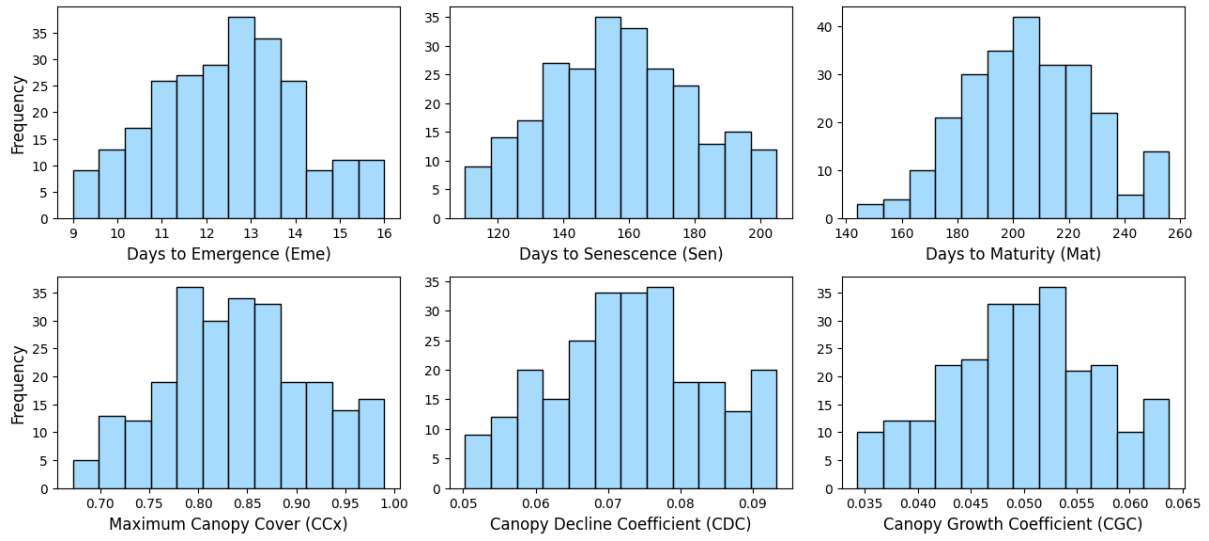


Figure 27 Histograms of AquaCrop sampled parameters.

4.2.5.3 Assimilation of remote sensed data into AquaCrop

Three data assimilation methods were used in this study: The ensemble Kalman Filter (ENKF) and Ensemble Transform Kalman Filter (ETKF) and a Bayesian assimilation (Bayes). The algorithms for these methods were incorporated into the Python version of AquaCrop. For both ENKF and ETKF, the ensemble mean ($\bar{A}$) and deviation of the ensemble members from this mean (A') are calculated. The forecast error covariance matrix, P_e , is then derived from these deviations using the formula:

$$P_e = \frac{1}{N-1} A' (A')^T \quad (4-10)$$

where N is the number of ensemble members. Subsequently, perturbed observations, D , are generated based on predicted canopy cover values and their associated uncertainties (i.e., standard deviations).

In the ENKF, the Kalman gain, K , is then computed using the equation:

$$K = P_e H^T (H P_e H^T + R_e)^{-1} \quad (4-11)$$

where H is the observation matrix and R_e is the observation error covariance matrix. This gain adjusts the forecast of the ensemble to better match the satellite imagery observations. Following this, the ensemble is then updated using the equation:

$$A_a = A + K (D - H A) \quad (4-12)$$

where A_a is the updated (analysed) ensemble. In the ETKF, the deviations of the observations from this mean, D' , are computed using the Cholesky decomposition of the observation error covariance matrix, $\sqrt{R_e}$ using the formula:

$$D' = \sqrt{R_e}^{-1} (D - H \bar{A}) \quad (4-13)$$

The deviations are then transformed using the equation:

$$\hat{D} = \sqrt{P_e} A' \quad (4-14)$$

Next, the model is updated using the formula:

$$A_a = \bar{A} + \sqrt{P_e} (\hat{D} + D') \quad (4-15)$$

For both ENKF and ETKF, the model is run again until the next satellite observation is available, and the analysis step is repeated until no satellite imagery is available. Finally, the model is run till termination, the final yield for each ensemble member is computed and the ensemble mean yield is taken as the final yield estimate.

For the Bayesian assimilation (Bayes), the process involves integrating observed satellite imagery with model simulations to estimate the posterior distribution $p(\theta|D)$ of the model parameters using Bayes' theorem:

$$p(\theta|D) = \frac{p(\theta) \cdot p(D|\theta)}{p(D)} \quad (4-16)$$

where $p(\theta)$ represents the prior distribution of the AquaCrop model parameters. $p(D|\theta)$ is the likelihood function, which quantifies the agreement between the

models' predictions and the observed data. $p(D)$ is the marginal likelihood, a normalizing constant ensuring the posterior distribution sums to one.

Assuming normally distributed errors, the likelihood is defined as :

$$p(D|\theta, \sigma_{obs}^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_{obs}^2}} \exp\left(-\frac{(cc_{obs,i} - cc_{sim,i}(\theta))^2}{2\sigma_{obs}^2}\right) \quad (4-17)$$

where $cc_{sim,i}(\theta)$ is the model's prediction for the i th observation given the parameters θ , σ_{obs} is the standard deviation of the observational errors. Through this process, the posterior distribution is derived, representing updated beliefs about model parameters after considering the observed data.

4.2.6 Model evaluation

Model evaluations were carried out to assess the performances of the three data-assimilation methods using the two sensors and at different crop growth stages and. The coefficient of determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (4-18)$$

quantifies the goodness of fit of the assimilated data to the observed data, where n is the number of observations, O_i , $\bar{O}$ and P_i are the observed, mean of observed and predicted rainfed wheat yield values, respectively.

The Root Mean Square Error (RMSE) was used to assess the overall discrepancy between the actual and the predicted yield according to

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (4-19)$$

and the Mean Absolute Percentage Error (MAPE) was calculated to express errors as a percentage of the actual yield values

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{O_i - P_i}{O_i} \right| \quad (4-20)$$

4.3 Results

4.3.1 Canopy cover estimation from remote sensed data

The lasso regression model trained and validated using Landsat 7 data provided robust estimated for canopy cover. Table 15 summarises the performance metrics of the fitted model. The model, evaluated using a leave-one-out cross-validation (LOOCV), showed an R^2 of 0.8, an RMSE of 0.13% and a MAE of 11% indicating a robust correlation between the predicted and observed CC values (Figure 28). The residuals show that the model does not systematically overestimate or underestimate the canopy cover (Figure 28.c). Notably, the higher coefficient for NDMI (0.79) highlighted its strong association with moisture conditions, which is particular for rainfed wheat in arid and semi-arid regions (Table 15 and Figure 28.a).

Table 15 Fitted model and performance metrics for Lasso model trained and validated using Landsat 7 data.

Fitted model for canopy cover	Sample size	R^2	RMSE	MAE
0.5274 NDVI _{L7} + 0.7985 NDMI _{L7} + 0.6077 SAVI _{L7} - 0.0595	48	0.80	0.13	0.11

The fitted linear relationships between the Landsat and Sentinel derived metrics achieved high performance, with R^2 values all exceeding 0.9 (Table 16, Figure 29).

Figure 29 illustrates the regression lines along with the 95% confidence and prediction intervals for the NDVI, NDMI and SAVI transformations between Landsat 7 and sentinel 2. The narrow confidence and prediction intervals indicate prediction and reliability of the transformation models. However, there were some outliers present in plots, particularly in the lower values of the vegetation indices which may be due to differences atmospheric or illumination conditions at the time of image acquisition.

Table 16 Performance metrics for vegetation indices transformation.

Fitted model	Sample size	R2	RMSE	MAE
$NDVI_{L7} = 0.91 NDVI_{S2} + 0.006$	2598	0.93	0.06	0.04
$NDMI_{L7} = 0.9 NDMI_{S2} + 0.033$	2598	0.95	0.05	0.03
$SAVI_{L7} = 0.89 SAVI_{S2} + 0.004$	2598	0.945	0.04	0.02

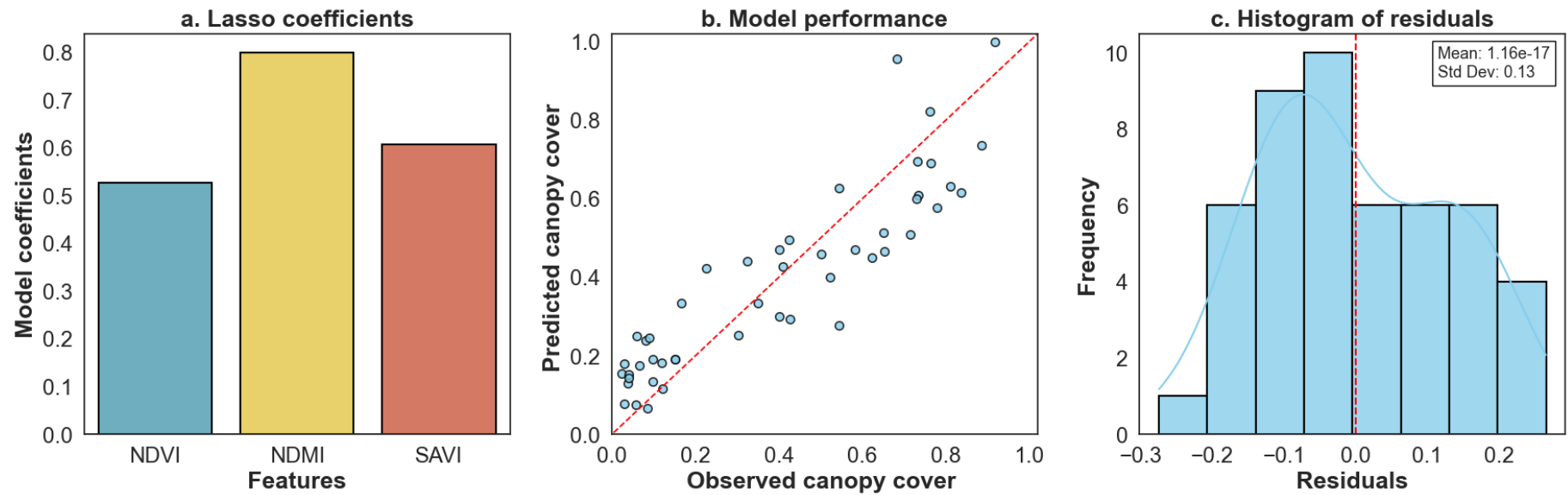


Figure 28 Analysis of the Lasso regression model for canopy prediction using Landsat 7 data.

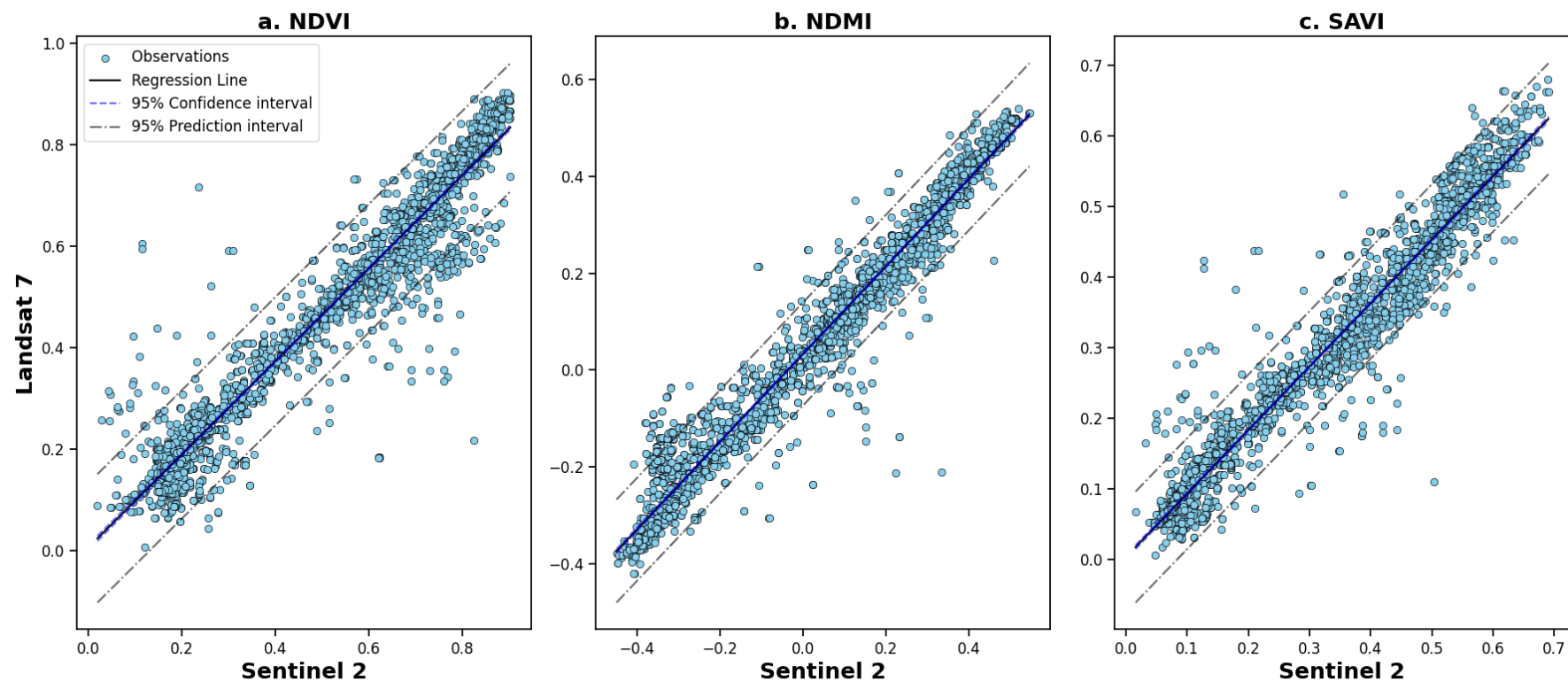


Figure 29 Regression lines with 95% confidence and prediction intervals for NDVI, NDMI and SAVI transformation between Landsat7 and Sentinel 2.

4.3.2 Canopy cover assimilation

The analysis of the temporal distribution of satellite observations (Figure 30.a) shows that Sentinel 2 provided more frequent observations ($n=3337$) compared to Landsat 7 ($n=2598$), especially during the peak growth period from January to March 2021 (Fig 7a). This is not surprising given the higher temporal resolution from Sentinel 2, with a 5-day revisit cycle, compared to Landsat 7's 16-day revisit cycle. The distribution of canopy cover values (Figure 30.b and Figure 30.c) highlights differences between the two sensors. Landsat 7 data shows a higher frequency of lower canopy cover values (0.0 to 0.4), while Sentinel 2 data exhibits a more uniform distribution of canopy cover values across the range, with a notable increase in frequency towards higher canopy cover values (0.8 to 1.0).

The assimilation of remote sensing data into the AquaCrop model using the three approaches revealed two general patterns in canopy cover prediction. Figure 31 illustrates these. The first general pattern that we observed was where the unassimilated ensemble showed the simulated canopy to emerge in line with that observed, yet the decline was predicted to be generally more protracted than observed (e.g., Site 1, see Fig. 8a). In these cases, ETKF adjusted the simulated canopy cover to align with the remote sensed observations (red dots), resulting in a pronounced peak achieved earlier in the season and a rapid decline that closely mirrored the observed data. This aggressive adjustment in the increase and decrease of the canopy led to a shorter period of canopy cover, which limited photosynthesis and biomass accumulation, resulting in lower simulated yields compared to the baseline simulations. EnKF tended to provide a more moderate adjustment, showing a gradual peak and decline. The Bayesian assimilation, tended to result in a canopy cover more closely following the observed data with minimal deviation. This often meant that the Bayesian assimilation gave a better yield prediction, but this was not always the case.

The second general pattern that we observed was where the observed canopy emergence was much faster than that simulated (e.g., Sites 18 and 117, see Figure 31). Again, the KF methods (and particularly ETKF) did a good job of

adjusting for this over the period that the RS data were available but again this often resulted in a protracted senescence period that was not realistic. The Bayesian assimilation retained a sensible canopy profile but was not always able to adjust sufficiently to match observations as the parameter ranges in the prior did not encompass what was necessary to capture the observations. These posterior parameter distributions did tend to give a less protracted canopy, however, usually improving yield prediction despite obvious large errors in the canopy profile (e.g., Site 18) but not always (Site 117).

The Bayesian assimilation consistently provided the highest precision and confidence in canopy cover predictions (narrow shaded blue area). The presence of observations significantly reduced uncertainties in simulated canopy cover, as compared to the absence of observations specifically at the end of the growing season.

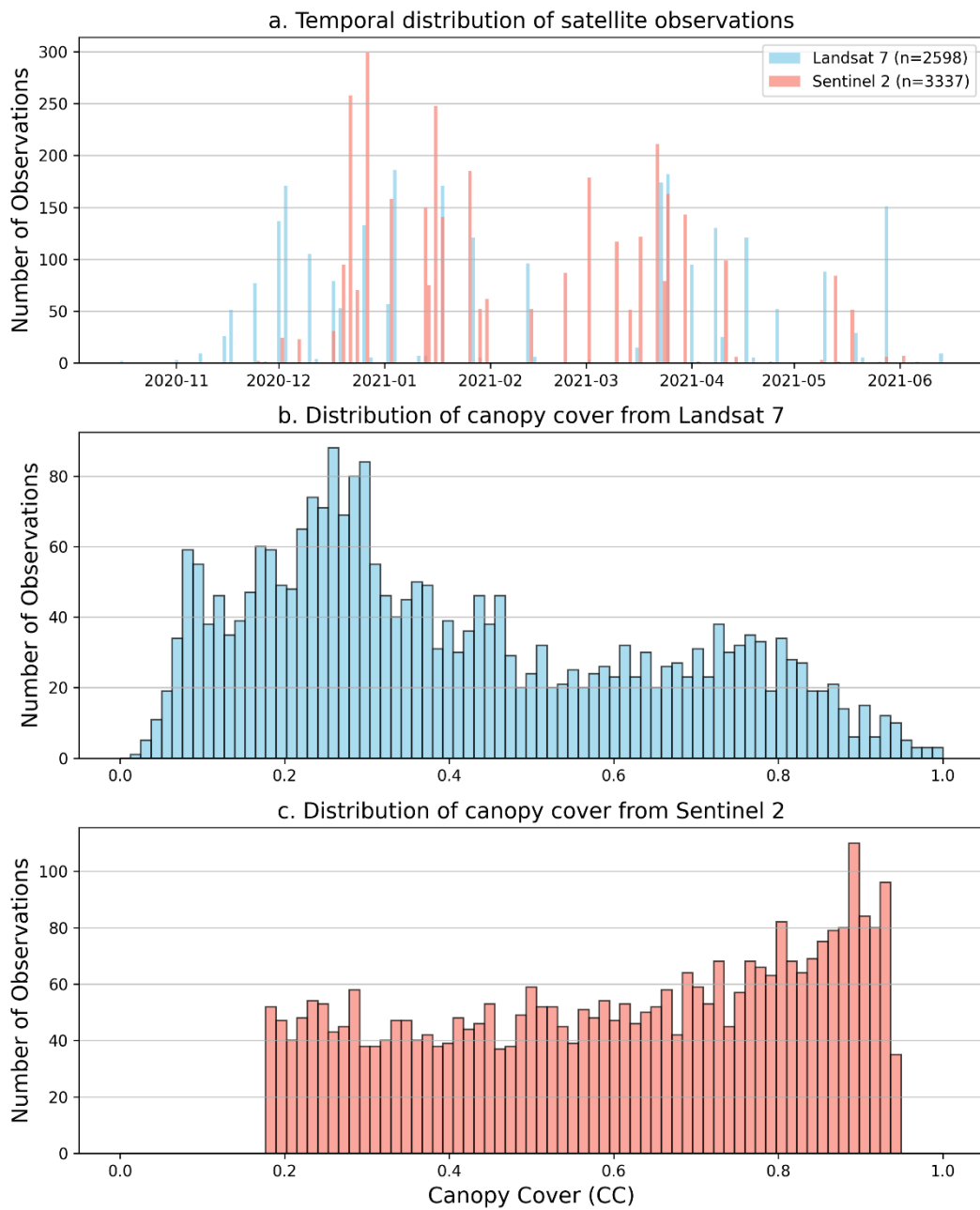


Figure 30 Temporal distribution of satellite imagery and canopy cover distribution from Landsat7 and Sentinel2 used for data assimilation into AquaCrop.

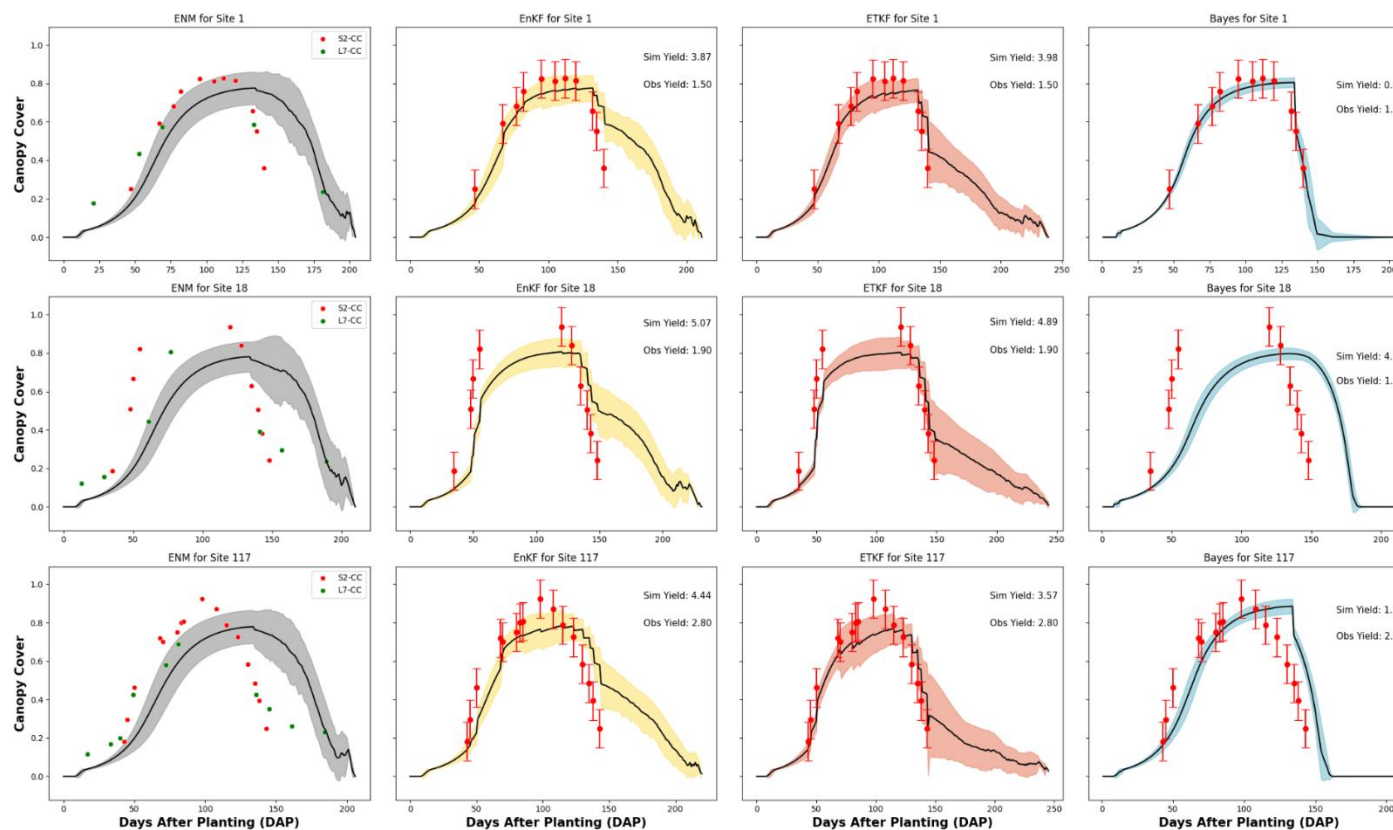


Figure 31 Comparison of data assimilation techniques for simulated canopy cover using Sentinel 2 observations. ENM shows AquaCrop canopy cover ensemble with the sampled parameters. EnKF, ETKF and Bayes denote the Ensemble Kalman Filter, Transform Kalman Filters and Bayesian assimilations, respectively. The black line and shaded area indicate the mean and standard deviation of the assimilated canopy cover. The red dots with errors represent the sentinel2-based canopy cover with its standard deviation.

4.3.3 Yield predictions after assimilations

The full-season assimilation showed Bayesian assimilation to consistently achieve higher performance across all datasets (Figure 32 a-c). For Sentinel 2, Bayesian assimilation achieved the highest R^2 of 0.51 and the lowest RMSE (0.64 t. ha⁻¹) and MAPE (16.72%). Similarly, for Landsat 7, Bayes resulted in an RMSE of 0.66 t. ha⁻¹ and a MAPE of 17.9%, with an R^2 of 0.48. When combining both Sentinel 2 and Landsat 7, Bayesian assimilation showed an RMSE of 0.66 t. ha⁻¹ and a MAPE of 17.51% with an R^2 of 0.49. In contrast, the no assimilation benchmark approach, despite having a similar R^2 of 0.49, significantly overestimated yields, as indicated by an RMSE of 0.88 t. ha⁻¹ and a MAPE of 74.53%. This demonstrates a systematic bias in the predictions without data assimilation. Bayesian assimilation reduced the prediction errors, decreasing RMSE and MAPE by 25% and 76.5%, respectively, compared to no assimilation approach. The differences in performance ENKF or ETKF were marginal with no consistent trend (Figure 32). When comparing sensor, the full-season assimilation showed that Sentinel 2 provided the best performance across all configurations, with the lowest prediction errors and the highest R^2 . The combined dataset generally did not outperform the individual sensors.

The evaluation of the assimilation methods across different time periods using Sentinel 2 (Figure 33) revealed notable variations in yield prediction performance. The Bayes method demonstrated the most accurate results during when remotely sensed canopy cover acquired across the full growing season is used, achieving the lowest prediction errors of 0.64 t. ha⁻¹ RMSE and R^2 of 0.51. Comparatively, Kalman Filter variations (ENKF and ETKF) exhibited optimal performance in the emergence to anthesis stage with an RMSE of 0.68 t. ha⁻¹ and an R^2 of 0.45. A similar pattern was observed while assimilating using Landsat 7 and the combined dataset, where ETKF and ENKF again performed best during the emergence-anthesis stage. However, Bayesian method demonstrated the best performance during the anthesis-maturity stage for these two sensor configurations. For detailed performance metrics of Landsat7 and the combined dataset, refer to the appendix D.2.

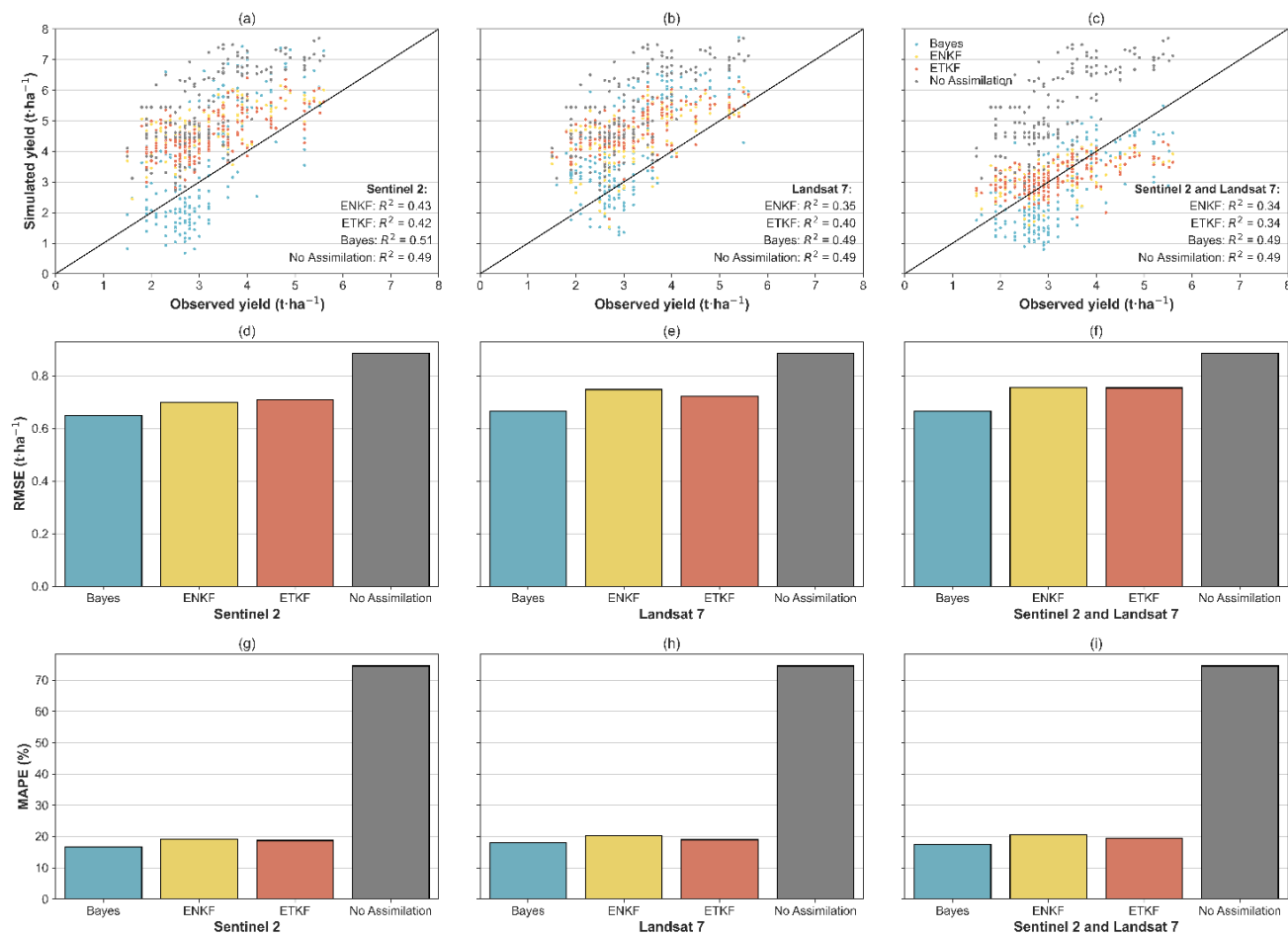


Figure 32 Performance evaluation of data assimilation techniques using Sentinel2, Landsat 7 and a combined dataset of Landsat 7 and Sentinel2 for yield prediction.

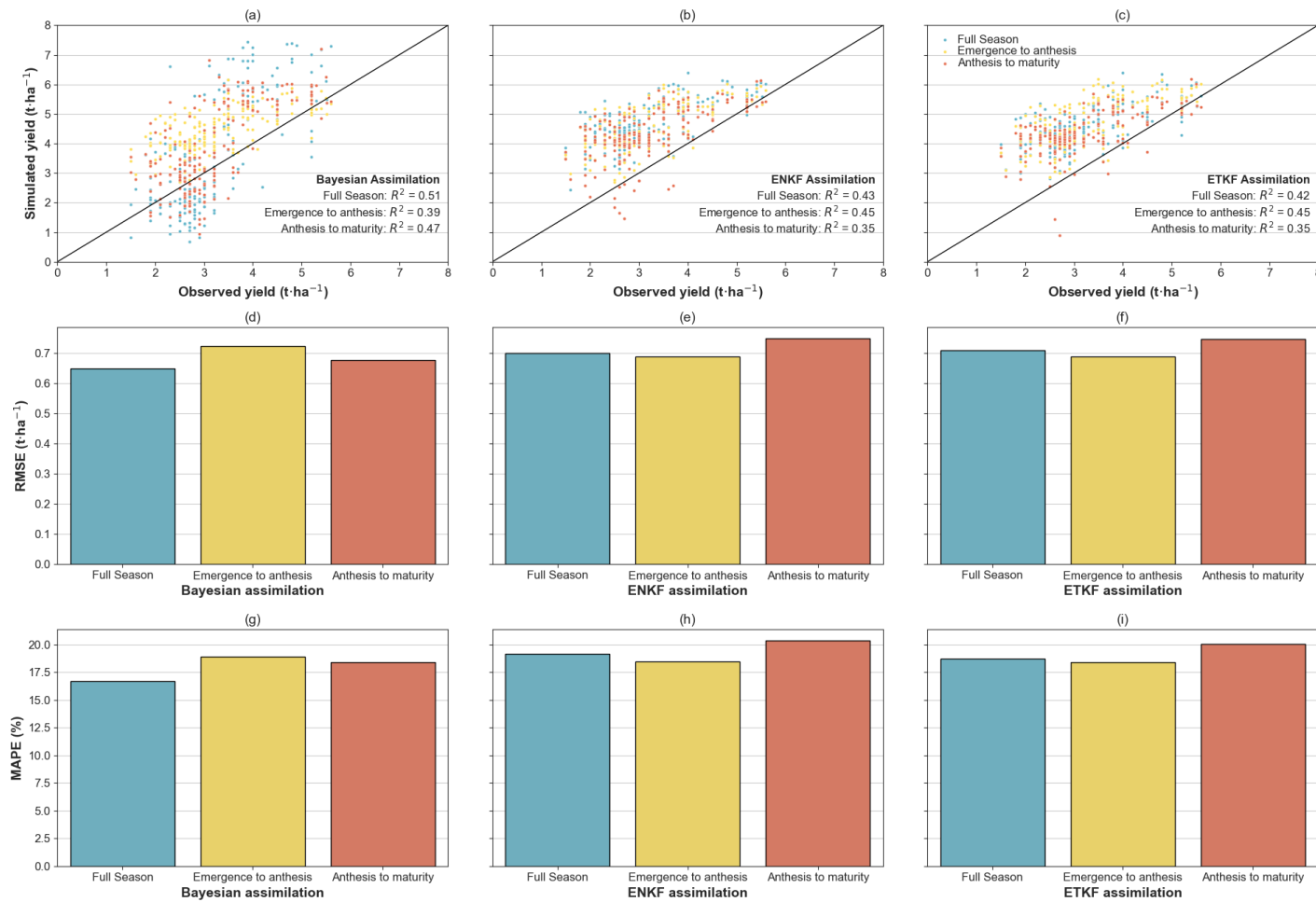


Figure 33 Performance evaluation of data assimilation techniques using Sentinel 2 for yield prediction under different assimilation timings

4.4 Discussion

4.4.1 Remotely sensed state variables in crop models

Precise observations of state variables, such as canopy cover, from remote sensing data can be helpful for improving crop yield predictions. This study used a Lasso regression model to estimate canopy cover from satellite-derived vegetation indices. The Lasso regression model, trained and validated with Landsat 7 demonstrated robust performance even with a relatively small sample size of 48 observations. The use of leave-on-out cross-validation (LOOCV) further validated its performance across all individual data points, ensuring the model generalizes well to new, unseen data. Previous research has demonstrated the effectiveness of using multiple vegetation indices to capture the development of canopy cover over the growing season. For instance, (Tenreiro et al., 2021) assessed the use NDVI as a proxy to estimate canopy cover and found that the correlations are adequate for the majority of crops as ($R^2 > 0.57$) for most of the crops including wheat. Badrul et al. (2022) demonstrated a positive correlation between high NDMI values and increased canopy cover estimates. Among the vegetation indices used in this study, the Normalized Difference Moisture Index (NDMI), which capture moisture content in vegetation, emerged as the best estimator of canopy cover. The NDMI had the largest fitted coefficient (0.79) in the Lasso model, highlighting the strong association of moisture content with canopy cover, particularly in rainfed wheat systems. This result aligns with the findings of Andreatta et al., (2022) who demonstrated NDMI'S effectiveness as a predictor of canopy cover in water-limited areas. Inversion methods are different approaches of estimating state variables from remotely sensed data. Upreti et al., (2020) used an empirical relationship to derive the canopy cover from leaf area index (LAI) based on Venus imagery. While this method is straightforward, applying the LAI-CC inversion to ecosystems with different vegetation types and canopy structures may lead to significant errors (C. Li et al., 2022; Raj et al., 2021). The Lasso regression model used to estimate canopy cover is bound by several assumptions and limitations

that impact its broader applicability. The model assumes linear relationships between vegetation indices and canopy cover, which may oversimplify the complex, nonlinear interactions that characterize canopy dynamics, particularly under extreme environmental conditions or in mixed cropping systems. Additionally, the sparsity assumption inherent in Lasso regression prioritizes a subset of features, potentially excluding weaker but biologically relevant predictors. This feature selection process, while improving interpretability, could limit the model's robustness when applied to regions with differing environmental conditions or crop management practices. The regression coefficients derived from this study reflect the specific agro-climatic conditions of semi-arid and arid regions and may not generalize to regions with contrasting climatic dynamics. Future research should explore integrating more flexible machine learning models, such as random forests or neural networks, to capture nonlinear relationships and interactions, while retaining interpretability through techniques similar to SHAP analysis.

One notable observation from this study is that Sentinel2 data often predicted the cover to be larger than the predicted by Landsat 7, particularly in the middle of the season. This is despite both sets of data being processed to have the same spatial resolution and the acquisition dates being filtered to ensure a maximum difference of +/- 2 days. Several factors may contribute to the larger canopy cover values observed in Sentinel 2 data. Firstly, Sentinel 2 has finer radiometric resolution (12 bits) compared to Landsat7 (8 bits). This characteristic allows for greater distinction in vegetation reflectance, which may lead to capturing more detailed variations in the vegetation. The difference in the atmospheric correction algorithms could further contribute to the discrepancies of estimated canopy cover by the two sensors. Sentinel 2 data are processed with the Sen2Cor algorithm (Main-Knorn et al., 2017), which includes cloud screening, atmospheric and terrain correction, while Landsat 7 is processed using LEDAPS (Vermote, 2007), which focuses primarily on atmospheric correction without terrain.

4.4.2 Effectiveness of data assimilation techniques and periods

In this study, Bayesian assimilation consistently demonstrated better performance across all datasets (Landsat 7, Sentinel 2 and the combined dataset) when applied throughout the entire growing season. The Bayesian method resulted in a significant reduction of prediction errors, with a 25% decrease in RMSE and a 76.5% decrease in MAPE compared to the no assimilation benchmark. EnKF and ETKF also showed improvements over the no assimilation scenario but were less consistent compared to Bayesian assimilation. To our knowledge, no prior research has conducted a comparative assessment of the performance of three methods in general, or tested the Ensemble Transform Kalman filter on AquaCrop, specifically. However, the overall trend of assimilation methods enhancing the accuracy of yield predictions aligns with previous research. De Wit .(2007) highlighted that assimilating soil moisture into the WOFOST model improved the yield predictions for over 66% of the studied regions in Spain, France, Italy and Germany. Kivi et al. (2023) conducted a comparative analysis of the impact of various sources of state variables on EnKF assimilation performance. The study demonstrated that in-situ based soil moisture measurements provided better EnKF assimilation outcomes compared to those derived from remote sensing estimates. Furthermore, the assimilation results were notably more significant in regions experiencing water stress.

The data assimilation methods in this study are subject to several model-related boundary conditions that influence their performance and generalizability. Bayesian assimilation assumes accurate prior distributions for crop growth parameters, which may fail to reflect real-world variability, particularly under extreme conditions like droughts or heatwaves. EnKF relies on the independence of ensemble members, which can lead to systematic biases if shared errors accumulate, such as during prolonged senescence. ETKF's linear adjustments to the ensemble mean can result in unrealistic canopy fluctuations when observational data diverges significantly from model predictions. Additionally, all three methods presume that AquaCrop adequately captures key crop processes,

which may not hold in scenarios involving complex interactions, such as multi-stressor environments with nutrient limitations or pest infestations. The methods also depend on consistent and frequent observational data; any gaps or inaccuracies in these data can reduce assimilation accuracy, especially during critical phenological stages like senescence or reproductive growth. Furthermore, the fixed parameterizations in AquaCrop may not account for spatial heterogeneity, such as variations in soil properties or microclimates, which could limit the effectiveness of assimilation techniques in different landscapes. These methods were tested in a specific agro-climatic context, using a single crop model, and may not perform similarly in other regions, cropping systems, or under variable management practices. Future work should address these limitations by improving model flexibility, integrating nonlinear interactions, and validating these techniques across diverse conditions to enhance their robustness and applicability.

The assimilation results were linked to the behavior of the assimilated canopy cover using the three methods. Bayesian assimilation adjusts the parameter space but does not update the state variable directly. This results in more realistic canopy cover profiles than ENKF and ETKF (Figure 31). Despite not always aligning perfectly with the remote sensing observations, Bayesian assimilation produced canopy shapes that closely resembled the expected crop growth patterns. This indicates that Bayesian methods effectively leveraged the model's sampled parameters and incorporated uncertainties in both the model and observational data. ETKF generally provided canopy cover estimates closer to the remote sensing observations compared to ENKF. Its more aggressive adjustments resulted in better alignment with observed data during specific periods of the growing season, particularly during the initial stages of canopy development and peak growth. This can be attributed to the transformation matrix used in the Ensemble Transform Kalman Filter (ETKF), which facilitates collective ensemble updates towards the observations, thereby effecting significant adjustments to the ensemble mean. This contrasts with the Ensemble Kalman Filter (EnKF), which updates the ensemble members on an individual basis. The

pronounced adjustments by ETKF occasionally led to unrealistic fluctuations in canopy cover, indicating a potential over-reliance on the observational data at the expense of the model's inherent dynamics. In addition, both ETKF and EnKF tended to prolong the senescence period, leading to less accurate yield predictions. This extended senescence period can be attributed to the absence of observational data, which prevents the correction of the model state and allows model errors to accumulate. Consequently, the ensemble forecasts fail to capture the timing and progression of senescence accurately, reducing the precision of yield predictions. Consequently, the reliance on sampled model parameters only could have led to an overestimation of the senescence period.

The timing of data assimilation was found to be essential in determining the effectiveness of the methods. Bayesian methods performed best when integrating canopy cover data throughout the entire growing season, providing a comprehensive update of the model states at each stage of the crop cycle. This continuous updating process allows for a more accurate reflection of the crop's development and response to environmental conditions. Conversely, the Kalman filter-based methods were most effective during the emergence-to-anthesis stage. This early stage of growth is critical for determining the overall productivity of the crop, as it covers key phenological phases such as leaf area expansion, canopy establishment, and the beginning of reproductive development. Accurate assimilations during this period can significantly enhance yield forecasts by ensuring that the model closely follows the observed early growth patterns. The effectiveness of EnKF and ETKF during this stage highlights their ability to capture early growth dynamics by adjusting the canopy cover to align with observations, thus improving the representation of initial crop canopy conditions and subsequent growth trajectories. These results align with previous findings from Huang et al., (2016) who highlighted the importance of pre-heading stage assimilation and noted the challenges associated with post-heading assimilations, which are related to crop drying and the decreasing LAI estimates.

The datasets used in this study were focused on semi-arid and arid regions of Morocco, which limits their representativeness for other agro-climatic zones.

In particular, the reliance on Landsat 7 and Sentinel-2 data at moderate spatial resolutions (30m and 10m, respectively) may mask smaller-scale heterogeneity within fields, especially in highly fragmented agricultural landscapes. This could influence the effectiveness of data assimilation, particularly for capturing localized variations in canopy dynamics and water stress responses. Furthermore, the single-season nature of the dataset restricts the ability to evaluate the methods' robustness under varying climatic conditions, such as extreme drought or wetter-than-average years. While the findings provide valuable insights for regions with similar climatic and management conditions, broader applicability may require datasets covering diverse cropping systems, environmental conditions, and multi-year observations. Additionally, differences in management practices, such as irrigation timing or fertilization strategies, can further constrain the transferability of these results. Future studies should address these gaps by integrating high-resolution remote sensing data and expanding the temporal and geographic scope of the datasets to validate these methods across varying contexts.

4.5 Conclusions

The research focused on 185 wheat sites scattered across four regions in Morocco, characterized by semi-arid and arid conditions, which present unique challenges for crop management and yield prediction. Our key findings indicate that Bayesian assimilation consistently provided the highest accuracy in yield predictions across all datasets, with the lowest prediction errors. This method's ability to probabilistically integrate prior knowledge and observational data allows for more robust handling of uncertainties, particularly in data-limited and variable environments. The comparative performance of ENKF and ETKF showed slight variations depending on the sensor used, highlighting the importance of selecting appropriate assimilation methods based on data characteristics and computational resources. Additionally, we investigated whether coupling data from both sensors improves assimilation accuracy. Interestingly, the combined use of Sentinel-2 and Landsat 7 in data assimilation did not consistently

outperform the individual sensors, highlighting that increasing temporal resolution through sensor fusion does not necessarily enhance prediction accuracy. The study also investigated the impact of data assimilation timing, with Bayesian methods performing best when integrating canopy cover data throughout the entire growing season. In contrast, Kalman filter-based methods were most effective during the emergence-to-anthesis stage, capturing early growth dynamics essential for accurate yield forecasting. By leveraging remote sensing data from Landsat 7 and Sentinel-2, this research demonstrated the significant advantages of RS-based canopy cover retrieval in improving crop model accuracy. Bayesian assimilation reduced RMSE by 25% and MAPE by 76.5% compared to the no assimilation benchmark.

Credit authorship contribution statement

Bader Oulaid: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Alice E. Milne:** Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Toby Waine:** Writing – review & editing. **Rafiq El Alami:** Resources, Writing –review & editing, Supervision, Project administration, Funding acquisition. **Tadesse Gashaw Asrat:** Resources Data curation, Writing – review & editing. **Stephan Haefele:** Writing – review & editing. **Abdellah Hamma:** Resources, review & editing. **Ron Corstanje:** Resources, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

4.6 References

- Andreatta, D., Gianelle, D., Scotton, M., & Dalponte, M. (2022). Estimating grassland vegetation cover with remote sensing: A comparison between Landsat-8, Sentinel-2 and PlanetScope imagery. *Ecological Indicators*, 141. <https://doi.org/10.1016/j.ecolind.2022.109102>
- Ansarifar, J., Wang, L., & Archontoulis, S. V. (2021). An interaction regression model for crop yield prediction. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-97221-7>
- Badrul Hisham, N. H., Hashim, N., Saraf, N. M., & Talib, N. (2022). Monitoring of Rice Growth Phases Using Multi-Temporal Sentinel-2 Satellite Image. *IOP Conference Series: Earth and Environmental Science*, 1051(1). <https://doi.org/10.1088/1755-1315/1051/1/012021>
- Bowler, N. E., & Mylne, K. R. (2009). Ensemble transform Kalman filter perturbations for a regional ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 135(640), 757–766. <https://doi.org/https://doi.org/10.1002/qj.404>
- Chang, Y., Latham, J., Licht, M., & Wang, L. (2023). A data-driven crop model for maize yield prediction. *Communications Biology*, 6(1). <https://doi.org/10.1038/s42003-023-04833-y>
- Coleman, K., Muhammed, S. E., Milne, A. E., Todman, L. C., Dailey, A. G., Glendining, M. J., & Whitmore, A. P. (2017). The landscape model: A model for exploring trade-offs between agricultural production and the environment. *Science of The Total Environment*, 609, 1483–1499. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2017.07.193>
- Coleman, K., Whitmore, A. P., Hassall, K. L., Shield, I., Semenov, M. A., Dobermann, A., Bourhis, Y., Eskandary, A., & Milne, A. E. (2021). The potential for soybean to diversify the production of plant-based protein in the UK. *Science of The Total Environment*, 767, 144903. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.144903>

- Curnel, Y., de Wit, A. J. W., Duveiller, G., & Defourny, P. (2011). Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. *Agricultural and Forest Meteorology*, 151(12), 1843–1855. <https://doi.org/https://doi.org/10.1016/j.agrformet.2011.08.002>
- de Wit, A. J. W., & van Diepen, C. A. (2007a). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146(1–2), 38–56. <https://doi.org/10.1016/j.agrformet.2007.05.004>
- de Wit, A. J. W., & van Diepen, C. A. (2007b). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146(1), 38–56. <https://doi.org/https://doi.org/10.1016/j.agrformet.2007.05.004>
- Dlamini, L., Crespo, O., van Dam, J., & Kooistra, L. (2023). A Global Systematic Review of Improving Crop Model Estimations by Assimilating Remote Sensing Data: Implications for Small-Scale Agricultural Systems. In *Remote Sensing* (Vol. 15, Issue 16). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/rs15164066>
- Domingues, T., Brandão, T., & Ferreira, J. C. (2022). Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey. In *Agriculture (Switzerland)* (Vol. 12, Issue 9). MDPI. <https://doi.org/10.3390/agriculture12091350>
- Dong, X. L., Halevy, A., & Yu, C. (2009). Data integration with uncertainty. *VLDB Journal*, 18(2), 469–500. <https://doi.org/10.1007/s00778-008-0119-9>
- Engen, M., Sandø, E., Lucas, B., Sjølander, O., Arenberg, S., Gupta, R., & Goodwin, M. (2021). *Farm-Scale Crop Yield Prediction from Multi-Temporal Data Using Deep Hybrid Neural Networks*. <https://doi.org/10.3390/agronomy>

- Enting, I. G., Trudinger, C. M., & Etheridge, D. M. (2006). Propagating data uncertainty through smoothing spline fits. *Tellus, Series B: Chemical and Physical Meteorology*, 58(4), 305–309. <https://doi.org/10.1111/j.1600-0889.2006.00193.x>
- Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4), 343–367. <https://doi.org/10.1007/s10236-003-0036-9>
- Florence, A., Revill, A., Hoad, S., Rees, R., & Williams, M. (2021). The effect of antecedence on empirical model forecasts of crop yield from observations of canopy properties. *Agriculture (Switzerland)*, 11(3). <https://doi.org/10.3390/agriculture11030258>
- Hengl, T., Miller, M. A. E., Križan, J., Shepherd, K. D., Sila, A., Kilibarda, M., Antonijević, O., Glušica, L., Dobermann, A., Haefele, S. M., McGrath, S. P., Acquah, G. E., Collinson, J., Parente, L., Sheykhmousa, M., Saito, K., Johnson, J. M., Chamberlin, J., Silatsa, F. B. T., ... Crouch, J. (2021). African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-85639-y>
- Hu, S., & Mo, X. (2011). Interpreting spatial heterogeneity of crop yield with a process model and remote sensing. *Ecological Modelling*, 222, 2530–2541. <https://doi.org/10.1016/j.ecolmodel.2010.11.011>
- 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. <https://doi.org/10.1016/j.agrformet.2015.10.013>
- Ines, A. V. M., Das, N. N., Hansen, J. W., & Njoku, E. G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment*, 138, 149–164. <https://doi.org/https://doi.org/10.1016/j.rse.2013.07.018>

- 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. <https://doi.org/https://doi.org/10.1016/j.eja.2017.11.002>
- Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18(3), 235–265. [https://doi.org/https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/https://doi.org/10.1016/S1161-0301(02)00107-7)
- Kephe, P. N., Ayisi, K. K., & Petja, B. M. (2021). Challenges and opportunities in crop simulation modelling under seasonal and projected climate change scenarios for crop production in South Africa. In *Agriculture and Food Security* (Vol. 10, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s40066-020-00283-5>
- Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.00621>
- Kivi, M., Vergopolan, N., & Dokoohaki, H. (2023). A comprehensive assessment of in situ and remote sensing soil moisture data assimilation in the APSIM model for improving agricultural forecasting across the US Midwest. *Hydrology and Earth System Sciences*, 27(5), 1173–1199. <https://doi.org/10.5194/hess-27-1173-2023>
- Knowling, M. J., White, J. T., Grigg, D., Collins, C., Westra, S., Walker, R. R., Pellegrino, A., Ostendorf, B., Bennett, B., & Alzraiee, A. (2023). Operationalizing crop model data assimilation for improved on-farm situational awareness. *Agricultural and Forest Meteorology*, 338. <https://doi.org/10.1016/j.agrformet.2023.109502>
- Li, C., Zheng, Y., Zhang, X., Wu, F., Li, L., & Jiang, J. (2022). Comparison of Canopy Cover and Leaf Area Index Estimation from Airborne LiDAR and Digital Aerial Photogrammetry in Tropical Forests. *Applied Sciences (Switzerland)*, 12(19). <https://doi.org/10.3390/app12199882>
- Li, L., Wang, B., Feng, P., Wang, H., He, Q., Wang, Y., Liu, D. L., Li, Y., He, J., Feng, H., Yang, G., & Yu, Q. (2021). Crop yield forecasting and associated optimum

- lead time analysis based on multi-source environmental data across China. *Agricultural and Forest Meteorology*, 308–309, 108558. <https://doi.org/https://doi.org/10.1016/j.agrformet.2021.108558>
- Li, Y., Zhou, Q., Zhou, J., Zhang, G., Chen, C., & Wang, J. (2014). Assimilating remote sensing information into a coupled hydrology-crop growth model to estimate regional maize yield in arid regions. *Ecological Modelling*, 291, 15–27. <https://doi.org/https://doi.org/10.1016/j.ecolmodel.2014.07.013>
- Liman Harou, I., Whitney, C., Kung'u, J., & Luedeling, E. (2021). Crop modelling in data-poor environments – A knowledge-informed probabilistic approach to appreciate risks and uncertainties in flood-based farming systems. *Agricultural Systems*, 187, 103014. <https://doi.org/https://doi.org/10.1016/j.agry.2020.103014>
- Liu, X., Liu, C., Wang, P., & Xing, Y. (2010). Assimilation of ground measured wheat leaf area index into CERES-Wheat model based on Kalman Filter. *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, 26, 176–181. <https://doi.org/10.3969/j.issn.1002-6819.2010.z1.033>
- Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, 150(11), 1443–1452. <https://doi.org/https://doi.org/10.1016/j.agrformet.2010.07.008>
- Lu, Y., Chibarabada, T. P., Ziliani, M. G., Onema, J. M. K., McCabe, M. F., & Sheffield, J. (2021). Assimilation of soil moisture and canopy cover data improves maize simulation using an under-calibrated crop model. *Agricultural Water Management*, 252. <https://doi.org/10.1016/j.agwat.2021.106884>
- Maharjan, G. R., Hoffmann, H., Webber, H., Srivastava, A. K., Weihermüller, L., Villa, A., Coucheney, E., Lewan, E., Trombi, G., Moriondo, M., Bindi, M., Grosz, B., Dechow, R., Kuhnert, M., Doro, L., Kersebaum, K.-C., Stella, T., Specka, X., Nendel, C., ... Gaiser, T. (2019). Effects of input data aggregation on simulated crop yields in temperate and Mediterranean climates. *European Journal of Agronomy*, 103, 32–46. <https://doi.org/https://doi.org/10.1016/j.eja.2018.11.001>

- Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., & Gascon, F. (2017). *Sen2Cor for Sentinel-2*. 3. <https://doi.org/10.1117/12.2278218>
- Meroni, M., Waldner, F., Seguíni, L., Kerdiles, H., & Rembold, F. (2021). Yield forecasting with machine learning and small data: What gains for grains? *Agricultural and Forest Meteorology*, 308–309, 108555. <https://doi.org/https://doi.org/10.1016/j.agrformet.2021.108555>
- Nyéki, A., & Neményi, M. (2022). Crop Yield Prediction in Precision Agriculture. In *Agronomy* (Vol. 12, Issue 10). MDPI. <https://doi.org/10.3390/agronomy12102460>
- Oikonomidis, A., Catal, C., & Kassahun, A. (2023). Deep learning for crop yield prediction: a systematic literature review. In *New Zealand Journal of Crop and Horticultural Science* (Vol. 51, Issue 1, pp. 1–26). Taylor and Francis Ltd. <https://doi.org/10.1080/01140671.2022.2032213>
- Oulaid, B., Milne, A. E., Waine, T., El Alami, R., Rafiqi, M., & Corstanje, R. (2024a). Stepwise model parametrisation using satellite imagery and hemispherical photography: Tuning AquaCrop sensitive parameters for improved winter wheat yield predictions in semi-arid regions. *Field Crops Research*, 309. <https://doi.org/10.1016/j.fcr.2024.109327>
- Oulaid, B., Milne, A. E., Waine, T., El Alami, R., Rafiqi, M., & Corstanje, R. (2024b). Stepwise model parametrisation using satellite imagery and hemispherical photography: Tuning AquaCrop sensitive parameters for improved winter wheat yield predictions in semi-arid regions. *Field Crops Research*, 309. <https://doi.org/10.1016/j.fcr.2024.109327>
- Panigrahi, B., Kathala, K. C. R., & Sujatha, M. (2023). A Machine Learning-Based Comparative Approach to Predict the Crop Yield Using Supervised Learning With Regression Models. *Procedia Computer Science*, 218, 2684–2693. <https://doi.org/https://doi.org/10.1016/j.procs.2023.01.241>
- Raj, R., Walker, J. P., Pingale, R., Nandan, R., Naik, B., & Jagarlapudi, A. (2021). Leaf area index estimation using top-of-canopy airborne RGB images.

International Journal of Applied Earth Observation and Geoinformation, 96.
<https://doi.org/10.1016/j.jag.2020.102282>

Ren, S., Chen, H., Hou, J., Zhao, P., Dong, Q., & Feng, H. (2024). Based on historical weather data to predict summer field-scale maize yield: Assimilation of remote sensing data to WOFOST model by ensemble Kalman filter algorithm. *Computers and Electronics in Agriculture*, 219.
<https://doi.org/10.1016/j.compag.2024.108822>

Singels, A. (2013). Crop Models. In *Sugarcane: Physiology, Biochemistry, and Functional Biology* (pp. 541–577). John Wiley & Sons, Ltd.
<https://doi.org/https://doi.org/10.1002/9781118771280.ch20>

Sun, L., Seidou, O., Nistor, I., & Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. *Hydrological Sciences Journal*, 61(13), 2348–2366.
<https://doi.org/10.1080/02626667.2015.1127376>

Tao, F., Rötter, R. P., Palosuo, T., Gregorio Hernández Díaz-Ambrona, C., Mínguez, M. I., Semenov, M. A., Kersebaum, K. C., Nendel, C., Specka, X., Hoffmann, H., Ewert, F., Dambreville, A., Martre, P., Rodríguez, L., Ruiz-Ramos, M., Gaiser, T., Höhn, J. G., Salo, T., Ferrise, R., ... Schulman, A. H. (2018). Contribution of crop model structure, parameters and climate projections to uncertainty in climate change impact assessments. *Global Change Biology*, 24(3), 1291–1307.
<https://doi.org/10.1111/gcb.14019>

Tenreiro, T. R., García-Vila, M., Gómez, J. A., Jiménez-Berni, J. A., & Fereres, E. (2021). Using NDVI for the assessment of canopy cover in agricultural crops within modelling research. *Computers and Electronics in Agriculture*, 182.
<https://doi.org/10.1016/j.compag.2021.106038>

Thorp, K. R., DeJonge, K. C., Kaleita, A. L., Batchelor, W. D., & Paz, J. O. (2008). Methodology for the use of DSSAT models for precision agriculture decision support. *Computers and Electronics in Agriculture*, 64(2), 276–285.
<https://doi.org/https://doi.org/10.1016/j.compag.2008.05.022>

- Toumi, J., Er-Raki, S., Ezzahar, J., Khabba, S., Jarlan, L., & Chehbouni, A. (2016). Performance assessment of AquaCrop model for estimating evapotranspiration, soil water content and grain yield of winter wheat in Tensift Al Haouz (Morocco): Application to irrigation management. *Agricultural Water Management*, 163, 219–235. <https://doi.org/10.1016/j.agwat.2015.09.007>
- 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(16). <https://doi.org/10.3390/RS12162666>
- Vanuytrecht, E., Raes, D., Steduto, P., Hsiao, T. C., Fereres, E., Heng, L. K., Garcia Vila, M., & Mejias Moreno, P. (2014). AquaCrop: FAO's crop water productivity and yield response model. *Environmental Modelling & Software*, 62, 351–360. <https://doi.org/https://doi.org/10.1016/j.envsoft.2014.08.005>
- Vermote, E. (2007). *LEDAPS surface reflectance product description*.
- Wang, B., Sun, Z., Jiang, X., Zeng, J., & Liu, R. (2023). Kalman Filter and Its Application in Data Assimilation. In *Atmosphere* (Vol. 14, Issue 8). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/atmos14081319>
- Zhu, W. (2024). A Deep Factor Model for Crop Yield Forecasting and Insurance Ratemaking. *North American Actuarial Journal*, 28(1), 57–72. <https://doi.org/10.1080/10920277.2023.2182792>
- Ziliani, M. G., Altaf, M. U., Aragon, B., Houburg, R., Franz, T. E., Lu, Y., Sheffield, J., Hoteit, I., & McCabe, M. F. (2022a). Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model. *Agricultural and Forest Meteorology*, 313. <https://doi.org/10.1016/j.agrformet.2021.108736>
- Ziliani, M. G., Altaf, M. U., Aragon, B., Houburg, R., Franz, T. E., Lu, Y., Sheffield, J., Hoteit, I., & McCabe, M. F. (2022b). Early season prediction of within-field crop

yield variability by assimilating CubeSat data into a crop model. *Agricultural and Forest Meteorology*, 313. <https://doi.org/10.1016/j.agrformet.2021.108736>

5 Discussion

My research aims to advance scalable wheat yield prediction by integrating remote sensing data with process-based and data-driven models in arid and semi-arid regions. It addresses four interconnected research questions: (i) What are the key factors that significantly impact wheat yield prediction in arid and semi-arid regions, and how do they change over the season? (Chapter 2 and Chapter 3); (ii) What are the trade-offs associated with calibrating process-based models for yield prediction using ground-based data compared to remotely sensed data? (Chapter 2); (iii) How effective are early-season data-driven yield prediction models across different geographic regions? (Chapter 3) and (iv) To what extent does the integration of data-driven models into process-based models through data assimilation techniques enhance the precision of wheat yield predictions? (Chapter 4). Addressing these research questions will fill many gaps in the existing literature, such as insufficient understanding of the temporal sensitivity of key parameters impacting wheat yield predictions, unexplored performance of early-season data-driven models beyond areas of their implementation, and the limited understanding of the impact of data assimilation strategies, including sensor selection and assimilation timing on the overall assimilation accuracy. This study contributes to the body of knowledge by exploring methods for enhancing model precision and offering practical solutions for tactical and strategic decision-making in arid and semi-arid agricultural regions.

5.1 What are the key factors that significantly impact wheat yield prediction in arid and semi-arid regions and how does that change over the season?

5.1.1 Key findings

To accurately predict wheat yields, it is important to first identify the critical factors that influence yield variation. For process-based models like AquaCrop, this involves identifying the model parameters that most impact yield. This was

achieved by combining screening methods with Fourier Amplitude Sensitivity Test (FAST), a robust sensitivity approach. For data-driven models I identified the features that best explain the variation in yield. This was accomplished through a combination of correlation heatmaps, cross-validated forward feature selection and SHAP feature contribution analysis.

5.1.1.1 Process-based models

Similar to other crop models, in AquaCrop, quantifying the canopy cover is essential for accurately predicting yield outcomes, as it directly influences the plant's photosynthesis capacity and yield. In this study, the sensitivity analysis of canopy cover (CC), which is poorly understood in arid and semi-arid regions of Africa, indicated that the time from sowing to emergence (gEme), canopy growth coefficient (CGC), and the time from sowing to maturity (gMat) were consistently influential across different sites. These findings are consistent with the study of Lu et al. (2021) on wheat crops in semi-arid regions of United states and highlights the essential role of crop phenological stages in determining canopy cover development. My work has shown that gEme and CGC were critical at the beginning of the season, influencing early growth, while gMat and the canopy decline coefficient (CDC) became more influential towards the end of the season as the crop approached maturity. This temporal variation in parameter sensitivity highlights the dynamic nature of wheat crop growth and the importance of different parameters at various growth stages. In terms of agricultural decision making, these findings underscore the necessity of timely monitoring of crop phenological stages. Early canopy projections can inform adjustments in irrigation and nutrient management practices to optimise the canopy cover development. When contextualised within existing literature, these findings align with Upreti et al. (2020) who used both Fourier-based and PAWN methods and found similar sensitive parameter. Notably, Akbari et al. (2024) also observed that that CGC was responsive of most of the variance in the canopy cover response. In terms of agricultural decision making, these findings underscore the necessity of timely monitoring of crop phenological stages. Early identification of sensitive growth parameter can inform adjustments in irrigation and nutrient management

practices to optimise the canopy cover development. However, other studies have identified different sensitive parameters. For instance, Salemi et al. (2011) found that CGC displayed moderate sensitivity levels compared to the initial soil water content. As irrigation levels decreased to 60% of full irrigation, sensitivity increased significantly, highlighting the role of water management on parameters' sensitivity. Jin et al. (2018), noted maximum canopy cover (mcc) as most sensitive parameter to canopy cover for wheat crops in semi-humid continental climates, highlighting climate-specific influence of crop parameters on AquaCrop.

Yield prediction using AquaCrop was predominantly influenced by normalized water productivity (Wp) and the reference harvest Index (Hi0). These findings support the work of XING et al. (2017) who also identified these parameters as key determinants of wheat yield under different irrigation regimes. The sensitivity of these parameters varied across different sites, reflecting on the influence of local environmental conditions and management practices on yield outcomes. For instance, in fields which had a prolonged growing season and higher irrigation amounts, the sensitivity of Wp and Hi0 was more pronounced. Conversely, in field which experience water stress, the sensitivity of these parameters was lower, resulting in reduced yields. These comparisons indicate that while Wp and Hi0 are critical determinants of yield, their impact is significantly influenced by local conditions such as sowing dates and irrigation practices.

Understanding the factors that influence wheat yield in arid and semi-arid regions is essential due to the challenge posed by water scarcity and unpredictable weather patterns. This research provides a detailed temporal profile of parameter importance, highlighting the dynamic interactions between key parameters for AquaCrop and the main responses preceding yield formation. This refined climate-specific understanding supports a more accurate and adaptable use of AquaCrop in the region. The method used in this study could be applied to other crops and decision support systems. The combined use of Morris and FAST sensitivity analysis methods demonstrates a robust approach for quantifying key parameters in crop modelling. This methodology not only highlights critical factors but also captures when they are most influential.

5.1.1.2 Data-driven models

My research investigated several factors impacting yield prediction for data-driven models. Seasonal rainfall positively impacted wheat yield in both arid and semi-arid regions, underscoring the critical importance of water availability throughout the growing season. Specifically, early-season precipitation was found to have a significant positive impact on yield, particularly between the crop green-up and peak of the canopy. Rainfall during this stage ensures that the soil moisture levels are sufficient to support rapid expansion of the canopy. This period is essential for establishing a strong photosynthetic base, which supports the development of tillers and leaves, leading to greater biomass accumulation and higher potential yields. Adequate early-season rainfall also promotes root growth, enhancing the plant's ability to access deeper soil moisture and nutrients, which are important for sustaining growth during drier periods later in the season. Cumulative temperature, measured as growing degree days (GDD), exhibited varying impacts on wheat yield in arid and semi-arid regions. In arid regions, higher accumulated temperatures during the reproductive phase showed a negative impact on the yield. This is likely due to the fact that excessive heat during critical stages such as grain filling accelerated senescence, reduced grain filling duration, and ultimately led lower yields. This finding is consistent with Madhukar et al. (2022), who observed that high temperatures during reproductive stages adversely affects wheat yields in arid environments. In contrast, semi-arid regions showed a positive correlation between cumulative temperature during early growth stages and yield. Warmer temperatures during initial phases of growth promote seed germination (Dwyer & Erickson, 2016) and early seedling development (Khaeim et al., 2022), all of which are essential for maximising potential yield. These distinctions are particularly relevant in the context of Moroccan climate. The south arid regions of Morocco typically experience extreme temperatures, especially during late spring to early summer, which coincide with the reproductive stage of wheat and lead to heat stress and its negative effect on yield. Conversely, north semi-arid regions of Morocco benefit from more moderate temperatures during early growing season, which supports

healthy early wheat growth and establishes a better foundation for subsequent development. In both arid and semi-arid regions, the number of selected features for yield prediction models increased progressively across different prediction stages, reflecting the growing complexity of factors influencing wheat yield as the crop matures. In arid regions, the number of selected features increased moderately, suggesting that key determinants like slope, applied phosphorus, potassium and specific vegetation indices remain consistently influential throughout the growth cycle. This moderate increase highlights the stable yet critical role of these factors in environments characterised by extreme heat and limited water availability. In semi-arid regions, there was a substantial increase in the number of selected features from early to later growth stages. This increase reflects the more dynamic conditions in this region and highlights the necessity for more sophisticated models in these regions. The feature contribution analysis performed in this study addresses the black box nature of machine learning models, providing more interpretable results that can guide precise agricultural interventions. This knowledge supports the development of more accurate data-driven model by researchers in areas with similar climatic patterns. For agronomists in the investigated regions, the findings provide different insights to improve sustainable wheat production in Morocco. In arid regions characterised by water scarcity and high evaporation rates, management practices should focus on selecting optimal sowing dates to avoid peak temperature and adopting heat-tolerant wheat varieties. In semi-arid regions, the focus should be on optimising sowing dates to benefit from early growth warmth. Most of the features used to address this research question were remotely-sensed, offering a cost-effective, non-destructive, and human intervention-free method for monitoring environmental variables critical to wheat prediction.

5.1.2 Limitations and recommendations for further research

Chapter 2 provides a robust analysis of key factors influencing yield predictions using AquaCrop. However, a few limitations must be noted. The study was conducted across nine fields covering two growing seasons with specific management practices. Therefore, further research should focus on conducting

longitudinal studies to observe the long-term impacts of varying environmental conditions on yield outcomes over multiple growing seasons. Exploring the interactions between multiple stress factors, such as drought, nutrient deficiency, and pests, will provide knowledge about how these stresses interact and affect wheat productivity. Lastly, investigating the sensitivity of different wheat varieties to identify if certain cultivars respond differently to key parameters will enhance the selection of crop varieties better suited to specific regions.

The reliance on remote sensing data from Sentinel 2 to generate features impacting the yield for data driven models (Chapter 3), while advantageous for its cost-effectiveness and non-destructive nature, can be limited by its spatial resolution and the frequency of data acquisition. Further novel research could explore the integration of higher-resolution remote sensing data to capture finer spatial details, and Synthetic Aperture Radar (SAR) data to penetrate cloud cover and provide consistent observations regardless of weather conditions in the arid and semi-arid climates. Additionally, the phenology metrics extracted from remotely-sensed data should be cross-validated by fieldwork to ensure their agronomical relevance. This detail, while it might not have a direct impact on the performance of yield prediction models, is essential for making the models relevant to actual field conditions and so more interpretable.

5.2 What are the trade-offs associated with calibrating process-based models for yield prediction using ground-based data in comparison to remotely sensed data?

5.2.1 Key findings

In Chapter 2, I systematically compare the accuracy, reliability, and resource requirements of high-resolution ground-based data against lower-resolution remotely sensed data for calibrating AquaCrop model for semi-arid and arid regions. The primary objectives were to evaluate the effectiveness of these data sources in providing accurate and reliable model parameters and to assess the practical implications of their use in terms of resource investment. I demonstrated

that while ground-based data offer higher accuracy and reliability, remote sensing data provide advantages related to resource efficiency and scalability.

My study indicated that high-resolution ground-based hemispherical photography, provided accurate yield predictions during the calibration phase, with strong alignment between observed and simulated canopy cover. This method's reliability was evident as the accuracy was maintained across different fields in the validation phase. This accurate canopy cover estimation directly contributed to the precision of the yield predictions even though there was a drop of the performance in the validation dataset. The overall consistency of results underscored the reliability of ground-based data. Conversely, Landsat7-based parametrisation showed a lower accuracy and reliability in both canopy cover and yield predictions during calibration and validation phases. The limitations were partly due to the lower spatial resolution and spectral saturation issues in regions with dense vegetation. These findings aligned with a recent publication by Kivi et al. (2023) who assessed the performance of calibrated APSIM based on ground-based and remotely sensed data and highlighted the high accuracy of ground-based measurements for model calibration. The stepwise model parametrisation performed in Chapter 2 showed that hemispherical photography, while providing high accuracy for both canopy cover and yield predictions, required substantial labour and time, involving frequent field visits and destructive human intervention. Each field necessitated multiple photographs to ensure accurate representation, increasing the overall resource reliance. This is particularly challenging in Morocco, where fieldwork could be interrupted by difficult terrain, extreme weather conditions, and limited access to remote agricultural areas. The need for specialized equipment, such as hemispherical cameras and image processing software further added to the resource investment required. In contrast, remote sensing parametrisation using Landsat7 significantly reduced the need for on-site personnel and frequent field visits. Data could be freely collected and remotely processed, offering a practical solution for large-scale studies. These findings supported previous research by Huang et al. (2019) and Jin et al. (2018) which

highlighted the resource efficiency and the practicality of remote sensing data in calibrating crop models.

Findings from my research hold implications for both theoretical advancements and practical applications in agricultural modelling. Theoretically, this research enhances the understanding of integrating diverse data sources to improve crop model calibration. It introduces advancements in stepwise parameterisation by demonstrating the benefits of automating the calibration process, thereby reducing the need for manual adjustments, and enhancing calibration efficiency. This methodology addresses the complexity of crop models by lowering parameter dimensionality through sensitivity analysis and applying swarm optimization techniques, offering a robust framework for model parameterization. Practically, the research provides an automated framework for optimizing the calibration of AquaCrop, significantly improving its predictive accuracy and applicability by moving away from default parameters. The scalable calibration allows for efficient application over large areas, reducing the labour and time required for model calibration. The novelty of the work lies also in the integration of soil fertility component into AquaCrop. Specifically, the integrated code accounts for the influence of soil fertility on key model parameters related to canopy cover and aboveground biomass. This is particularly important in regions where soil fertility and availability of fertilizer varies widely across small spatial scales.

5.2.2 Limitations and recommendations for further research

One of the primary limitations identified in this study is the lower spatial resolution of Landsat 7 imagery. The spatial resolution of Landsat 7, while adequate for broad-scale observations, may not capture the finer details necessary for precise canopy cover and yield predictions at smaller scales. This limitation can lead to less accurate estimations of the crop parameters, particularly in heterogeneous landscapes where variations within fields are large. Another limitation is the issue of spectral saturation of Landsat 7 data in regions with dense vegetation (Mutanga et al., 2023). This saturation can result in an underestimation of canopy

density and biomass, thereby affecting the accuracy of the model predictions. To address these limitations and enhance the accuracy and applicability of the model, future research should investigate the use of high-resolution satellite or the combination of multiple remote sensing platforms to improve the spatial detail and capture finer variations within fields. Coupling ground-based and satellite imagery is another approach that can enhance the model reliability by combining the details local measurements with broad-scale observations, particularly during times when the identified influential are most critical.

5.3 How effective are early season data-driven yield prediction models across geographic regions?

5.3.1 Key findings

Chapter 3 evaluates the in-season predictive performance of three machine learning algorithms—Multiple Linear Regression (MLR), Support Vector Regression (SVR), and Random Forests (RF)—across four prediction stages in arid and semi-arid regions of Morocco. The SVR demonstrated the lowest prediction errors in arid climates, particularly at later stages, with an RMSE of 0.23 t/ha at the canopy senescence stage. In contrast, in semi-arid regions, RF consistently outperformed other models across all prediction stages, achieving an RMSE of 0.50 t/ha at the time of canopy senescence. When tuning models with both arid and semi-arid regions data, RF showed better performance, achieving an RMSE of 0.46 t/ha at the senescence stage.

These results highlight that the choice of model significantly impacts the accuracy of yield predictions. SVR and RF high performances in arid and semi-arid regions, respectively can be explained by the number of features selected prior to model implementation, in the arid regions the number of selected features ranged between 8 and 12. This smaller number of features aligns with SVR's ability to fit simpler, more focused datasets, effectively capturing key non-linear relationships in the data. These results align with Kumar et al. (2023) who implemented in-season yield prediction models for corn and highlighted SVR's ability to maintain

a relatively high accuracy despite using fewer features. In semi-arid regions, the influential features ranged between 9 and 20. The larger and more diverse feature set in these regions allowed RF to leverage its capability to manage high-dimensional and more complex interactions. This pattern was also observed by Ghosh & Cabrera (2022) where RF effectively managed to provide accurate predictions in high-dimensional datasets. However, they also noted that the performance of RF can deteriorate significantly when the number of features is disproportionately high compared to the number of samples.

The improvement in prediction accuracy as the season progresses reflects the increasing availability and relevance of phenological data. Early predictions, while essential for timely agricultural decisions, generally exhibit lower accuracy due to the limited data available at initial stages. As more data becomes available through the season, the models' predictive capabilities improve, demonstrating the trade-off between earliness and accuracy in yield predictions. These findings align with previous studies highlighting the trade-offs between accuracy and earliness in yield predictions. For instance, Han et al. (2020) highlighted the increasing accuracy of yield predictions with more comprehensive phenological data. Similarly, Zhou et al. (2023) demonstrated that using features extracted from the entire growing season provide more stable and accurate results as compared to using phenological data from limited time windows. While yield predictions at later stages of wheat growth stages are more accurate, they require more satellite observations and provide predictions at stages when human intervention is minimal. Early season predictions, although less accurate, are crucial for timely management decisions and interventions. The trade-off between early predictions and their accuracy is a recurring theme in agricultural modelling, necessitating a balance based on specific requirements and timelines for decision-making.

I also examined the transferability of models between different climatic zones and the generalization capability of global models. Transferring models between arid and semi-arid zones resulted in significantly higher prediction errors, indicating reduced accuracy. This decline in performance highlights the

challenges associated with using data-driven models beyond their areas of initial implementation, primarily due to distinct feature contributions and environmental conditions that differ across climates. For instance, the specific environmental conditions such as limited rainfall and higher temperatures in arid regions require tailored modelling approaches that may not translate well to semi-arid regions, where more varied rainfall patterns and different management practices are adopted. Generalised models, trained on combined data from both arid and semi-arid regions, demonstrated better performance than transferred models but were still outperformed by localised models. While generalised models lack the fine-tuned adjustments that localized models can provide, they benefit from a broader dataset that helps in capturing the overall data trends more robustly. Consequently, despite their overall lower performance compared to climate-specific models, generalised models offer a more reliable alternative to transferred models, providing a practical solution for large-scale applications that require adaptability to diverse climatic conditions.

The findings from this study have practical implications for farmers and policymakers. The ability to make early season predictions, despite their lower accuracy, is invaluable for tactical decision-making. Early predictions allow farmers to make timely interventions, such as adjusting irrigation schedules, applying fertilizers, or implementing pest control measures. These tactical decisions can have a significant impact on the growing season's outcomes by addressing issues as they arise, thereby preventing potential yield losses or wasted resources. In addition, accurate yield predictions at later stages of the growing season are crucial for strategic decision-making. These late predictions, enhanced by the availability of comprehensive phenological data, provide essential information for planning post-harvest activities, such as storage, marketing, and distribution and resource allocation. The improved accuracy of late-season predictions ensures that strategic decisions are based on reliable sources, ultimately contributing to better resource management and financial planning for farmers and agricultural stakeholders. The workshop held at UM6P (Appendix E), underscored the importance of stakeholder' engagement in

understanding and using the implemented predictive models. The adoption of early prediction models was favoured by all of the stakeholders at the workshop for their ability to enable timely and adaptative agricultural decisions. Outcomes from this research facilitated discussions on key agricultural decisions at different prediction stages and highlighted the trade-offs between prediction time and uncertainties in yield predictions. This initiative provided a platform for researchers, farmers, and policy advisors to collectively identify practical strategies to be coupled with my findings.

5.3.2 Limitations and recommendations for further research

The reliance on remote sensing data for feature extraction, while advantageous for its broad coverage and non-destructive nature, faces significant challenges due to atmospheric conditions. Clouds, specifically can degrade the quality of satellite imagery, introducing errors into the data. To counteract these issues, severe filters for data quality are often applied, which are necessary to ensure data accuracy but significantly reduce the number of available observations. This reduction can limit the development of robust predictive models. Future research should explore advanced atmospheric correction techniques to minimize these disturbances. Additionally, integrating multi-source remote sensing data, including data from different satellites or ground-based observations, could increase the frequency and reliability of the data. Combining various data sources may help fill in gaps caused by poor atmospheric conditions, enhancing the overall quality of the dataset. Future research could also explore the integration of higher-resolution satellite imagery to capture finer spatial details, enhancing model precision. Another limitation is the phenology-based metrics extracted through derivative methods. While this approach is well documented in literature (Bachoo & Archibald, 2007; Tan et al., 2008), cross-validating remote sensing-derived metrics with field data can improve the agronomical relevance of the models, ensuring that they accurately reflect field conditions.

5.4 To what extent does the integration of data-driven models into process-based models through data assimilation techniques enhance the precision of wheat yield predictions?

5.4.1 Key findings

In Chapter 4, I evaluated the impact of integrating satellite-derived canopy cover data into the AquaCrop model using different data assimilation techniques, sensors and periods to enhance the accuracy of wheat yield predictions. The study found that Bayesian assimilation provided the highest accuracy in yield predictions, reducing RMSE by 25% and MAPE by 76.5% compared with KF and the no assimilation benchmark. Bayesian methods use observational data to probabilistically update the distribution of model parameters, resulting in balanced and realistic canopy cover profiles. Ensemble Kalman Filter (ENKF) and Ensemble Transform Kalman Filter (ETKF) were also effective but showed varied performance depending on the growth stage. ETKF provided canopy cover estimates closer to the remote sensing observations during specific periods, particularly during the initial stages of canopy development and peak growth. However, the aggressive adjustments by ETKF led to unrealistic fluctuations of the canopy, indicating a potential reliance on observational data at the expense of the model's inherent dynamics. Overall, all the assimilation methods enhanced the prediction of wheat prediction as compared to the initial benchmark. This is aligning with previous research demonstrating the effectiveness of assimilation methods to correct the trajectories of state variables, leading to reduced errors in key crop responses such as biomass and yield. Lu et al. (2022) performed a Bayesian assimilation of in situ soil moisture, biomass and canopy cover into AquaCrop, and demonstrated that MCMC based assimilation could not only be used in enhancing yield simulation performance but also to estimate planting dates and quasi-calibrated phenological parameters of crop models. Similarly, soil moisture and canopy cover were assimilated into AquaCrop by Yang et al. (2023). The findings indicated that while both methods outperformed the no assimilation benchmark, joint assimilation led to the lowest yield prediction errors.

A recent trend in the assimilation methods is merging the strength of Kalman and Bayesian based method which demonstrated its efficiency in non-linear complex systems. Huang et al. (2023) applied a hybrid ENKF-Bayesian method to assimilate leaf area index extracted from the Global Land Surface Satellite (GLASS). The Bayesian posterior of LAI was incorporated into ENKF to better characterise the ensemble distribution of model errors. The results indicated that hybrid method outperformed the standard Gaussian perturbation-based ENKF in yield predictions at both point and regional scale.

The study revealed important findings on the impact of using different sensors, specifically Landsat 7 and Sentinel 2, on the accuracy of wheat yield predictions. Both sensors were sampled to the same spatial resolution of 30 meters to only focus on the impact of temporal resolution on prediction accuracy. Sentinel 2, with its higher temporal resolution and a 5-day revisit cycle, provided more frequent observations compared to Landsat 7's 16-day revisit cycle. This higher frequency of data collection, particularly during the peak growth period from January to March, allowed for more detailed monitoring of the canopy cover dynamics. The Sentinel 2 data generally led to better prediction performance, as reflected in the lower prediction errors. These results indicate that the higher temporal resolution of Sentinel 2 significantly enhances the accuracy of canopy cover estimates and, consequently, the yield predictions. Despite these advantages, the combined use of Sentinel 2 and Landsat 7 did not consistently outperform the individual use of Sentinel 2. This suggests that while increasing temporal resolution through sensor fusion is beneficial, it does not necessarily lead to better predictions if the sensor estimates are mismatching, or the integration process does not effectively handle the differences between the sensors. This discrepancy is particularly evident at the peak of the canopy cover, where differences between the two sensors' estimates were most pronounced. One possible explanation of the mismatch between canopy estimates from the two sensors could be the saturation effects of the underlying vegetation indices used for CC estimation. As vegetation density increases, the red band reflectance

becomes insensitive due to chlorophyll absorption leading to indices saturation and higher discrepancies in the estimated canopy cover estimates.

The use of Bayesian assimilation, which significantly reduced prediction errors, highlights its effectiveness in improving yield prediction accuracy, suggesting that probabilistic integration with crop models can provide reliable predictions beneficial for agricultural decision-making. The approach's reliance on default parameters and uncalibrated models underscores its broad applicability across diverse regions without extensive site-specific calibration, enhancing its utility. The enhanced performance of Sentinel 2 data, due to its higher temporal resolution, emphasizes the importance of frequent data collection for accurate crop monitoring. However, the mixed results from combining Sentinel-2 and Landsat 7 highlight the challenges in sensor fusion, calling for advanced integration methods to merge data from multiple sources effectively. The automation of the data assimilation process, capable of running on a set of sites in batch, demonstrates its scalability for large-scale agricultural monitoring, reducing labour and time for manual model calibrations. Using freely available remote sensing, weather, and soil data ensures accessibility to a wide range of users and enables a broader adoption.

5.4.2 Limitations and recommendations for further research

The research was constrained to data assimilation for a single growing season, limiting the generalisability of the results across different temporal scales and varying climatic conditions. Additionally, the study focused on one wheat variety, which may not adequately capture the response variability across different genotypes. The accuracy of the assimilation outcomes is dependent on the quality of weather data; inherent uncertainties and inaccuracies in this data can propagate through the model, potentially skewing yield predictions. Further research should explore integrating weather forecasting models that incorporate quantified uncertainties into the data assimilation process. This approach would enable crop models to provide early yield prediction, thereby enhancing their

usefulness. The automation of the data assimilation process, although beneficial for large-scale applications, demands robust computational systems to manage extensive datasets and complex calculations efficiently. Future research should address these limitations by incorporating multiple growing seasons, diverse wheat varieties, more accurate and reliable weather data sources.

5.5 What are the most promising approaches for wheat yield prediction in arid and semi-arid regions for strategic and tactical decision-making?

To my knowledge, this study is among the first to comprehensively compare multiple approaches for predicting crop yields in semi-arid and arid regions. While existing research often compares individual methods in terms of accuracy, this research examines four methodologies and compare their usefulness for tactical and strategic decisions. Notably, Chumbe et al., (2023) conducted a comparative study between in-situ calibrated AquaCrop and data-driven models and found that Adaptive Boosting model could provide similar results as AquaCrop for quinoa crops in Peru. Another study by Batool et al. (2022) on Tea concluded that machine learning models performed better in yield prediction using fewer data than AquaCrop. Silvestro et al. (2017) performed a comparative study between particle swarm calibration and ensemble Kalman filters assimilation both applied to predict yield for same fields but applied to AquaCrop and SAFY. Their findings suggested that EnKF was more accurate and suitable than PSO for predicting yields.

In crop yield modelling, tactical decisions involve short-term actions that directly impact day-to-day operations and immediate outcomes. They require up-to-date information to optimise profitability and respond to current conditions. Strategic decisions, however, involve long-term planning to achieve broader objectives, benefiting from scalable cost-effective methods that provide overall trends and precise end of season projections. My findings indicated that the Particle Swarm Optimization (PSO) calibration of AquaCrop using hemispherical photographs

demonstrated high accuracy in yield predictions due to its reliance on ground-based data. This method's reliability in providing accurate canopy cover estimation directly contributes to precise yield predictions. However, the labour-intensive and costly nature of this approach, which involves frequent field visits and the use of specialised equipment, limited its practicality for large-scale applications. In addition, the PSO calibration of AquaCrop using Landsat 7 data offered a cost-effective and scalable alternative. Landsat 7 data, while less detailed than ground-based measurements, provided sufficient accuracy for large-scale applications without the need for extensive fieldwork. This approach is beneficial in regions where ground-based data collection is impractical due to logistical or resource constraints. The moderate accuracy achieved through the remote sensing-based calibration is compensated by the method's scalability and cost-effectiveness, making it suitable for broader applications in arid and semi-arid regions. Both swarm using hemispherical photographs and remote sensing data optimisations provide locally relevant parametrisations, but they are not suitable for tactical decisions unless weather forecasts are available since they provide predictions at the end of the season.

Data-driven models, which are stage-specific and based on phenology metrics derived from remote sensing, presented another approach for yield prediction. These models used key phenological stages of wheat growth, such as crop green-up and senescence, to improve prediction accuracy. They rely on observations from sowing dates onwards. These models are straightforward, cost-effective, and capable of making tactical and strategic decisions. For tactical decisions, they offer rapid and reasonably accurate yield predictions that can inform immediate agricultural practices. For strategic planning, their cost-effectiveness makes them suitable for planning over larger scales and longer time horizons, especially in resource-limited settings. However, they often lack the nuanced understanding of the underlying biological processes that influence yield outcomes. Despite this limitation, data-driven models are valuable for rapid yield predictions and are particularly useful in areas where resources and human capacity for detailed process-based use are limited.

Data assimilation of remotely sensed data into AquaCrop emerged as a highly effective method for improving yield prediction accuracy. Techniques such as Bayesian assimilation significantly enhance prediction performance by integrating observational data with model forecasts and accounting for uncertainties. This approach uniquely addressed the uncertainties in both the model parameters and the remotely sensed data, reducing prediction errors and starting with uncalibrated crop model, thereby providing more accurate and reliable yield predictions. The use of higher temporal resolution data, like that from Sentinel 2, further improves the accuracy by allowing more frequent updates of the model states. However, this method requires sophisticated computational techniques to manage complex calculations and data integration, which can be a limitation in resource-constrained environments. Similar to process-based models, for in season yield prediction and associated tactical decisions this method would require forecasted weather data. The assimilation method leverages the detailed process-based understanding of crop growth provided by AquaCrop and the predictive power of data-driven models making it the most promising method for strategic decisions.

5.6 References

- Akbari, E., Darvishi Boloorani, A., Verrelst, J., Pignatti, S., Neysani Samany, N., Soufizadeh, S., & Hamzeh, S. (2024). How global sensitive is the AquaCrop model to input parameters? A case study of silage maize yield on a regional scale. *Frontiers in Agronomy*, 6. <https://doi.org/10.3389/fagro.2024.1304611>
- Bachoo, A., & Archibald, S. (2007). Influence of using date-specific values when extracting phenological metrics from 8-day composite NDVI data. *Proceedings of MultiTemp 2007 - 2007 International Workshop on the Analysis of Multi-Temporal Remote Sensing Images*. <https://doi.org/10.1109/MULTITEMP.2007.4293044>
- Batool, D., Shahbaz, M., Shahzad Asif, H., Shaukat, K., Alam, T. M., Hameed, I. A., Ramzan, Z., Waheed, A., Aljuaid, H., & Luo, S. (2022). A Hybrid Approach to Tea Crop Yield Prediction Using Simulation Models and Machine Learning. *Plants*, 11(15). <https://doi.org/10.3390/plants11151925>
- Chumbe-Llimpe, R. J., Silva-Paucar, S. D., & García-Lopez, Y. J. (2023). Comparison of the machine learning and AquaCrop models for quinoa crops. *Research in Agricultural Engineering*, 69(2), 65–75. <https://doi.org/10.17221/86/2021-RAE>
- Dwyer, J. M., & Erickson, T. E. (2016). Warmer seed environments increase germination fractions in Australian winter annual plant species. *Ecosphere*, 7(10). <https://doi.org/10.1002/ecs2.1497>
- Ghosh, D., & Cabrera, J. (2022). Enriched Random Forest for High Dimensional Genomic Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(5), 2817–2828. <https://doi.org/10.1109/TCBB.2021.3089417>
- Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., & Zhang, J. (2020). Prediction of winter wheat yield based on multi-source data and machine learning in China. *Remote Sensing*, 12(2). <https://doi.org/10.3390/rs12020236>
- Huang, H., Huang, J., Wu, Y., Zhuo, W., Song, J., Li, X., Li, L., Su, W., Ma, H., & Liang, S. (2023). The Improved Winter Wheat Yield Estimation by Assimilating GLASS LAI Into a Crop Growth Model With the Proposed Bayesian Posterior-

Based Ensemble Kalman Filter. *IEEE Transactions on Geoscience and Remote Sensing*, 61. <https://doi.org/10.1109/TGRS.2023.3259742>

Huang, J., Gómez-Dans, J. L., Huang, H., Ma, H., Wu, Q., Lewis, P. E., Liang, S., Chen, Z., Xue, J.-H., Wu, Y., Zhao, F., Wang, J., & Xie, X. (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agricultural and Forest Meteorology*, 276–277, 107609. <https://doi.org/https://doi.org/10.1016/j.agrformet.2019.06.008>

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. In *European Journal of Agronomy* (Vol. 92, pp. 141–152). Elsevier B.V. <https://doi.org/10.1016/j.eja.2017.11.002>

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. <https://doi.org/https://doi.org/10.1016/j.fcr.2018.07.002>

Khaeim, H., Kende, Z., Balla, I., Gyuricza, C., Eser, A., & Tarnawa, Á. (2022). The Effect of Temperature and Water Stresses on Seed Germination and Seedling Growth of Wheat (*Triticum aestivum* L.). *Sustainability (Switzerland)*, 14(7). <https://doi.org/10.3390/su14073887>

Kivi, M. S., Vergopolan, N., & Dokoochaki, H. (2023). A comprehensive assessment of in situ and remote sensing soil moisture data assimilation in the APSIM model for improving agricultural forecasting across the US Midwest. *Hydrology and Earth System Sciences*. <https://api.semanticscholar.org/CorpusID:253381675>

Kumar, C., Mubvumba, P., Huang, Y., Dhillon, J., & Reddy, K. (2023). Multi-Stage Corn Yield Prediction Using High-Resolution UAV Multispectral Data and Machine Learning Models. *Agronomy*, 13(5). <https://doi.org/10.3390/agronomy13051277>

- Lu, Y., Chibarabada, T. P., McCabe, M. F., De Lannoy, G. J. M., & Sheffield, J. (2021). Global sensitivity analysis of crop yield and transpiration from the FAO-AquaCrop model for dryland environments. *Field Crops Research*, 269, 108182. <https://doi.org/10.1016/j.fcr.2021.108182>
- Lu, Y., Wei, C., McCabe, M. F., & Sheffield, J. (2022). Multi-variable assimilation into a modified AquaCrop model for improved maize simulation without management or crop phenology information. *Agricultural Water Management*, 266, 107576. <https://doi.org/10.1016/j.agwat.2022.107576>
- Madhukar, A., Kumar, V., & Dashora, K. (2022). Temperature and precipitation are adversely affecting wheat yield in India. *Journal of Water and Climate Change*, 13(4), 1631–1656. <https://doi.org/10.2166/wcc.2022.443>
- Mutanga, O., Masenyama, A., & Sibanda, M. (2023). Spectral saturation in the remote sensing of high-density vegetation traits: A systematic review of progress, challenges, and prospects. In *ISPRS Journal of Photogrammetry and Remote Sensing* (Vol. 198, pp. 297–309). Elsevier B.V. <https://doi.org/10.1016/j.isprsjprs.2023.03.010>
- Salemi, H., & Amin, M. S. M. (2011). Application of AquaCrop model in deficit irrigation management of Winter wheat in arid region. *Article in African Journal of Agricultural Research*. <https://doi.org/10.5897/AJAR10.1009>
- Silvestro, P. C., 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(5). <https://doi.org/10.3390/rs9050509>
- Tan, B., Morisette, J. T., Wolfe, R. E., Gao, F., Ederer, G. A., Nightingale, J., & Pedelty, J. A. (2008). Vegetation Phenology Metrics Derived from Temporally Smoothed and Gap-Filled MODIS Data. *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, 3, III-593-III-596. <https://doi.org/10.1109/IGARSS.2008.4779417>

- 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(16). <https://doi.org/10.3390/RS12162666>
- 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(11), 2444–2458. [https://doi.org/https://doi.org/10.1016/S2095-3119\(16\)61626-X](https://doi.org/https://doi.org/10.1016/S2095-3119(16)61626-X)
- Yang, Q., Shi, L., Han, J., Zha, Y., Yu, J., Wu, W., & Huang, K. (2023). Regulating the time of the crop model clock: A data assimilation framework for regions with high phenological heterogeneity. *Field Crops Research*, 293, 108847. <https://doi.org/https://doi.org/10.1016/j.fcr.2023.108847>
- Zhou, H., Yang, J., Lou, W., Sheng, L., Li, D., & Hu, H. (2023). Improving grain yield prediction through fusion of multi-temporal spectral features and agronomic trait parameters derived from UAV imagery. *Frontiers in Plant Science*, 14. <https://doi.org/10.3389/fpls.2023.1217448>

6 Conclusion

The aim of my research was to advance the accuracy and scalability of wheat yield prediction models in arid and semi-arid regions through the integration of remote sensing data. There are several ways in which remote sensed data can be used to advantage in yield prediction. These include calibration of process-based models, data driven models and hybrid models, which combine process-based and data-driven approaches through data assimilation techniques. Hybrid models emerged as the most promising for yield prediction. These models leverage the detailed understanding of crop growth provided by process-based models and the predictive power of data-driven models, ensuring robust predictions by continuously updating model states with near real-time observational data. This approach addresses both the need for scalability and the accuracy required for effective agricultural planning and management.

High-resolution ground-based data, such as hemispherical photography, generally improves the accuracy in yield predictions but was labour-intensive and costly, limiting its practicality for large-scale applications. Lower-resolution remote sensing data, while less detailed, offered significant advantages in terms of resource efficiency and scalability, making it suitable for broader applications. These trade-offs highlight the need for balancing accuracy and practicality in model calibration to enhance the scalability of yield prediction models.

The integration of data-driven models with process-based models through data assimilation techniques significantly enhanced yield prediction accuracy. Bayesian assimilation techniques, in particular, improved prediction performance by integrating observational data with model forecasts and accounting for uncertainties. The use of higher temporal resolution data from Sentinel-2 further improved accuracy by providing frequent updates on canopy cover dynamics. This integration is beneficial for achieving precise and adaptable yield predictions in varied environmental conditions.

The training workshop conducted in Morocco provided valuable perspectives on the practical implication of the predictive yield models, demonstrating that early-season predictions are highly valued for enabling timely

management decisions. The workshop outcomes reinforced the importance of balancing lead time with prediction accuracy to support adaptative agricultural practices.

Early-season predictions, despite being less accurate due to limited initial data, are essential for timely management decisions. Models developed using climate-specific data from different regions of Morocco showed that early predictions could guide tactical decisions for agricultural management. However, the increased prediction errors when transferring these models across different climatic zones indicate the necessity for region-specific calibrations to maintain accuracy, reinforcing the importance of localized data in early-season predictions.

My research has demonstrated the potential of integrating remote sensing data with process-based and data-driven models to improve wheat yield predictions. By addressing challenges related to data quality, model calibration, and scalability, this research contributes to advancing agricultural modelling and decision-making, ultimately supporting food security in vulnerable regions.

Appendices

Appendix A Ethical approval letter



16 April 2020

Dear Mr Oulaid ,

Reference: CURES/10823/2020

Title: The economic and environmental value of agricultural model-based decision support tools for wheat crops in Morocco

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/10823/2020 has been reviewed. You may now proceed with the research activities you have sought approval for.

If you have any queries, please contact CURES Support.

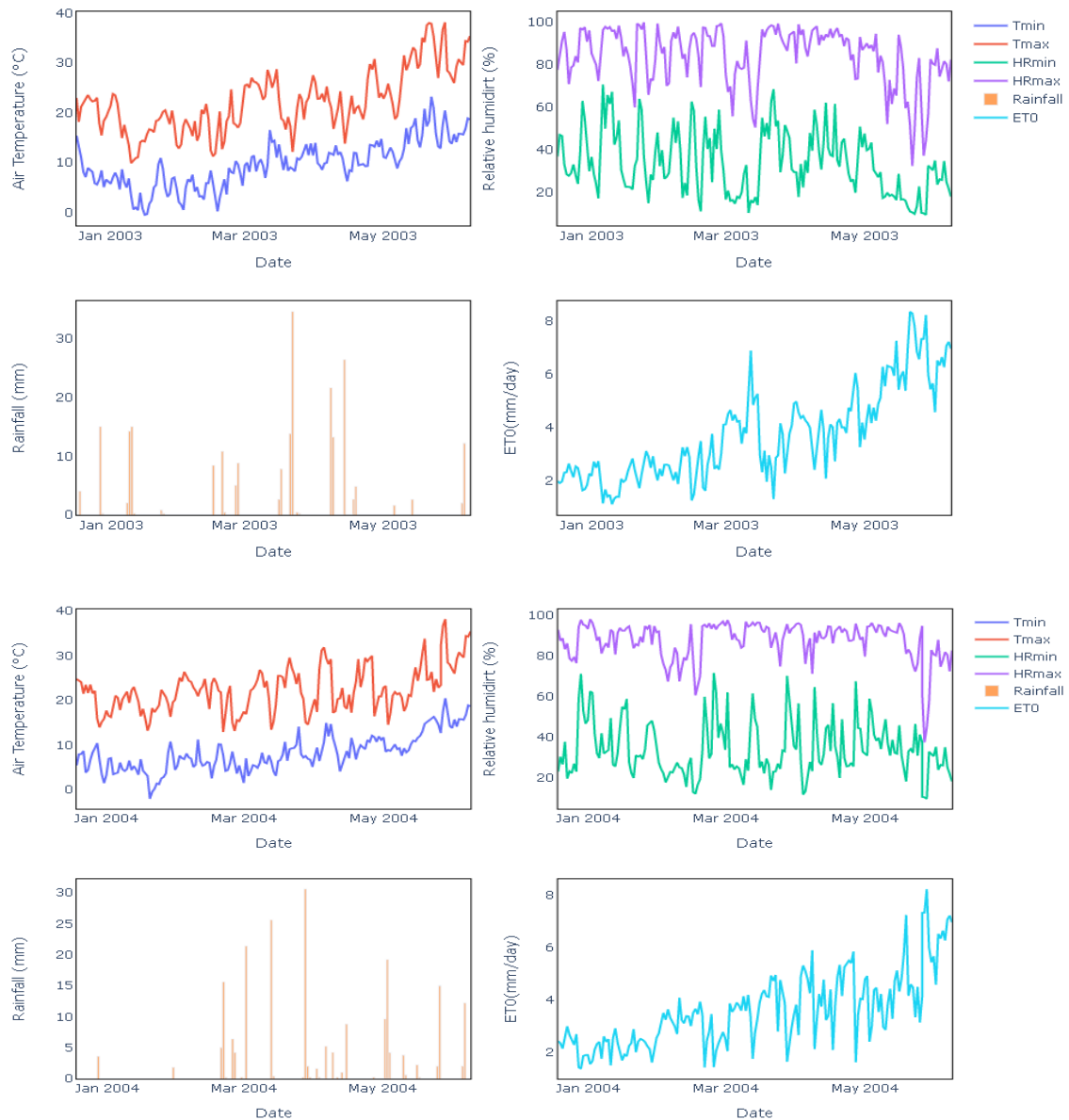
We wish you every success with your project.

Regards,

CURES Team

Appendix B – Supplementary Material to Chapter 2

B.1 Daily meteorological data



Figure_Apx B.1-1 Daily meteorological data on the experimental site for the 2002–2003 and 2002–2003 growing seasons.

B.2 AquaCrop model main parameter list

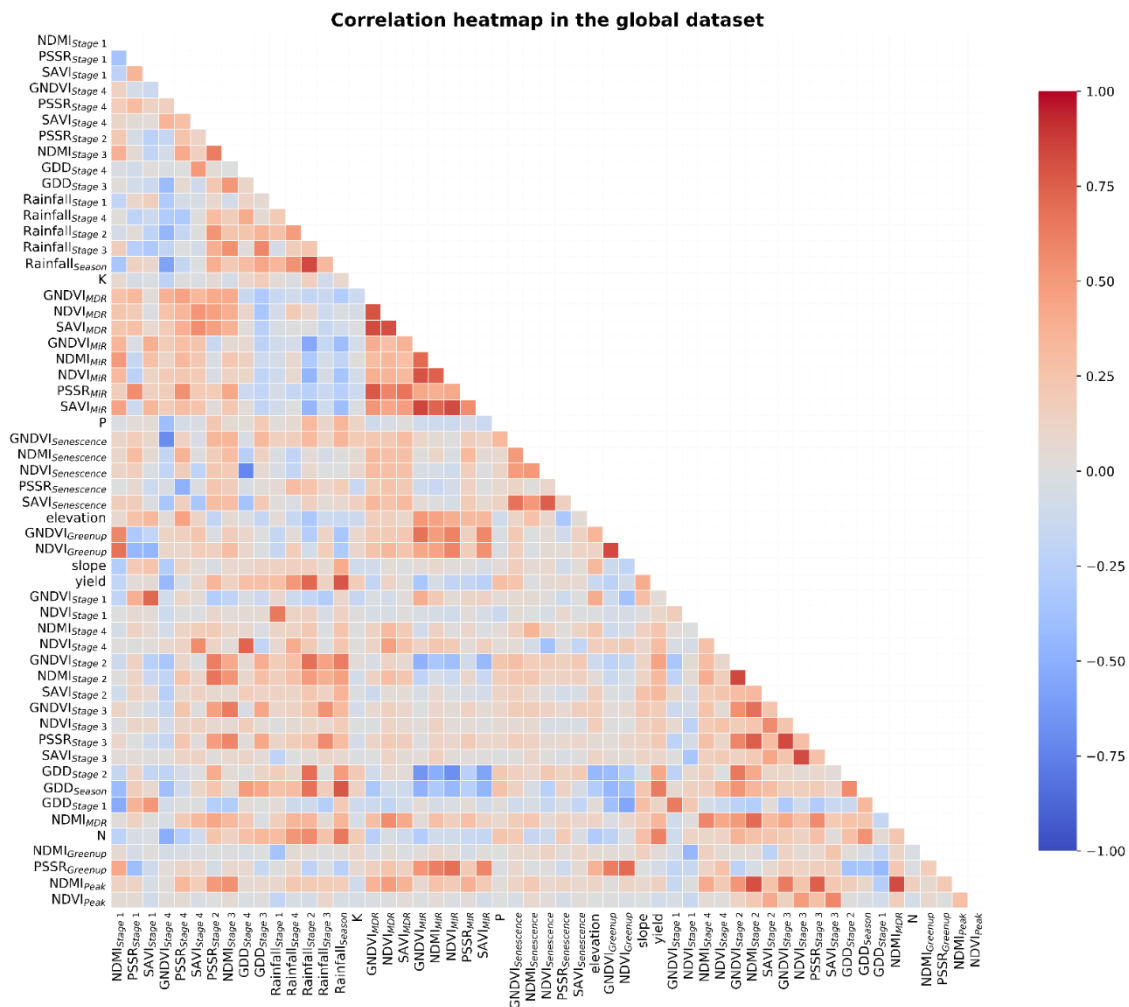
Table_Apx B.2-1 AquaCrop main sensitive model parameters values obtained using the two studied approaches

Input Parameters	Parametrization	
	PSO _{L7}	PSO _{HP}
gEme	47.250	98.56
gMat	1500	1472.10
gSen	610.99	817.63
CDC	0.004008	0.004929
CGC	0.011127	0.010551
CCX	0.97	0.95
WxTopZ	0.014176	0.05881
WxBotZ	0.002124	0.058738
pExpLo	0.719999	0.719994
pStmLo	0.741239	0.699532
pStmUp	0.354651	0.34125
pSenUp	0.946777	0.858135

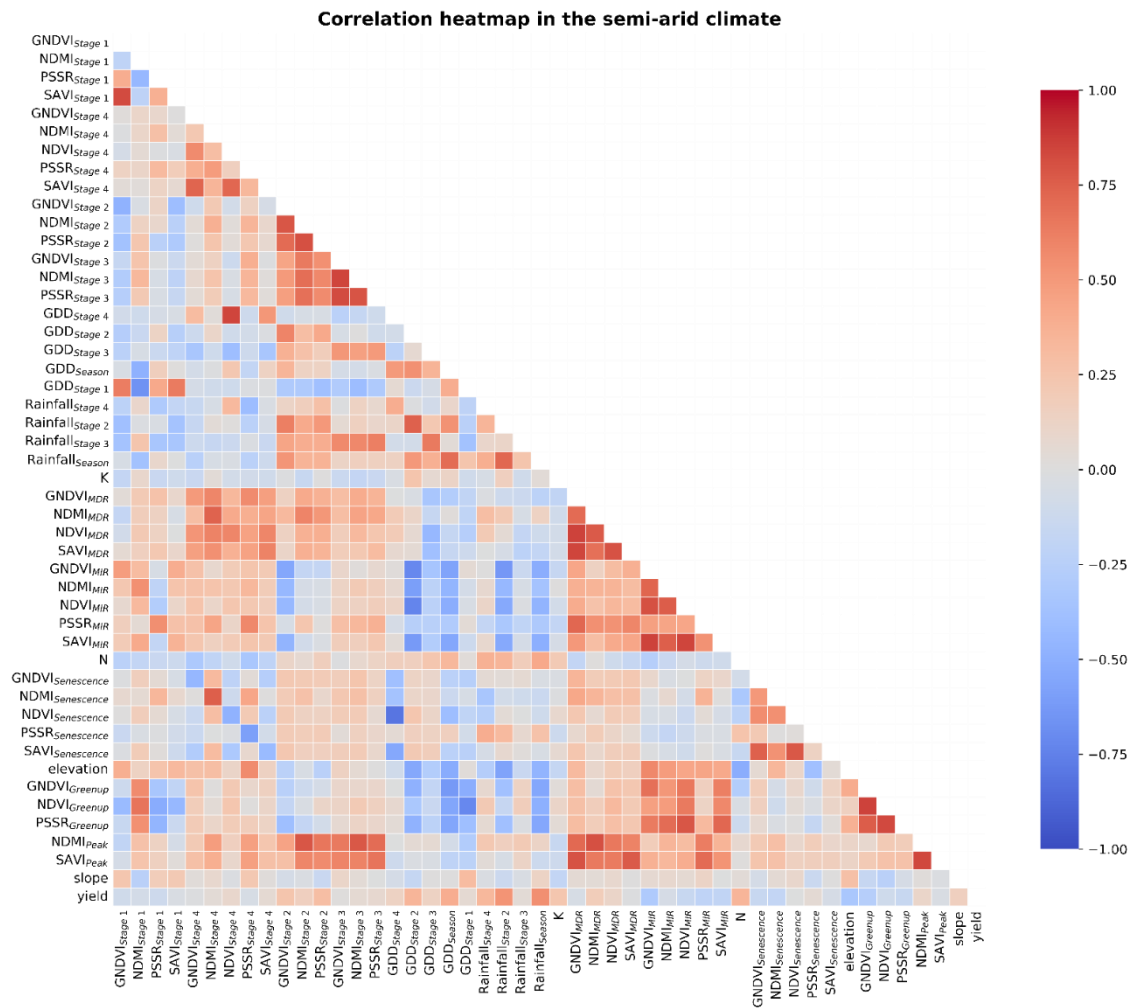
sExp	4.874926	4.87476
gYld	455.72	641.55
gYldDur	748.94	669
Kcb	1.785033	0.579192
Wp	27.5	20
Wpy	56.25	156.25
Hi0	0.31596	0.249975

Appendix C – Supplementary Material to Chapter 3

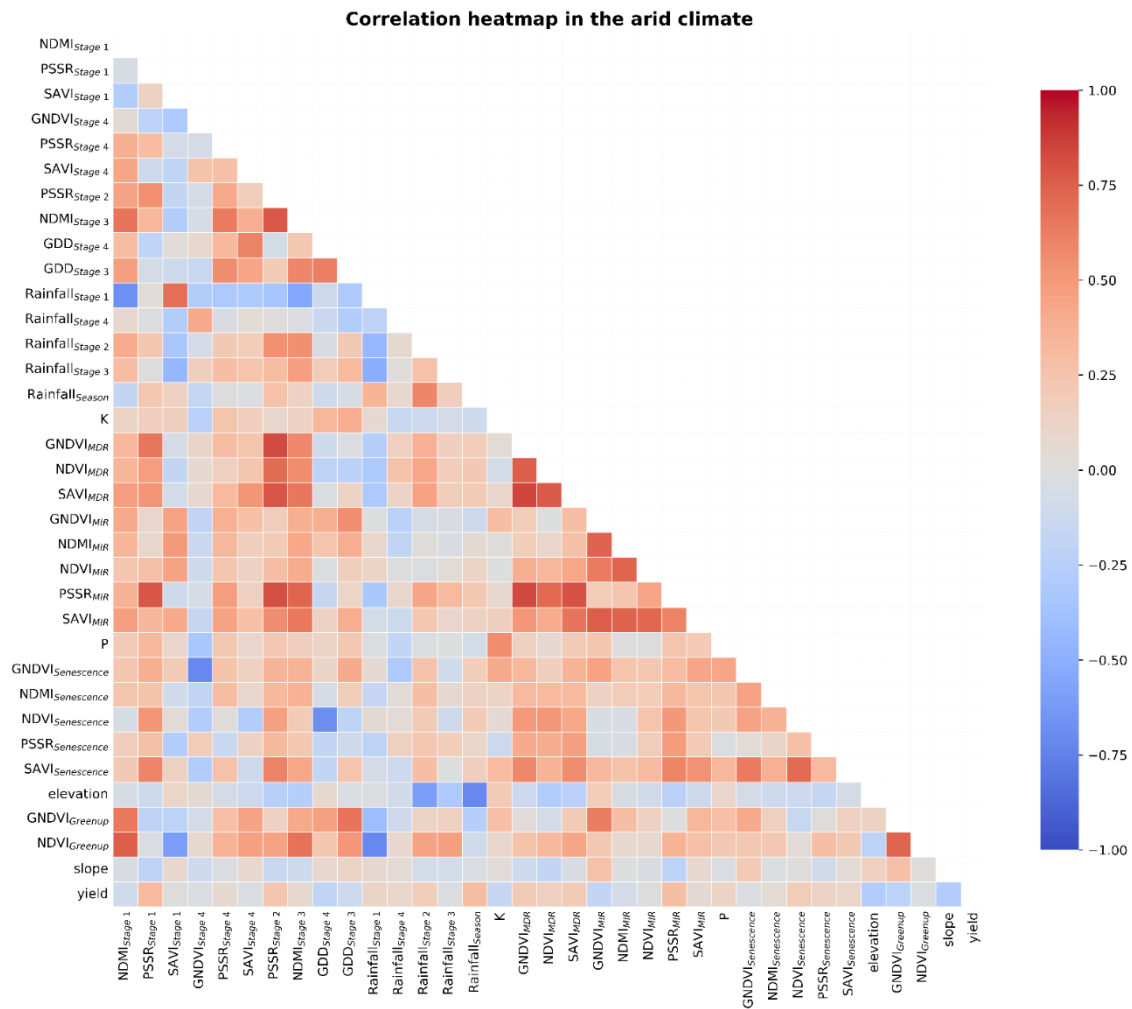
C.1 Yield correlation heatmaps



Figure_Apx C.1-1 Correlation heatmap in the global dataset.



Figure_Apx C.1-2 Correlation heatmap in the semi-arid climate dataset.



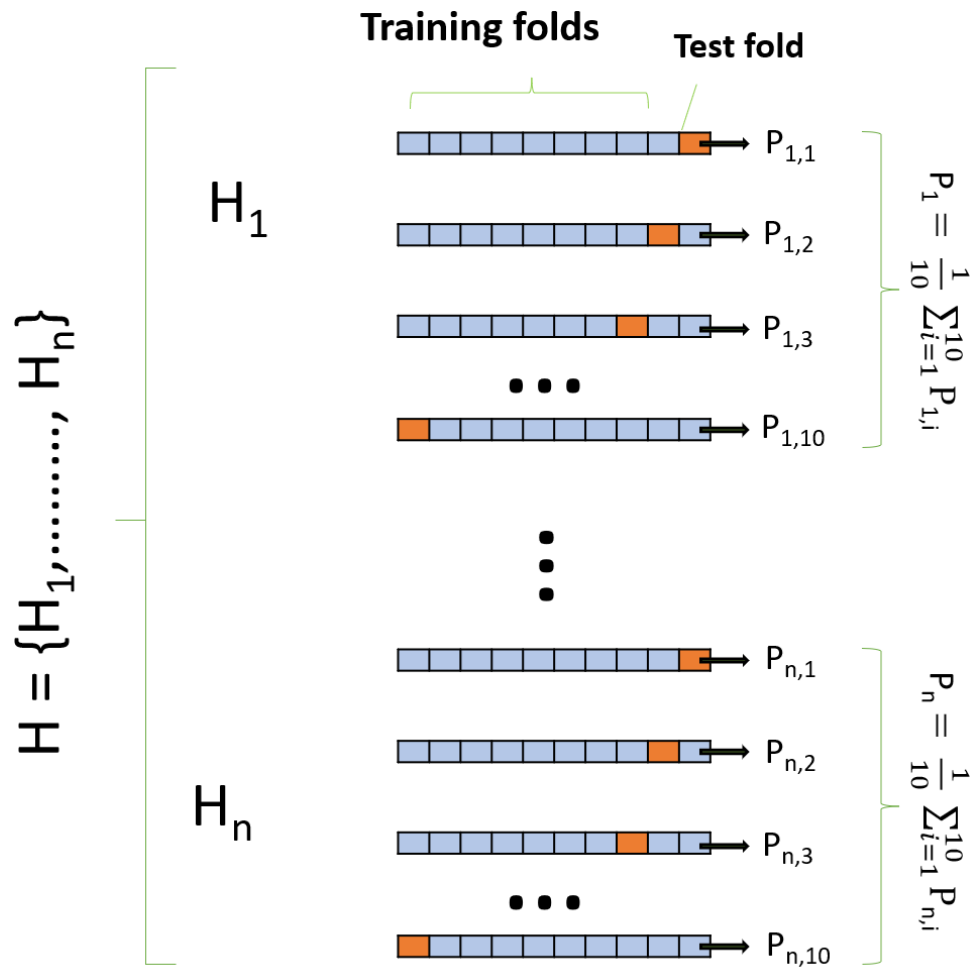
Figure_Apx C.1-3 Correlation heatmap in the arid climate dataset.

C.2 List of hyperparameters for model development.

Table_Apx C.2-1 Detailed list of hyperparameters used for model development.
Default values are based on the python Scikit-learn (version 0.22)

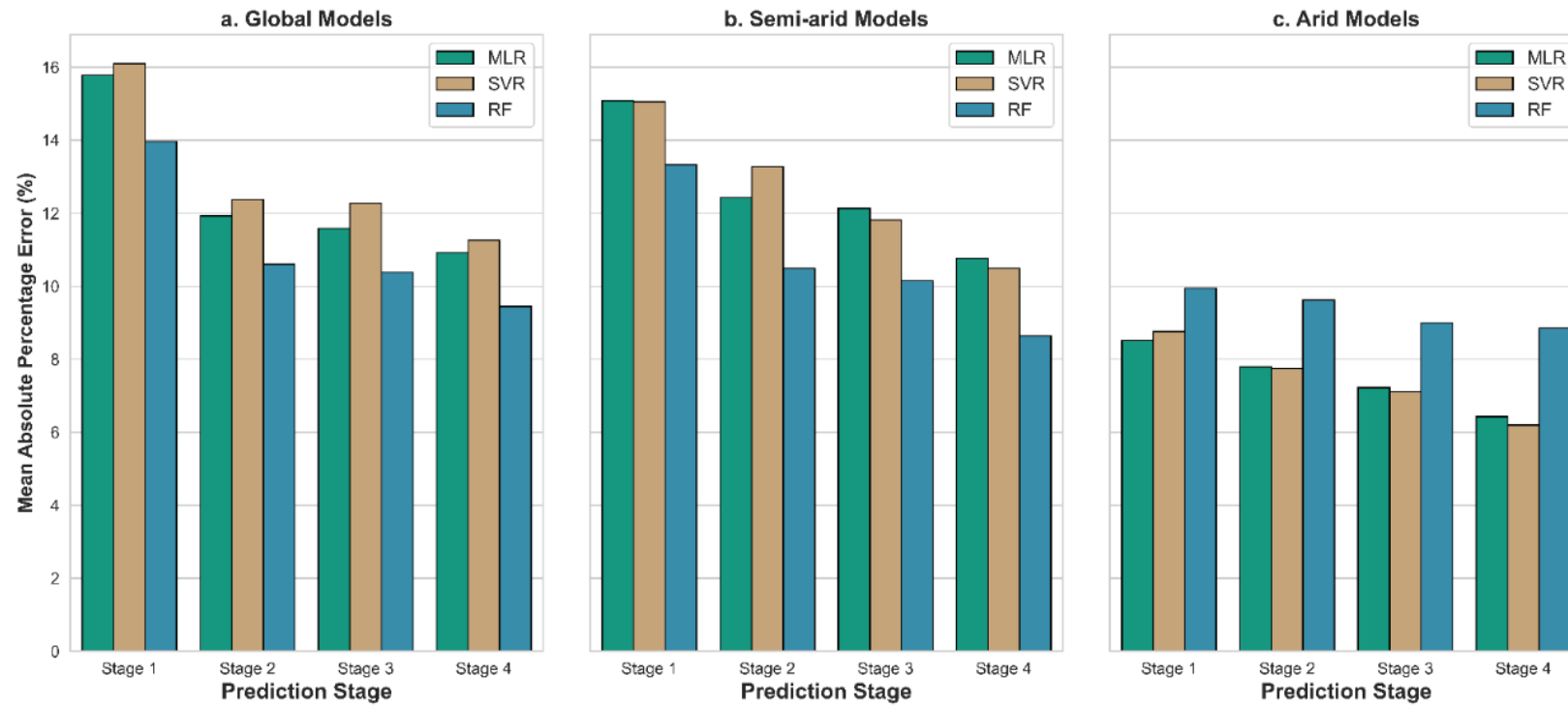
Model	Hyper-parameter	Search space	Default value
Support Vector Regressor	Kernel type	['linear', 'rbf', 'poly', 'sigmoid']	'Rbf'
	Regularisation parameter	$[10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1]$	1
	Epsilon	$[10^{-10}, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1]$	0.1
Random forest	Number of estimators	[10, 20, 30, 40, 50, 60, 70, 80, 90, 100]	100
	Maximum depth of the tree	[None, 5, 10, 15, 20]	None
	Minimum sample split	[2, 3, 4, 5, 6, 7, 8, 9, 10]	2
	Minimum leaf split	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	1

C.3 Hyperparameter tuning

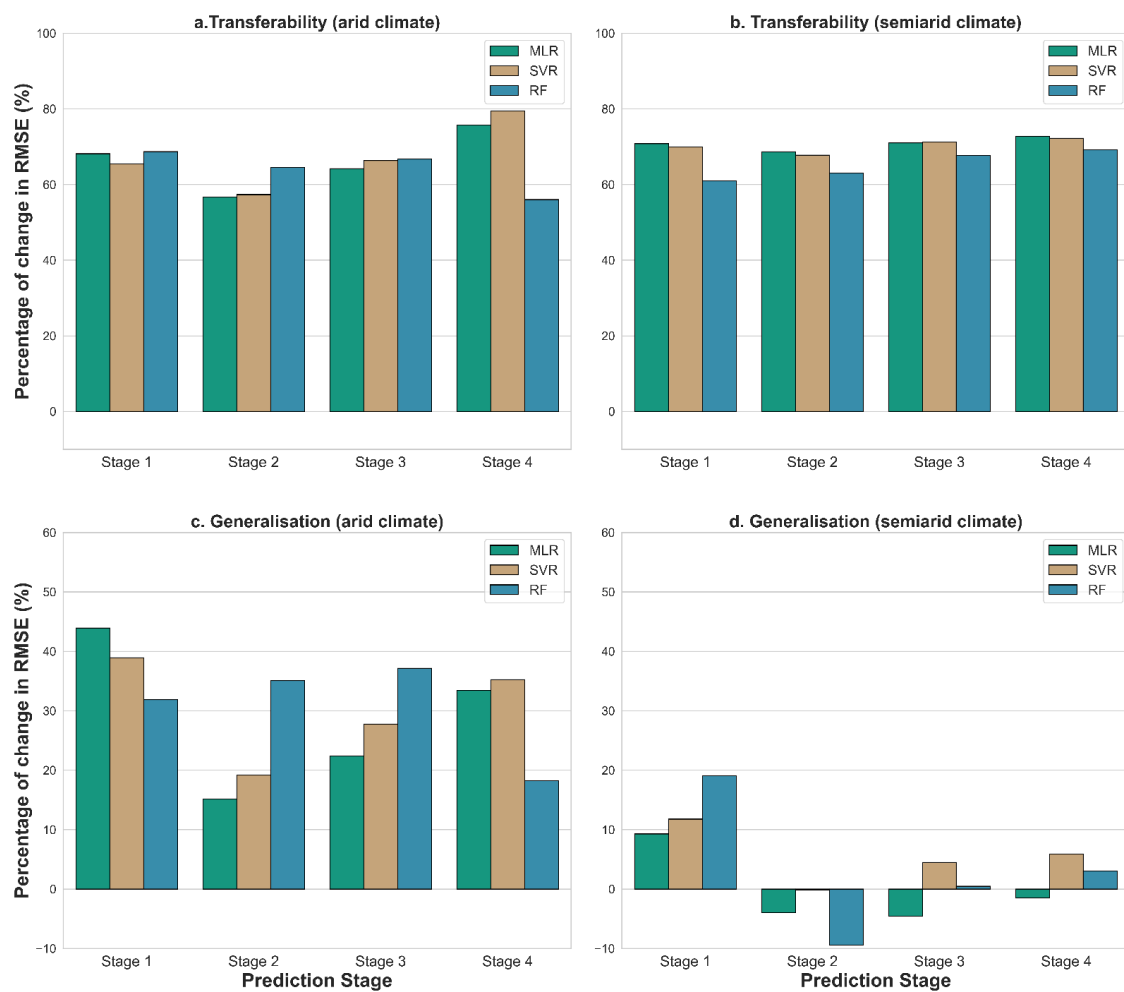


Figure_Apx C.3-1 Hyperparameter tuning using grid-search and cross validation
 H represent the set of hyperparameters, P_i is the performance of the model using hyperparameter combination H_i .

C.4 Evaluation metrics for stage-specific models



Figure_Apx C.4-1 Stage-specific Mean Absolute Percentage Error (MAPE) of the global (a) and localised models (i.e., b and c).



Figure_Apx C.4-2 Percentage of change in RMSE of transferred and generalised models across the four prediction stages.

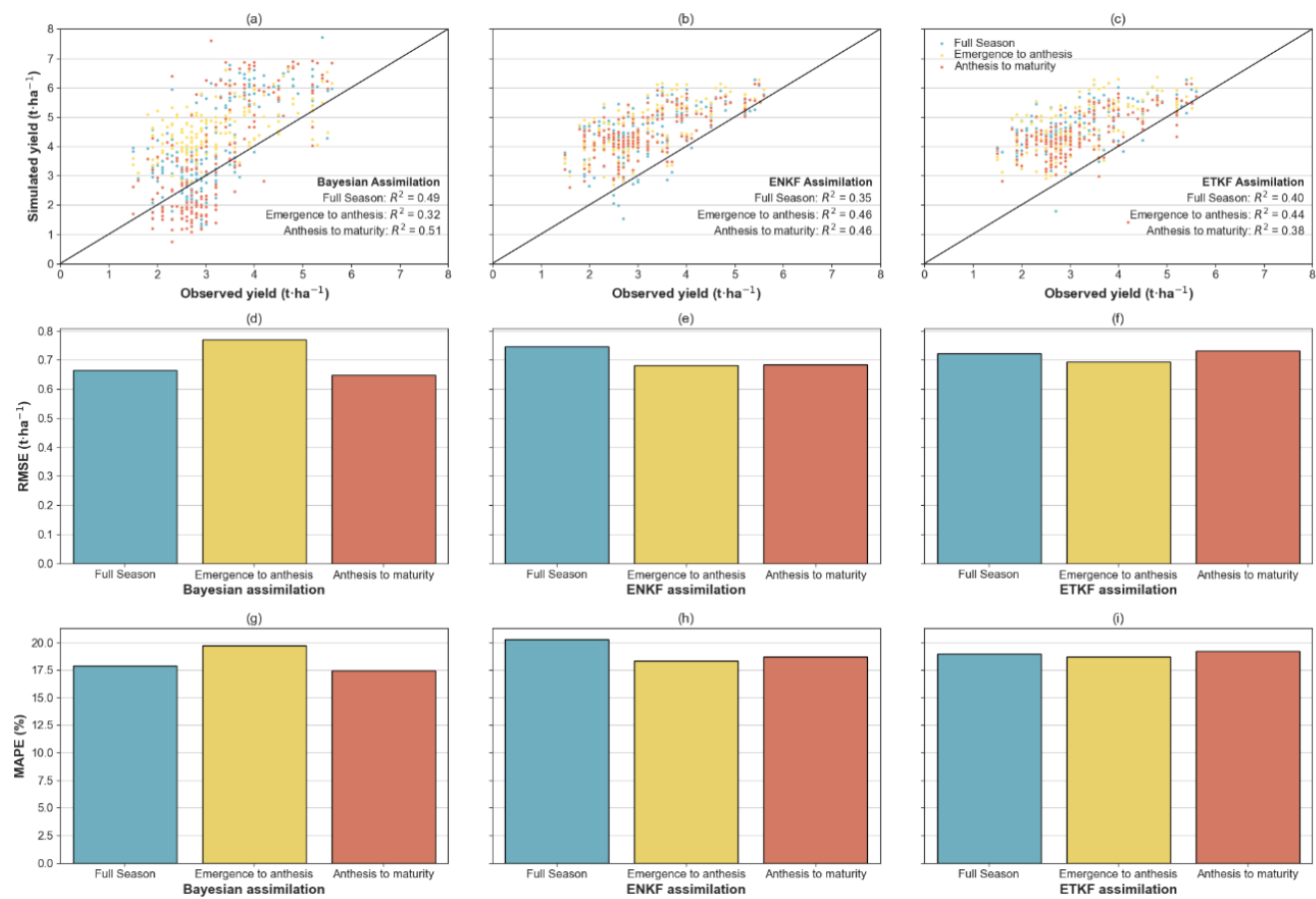
Appendix D – Supplementary Material to Chapter 4

D.1 Full list of vegetation indices

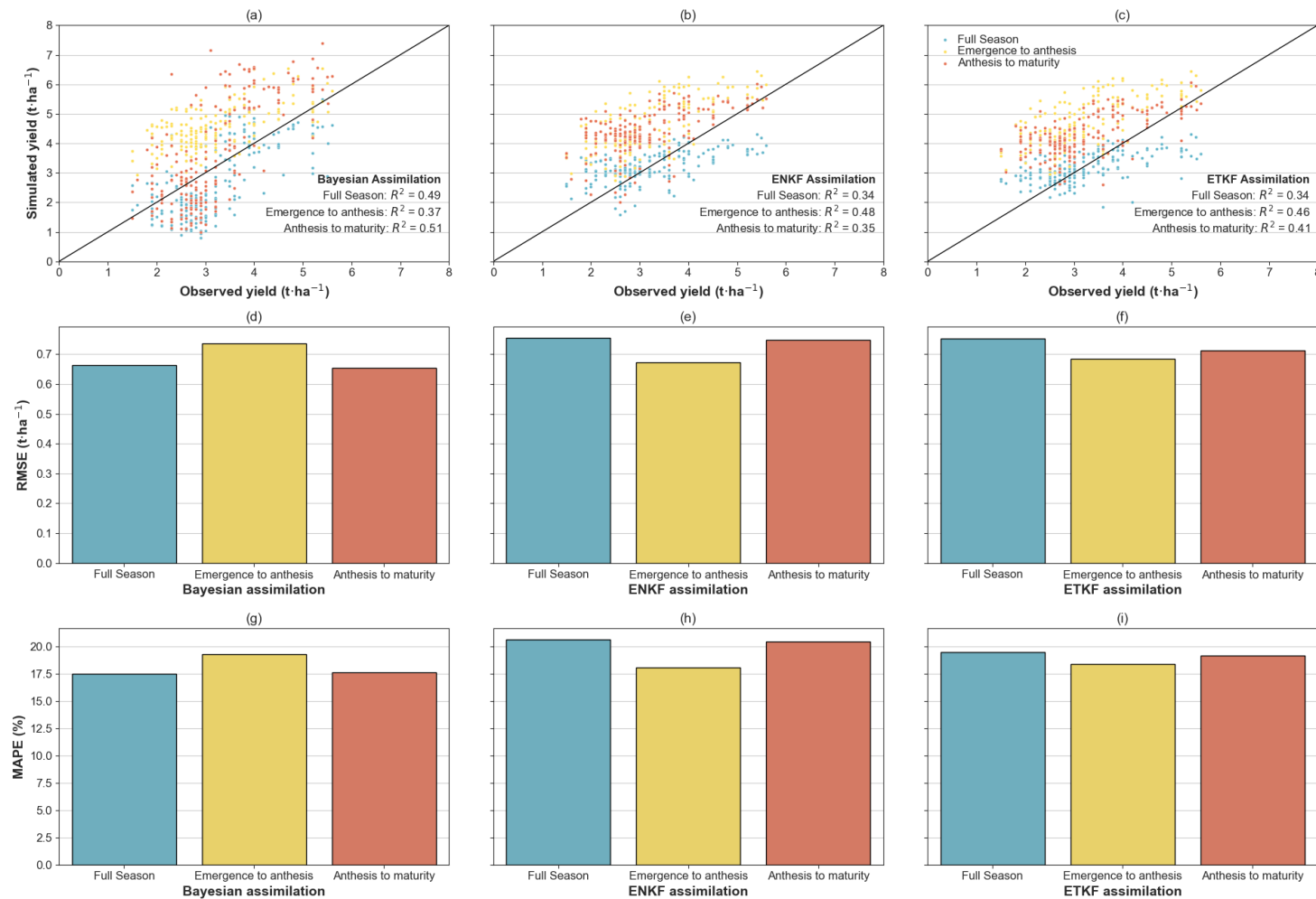
Table_Apx D.1-1 Full list of vegetation indices used to fit the Lasso model

Vegetation index	Equation
Normalized Difference Vegetation Index	$\frac{NIR - RED}{NIR + RED}$
Normalized Difference Moisture Index	$\frac{NIR - SWIR}{NIR + SWIR}$
Soil Adjusted Vegetation index	$\frac{1.5(NIR - RED)}{NIR + RED + 0.5}$
Modified Soil Adjusted Vegetation index	$\frac{2 NIR + 1 - \sqrt{(2 NIR + 1)^2 - 8(NIR - RED)}}{2}$
Infrared Percentage Vegetation Index	$\frac{NIR}{NIR + RED}$
Green Leaf Index	$\frac{2 NIR - RED + SWIR}{2 NIR + RED + SWIR}$
Shortwave Infrared Ratio	$\frac{SWIR}{NIR}$
Water Band Index	$\frac{NIR}{SWIR}$

D.2 In-season assimilation results using Landsat 7 and the combined dataset



Figure_Apx D.2-1 In-season assimilation accuracy metrics for Landsat 7



Figure_Apx D.2-2 n-season assimilation accuracy metrics for the combined dataset

Appendix E – Supplementary Material to Chapter 5

In-season wheat yield predictions and associated agricultural decisions:

Workshop outcomes

On the 11th of May 2023, a workshop was held at University Mohamed VI Polytechnic (UM6P) and brought together a diverse group of stakeholders, including researchers, farmers, and irrigation advisors. All engaged in fruitful discussions on two themes: i) the key decisions that could be supported the different stages of wheat production, ii) the trade-offs between the lead time and the uncertainties in yield predictions. This initiative is part of the broader SAFA (Sustainable Agriculture for Africa) project. Participants were split into two tables, namely Table A and Table B, each focusing on specific aspects of wheat cultivation.

Table A primarily addressed irrigated wheat and covered the different stages of the crop's growth cycle. Prior to stage 1, the stakeholders emphasized the importance of addressing soil surface crusting by implementing two successive irrigation events. They also highlighted the need to apply adequate amounts of potassium (K), phosphorus (P), and one-third of the required nitrogen (N) to prevent lodging. As stage 1 approaches, the participants emphasized the need to continue managing soil surface crusting. In stage 2, the discussions revolved around the application of the second and third doses of nitrogen (N) to meet the crop's nutritional needs. Additionally, the stakeholders recognized the importance of implementing phytosanitary treatments, such as pesticides or fungicides, to protect the wheat plants from potential diseases and pests. As the crop progressed to stage 3, manual weeding was highlighted as essential for weed control and overall crop health. The stakeholders emphasized the continuation of phytosanitary treatments to mitigate risks associated with pests

and diseases. In stage 4, targeted interventions to address late-season cryptogamic diseases were emphasized, including treatment of affected plants and removal of infected residues to minimize disease spread. When discussing rainfed wheat in Table A, the attendees emphasized the importance of applying adequate amounts of K, P, and one-third of N as previously advised for irrigated wheat before stage 1. For stage 1, the decision on the second dose of N should be based on rainfall predictions. In stage 2, if a low yield is expected, it is strongly advised not to apply additional nitrogen. In stages 3 and 4, when a low yield or poor-quality wheat is anticipated, an alternative approach of using the crop as fodder without costly phytosanitary treatments was suggested.

Turning to Table B and focusing on irrigated wheat, the attendees provided the following recommendations: Before stage 1, The attendees reiterated the identical recommendations put forth by participants in Table A. However, they introduced two additional decisions, namely conducting a comprehensive soil analysis and considering a crop insurance. Moving to stage 1, a second irrigation was suggested. In stage 2, a third irrigation should be implemented, along with the second and third doses of N, and the use of herbicides. For stage 3, the participants recommended conducting a fourth irrigation event and applying fungicides. Shifting to rainfed wheat in Table B, the participants outlined the following decisions for farmers: Before stage 1, they advised conducting soil analysis, applying full doses of P and K, one-third of N, and considering crop insurance based on factors such as farm size, farmer, and location in Morocco. For stage 1, the application of N should be based on rainfall predictions. In stage 2, herbicides were recommended. In stage 3, insecticides should be applied. Stage 4 discussions left the recommendations open-ended, indicating the need for further consideration and exploration of suitable practices specific to rainfed wheat cultivation.

When the stakeholders in both tables were asked if they are willing to sacrifice yield prediction accuracy over the lead time of predictions, attendees in Table A expressed a preference for Model 1 describing it as the most beneficial for their decision-making process. They emphasized that earlier yield predictions

using this model would allow them to adapt their decisions, such as adjusting irrigation and implementing disease treatments, based on the predicted yield. Model 4, on the other hand, is considered less favourable because its predictions come too late in the growth cycle, limiting the farmer's ability to make important decisions in a timely manner. Additionally, an eco-friendly farmer in Table A suggested that adopting sustainable practices that promote good yields regardless of climatic conditions is crucial.

In Table B, for irrigated wheat, the consensus among farmers was that Model 1 is the best, since it will allow them to have more rational control over irrigation. Models 2 and 3 were also considered useful as they provide insights into the need for investing in phytosanitary treatments. However, Model 4 was perceived as less valuable since its late predictions do not offer practical benefits to the farmers.

Overall, the workshop served as a valuable platform for stakeholders to exchange knowledge, share experiences, and collectively identify practices and strategies for sustainable wheat production across different growth stages, considering the unique characteristics of irrigated and rainfed systems. The opinions of farmers in both tables align with the idea that models providing earlier predictions are more advantageous for making timely and informed decisions. The specific preferences may vary based on the farmers' level of control over irrigation and their overall focus on sustainability.