TU WIEN Informatics

Kombination von Fachwissen und KI zur robusten Parametrisierung von Intercrop Modellen mit begrenztem Datenbestand

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Data Science

eingereicht von

Johannes Pfennigbauer, BSc
Matrikelnummer 11902046

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Univ.Prof. Dr.techn. A Min Tjoa
Mitwirkung: Dr.techn. Thomas Neubauer
Assoc. Prof. Dr. Ahmad M. Manschadi

Wien, 24. August 2026

Johannes Pfennigbauer	A Min Tjoa

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Wien Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

TU WIEN Informatics

Combining Domain Knowledge and AI for Robust Parameterisation of Agricultural Intercrop Models with Limited Data

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Data Science

by

Johannes Pfennigbauer, BSc
Registration Number 11902046

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.Prof. Dr.techn. A Min Tjoa
Assistance: Dr.techn. Thomas Neubauer
Assoc. Prof. Dr. Ahmad M. Manschadi

Vienna, August 24, 2026

Johannes Pfennigbauer	A Min Tjoa

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Declaration of Authorship

Johannes Pfennigbauer, BSc

I hereby declare that I have written this thesis independently, that I have completely specified the utilized sources and resources and that I have definitely marked all parts of the work – including tables, maps and figures – which belong to other works or to the internet, literally or extracted, by referencing the source as borrowed.
I further declare that I have used generative AI tools only as an aid, and that my own intellectual and creative efforts predominate in this work. In the appendix "Overview of Generative AI Tools Used" I have listed all generative AI tools that were used in the creation of this work, and indicated where in the work they were used. If whole passages of text were used without substantial changes, I have indicated the input (prompts) I formulated and the IT application used with its product name and version number/date.

Vienna, August 24, 2026

Johannes Pfennigbauer

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Danksagung

Mein herzlicher Dank gilt meinem Betreuer, Prof. Tjoa und Dr. Neubauer für ihre fachliche Begleitung, die hilfreichen Rückmeldungen und die Unterstützung während der Entstehung dieser Arbeit.

Ein besonderer Dank gilt Prof. Manschadi. Sein umfassendes Fachwissen und seine Unterstützung bei der Modellierung waren für diese Arbeit von großem Wert. In zahlreichen Gesprächen nahm er sich Zeit, die Simulationsergebnisse gemeinsam mit mir zu interpretieren und offene Fragen zu klären. Dadurch erhielt ich einen tiefen und für mich besonders wertvollen Einblick in die pflanzenbauliche Modellierung.

Abschließend danke ich allen Menschen, die mich auf diesem Weg begleitet haben und vor allem meinen Eltern, die mir dieses Studium ermöglicht haben.

Für Omi.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Kurzfassung

Prozessbasierte Modelle für Pflanzenwachstum leisten einen wertvollen Beitrag zur Evaluierung von Mischkulturen, setzen jedoch eine zuverlässige Parametrisierung voraus. Bei *InterCrop* ist dieser Prozess arbeitsintensiv und benötigt eine entsprechende Datengrundlage. Diese Arbeit entwickelt einen modularen Arbeitsablauf, der datenbasierte Methoden, agronomisches Fachwissen, Bayes'sche Optimierung und Sensitivitätsanalyse kombiniert und exemplarisch anhand einer Hartweizen–Ackerbohnen–Mischkultur untersucht.

Zuerst wurden einzelne Feldmessungen mithilfe von Satellitenbildern und Random-Forest-Modellen zu lückenlosen Zeitreihen vervollständigt, wodurch der Mittelwert von 4,12 Beobachtungen je Variable und Parzelle auf etwa 76 Werte stieg. Durch Feldmessungen gestützte Vorhersagen erreichten nRMSE-Werte von 5,7–16,7% für Pflanzenhöhe, oberirdische Biomasse und Stickstoffaufnahme. Für den Leaf Area Index (LAI) betrug der Wert 31,1%. Vorhersagen ohne unterstützende Feldmessungen waren, insbesondere bei unterrepräsentierten Kulturen und dem LAI, weniger zuverlässig.

Anschließend wurde eine Bayes'sche Optimierung mit biologischen Nebenbedingungen kombiniert. In den jeweils einmal durchgeführten Läufen erzielten die bedingte und die uneingeschränkte Konfiguration nach 200 Iterationen nahezu identische Endwerte; die geringen Unterschiede im Suchverlauf belegen keinen Konvergenzvorteil. Die Nebenbedingungen sicherten biologisch plausible Parameterbeziehungen.

Abschließend zeigte eine Variogramm-Analyse über 81 Kombinationen aus Wetter, Boden, Aussaattermin und Stickstoffversorgung, dass die Parameterbedeutung vom Umweltkontext und der Perturbationsskala abhing. Die Tiefe der Wasseraufnahme durch Wurzeln gewann unter trockenen Bedingungen an Bedeutung, während sonst Pflanzenhöhe und Entwicklungsparameter dominierten. Sensitivitätswerte der Pflanzenhöhe machten zudem Wechselwirkungen zwischen den Kulturen sichtbar. Das Wissen darüber, welche Parameter unter welchen Bedingungen entscheidend sind, zeigt zugleich, wo sich der Messaufwand lohnt, beispielsweise bei der Wurzeltiefe an trockenen Standorten mit leichten Böden. Insgesamt verringert der entwickelte Arbeitsablauf dadurch den manuellen Aufwand der *InterCrop*-Parametrisierung, während die kontextspezifische Sensitivitätsanalyse die Modelltransparenz erhöht. Weiterreichende Schlussfolgerungen erfordern wiederholte Optimierung und unabhängige Validierung über mehrere Saisonen und Kulturkombinationen.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Abstract

Process-based crop models can support the evaluation of intercropping systems, but their practical application depends on reliable parameterisation. For the *InterCrop* model, this process is currently labour-intensive and is further complicated by sparse field observations. This thesis develops a modular workflow that combines data-driven methods, agronomic domain knowledge, Bayesian optimisation, and sensitivity analysis to partially automate parameterisation and evaluates it in an exemplary durum wheat–faba bean intercropping setting.

First, field observations were supplemented with Landsat-derived time series. Physical bounds, directional outlier detection, biologically constrained seasonal curve fitting, phenological information, and Random Forest models were combined to reconstruct crop trajectories, raising the initial average of 4.12 observations per variable and plot to roughly 76 values per season across 221 plots. Prediction anchored by surrounding field measurements achieved validation nRMSE values of 5.7–16.7% for plant height, above-ground biomass, and crop nitrogen uptake, while leaf area index (LAI) reached 31.1%. Fully predictive reconstruction without sufficient field anchors was less reliable, particularly for underrepresented crops and LAI.

Second, Trust Region Bayesian Optimisation was combined with directly derived cultivar information and biological constraints. In the single runs conducted, the constrained and the unconstrained configuration attained nearly identical final errors after 200 iterations, and the small difference in their search behaviour does not establish a convergence advantage. The constraints additionally ensured biologically plausible parameter relations.

Finally, a scenario-resolved Variogram Analysis of Response Surfaces was applied across 81 combinations of weather, soil, sowing date, and nitrogen supply. IVARS-10, IVARS-30, and IVARS-50 showed that parameter importance depended on environmental context and perturbation scale. Root water-extraction depth became more influential under dry conditions, while plant height and development parameters dominated other settings. Cross-component height sensitivities further exposed intercropping interactions represented within the model. Identifying which parameters matter under which conditions also shows where measurement effort pays off. Overall, the workflow lowers the manual effort required to parameterise *InterCrop* and increases model transparency. Broader conclusions require repeated optimisation, independent validation, additional weather years, and further crop combinations.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Contents

Kurzfassung **ix**

Abstract **xi**

Contents **xiii**

1 Introduction **1**
1.1 Motivation . . . 1
1.2 Problem Statement . . . 2
1.3 Aim of the Work . . . 4
1.4 Methodological Approach . . . 5
1.5 Thesis Structure . . . 7

2 Literature Review and Related Work **9**
2.1 Process-Based Intercropping Models . . . 9
2.2 Data Processing . . . 10
2.3 Optimisation Techniques in Model Parameterisation . . . 15
2.4 Sensitivity Analysis and Explainability . . . 16

3 The InterCrop Model **21**
3.1 Plant-Related Processes . . . 21
3.2 Soil-Related Processes . . . 22
3.3 Extensions for Intercropping . . . 23

4 Data and Study Area **25**
4.1 Field Trials . . . 25
4.2 InterCrop Variables . . . 28
4.3 Satellite Data . . . 30

5 Methodology **33**
5.1 Data Processing . . . 33
5.2 Parameterisation of the InterCrop Model . . . 41
5.3 Combined Sensitivity Analysis and Explainability . . . 48

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6 Results **53**
6.1 Dataset Enrichment and Synthetic Data Generation 53
6.2 Model Optimisation Evaluation 58
6.3 Sensitivity Analysis Findings 61

7 Discussion and Conclusion **69**
7.1 Answering the Research Questions 69
7.2 Conclusion 77
7.3 Limitations of the Study 78
7.4 Future Research Directions 79

A Appendix **81**

Overview of Generative AI Tools Used **85**

List of Figures **87**

List of Tables **91**

Bibliography **93**

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 1

Introduction

1.1 Motivation

In the face of climate change, soil degradation, and the growing need for resilient food systems, the importance of sustainable and efficient agricultural practices has become increasingly apparent. One promising strategy is intercropping, a form of polyculture involving the simultaneous cultivation of two or more crop species on the same field. Intercropping arrangements may take various forms, including casual seed spreading, alternating crop rows, and alternating strips [SAB+25]. Studies and reviews associated with this agroecological practice have consistently reported overall yield-per-area increases, reductions in fertiliser and pesticide inputs, and a diverse range of soil health benefits. Beyond yield improvements, intercropping strengthens crop resilience against both biological and environmental stressors, which in turn stabilises yields, while also improving biological productivity, nutrient use efficiency, and pest and weed suppression. At a broader scale, intercropping systems can sequester carbon, reduce agrochemical and fossil fuel usage, and improve economic performance at the farm system scale while lowering environmental impacts [KSS23, JCD+20, SAB+25, MHAN25].

A well-studied example is the intercropping of cereals and legumes, where the two species complement each other in nitrogen utilisation: atmospheric nitrogen fixed by legumes becomes available to neighbouring cereals, thereby reducing inter-specific competition. Such complementary associations frequently lead to higher yields, as the land area needed to achieve a given output from both species is smaller than when each is grown as a separate monoculture [MHAN25].

Despite these advantages, several barriers impede the widespread adoption of intercropping. Key challenges include the lack of established markets, difficulties in sorting mixed grains, insufficient processing facilities, and stringent food safety requirements — particularly when crops are sown together and grains of the component species are

harvested simultaneously [SAB+25]. A multi-country analysis of Strengths, Weaknesses, Opportunities and Threats (SWOT analysis) conducted by Stone et al. [SAB+25] across ten European focus groups supports these findings: the most widely recognised strengths were the reduction in fertiliser and pesticide inputs, soil health improvements, and increased yield stability under climate change conditions. Conversely, the most frequently reported weaknesses related to selling intercrop products profitably and managing crop interactions, compounded by a lack of tested markets and pricing structures for mixed legume–cereal grains. The most prominent opportunities centred on reducing fertiliser use and leveraging governmental support, while the principal threats concerned insufficient knowledge and technology as well as the risk of new weed, disease, and pest control challenges [SAB+25]. This last point — the lack of knowledge and adequate technological tools — underscores the need for improved modelling and decision-support frameworks, a challenge addressed in the following section.

1.2 Problem Statement

Building upon these premises, predicting crop growth and yield remains a formidable challenge, as these outcomes vary across both time and space depending on the prevailing production environment. At a global level, inter-seasonal fluctuations in precipitation and temperature alone account for approximately one-third of the observed yield variability in major crops such as wheat, maize, and soybean [MPF+22]. Moreover, the absence of transparent, interactive, and accessible software tools for translating agricultural data into informed management decisions has been identified as a key barrier to the uptake of smart farming technologies by practitioners [MPF+22]. The inherent complexity of intercropping systems poses an additional challenge, particularly regarding their adaptation to specific production conditions. In this context, process-based crop growth models are invaluable, as they can quantitatively integrate the interactions among intercropped species and predict resulting production outcomes. Such models enable the systematic exploration of how different factors affect the growth and yield of intercrop species, and support the selection of appropriate crop species, cultivars, and management practices for specific sites. To this end, several well-known crop growth models have been adapted for simulating intercropping systems, including APSIM, DSSAT, STICS, FASSET, CROPSYST, SWAP, and WaNulCAS [MHAN25].

One such model builds on the SSM-iCrop framework [SMS13], a single-crop, process-based model originally developed for wheat. SSM-iCrop employs identical sub-models for simulating crop processes across all grain species, with the exception of phenology, which may use species-specific sub-models. It further features a layered soil representation, nitrogen fixation in grain legumes, and several physiological concepts for evaluating genetic improvement [MEB+21]. As part of the EU-funded LEGUMINOSE project, the University of Natural Resources and Life Sciences, Vienna (BOKU) and the Vienna University of Technology (TU Wien) extended this single-crop foundation into *InterCrop*, a process-based model capable of simulating the development and growth of two crop species simultaneously. *InterCrop* accounts for climatic conditions, soil properties, management

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

practices, and crop-specific genetic traits, and has been integrated into a web-based decision support system designed to optimise the performance and sustainability of intercropping systems across different European regions.

The effectiveness of any such model depends to a great extent on how it is set up. This is known as parameterisation, which is the process of assigning numerical values to describe the physiological and phenological characteristics of crops as well as relevant soil properties. As agronomic outcomes such as yield and biomass result from the interaction between these parameters and external conditions, parameterisation is inherently challenging and resource-intensive. Manschadi et al. [MEB+21] demonstrated that full parameterisation is crucial for improving the accuracy of crop model predictions regardless of model complexity, a conclusion supported by earlier studies [HWWR17]. Notably, simpler models may benefit more from thorough parameterisation than complex ones: in the same study, the single-crop SSM-iCrop outperformed APSIM in predicting final nitrogen uptake and grain yield, with root mean square errors more than 40% lower [MEB+21]. A large-scale survey of calibration practices across over 200 modelling teams further revealed substantial variability in approaches, underscoring the need for standardised parameterisation guidelines [SPTW18]. Achieving robust parameterisation also demands high-quality, comprehensive datasets that cover not only production-related variables such as crop yield but also temporal patterns of crop growth and the dynamics of soil water, carbon, and nitrogen [MEB+21, GVBVW+15].

In this context, it is useful to distinguish parameterisation from calibration. Parameterisation in the strict sense refers to deriving parameter values directly from measured or observed data, whereas calibration is an indirect approach in which parameter values are iteratively adjusted until simulated outputs, such as yield or leaf area, match observed values [MEB+21]. A key design principle of SSM-iCrop, inherited by the *InterCrop* extension, is that its input parameters are predominantly based on directly measurable plant and soil properties. By relying on a relatively small number of such parameters, with the exception of a few phenology-related coefficients, the model avoids the pitfalls of unconstrained calibration, which tends to produce an empirical fit to the calibration data and limits applicability beyond the calibration environment [MPF+22]. Furthermore, keeping the number of input coefficients to a minimum and ensuring they can be independently measured or derived from data enhances model transparency, which generally diminishes as model complexity increases [MPF+22]. Nonetheless, directly deriving all parameters from measured data is not always feasible in practice, particularly under data-scarce conditions or when field observations are incomplete. In such settings, optimisation techniques can complement the measurement-based approach by searching for suitable parameter values within physically meaningful bounds informed by domain knowledge. Unlike unconstrained calibration, this knowledge-guided optimisation preserves the interpretability of the model parameters and avoids overfitting to a narrow calibration environment, while still enabling systematic and reproducible parameterisation where direct measurement alone is insufficient.

Currently, the parameterisation of the *InterCrop* model is conducted manually through

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

several labour-intensive steps: analysing yield and crop performance variations across field trials, preparing datasets for simulation, extracting relevant crop- and soil-related parameters from observed data, configuring the model using site-specific climate, soil, and management information, and comparing simulated outcomes with observed measurements as part of a manual calibration process. This workflow is time-consuming, requires substantial domain expertise, and poses a significant barrier to the wider adoption of the *InterCrop* model in practical agricultural decision-making. These limitations hinder the scalability of intercropping strategies, despite their recognised potential to promote sustainable farming practices and increase resilience to climate change. Addressing these challenges requires a more systematic and partially automated approach to parameterisation, as outlined in the following section.

1.3 Aim of the Work

To address these challenges, this thesis aims to improve and partially automate the parameterisation process of the *InterCrop* model through the combination of artificial intelligence (AI), advanced computational techniques, and domain expertise. The objective is to replace the current manual workflow with a modular and scalable pipeline that improves data quality, modelling robustness, and practical usability. The work focuses on three core areas:

1. **Data preprocessing:** Developing a data pipeline capable of detecting anomalies, handling missing values, and generating synthetic data to fill critical gaps and to improve the quantity and quality of model inputs.

2. **Model parameterisation:** Applying state-of-the-art optimisation techniques to identify site- and crop-specific model parameters in an efficient and reproducible manner.

3. **Sensitivity analysis and model robustness:** Conducting sensitivity analysis to evaluate the influence of key parameters on model outputs, thereby increasing trust in the simulation results and identifying the variables that drive model behaviour.

In addition, the simulations produced with the identified parameters are compared with the field trial measurements and the reconstructed trajectories to assess how closely the parameterised model reproduces them. Since both served as targets during parameterisation, this comparison describes the achieved fit rather than predictive accuracy on independent data. The sensitivity analysis is furthermore related to explainable AI (XAI) at the methodological level. Section 2.4 compares sensitivity analysis with post-hoc explanation methods, specifically SHAP and the XVARS extension, and derives from that comparison which explanatory requirements can be met within a sensitivity framework.

The research questions guiding this work are intentionally broad, and answering them in full generality lies beyond the reach of a single thesis. Instead, they are addressed within a

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

deliberately bounded, exemplary setting: the *InterCrop* model together with the specific intercropping dataset introduced in Chapter 4. Within this setting, promising methods are identified through a review of the state of the art, implemented, and evaluated. The findings reported throughout this work should therefore be read as evidence gathered from one carefully studied case rather than as universally valid statements. The chosen example is intended to be representative enough for the findings to inform comparable settings, but any such transfer remains a hypothesis to be tested rather than a result established here.

The term *robust* in the title and in the research questions is used in the same bounded sense. It describes how the workflow is constructed, namely through directional cleaning rules, biologically bounded seasonal curve shapes, anchored interpolation, and constraints that keep the parameter search within plausible regions. It does not assert that the estimated parameter values are shown to be statistically stable. Demonstrating the latter would require repeated optimisation runs, uncertainty intervals for the estimated parameters, and evaluation against observations held out from the parameterisation, none of which is part of the design adopted here. Section 7.3 returns to what this leaves open.

The central research questions (RQ) guiding this work are:

RQ1: **In conditions of extreme data scarcity, how can qualitative knowledge constraints be integrated to guide robust preprocessing and generate sufficient synthetic data for parameterising process-based intercropping models?**

RQ2: **In what ways can hybrid optimisation strategies - combining data-driven, heuristic, and domain knowledge - enhance parameter estimation in underdetermined process-based models?**

RQ3: **How can sensitivity analysis be combined with explainability techniques to improve model transparency and foster trust in the simulation results, especially when working with limited or partially synthetic data?**

1.4 Methodological Approach

This thesis adopts a modular, data-driven approach to enhance and partially automate the parameterisation of the *InterCrop* model by combining AI-based techniques with domain knowledge. Rather than committing to specific methods *a priori*, the work begins with a targeted and iterative review of the current state of the art (Chapter 2) to identify promising techniques for optimising and analysing process-based crop models, with particular emphasis on methods that remain effective under data-scarce and heterogeneous conditions. The concrete techniques applied in each phase are therefore outcomes of this review process. The overall workflow is organised into three phases, which are discussed in detail below.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Phase 1: Data Preprocessing. A robust data preprocessing pipeline is developed to extract, clean, and structure the required model inputs. These inputs encompass field trial measurements (e.g. yield, biomass and phenological stages), management practice records, and spectral information derived from satellite imagery. Techniques for anomaly detection, missing-data imputation, and the generation of partly synthetic training data—whether through interpolation or AI-based methods guided by agronomic constraints—are evaluated during the literature review. The approach considered most suitable for the requirements of this study is subsequently applied to produce a consistent dataset for model parameterisation.

Phase 2: Model Parameterisation and Evaluation. The second phase focuses on deriving crop-specific parameters through a hybrid strategy that couples advanced optimisation algorithms with knowledge-based constraints. The literature review assesses the suitability of algorithms, especially focussing on Bayesian optimisation. The reviewed algorithms are discussed regarding their capability to leverage domain knowledge and to operate efficiently in high-dimensional parameter spaces with sparse input data. The selected algorithm minimises a custom loss function that encodes the discrepancy between simulated and observed outputs together with domain-knowledge constraints, aiming for robust parameter estimates across the intercropping configurations available in the dataset. To isolate the contributions of the synthetic data and of the knowledge-based guidance, the model is parameterised three times: once on the raw observed data, once on the cleaned synthetic data, and once again on the cleaned synthetic data but with the domain-knowledge guidance of the custom loss function removed. These three runs are compared with respect to their convergence behaviour, alongside their overall error. Each run is carried out for a fixed number of iterations, irrespective of whether full convergence is reached, since exhaustive calibration is neither the primary goal nor within the scope of this study. The main focus lies on comparing the informed parameterisation procedure against a standard baseline workflow.

Phase 3: Sensitivity Analysis and Explainability. Following model evaluation, a sensitivity analysis is conducted to identify the influence of variations in individual input variables on the simulated outputs. This analysis reveals the most critical parameters and clarifies whether their influence shifts under specific environmental conditions. The comparison with explainable AI techniques is carried out in the literature review rather than by running a second attribution method on the model. Its purpose is to establish which properties make an explanation informative, in particular conditional resolution by context and resolution across perturbation scales, and to carry these properties into the design of the analysis. The sensitivity analysis is accordingly structured as repeated scenario-wise evaluations at several perturbation scales instead of a single global ranking, with the aim of enhancing the transparency of the model and fostering user trust in the simulation results.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

1.5 Thesis Structure

The remainder of this thesis is organised as follows.

Chapter 2 – Literature Review and Related Work surveys the state of the art across the areas relevant to this work: process-based crop models and data assimilation techniques, data processing methods for anomaly detection and synthetic data generation, optimisation algorithms for model parameterisation including the integration of domain-knowledge, and approaches to combine sensitivity analysis and explainability techniques.

Chapter 3 – The InterCrop Model provides a conceptual overview of the *InterCrop* model and the SSM-iCrop framework on which *InterCrop* was built. It describes the key plant- and soil-related processes simulated in SSM-iCrop and explains how *InterCrop* extends these by taking competition and facilitation between two intercropped species with respect to light, water, and nitrogen into account.

Chapter 4 – Data and Study Area describes the datasets and study context underlying this work. It presents the intercropping field trial data sourced from an EU-wide project, defines the model input variables, optimisation objectives and parameters, and introduces the satellite-derived spectral bands and vegetation indices used to supplement the ground-based observations.

Chapter 5 – Methodology discusses the technical implementation of the three-phase workflow outlined in Section 1.4. It covers the data preprocessing pipeline including outlier detection, time-series smoothing, curve fitting, and point-data prediction, the parameterisation procedure via optimisation algorithms with an informed loss function, the model evaluation framework, and the combined sensitivity analysis and explainability approach.

Chapter 6 – Results reports the outcomes of each phase. It compares the performance of the optimised model under different configurations: optimisation with partly synthetic data, optimisation with observed data only, and optimisation without domain-knowledge constraints. The findings of the sensitivity analysis regarding key variables and model robustness conclude the chapter.

Chapter 7 – Discussion interprets the results with respect to the three research questions. Specifically, it examines to what extent domain-knowledge-guided preprocessing and synthetic data generation improve model parameterisation (RQ1), examines how the hybrid optimisation strategy affects the search behaviour and the biological plausibility of the resulting parameter estimates compared with a standard procedure (RQ2), and assesses how combining sensitivity analysis with explainability techniques contributes to model transparency and stakeholder trust (RQ3). The chapter closes by reflecting on the limitations of the study and outlining directions for future research.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 2

Literature Review and Related Work

To identify suitable methods for the methodological development and exploratory study undertaken in this thesis, a targeted and iterative literature review was conducted. The search followed a top-down methodological funnel. It started with broad review studies that were first used to map the principal approaches available for data reconstruction, parameter optimisation, and sensitivity analysis. The identified methodological families were then assessed against the requirements of the present problem, including data scarcity, nonlinear model behaviour, biological constraints, and the computational cost of repeated *InterCrop* simulations. The review was subsequently refined towards specific methods and applications in crop modelling or comparable environmental modelling contexts. This iterative refinement supported methodological convergence on the approaches implemented in the exemplary study, rather than aiming to provide an exhaustive mapping of all publications in the three research areas.

2.1 Process-Based Intercropping Models

While a range of process-based crop models have been successfully developed to simulate the growth and yield of monocultures across various environments [BWVDW+21], extending these frameworks to simulate the competitive and facilitative interactions of intercropping remains a formidable challenge. Several established models, such as FASSET, STICS, and APSIM, have been adapted to address these more complex multi-species systems [BWVDW+21]. However, it is difficult to design a system that accurately simulates the dynamic interactions of intercrops while keeping the number of parameters manageable and measurable. For instance, it is preferable to rely on crop height as a proxy for light interception rather than on more complex leaf angle distributions [VMG+23]. Furthermore, thorough parameterisation and validation of these models

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

for cereal–legume intercrops under European conditions remain critically sparse in the literature [BWVDW+21].

Consequently, attempts to model intercropping systems often expose structural limitations within existing architectures. For instance, while APSIM offers broad applicability across diverse pedoclimatic conditions, Berghuijs et al. [BWVDW+21] demonstrated that it systematically underestimated wheat dry matter and overestimated faba bean dry matter when simulating them together as intercrops. This discrepancy was partly attributed to the model linking plant height growth directly to stem weight per plant, thereby compromising its ability to accurately simulate light competition between the taller and shorter crop canopies [BWVDW+21]. By revisiting the formalisms for light interception, microclimatic changes, and resource acquisition in the STICS model, Vezy et al. [VMG+23] achieved relatively high consistency in modelling intercrops. They reported a normalised root mean square error (nRMSE) of 18% for yield and found evidence for the transfer of nitrogen in legume mixtures [VMG+23].

To address the complexities of interspecific competition while preserving model transparency, the *InterCrop* model was developed as a dedicated extension of the monocultural SSM-iCrop framework [SMS13, MHAN25]. By deliberately keeping the number of input parameters to a minimum and relying on predominantly measurable plant and soil properties, *InterCrop* mitigates the risks of unconstrained calibration, allowing it to systematically predict complex agronomic outcomes without sacrificing interpretability [MPF+22]. This transparent and constrained design makes *InterCrop* uniquely well-suited for the objectives of this study. A detailed explanation of the underlying simulation processes and their concurrent execution for interacting species is provided in Chapter 3.

While pathways exist to adapt process-based models for intercropping, progress is mainly constrained by the scarcity of field data suitable for parameterisation. If sufficient datasets were available, data-driven approaches could substantially improve both applicability and accuracy.

2.2 Data Processing

Generative Adversarial Networks (GANs) can capture and reproduce complex, high-dimensional data distributions more effectively than classical sampling techniques [ABV24], and have found broad application in synthetic data generation and imputation. Yoon et al. [YJvdS18] introduced the Generative Adversarial Imputation Net (GAIN), which focuses on imputing missing data components conditioned on observed values, utilising an additional 'hint vector' to force the generator to learn desired distributions. Arjovsky et al. [ACB17] proposed an alternative training algorithm, the Wasserstein-GAN (WGAN), which improved learning stability and addressed persistent GAN challenges such as mode collapse. For tabular data encompassing both continuous and discrete variables, Xu et al. [XSCV19] developed the Conditional Tabular GAN (CTGAN). By incorporating mode-specific normalisation and a conditional generator, CTGAN successfully addressed data imbalance and outperformed competing Bayesian networks [XSCV19]. Neves et

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

al. [NAN+22] subsequently extended these foundations by combining GAIN with WGAN principles to create robust variants (e.g. WSGAIN-CP and WSGAIN-GP) under a novel meta-method framework called 'tabulator', which demonstrated empirical superiority over CTGAN in generating synthetic tabular datasets.

However, the efficacy of GANs is fundamentally tied to dataset scale. They are notoriously unstable to train and demand a significant volume of data—often scaling from thousands to millions of observations [ABV24]. In the high-variance, low-quantity environment characteristic of agricultural field trials (frequently operating with $n < 1000$), these generative approaches are fundamentally risky. They face severe challenges in accurately generalising across diverse geographic regions, climates, and soil types without succumbing to mode collapse [ABV24]. Consequently, methods such as the otherwise superior 'tabulator' framework [NAN+22] become unfeasible for highly scarce settings.

Class-imbalance methods narrow the problem: rather than learning the full data distribution, they reshape the sampling density of what already exists. Chawla et al. [CBHK02] introduced SMOTE as a combination of synthetic minority over-sampling and random majority under-sampling, reporting stronger classifier behaviour in ROC space than majority under-sampling alone. Instead of duplicating minority observations, SMOTE creates new samples through interpolation between a minority point and its nearest minority neighbours, thereby broadening the minority decision region [CBHK02]. In parallel, the majority class is reduced by randomly removing observations until a defined class ratio is reached [CBHK02]. For time-series applications, however, Qu and Zhang [QZ24] highlight limitations of standard SMOTE, including sensitivity to manually chosen neighbourhood size, boundary blurring at edge points, and the mismatch between Euclidean distance and temporal similarity. Their M-SMOTE variant addresses these weaknesses by partitioning minority samples into safe and noise points, generating synthetic observations only through interpolation between safe points and the minority centroid, and replacing Euclidean neighbourhood matching with Dynamic Time Warping to better capture phase-shifted temporal patterns [QZ24].

Although M-SMOTE introduces useful safeguards against time-series imbalance, it is unclear whether it can be applied directly to sparse agronomic field trial datasets. In such settings, local feature-space interpolation is likely to be problematic due to the way observations are structured by crop type, seasonality, and management constraints. Therefore, synthetic points may become physically implausible and drift off, improving the apparent class balance without necessarily adding agronomically valid information. In addition, within the literature reviewed, no previous study was identified that successfully applied M-SMOTE under comparable conditions. These constraints demonstrate the limitations of pure data augmentation. When samples are too few and their structure is too complex, it is more productive to extend the observational record within field trials than to manufacture synthetic additions. Remote sensing offers the most practical means of extending the observational record by integrating canopy state variables into crop simulations [JKL+18]. Reviewing recent developments in this field, Jin et al. [JKL+18] categorise data assimilation into three primary approaches: forcing, updating, and cal-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

ibration. Forcing methods directly replace model simulated data with remote sensing observations. This inherently introduces raw sensor errors directly into the model dynamics. Updating methods continuously correct the simulation state when observational data are available, relying on the assumption that a more accurate current state improves subsequent predictions. While the updating method mitigates some error propagation, Jin et al. [JKL+18] argue that the calibration method is theoretically superior. This approach employs optimisation algorithms to iteratively adjust the initial model parameters until the difference between remote sensing data and simulated state variables is minimised. By implicitly relying on mathematical boundaries within seasonal dynamics, calibration effectively utilises structural knowledge to fill observational gaps. Although computationally demanding due to the high volume of required iterations, this flexibility systematically restricts observation errors from dictating the simulation, presenting a vital constraint strategy for environments hampered by data scarcity [JKL+18]. Jin et al. [JKL+18] identify the joint application of optimisation algorithms and sensitivity methods to crop model parameterisation as an open research gap. This thesis addresses exactly this gap.

Raw satellite time series are noisy and require preprocessing to separate true biological signals from sensor artifacts. Effective preprocessing requires reliable outlier detection. The challenge of defining what constitutes an outlier is complex, as distinguishing between regular and irregular behaviour is fundamentally tied to the subject area [FG20]. Filzmoser et al. [FG20] argue that reliable outlier detection in high-dimensional 'flat' data necessitates robust statistical estimators. Crucially, they emphasise that traditional procedures which discard entire rows due to a single faulty variable must be avoided as they discard existing information. This is an especially relevant point when sample sizes are small [FG20].

In a review of 290 papers on machine learning for anomaly detection, Nassif et al. [NTND21] found that unsupervised methods are the most widely adopted (27%) for identifying outliers. However, modern approaches increasingly explore hybrid architectures and deep learning to assess data consistency. For example, He et al. [HLJ25] introduced SADF-Net, a deep learning framework designed to fuse diverse data sources, including satellite imagery and IoT sensors. Within this framework, attention mechanisms function essentially as data-driven outlier detectors, actively suppressing noisy or incomplete inputs during the data fusion process [HLJ25].

In practice, time series reconstruction for crop modelling reduces to a two-step pipeline, consisting of statistical rejection of gross outliers, followed by smooth interpolation of the retained observations [TGMH18]. As a robust first step to isolate major observation errors, Torres et al. [TGMH18] bridge the gap between outlier detection and data smoothing by practically applying the Hampel identifier. This method relies on local estimations of the median absolute deviation to detect and replace spurious data points with a local median [TGMH18]. Once this initial statistical noise is isolated and removed, the remaining point-cloud of satellite data must be reconstructed into a continuous growth trajectory. For this, mathematical smoothing techniques such as the Savitzky-Golay

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

filter are applied. The Savitzky-Golay method performs a simplified least-squares-fit convolution, essentially acting as a weighted moving average with polynomial coefficients that preserve higher data moments while minimising filter bias [CJT+04]. Because noise, such as cloud cover, typically depresses vegetation index observations, Chen et al. [CJT+04] introduced an iterative adaptation of this filter. Under the premise that vegetation indices naturally follow an annual cycle, their algorithm first eliminates physically implausible index spikes—such as a random increase greater than 0.4 within 20 days—and applies Savitzky-Golay smoothing to establish a long-term trend. By assigning higher weights to observations nearing or exceeding this curve in a secondary refitting phase, the filter iteratively approaches the true upper vegetation envelope [CJT+04]. Leveraging similar smoothing principles, Pan et al. [PHC17] combined a baseline Savitzky-Golay filter with locally adaptive linear interpolation. This combination successfully generated smooth, phenology-preserving daily composites from the 30-metre resolution, two-day revisit cycle of HJ-1 A/B satellites [PHC17]. All three approaches, however, tune their parameters statistically: window sizes, weight schedules, and polynomial orders are chosen for smoothness, not derived from any model of crop physiology.

Imposing explicit biological shape on the trajectory, representing a growing season with defined onset, peak, and senescence, calls for parametric curve-fitting rather than smoothing. TIMESAT [JE04] formalised this approach and remains the reference framework for parametric vegetation time-series modelling. In addition to providing an adaptive Savitzky-Golay filter, TIMESAT introduced two ordinary least-squares methods that fit data to either harmonic or asymmetric Gaussian functions [JE04]. It is worth mentioning that these fits rely on a predefined uni-modal or bi-modal seasonality and assign higher weights to data points above the initial model function to construct the upper envelope of the trajectory [JE04].

Beck et al. [BAH+06] proposed applying a double logistic (double-sigmoid) function to uniquely describe yearly time series using six specific parameters: winter index, maximum index, and two inflection points with their corresponding rates of change. Similar to TIMESAT, this model utilizes an iterative weighting scheme that assigns lower weights to overestimated observations, operating on the assumption that anomalies are typically negatively biased [BAH+06]. Crucially, Beck et al. [BAH+06] demonstrated that this double logistic function captures trajectories outside and at the very end of the growing season better than asymmetric Gaussian functions.

Gobin et al. [GSC+23] successfully extended the application of double-sigmoid models to within-season fitting for various crops. They evaluated biophysical variables, specifically fAPAR and fCover, derived from high-resolution satellite sensors via neural networks. Finding that fAPAR exhibited less variability during critical transition stages such as crop maturity, they utilised it to describe phenological development [GSC+23]. Their work proves that simultaneously fitting upward and downward sigmoid models reliably defines inflection points that correspond directly to stem elongation and senescence, yielding results that have direct implications for improving crop yield forecasting and parameterising mechanistic models via data assimilation [GSC+23].

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Where curve-fitting extracts phenological milestones through fixed shape parameters, machine learning can learn the mapping from spectral time series to those milestones directly from labelled observations. As noted by Belgiu and Drăguţ [BD16], Random Forest (RF) classification has emerged as a leading non-parametric tool tailored for remote sensing tasks. The RF algorithm can handle high data dimensionality and multicollinearity while remaining fast and largely insensitive to overfitting due to its ensemble structure [BD16]. To process highly correlated multisource data effectively, an iterative backward feature elimination procedure, guided by the internal feature importance metric, is recommended to isolate the most relevant predictors without compromising accuracy [BD16].

The same properties that make RF effective for classification, such as tolerance for multicollinearity and low sensitivity to overfitting, carry over equally to regression tasks. Lu et al. [LZH+19] evaluated multiple ML algorithms against stepwise multiple linear regression to estimate aboveground biomass. They found that combining vegetation indices with canopy height metrics improved predictions, with RF yielding the highest accuracy among all tested methods [LZH+19]. Similarly, Wei et al. [WHM+17] used RF to estimate leaf area index (LAI) for winter oilseed rape using high-resolution optical satellite data, while Kumar et al. [KPG+18] employed RF alongside Sentinel-1A SAR data to predict winter wheat LAI, biomass, and plant height.

Beyond continuous parameter estimation, Kölbing et al. [NKM25] leveraged RF to predict discrete crop phenological BBCH stages directly from satellite observations, evaluating their model based on the temporal difference of occurrence in days [NKM25]. In the context of crop modelling, such accurate prediction of BBCH stages operates as a mechanism for "gating" subsequent growth models. To further refine these models by explicitly incorporating temporal dynamics, researchers can leverage the underlying continuous biological outcomes. Vicente-Guijalba et al. [VML15] presented a different approach by incorporating multi-temporal remote sensing data in a dynamical system. Interpreting crop phenology as a dynamic evolution responding to an unknown analytical model of environmental variables, they used state-space techniques to extract an evolution model in an initial stage. This model is subsequently exploited in a real-time estimation phase [VML15]. Such increment-based time series prediction allows the estimation to be continuously updated with each new satellite acquisition, eliminating the need to recalculate the entire temporal series off-line [VML15].

The common ground across all reviewed methods is that the biological structure of the growing season must be encoded explicitly. This can be achieved using a parametric curve shape, upper-envelope weighting, or mechanistic calibration constraints. The same principle is also evident in the literature from other domains. Embedding domain knowledge improves generalisation precisely because structural assumptions substitute for a large quantity of observations [**?**]. For cereal–legume intercrops, where field data is sparse, applying such a methodology is crucial for any reliable reconstructing of growth trajectories.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

2.3 Optimisation Techniques in Model Parameterisation

When parameterising complex process-based models, limited computational budgets often necessitate the use of surrogate models. These approximations mimic the behaviour of an expensive objective function while remaining computationally cheap to evaluate [SRLD24]. The most widely used of these is Bayesian Optimisation (BO). The procedure begins with an initial set of evaluated points to fit a Gaussian process prior probability distribution. As the optimisation progresses, the posterior distribution is updated with each new evaluation and an acquisition function selects the next candidate to evaluate [SRLD24].

Despite its efficiency in limited-budget scenarios, BO faces significant challenges when applied to crop modelling problems. Akhavizadegan et al. [AAW+21] highlight that BO is highly sensitive to the choice of the acquisition function, kernel types, and associated hyperparameters. Furthermore, the algorithm struggles with high dimensionality. As the number of variables increases typically beyond 15, the computational complexity required to cover the search space and optimise the acquisition function grows exponentially [SRLD24]. Time-dependent crop parameters, such as cultivar-specific traits with multi-year lifespans, require this dimensional expansion, making traditional parameter estimation exceedingly difficult [AAW+21]. An alternative for such noisy, non-convex, or deceptive search landscapes is the Covariance Matrix Adaptation Evolution Strategy (CMA-ES), a derivative-free method that iteratively adapts a multivariate normal distribution to sample candidate solutions [HO01, SRLD24]. However, for small dimensions and low evaluation budgets, standard BO generally outperforms CMA-ES [SRLD24].

Several algorithmic responses to this dimensionality bottleneck have been proposed, including variable selection, additive decompositions, low-dimensional embeddings, and trust region constriction [SRLD24]. Comprehensive benchmarking across the 24 BBOB test functions shows that trust-region variants consistently outperform the alternatives [SRLD24]. A prominent algorithm in this category is Trust Region Bayesian Optimisation (TuRBO), introduced by Eriksson et al. [EPG+19]. Instead of fitting a single global surrogate model, TuRBO utilises a collection of independent local models bounded within trust regions. A trust region is defined as a hyperrectangle centred around the best solution found thus far. TuRBO handles the trade-off between local exploitation and global exploration by maintaining multiple trust regions simultaneously and allocating the evaluation budget across them via a multi-armed bandit strategy [EPG+19]. Independent restarts allow the algorithm to escape local optima, accounting for its strong performance on high-dimensional non-convex problems [SRLD24].

Algorithm choice is only part of the problem: the objective function itself can encode agronomic prior knowledge to prevent the optimiser from exploring biologically implausible parameter combinations. Akhavizadegan et al. [AAW+21] integrated prior agronomic knowledge into a constrained optimisation framework, effectively preventing the algorithm from over-fitting to observed data. This hybrid approach significantly outperformed both standard BO and manual calibration, reducing prediction errors by 11.6% and 52.1%,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

respectively [AAW+21]. Similarly, Mitra et al. [MDPH18] successfully incorporated qualitative biological data into the objective function by converting general domain knowledge into inequality constraints on the model outputs. Because constraints derived from biological experiments inherently carry a degree of uncertainty, they propose the use of static penalty functions to treat them as soft constraints. This technique applies a cost proportional to the magnitude of the constraint violation, thereby converting the constrained problem into a scalar minimisation task suitable for metaheuristic algorithms [MDPH18].

TuRBO thus provides the algorithmic basis for the high-dimensional calibration pursued in this thesis. Encoding domain knowledge as constraints, as demonstrated above, prevents the optimiser from exploring biologically implausible parameter regions. This thesis therefore adapts the approach of Mitra et al. [MDPH18] by incorporating domain-informed inequality constraints directly into the TuRBO objective function.

2.4 Sensitivity Analysis and Explainability

Sensitivity analysis (SA) investigates how the outputs of a mathematical model depend on its inputs. The models themselves simulate real-world systems, and their inputs—frequently termed factors—encompass a wide variety of elements, including model parameters, initial and boundary conditions, or structural assumptions. Similarly, outputs can range from spatio-temporal predictions to objective and error functions [RJS+21]. Central to SA is a practical heuristic:

> In general, and regardless of any specific purpose, SA aims to exploit the 'sparsity of factors' principle, a heuristic stating that, very often, only a small subset of factors in a system have a significant influence on a specific system output [RJS+21].

The modern era of this field is predominantly focused on Global Sensitivity Analysis (GSA), which can be broadly categorised into four methodological approaches [RJS+21]. **Derivative-based** methods extend classical local sensitivity analysis by evaluating numerical permutations at multiple base points across the factor space. In contrast, **distribution-based** methods, such as the variance-based framework formalised by Sobol' in 1993, decompose the total variance of the model output and attribute it to individual inputs or groups of inputs. A more recent advancement is the **variogram-based** approach, which bridges derivative and variance methodologies by recognising that model outputs possess a spatially ordered covariance structure. By characterising changes across a comprehensive range of perturbation scales, variogram-based methods offer high computational efficiency, particularly by extracting information from all pairs of sample points [RJS+21]. Finally, **regression-based** approaches infer sensitivity from fitted models. While early linear applications were often criticised for their strict structural assumptions, this paradigm has been strongly revitalised by the machine

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

learning (ML) community. Modern regression-based SA focuses on assessing feature importance, frequently by evaluating how prediction quality degrades when specific inputs are permuted [RJS+21].

Two practical obstacles cut across all four methodological families. The first is the near-universal assumption of input independence: when parameters are perturbed in isolation despite being structurally coupled in the model, attributed sensitivities can be seriously misleading [RJS+21]. The second is scale: nearly 70% of SA studies in environmental modelling are confined to 20 or fewer factors, and evaluation budgets are routinely set on an ad hoc basis rather than by formal convergence criteria [RJS+21].

These computational and methodological limitations are illustrated by recent applications of variance-based GSA to process-based crop models. Both Hao et al. [HRW+24] and Shen et al. [SZZ+25] explicitly rejected local sensitivity analysis in favour of the Sobol' method, arguing that variance-based approaches are essential for capturing complex parameter interactions without relying on structural assumptions. To comprehensively evaluate the APSIM model, Hao et al. [HRW+24] constructed 18 distinct scenarios combining varying weather conditions, soil types, and nitrogen applications. Similarly, Shen et al. [SZZ+25] tested the STICS model across five field experiments and four engineered nitrogen-water stress configurations. In doing so, Shen et al. [SZZ+25] deliberately excluded initialisation parameters to prevent them from obscuring the relative importance of crop and soil traits, sampling the remaining variables from uniform distributions with a ±15% perturbation around their baselines.

However, the inherent mechanics of the Sobol' approach impose severe computational demands on these experimental designs. For an initial sample size of $n = 10,000$ and $k = 6$ analysed parameters, the standard sampling formula $(n \cdot (k + 2))$ yields 80,000 model realisations per scenario [HRW+24]. To mitigate this computational bottleneck across their scenario matrices, both studies were forced to evaluate parameter sensitivity using single representative statistical years [HRW+24, SZZ+25]. Consequently, faster-converging alternatives like the Fourier Amplitude Sensitivity Test (FAST) have been recommended to optimise computational efficiency in future work [HRW+24].

Despite these structural compromises, the agronomic insights from both analyses are highly consistent. Under rainfed or stressed conditions, parameters controlling water availability—such as the coefficient for water requirements in STICS or maximum root depth in APSIM—act as the overwhelming drivers of yield variance [HRW+24, SZZ+25]. Therefore, future data assimilation strategies must primarily focus on addressing uncertainty in water supply when simulating rainfed systems [HRW+24]. It is only when baseline water requirements are fully met that nitrogen limitations and photosynthetic parameters dominate the simulated outcomes [HRW+24, SZZ+25].

Shen et al. [SZZ+25] highlight a different issue when it comes to utilising the knowledge of which parameters matter for calibration. While parameterising highly sensitive factors from end-of-season yield data alone is a widely adopted approach that successfully ensures accurate final predictions, it carries an inherent risk of generating physically unrealistic

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

intermediate biomass and leaf area trajectories if left entirely unconstrained.

The same variability analysis that identifies influential factors in a SA also underpins uncertainty quantification and model explainability [RJS+21]. In both mechanistic and data-driven settings, recent work has shown that explainability methods are most effective when they are grounded in sensitivity concepts and when their fidelity to the underlying model is explicitly assessed. The SAGE project, for example, demonstrates that common explanation approaches can become unreliable when they rely on strong independence or linearity assumptions, since these may generate out-of-distribution perturbations and fail to capture nonlinear feature interactions [SAA+21]. From this perspective, SA is not only a tool for ranking inputs, but also a means of evaluating whether an explanation is faithful, whether a prediction is sufficiently supported by the available data, and whether the resulting interpretation can be trusted in practice [RJS+21, SAA+21].

This line of reasoning points to two concrete methodological requirements. First, the computational bottleneck of Sobol'-type variance decomposition—which, as demonstrated above, demands tens of thousands of model evaluations even for modest parameter sets—must be overcome. Second, and more fundamentally, the response surface of a process-based crop model is inherently nonlinear: the influence of a parameter may emerge at a particular magnitude of perturbation and be entirely absent at another, so methods that assess sensitivity at a single, fixed scale inevitably provide only a partial picture. A suitable approach must therefore characterise output variability globally, across the full spectrum of perturbation scales, while remaining computationally tractable for high-dimensional parameter spaces. Variogram-based sensitivity analysis meets both requirements, combining scale-resolved characterisation of the response surface with high sampling efficiency.

Razavi and Gupta [RG16] proposed the Variogram Analysis of Response Surfaces (VARS), a framework that characterises the model response surface by integrating directional variograms over the entire input space and across a full spectrum of perturbation scales. In doing so, VARS simultaneously delivers derivative-based, variance-based, and integrated variogram sensitivity metrics as inherent products of a single analysis, with the IVARS-50 index—the integrated variogram evaluated at half the input range—serving as the primary measure of global sensitivity. To make this theoretical framework practically applicable, Razavi and Gupta [RG16] developed STAR-VARS, a star-based sampling strategy in which Latin Hypercube Sampling generates a set of star centres, each radiating cross-sectional lines across the parameter space. This design enables VARS to fully characterise global sensitivity with substantially fewer model evaluations: applied to hydrological models ranging from 5 to 45 parameters, STAR-VARS produced reliable and stable factor rankings while requiring one to two orders of magnitude fewer model runs than either the Morris or Sobol' method [RG16]. Furthermore, bootstrap resampling provides confidence intervals on the resulting sensitivity indices, allowing the reliability of factor rankings to be assessed without additional model evaluations.

Building on this foundation, Panigrahi et al. [PRD+25] introduced XVARS, an extension of VARS that additionally incorporates explainability for machine learning models.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

XVARS addresses the limitation of assumed input independence by explicitly accounting for the complex, multivariate distributional properties inherent to environmental data [PRD+25]. Rather than relying on assumed parametric forms, it infers empirical marginal distributions directly from observed data samples, enabling a global assessment of the input–output response surface of an ML model. Technically, XVARS builds upon the Generalised VARS (G-VARS) extension, which maps the input space onto a standard normal space via the Nataf Isoprobabilistic Transformation to account for pairwise correlations. XVARS then replaces these parametric marginals with empirical distributions approximated from available data samples [PRD+25]. Applying this framework to water quality prediction models revealed that subtle alterations in model design, such as changes to the random seed or hyperparameters, can lead to entirely different interpretations of feature importance [PRD+25]. While XVARS is a powerful tool for post-hoc explainability of data-driven models, its empirical distribution approach requires observed joint input–output samples from which the input distribution can be inferred. In the sensitivity analysis presented in this thesis, however, inputs are treated as independent with predefined bounded uniform priors over a controlled parameter space, for which no observed joint distribution exists. Under these conditions, the added complexity of XVARS is unnecessary, and the simpler VARS framework is the more appropriate choice.

Panigrahi et al. [PRD+26] subsequently frame SA and XAI as complementary rather than competing approaches to system understanding. SA provides the global foundation by characterising how changes in the inputs propagate through the response surface, whereas XAI extends this understanding towards the reasons for model predictions. XVARS implements this connection by transferring a global sensitivity framework into model-agnostic post-hoc explanation. Its additional contribution to transparency, compared with traditional SA, is therefore not another parameter ranking alone, but an explanation that relates feature importance to the model's behaviour across the full input–output space. In a systematic comparison across connectionist, kernel, and tree-based ML models, XVARS required less than 10% of the computational time of Kernel SHAP and Tree SHAP for typical problems, and its computational demand was essentially independent of training sample size [PRD+26]. It was also less sensitive to sampling variability and produced more stable feature-importance estimates because it characterised the response surface globally, while the SHAP approaches concentrated on local neighbourhoods around observed samples [PRD+26]. This comparison shows that SA can be extended to meet XAI requirements when the global and scale-dependent information supplied by SA is retained, while the resulting importance measures are used to make the behaviour of individual model predictions more transparent.

Sensitivity analysis is essential for understanding complex simulators, but its application is frequently hindered by prohibitive computational costs and restrictive assumptions. Variance-based methods such as Sobol' demand tens of thousands of model evaluations, while more recent frameworks such as XVARS, although powerful, require empirically observed joint input distributions that are unavailable when inputs follow predefined uniform priors. VARS, via the STAR-VARS sampling strategy, bridges this gap by

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

delivering comprehensive multi-scale sensitivity assessments at a fraction of the computational cost. The IVARS indices not only rank parameter importance but also reveal the perturbation scale at which individual parameters exert their influence, providing the structural interpretability needed to validate the calibrated *InterCrop* parameters.

This section therefore settles the explainability half of RQ3 at the methodological level. The comparison between sensitivity analysis and post-hoc explanation belongs here rather than in a second empirical analysis, and its purpose was to establish what an explanation of a crop model would have to deliver. What the XAI literature contributes to the present design is consequently not a second attribution algorithm but two requirements on the explanation itself. Importance must be resolved conditionally, so that a factor can be described as gaining or losing influence when the environmental context changes, and it must be resolved across perturbation scales, so that effects appearing only over part of a parameter range are not concealed by a single global index. This thesis adopts the VARS methodology and organises it along these two requirements, thereby addressing how sensitivity analysis and explainability techniques can together improve model transparency, validate parameter constraints, and increase confidence in the simulated results.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 3

The InterCrop Model

The *InterCrop* model, introduced in Section 1.2, extends the SSM-iCrop framework for monocultures to simulate intercropping systems. It accounts for both competition and facilitation processes between the component species, including the partitioning of above-ground light and below-ground soil resources such as water and nitrogen [MHAN25]. Because *InterCrop* inherits its core process representations from SSM-iCrop, an understanding of the underlying single-crop model is essential. The following sections therefore first describe the key plant- and soil-related processes simulated by SSM-iCrop before presenting the conceptual extensions introduced for intercropping. For the scope of this thesis a general understanding of the underlying processes is sufficient; for detailed mathematical formulations the reader is referred to Soltani et al. [SMS13] and Manschadi et al. [MPF+22].

3.1 Plant-Related Processes

SSM-iCrop simulates phenological development, leaf development and senescence, crop mass production and partitioning, plant nitrogen balance, yield formation and soil water and nitrogen balances of individual grain crops [SMS13]. With the exception of phenology, which may employ species-specific parameters, these sub-models are shared across all grain crop species [MEB+21]. The following description uses wheat as an illustrative example, as it was the species for which SSM-iCrop was originally developed.

Phenology. The model tracks a sequence of developmental stages—emergence, vegetative growth, anthesis, grain filling, physiological maturity, and harvest—using the *biological day* (BD) concept [SMS13, MPF+22]. A biological day represents a day under optimal conditions for plant development. Each day from sowing to maturity, the BD value is computed by a function capturing the effects of temperature, photoperiod, vernalisation, and water deficit. These daily values are accumulated into cumulative

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

biological days (CBD), and the CBD thresholds required to reach successive phenological stages are cultivar-specific input parameters [MPF+22].

Leaf Area Development and Senescence. Canopy development proceeds through several phases. Early leaf area expansion is driven by the phyllochron—the thermal time interval between successive leaf emergences on the main stem—and plant leaf area is derived from the main-stem leaf number, adjusted for water and nitrogen availability [SMS13, MPF+22]. After main-stem leaf production ceases, leaf area may continue to expand on tillers until the beginning of seed growth, at which point LAI development stops. The model also accounts for frost damage and nitrogen-limited leaf expansion [SMS13].

Crop Mass Production and Partitioning. Daily above-ground dry mass accumulation is driven by the radiation use efficiency (RUE) concept: available photosynthetic active radiation (PAR) is multiplied by the fraction intercepted by the crop (FINT) and the crop's efficiency in converting that intercepted PAR into biomass, with adjustments for temperature, water deficit, and atmospheric CO_2 concentration [SMS13, MPF+22]. This daily production is partitioned among three sinks—leaves, stems, and grains—using cultivar-specific partitioning coefficients that vary with the developmental stage [MPF+22].

Yield Formation. Yield formation is modelled through a linear increase in harvest index, where the potential rate of increase is a cultivar-specific input parameter. The actual daily seed growth rate is constrained by current photosynthetic production plus dry matter translocated from vegetative organs when production alone is insufficient [SMS13, MPF+22].

Plant Nitrogen Balance. The plant nitrogen balance is simulated differently before and after the onset of seed growth [SMS13, MPF+22]. During the vegetative phase, nitrogen demand is driven by leaf area and stem growth. When uptake falls short, a cascade of responses follows: stem nitrogen concentration is reduced, then leaf expansion is inhibited, and under extreme limitation existing leaves are senesced to remobilise nitrogen. Once seed growth begins, grains become the primary nitrogen sink; if soil uptake is insufficient, nitrogen is translocated from vegetative tissues [MPF+22].

3.2 Soil-Related Processes

In addition to the plant-level sub-models described above, SSM-iCrop simulates the soil environment in multiple layers using a one-dimensional approach [SMS13, MPF+22].

Soil Water Balance. Soil water dynamics are modelled using a cascade method. The model tracks water additions from rainfall and irrigation as well as water removal through

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

run-off, soil evaporation, crop transpiration, and drainage. Each soil layer is characterised by volumetric water content at saturation, the drained upper limit, the crop lower limit, and the air-dried value. A fraction of water above the drained upper limit is allowed to drain from each layer per day. The rate of crop dry mass production is used to calculate daily transpiration using a transpiration efficiency coefficient and vapour pressure deficit [MPF+22]. The model uses the fraction of available soil water in the root zone to derive three water-deficit factors, which are then used to adjust dry mass production, leaf area expansion, and phenological development under water-limited conditions [MPF+22].

Root growth is simulated through a potential daily increase in root depth, scaled by the biological day value. Root penetration ceases when daily dry mass production is zero, the current soil layer is dry, or after the beginning of seed growth [MPF+22].

Soil Nitrogen Balance. Analogous to the water balance, soil nitrogen is simulated layer by layer as the net result of processes that add nitrogen to the soil—mineralisation of organic matter and fertilisation—and those that remove it, namely volatilisation, denitrification, leaching, and crop nitrogen uptake [SMS13, MPF+22]. The daily rate of crop nitrogen uptake is determined as the minimum of crop demand and the potential supply from the soil, subject to a maximum uptake rate. Uptake from each layer depends on the fraction of the layer explored by roots and the proportion of available nitrogen in that layer relative to the total in the root zone [MPF+22].

3.3 Extensions for Intercropping

In the *InterCrop* model, the single-crop sub-models described above are executed simultaneously for two species, and their resource acquisition is mutually adjusted at every daily time step [MHAN25]. *InterCrop* does not alter the crop parameters or the underlying physiological formalisms of the individual species compared to the baseline SSM-iCrop model. Instead, it simulates the intercrop interactions on top of this established single-crop foundation. The following paragraphs describe how above-ground light partitioning and below-ground competition for water and nitrogen are handled.

Light Interception and Plant Height. Light competition is governed by the planting configuration (e.g. row number and spacing), the LAI, and the plant height of both species. Each day the model classifies the two species as tall and short and computes the fraction of intercepted radiation for each, which then drives potential daily dry matter production and transpiration demand [MHAN25]. The canopy of the taller species is subdivided into an upper portion that extends above the shorter species and a lower portion whose height equals that of the shorter species, allowing radiation to be partitioned accordingly [MHAN25]. Daily height increase for each species follows a logistic function of cumulative biological days, the species-specific maximum height, and a potential height growth rate. Height growth terminates at species-specific developmental stages—at silking for maize,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

at the beginning of seed growth for legumes, and at flowering for wheat and potato [MHAN25].

Water Uptake. Each species takes up water based on its transpirational demand and rooting depth, and the combined uptake is used together with the other water-balance components (Section 3.2) to update the water content of each soil layer [MHAN25]. Available water in a layer is then allocated to each species using root depth proportionality factors: when both species have fully rooted a layer, each receives half of the available water; when only one species is present, it receives the full amount. The intercropped species also jointly cover a greater proportion of the soil surface, thereby reducing soil evaporation relative to monocultures [MHAN25].

Nitrogen Uptake. Whereas soil water may redistribute horizontally in response to differential demand, soluble nitrogen in the soil solution is tracked separately for each species [MHAN25]. This crop-specific accounting permits differentiated fertilisation of the two components. Nitrogen inputs and outputs are computed individually based on each species' root depth and the previous day's soluble nitrogen, and available nitrogen in each layer is allocated using the same root depth proportionality factors employed for water [MHAN25]. At the boundary rows between the two intercrop strips, intra-specific competition gives way to inter-specific competition; in cereal–legume intercrops, biologically fixed nitrogen can facilitate cereal growth, particularly in these boundary rows [MHAN25].

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 4

Data and Study Area

4.1 Field Trials

The field trial data used in this thesis were compiled by Gaudio et al. [GMB+24] into a global dataset aggregating results from 37 field trials conducted across five European countries—France, Denmark, Italy, Germany, and England—between 2001 and 2017, as shown in Figure 4.1. The trials covered cereal–legume intercropping systems alongside their respective sole crops, comprising five legume species (chickpea, faba bean, lentil, lupin, and field pea) and three cereal species (barley, durum wheat, and soft wheat). In total, the dataset contains 299 sole crop and 308 intercrop plots, where each corresponds to a unique combination of site, year, and crop management, including species and cultivar selection as well as agricultural interventions such as sowing conditions and fertilisation levels [GMB+24]. The dataset is publicly available at `https://doi.org/10.5281/zenodo.14160885`. Climate variables for all trials were sourced from the NASA POWER API [GMB+24].

Across the 37 trials, a total of 26 plant variables were recorded, though not all were measured in every trial [GMB+24]. For the parameterisation carried out in this thesis, seven of these variables were selected on the basis of their availability and agronomic relevance to intercropping systems (Table 4.1). Grain biomass (WGRN) and grain nitrogen content (NGRN) constitute the primary outputs of interest for practitioners and were therefore included despite being measurable only at the end of the season. To represent in-season crop dynamics, above-ground biomass (WTOP), cumulative nitrogen uptake (CNUP), and phenological stage (BBCH) were selected, as they collectively capture the main trajectories of biomass accumulation, nitrogen dynamics, and developmental progress. Plant height (PLHGT) and leaf area index (LAI) were also included, given their particular importance for the light partitioning dynamics specific to intercropping systems, as described in Section 3.3. The number of trials reporting each variable is summarised in Table 4.1.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

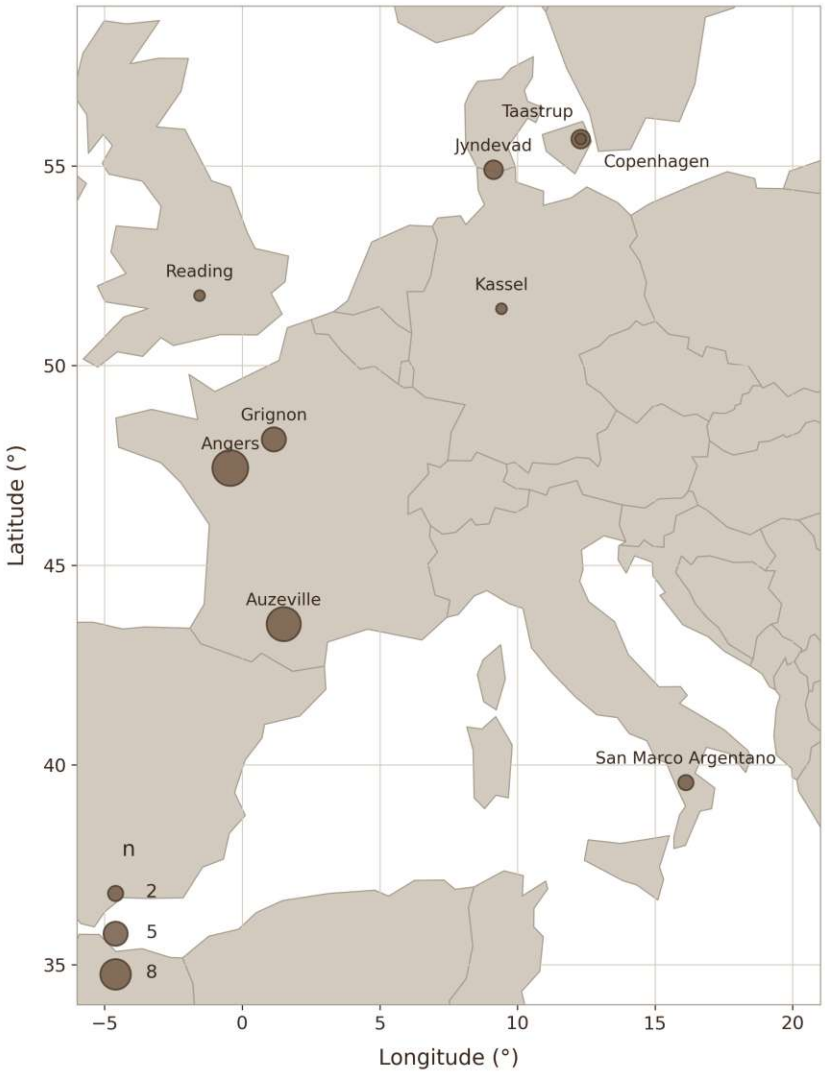


Figure 4.1: Geographic distribution of the 37 field trial sites across Europe. Circle size is proportional to the number of trials at each location. Adapted from [GMB+24].

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Table 4.1: Number of trials in the full dataset of Gaudio et al. [GMB+24] that reported each of the seven selected plant variables.

Variable	InterCrop name	Unit	No. trials
Grain biomass	WGRN	$g\,m^{-2}$	37
Above-ground biomass	WTOP	$g\,m^{-2}$	37
Phenological stage	BBCH	–	35
Plant height	PLHGT	cm	31
Cumulative nitrogen uptake	CNUP	$g\,N\,m^{-2}$	30
Grain nitrogen content	NGRN	$g\,N\,m^{-2}$	24
Leaf area index	LAI	$m^2\,m^{-2}$	18

Prior to this work, *InterCrop* was parameterised manually by Manschadi et al. [MHPN26] for two of the eight available intercrop combinations: durum wheat–faba bean (*Triticum turgidum–Vicia faba*) and spring barley–field pea (*Hordeum vulgare–Pisum sativum*). For this manual parameterisation, the following cultivars were used.

- **Durum wheat**: L1823, Nefer, Sculptur
- **Faba bean**: Irena, Castel, Diver, Nordica
- **Spring barley**: Scarlett, Lysiba, Otira
- **Field pea**: Baccara, Bohatyr, Agadir

Besides these cultivars, the dataset also contained four additional durum wheat cultivars (Acalou, Isildur, Neodur, Orjaune), one further faba bean cultivar (Columbo), and 13 additional field pea cultivars. To assess whether these could be incorporated into the parameterisation, the length of the growing period was examined for each cultivar, with the distributions visualised in Figure 4.2. Columbo and the 13 field pea cultivars were excluded from this study because their growing period lengths differed substantially from the above, suggesting they represent distinct physiological types whose inclusion would introduce noise rather than useful information. The four additional durum wheat cultivars, in contrast, had comparable growing periods and were therefore retained, as they can provide additional data points for the modelling process without compromising consistency.

After applying the cultivar filter, the working dataset comprised 167 sole crop and 83 intercrop plots out of the original 299 and 308, respectively. Table 4.2 provides a breakdown by crop group.

Within this filtered dataset of 250 plots, the availability of observations varies by variable, as depicted in Figure 4.3. WTOP, CNUP, and BBCH were recorded in nearly all plots. PLHGT and LAI were less consistently measured, with 189 and 154 plots reporting them respectively. However, PLHGT stands out for having by far the highest number of

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

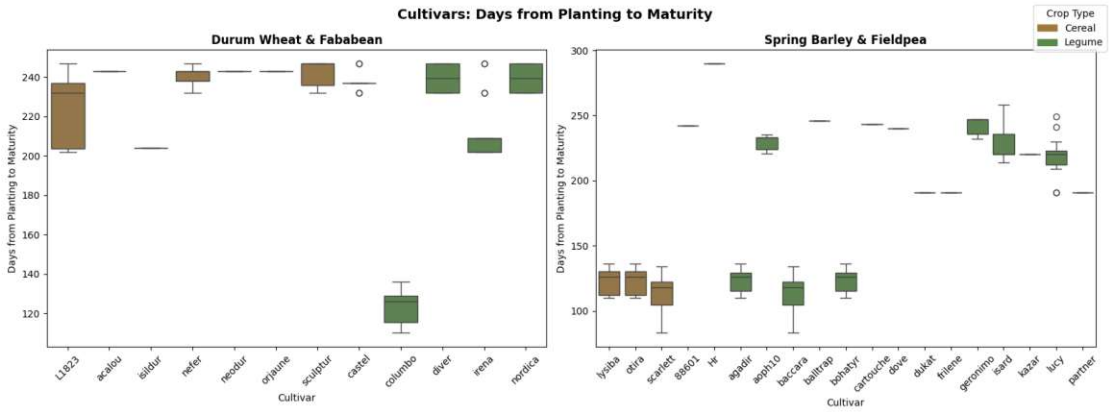


Figure 4.2: Boxplots of growing period length (days from sowing to harvest) for each cultivar. Left: durum wheat and faba bean; right: spring barley and field pea. Cultivars with markedly different growing period distributions from those used in the manual parameterisation were excluded from the dataset.

Table 4.2: Number of unique plots per crop group in the filtered dataset.

Crop group	**Unique plots**
Durum wheat	73
Faba bean	38
Durum wheat × faba bean	42
Spring barley	34
Field pea	22
Spring barley × field pea	41
Total	**250**

individual observations per plot. The end-of-season variable WGRN was available from all 250, while grain nitrogen content NGRN was reported from 188 plots.

4.2 InterCrop Variables

The *InterCrop* model requires data across four main domains to simulate crop growth and resource competition: soil conditions, weather data, management practices, and crop-specific parameters. For the observed data [GMB+24], both soil and weather measurements are included within the dataset and were similarly employed during the manual parameterisation process. To provide a clearer understanding of the inputs driving the simulations throughout this study, the management data and phenology parameters are detailed in the following.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

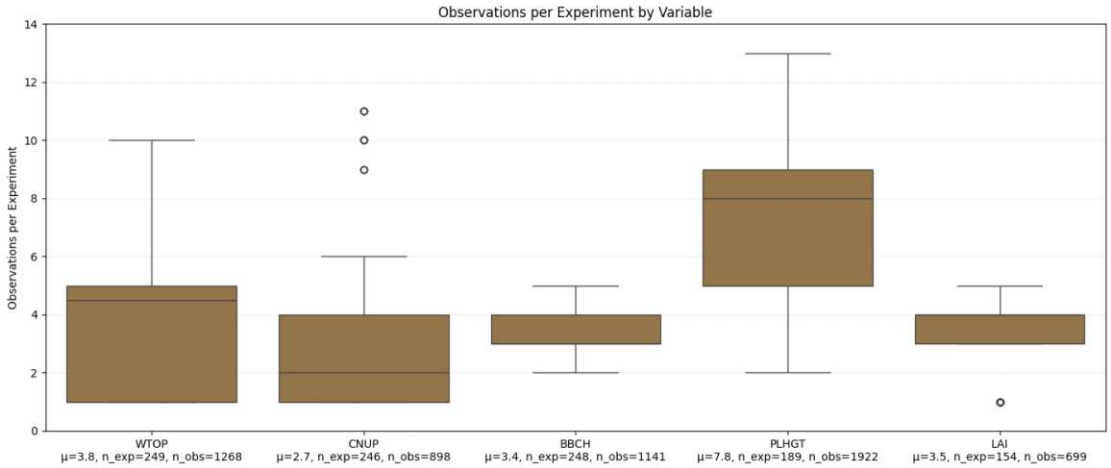


Figure 4.3: Distribution of the number of observations per plot for each of the seven selected variables in the filtered dataset. Boxes show the interquartile range; annotations report the mean number of observations and the total count of plots with at least one observation.

4.2.1 Management Data

Management data define the agricultural interventions applied to each plot and establish the competitive environment. In the dataset, each row corresponds to a single component species within a given plot. Intercrop treatments are represented by two paired rows detailing the respective crop partners. The key management variables extracted from the global dataset include:

- **experiment_id**: A unique identifier for the field trial.
- **crop_type**: A binary indicator distinguishing between intercrop and monocrop setups.
- **species**: The crop species (e.g., faba bean or durum wheat).
- **cultivar_mix**: The specific combination of cultivars used (e.g., nefer_castel), conventionally listing the cereal first.
- **sowing_date**: The date on which the seeds were sown.
- **density_factor**: The actual plant density expressed relative to the locally recommended full density of monocrops (where a value of 1 equals full density) in plants per square metre.
- **interrow**: The distance between two planting rows, measured in centimetres.
- **mixing_pattern**: The spatial arrangement of the plants, categorised as either fully mixed (MX) or planted in alternating rows (AR).

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

- **code_row**: For alternating-row setups, this indicates the specific row proportion for each species (e.g., *2_1* designates two rows of cereal alongside one row of legume). From this, the effective row space (ROWSP) and inter-row space (IROWSP) parameters are derived.
- **N_amount** and **N_type**: The total amount of nitrogen applied in kilogrammes per hectare and the form of the input (e.g., mineral, urea).
- **N_dose_x** and **N_date_x**: The specific quantity and timing of individual nitrogen split applications, where x reaches up to the total number of applications.
- **P_amount**: The total applied phosphorus in kilogrammes per hectare.
- **irrigation_amount**: The cumulative volume of supplemental water provided via irrigation, recorded in millimetres.

4.2.2 Phenology Parameters

While management data are experimentally observed inputs, phenological parameters act as internal model controls dictating crop development and physiological traits. These parameters are crucial when the model is being parameterised for specific cultivars. *InterCrop* features approximately 60 adjustable crop parameters; however, only a small subset of these parameters is adjusted for a single cultivar during parameterisation. For the optimisation carried out in this thesis, this subset was selected based on the parameters' general importance to the model's overall functionality and their specific role in capturing interspecific interactions, such as light perception and nitrogen and water uptake (as described in Chapter 3).

Table 4.3 summarises the parameters targeted during the optimisation phase (highlighted in bold). The table also includes closely associated parameters that were not part of the active optimisation space but provide context for the complete developmental cycle and morphological limits.

4.3 Satellite Data

The field trial observations were supplemented with remotely sensed optical imagery from the Landsat programme, covering Landsat 5 (TM), Landsat 7 (ETM+), and Landsat 8 (OLI). Landsat was selected because its historical archive spans the full 2001–2017 period of the field trials, as outlined in Section 4.1. Across these different sensor platforms, six harmonised spectral reflectance bands were extracted: the visible blue, green, and red bands, along with near-infrared (NIR) and two short-wave infrared bands (SWIR1 and SWIR2).

From these bands, ten vegetation indices were calculated and simultaneously used in the estimation of key crop variables: BBCH, WTOP, PLHGT, CNUP, and LAI. The

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Table 4.3: Overview of selected *InterCrop* parameters relevant for the parameterisation process. Parameters actively adjusted during the optimisation are indicated in **bold**, whereas the remaining parameters provide context or set upper boundaries.

Parameter	Unit	Meaning	Notes
GNCmin	g/g	Minimum grain nitrogen concentration	Crop quality classifications
GNCmax	g/g	Maximum grain nitrogen concentration	Crop quality classifications
heightMx	cm	Maximum plant height	Intercrop light perception
rgrHgt	–	Height growth rate	
MEED	mm	Maximum effective depth of water extraction by roots	Intercrop water and N uptake
GRTDP	mm/bd	Root depth growth rate	
bdSOWEMR	bd	Sowing to emergence	
bdEMRR1	bd	Emergence to anthesis	Legume specific
bdR1R3	bd	Anthesis to early pod	Legume specific
bdR3R5	bd	Early pod to beginning of seed growth	Legume specific
bdR5R7	bd	Beginning to termination of seed growth	Legume specific
bdR7R8	bd	Termination of seed growth to full maturity	Legume specific
bdTILSEL	bd	Tillering to stem elongation	Cereal specific
bdSELBOT	bd	Stem elongation to booting	Cereal specific
bdBOTEAR	bd	Booting to ear emergence	Cereal specific
bdEARANT	bd	Ear emergence to anthesis	Cereal specific
bdANTPM	bd	Anthesis to physiological maturity	Cereal specific
bdPMHM	bd	Physiological maturity to harvest maturity	Cereal specific

collective use of these indices is justified by their complementary strengths in capturing distinct physiological, biochemical, and structural traits (Table 4.4).

For instance, widely used metrics such as the Normalized Difference Vegetation Index (NDVI) and Green NDVI (gNDVI) are highly sensitive at lower LAIs but typically saturate at medium to high levels. However, using high-resolution RapidEye data this saturation effect becomes apparent only when the LAI exceeds approximately 6 m^2/m^2, compared to a threshold of 4 m^2/m^2 when using lower-resolution data [KML+15]. To counteract these saturation effects, standard indices are commonly supplemented with alternatives that remain sensitive at higher LAIs. MTVI2, for instance, shows reliable performance with in situ observations at higher LAIs. When evaluated against total biomass rather than LAI, however, MTVI2, NDVI, and gNDVI all exhibit similar saturation effects once biomass exceeds approximately 800 g/m^2 [KML+15]. The Simple Ratio (SR) maintains a continuous linear response to leaf biomass across its entire range and continues to increase

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

even after LAI has plateaued, making it a useful indicator of both canopy structure and chlorophyll concentration [KML+15].

The Normalized Difference Water Index (NDWI) and Moisture Stress Index (MSI) were included to track canopy moisture status. Four further indices specifically evaluated for nitrogen variability were also added: the Chlorophyll Index (CI), Chlorophyll Vegetation Index (CVI), Enhanced Vegetation Index (EVI), and the Modified Chlorophyll Absorption Ratio Index 2 (MCARI2) [PDLF+22].

Table 4.4: Summary of the calculated vegetation indices and their corresponding formulas. The cited secondary literature outlines the rationale for their selection, containing references to the foundational primary studies that introduced these formulas.

Index	**Full Name**	**Formula**	**Reference**
gNDVI	Green NDVI	$\frac{\text{NIR}-\text{Green}}{\text{NIR}+\text{Green}}$	
NDVI	Normalized Difference Vegetation Index	$\frac{\text{NIR}-\text{Red}}{\text{NIR}+\text{Red}}$	
SR	Simple Ratio	$\frac{\text{NIR}}{\text{Red}}$	[KML+15]
MTVI2	Modified Triangular Vegetation Index 2	$\frac{1.5(1.2(\text{NIR}-\text{Green})-2.5(\text{Red}-\text{Green}))}{\sqrt{(2\text{NIR}+1)^2-(6\text{NIR}-5\sqrt{\text{Red}})-0.5}}$	
NDWI	Normalized Difference Water Index	$\frac{\text{NIR}-\text{SWIR1}}{\text{NIR}+\text{SWIR1}}$	
MSI	Moisture Stress Index	$\frac{\text{SWIR1}}{\text{NIR}}$	
CI	Chlorophyll Index	$\frac{\text{NIR}}{\text{Green}}-1$	
CVI	Chlorophyll Vegetation Index	$\text{NIR}\cdot\left(\frac{\text{Red}}{\text{Green}^2}\right)$	[PDLF+22]
EVI	Enhanced Vegetation Index	$2.5\frac{\text{NIR}-\text{Red}}{\text{NIR}+6\text{Red}-7.5\text{Blue}+1}$	
MCARI2	Modified Chlorophyll Absorption Ratio Index 2	$\frac{1.5(2.5(\text{NIR}-\text{Red})-1.3(\text{NIR}-\text{Green}))}{\sqrt{(2\text{NIR}+1)^2-(6\text{NIR}-5\sqrt{\text{Red}})-0.5}}$	

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 5

Methodology

5.1 Data Processing

5.1.1 Data Acquisition

The analysis was based on a sparse dataset, comprising 250 experimental plots distributed across 23 distinct sites. The cumulative sample volume ranged from 699 to 1922 observations per variable, with an average of 4.12 temporal observations per variable per plot. This limited data density presented a substantial challenge, as the primary objective of this study is the parameterisation of the process-based *InterCrop* model through data-driven techniques—such as Bayesian optimisation. To overcome this limitation, the dataset was synthetically enriched through data assimilation using remote sensing data, a method identified as highly effective in the literature review.

Since the experimental sites were identified only by latitude and longitude coordinates without accompanying field polygons, satellite data extraction required manual field delineation. To avoid potential uncertainty regarding the influence of mixed pixels and edge effects, field boundaries were delineated conservatively to ensure that the resulting polygons were entirely contained within the actual plot boundaries, thereby minimising spectral contamination from adjacent land cover. Additionally, a single set of remote sensing data per site was applied to all plots conducted at that location within the same season. Consequently, the satellite signal represents general site conditions rather than plot-specific management impacts, such as varying nitrogen rates or planting densities. This approach is consistent with the soil and weather data in the observed dataset, which are likewise recorded on a site-specific basis. By providing the model with plot-specific management data as additional features, it can distinguish variations based on management expectations, even when the spectral and environmental inputs remain uniform across the site.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Following delineation, multispectral data were retrieved encompassing six spectral bands and ten derived vegetation indices (see Section 4.3). To cover the experimental period from 2001 to 2017, data from a combination of Landsat satellites was used, maintaining a spatial resolution of 30 m and a nominal revisit time of 16 days. Landsat 5 (TM) provided data before 2013, and Landsat 8 (OLI) was used from February 2013 onward. Landsat 7 (ETM+) served mainly to fill gaps between these missions, as its data quality was reduced by the Scan Line Corrector (SLC-off) error after 2003 [PJO22].

The satellite data include all available images from sowing to harvest, incorporating both Tier 1 and Tier 2 products. Surface reflectance values were standardised by applying a scaling factor of 0.0000275 and an offset of -0.2 [U.S21]. To ensure data quality, a clear-sky mask was generated using the Landsat QA_Bits. Due to the updated sensor design of Landsat 8 (OLI), a specific band mapping was applied to maintain consistency with the earlier TM and ETM+ sensors, as detailed in Table 5.1.

Table 5.1: Spectral band correspondence across Landsat sensor generations (TM, ETM+, and OLI).

Spectral Band	**Landsat 5 (TM)**	**Landsat 7 (ETM+)**	**Landsat 8 (OLI)**
Blue	B1	B1	B2
Green	B2	B2	B3
Red	B3	B3	B4
NIR	B4	B4	B5
SWIR1	B5	B5	B6
SWIR2	B7	B7	B7

The corrected spectral bands were used to calculate vegetation indices according to the equations listed in Table 4.4. For each acquisition, the clear-masked pixels were aggregated into two statistical metrics: the mean and the standard deviation. A *clear_pixel_percentage* was also calculated for each image to serve as a quality filter. Images with less than 5% clear pixels were excluded to limit the influence of extremely small valid-pixel samples. The threshold was used as a practical compromise between data quality and seasonal coverage. Because only pixels identified by the cloud mask were used for these statistical calculations, an additional cloud cover threshold was not required at this stage.

These cloud masking and availability constraints reduced the number of viable sites to 20, resulting in a condensed working dataset of 221 plots with 1027 observations per variable on average. While temporal resolution was generally adequate, it varied significantly across the dataset. Eight sites faced critical limitations with fewer than three valid images per season, whereas the remaining sites provided up to 21 records, offering sufficient density for temporal reconstruction.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.1.2 Data Cleaning and Trajectory Reconstruction

Given the inherent noise in optical remote sensing data, a multi-stage outlier detection process was implemented. First, an a priori spectral screening enforced physical range constraints on reflectance values to exclude processing artifacts from subsequent analyses.

A second outlier detection step then removed spurious temporal values using the Hampel Identifier [Ham74], which was applied independently to each spectral band and vegetation index [TGMH18]. Because vegetation indices exhibit characteristic annual cycles of growth followed by decline, sudden drops in these values are generally incompatible with gradual phenological changes and can be considered noise caused by clouds, haze or poor atmospheric conditions [CJT+04]. Therefore, the filter specifically targeted drops in the vegetation indices and sudden increases in the spectral band values, reflecting their distinct contamination signatures.

The Hampel Identifier calculates the median and Median Absolute Deviation (MAD) within a moving temporal window. An observation is flagged as an outlier if its absolute deviation from the local median exceeds a threshold of $2 \times 1.4826 \times \text{MAD}$, where the consistency factor 1.4826 aligns the MAD scale with two standard deviation of Gaussian-distributed data [RC93]. On average, 5.5 (1.5%) observations per spectral band or vegetation index were identified as outliers and excluded from the subsequent analysis, thereby improving the overall signal-to-noise ratio of the dataset.

Due to the high uncertainty and uneven temporal distribution of the remaining satellite observations, vegetative trajectories were reconstructed via double-sigmoid curve fitting to represent continuous crop phenology. As demonstrated by Beck et al. [BAH+06], double logistic functions excel at capturing temporal trajectories throughout the entire growing season, reliably handling critical transition periods such as crop maturity.

To mitigate residual noise during fitting, quality weights derived from the previously calculated *clear_pixel_percentage* were applied, with Tier 2 products receiving reduced weight compared to Tier 1. A minimum weight threshold of 0.1 ensured all valid observations influenced the fit.

The mathematical formulation of the double-sigmoid model used is defined as:

$$V(t) = v_{\text{base}} + \frac{v_{\text{peak}} - v_{\text{base}}}{1 + \exp(-r_{\text{gu}}(t - c_{\text{gu}}))} - \frac{v_{\text{peak}} - v_{\text{base}}}{1 + \exp(r_{\text{sen}}(t - c_{\text{sen}}))} \tag{5.1}$$

where $V(t)$ is the vegetative metric at time t, expressed as a relative season progress percentage. The parameter v_{base} represents the baseline minimum (bare soil), v_{peak} is the maximum peak greenness, c_{gu} and c_{sen} define the inflection points for green-up and senescence respectively, and r_{gu} and r_{sen} indicate their respective rates of change [BAH+06].

To ensure convergence and biological realism, particularly in sparse data scenarios, dynamic, data-driven parameter bounds were enforced. Regardless of the expected phenological trajectory, the rate of change parameters r_{gu} and r_{sen} were consistently

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

constrained to positive values between 0.001 and 0.2 [CJT+04, BAH+06]. Note that in Equation 5.1, the subtracted senescence term uses $+r_{\text{sen}}$ in the exponent, which intrinsically governs the transition dynamics without requiring negative rates.

Instead, the distinction between standard vegetation phenology (low→high→low, applicable to indices and the NIR band) and the inverse pattern (high→low→high, seen in other spectral bands) was controlled entirely by the relationship between the baseline (v_{base}) and peak (v_{peak}) parameters. Baseline bounds were mapped dynamically to the observed data range and clamped within physical sensor limits. To model standard positive phenology, v_{peak} was constrained strictly above v_{base} (scaled relative to the observed maximum), whereas for inverse trajectories, v_{peak} was forced strictly below v_{base} to correctly invert the curve. Finally, to enforce chronological logic, the temporal inflection points for green-up (c_{gu}) and senescence (c_{sen}) were restricted to relative season progress intervals of 10–80% and 60–100%, respectively.

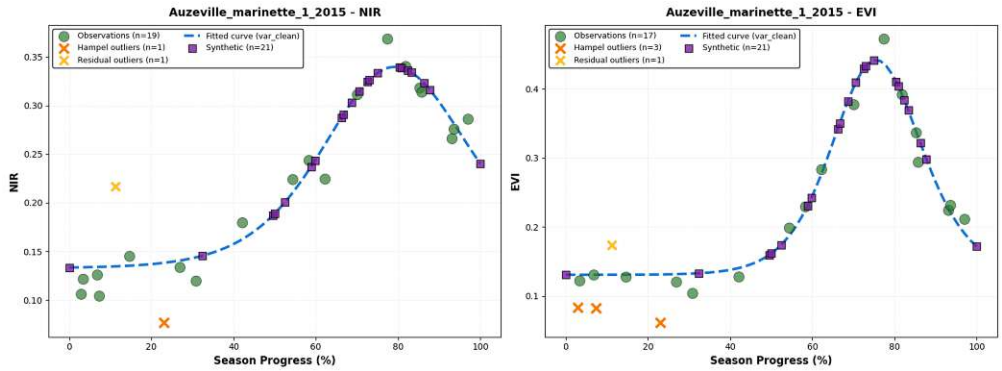


Figure 5.1: Fitted double sigmoid curve for a selected location representing the smoothing pipeline. Green dots are observed satellite values, and their size is relative to their quality weight. Red crosses indicate outliers identified by the Hampel filter. The blue line represents the fitted curve, and the violet squares are derived synthetic values at the dates of observed field measurements.

Where the double-sigmoid fit failed to converge appropriately, a fallback method combining Savitzky-Golay smoothing [SG64] with linear interpolation was employed, adapted from Pan et al. [PHC17]. Because remote sensing trajectories are fundamentally continuous, the Savitzky-Golay filter provides a robust non-parametric alternative. It performs a simplified least-squares-fit convolution that functions as a weighted moving average. By applying polynomial coefficients, the filter preserves higher data moments and minimises the bias typically introduced by standard moving averages [CJT+04]. The smoothing operation is defined mathematically as:

$$Y_j^* = \frac{\sum_{i=-m}^{m} C_i Y_{j+i}}{N} \tag{5.2}$$

where Y_j^* is the resultant smoothed value, Y represents the original observation, C_i is the polynomial coefficient for the i-th value within the filter window of size $2m+1$, and

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

$N = \sum_{i=-m}^{m} C_i$ is the corresponding normalisation factor [SG64].

Although the smoothing process effectively generated continuous, daily-interpolated trajectories naturally capable of providing values for any given day, retaining data at a strictly daily resolution introduces considerable redundancy for the subsequent modelling steps. Instead, the dataset fed into the machine learning models was structured to include a record at least every three days, in addition to all specific dates on which actual field or satellite observations occurred. This three-day interval was selected to balance the capture of necessary temporal dynamics with computational efficiency, an approach consistent with the methodology employed by Kölbing et al. [NKM25].

5.1.3 Machine Learning Pipeline

The core of the prediction phase leverages supervised machine learning to map satellite-derived trajectories to ground-based observations. Across all modelling targets—phenology, time-series dynamics, and end-of-season (EOS) endpoints—Random Forest (RF) regressors were uniformly adopted. This fundamental architectural choice was extensively supported by the literature review, which highlights the algorithm's pervasive success and high efficacy in agricultural remote sensing tasks [BD16, LZH+19, WHM+17, KPG+18, NKM25]. RF is an ensemble learning method that constructs multiple decision trees during training and outputs their averaged predictions. By combining bootstrap aggregation with a random subset of features at each split, the ensemble handles high data dimensionality and complex non-linear relationships efficiently. Consequently, it requires minimal preprocessing and operates robustly against the noise commonly found in agricultural datasets [BD16, NKM25].

Model performance is evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), normalised RMSE (nRMSE), and the coefficient of determination (R^2). For observed values y_i, predictions $\hat{y}_i$, sample size n, and mean of observations $\bar{y}$, the metrics are defined as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5.3}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5.4}$$

$$\text{nRMSE} = 100 \cdot \frac{\text{RMSE}}{\bar{y}} \tag{5.5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5.6}$$

where lower MAE, RMSE, and nRMSE indicate better predictive accuracy, while higher R^2 indicates stronger explained variance. In the following subsections, nRMSE is prioritised for cross-variable comparison because it provides a scale-independent relative error measure.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

To provide these models with comprehensive temporal context, a substantial feature engineering step was introduced. Beyond the raw values of the six spectral bands and ten vegetation indices, two additional transformations were computed over the smoothed trajectories: the temporal derivative (representing the daily rate of change or growth momentum) and the integrated version (representing the accumulated history of crop development), computed as the cumulative sum using the trapezoidal rule[1]:

$$I(s_n) = \sum_{i=1}^{n} \frac{V(s_{i-1}) + V(s_i)}{2}(s_i - s_{i-1}) \tag{5.7}$$

where $I(s_n)$ is the integrated version at seasonal day s_n, and $V(s)$ represents the value of the spectral band or vegetation index. Furthermore, management data (such as sowing dates and fertilisation schedules) were integrated into the feature space. This aligns with results from recent studies emphasising that the model's generalisability benefits significantly when agricultural practices are provided alongside spectral inputs [NKM25].

To iteratively manage the expanded dimensionality and ensure model reliability, a standardised pipeline was implemented for all prediction variables. First, to avoid overfitting to specific field conditions, trials were strictly grouped, and a leave-one-group-out cross-validation framework was applied. Second, because the predictive power of individual satellite observations shifts depending on the variable and temporal stage of the season [NKM25], a backward feature selection process was employed. The RF model was initially trained using the full feature set. The least important features were then iteratively removed based on the algorithm's built-in feature importance metric to successfully isolate the most informative subset of predictors [BD16]. Following feature selection, the Python framework *Optuna*[2] was utilised to systematically search the predefined hyperparameter space for the optimal configuration. Finally, the tuned configuration was refitted 25 times with different random seeds, so that the reported metrics average over the bootstrap resampling and the random feature subsets that the forest draws internally. The spread across these repetitions shows how much of the measured performance is attributable to the stochastic components of the ensemble.

5.1.4 Phenological (BBCH) Prediction

The objective was to complete the sparse ground observations by applying this generalised ML pipeline to phenological scaling. The application of remote sensing data provides free, recent information on the actual state of the crop across spatial ranges, reducing the need for physical observational units and facilitating valid comparisons across locations [NKM25].

Figure 5.2 highlights the consistent record of specific developmental milestones within the observed dataset, namely stages 0, 9, 30 (for cereals only), 65, and 89.

[1] https://math.libretexts.org/Workbench/Numerical_Methods_with_Applications_(Kaw)/7:_Integration/7.02:_Trapezoidal_Rule_of_Integration

[2] https://optuna.org/

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

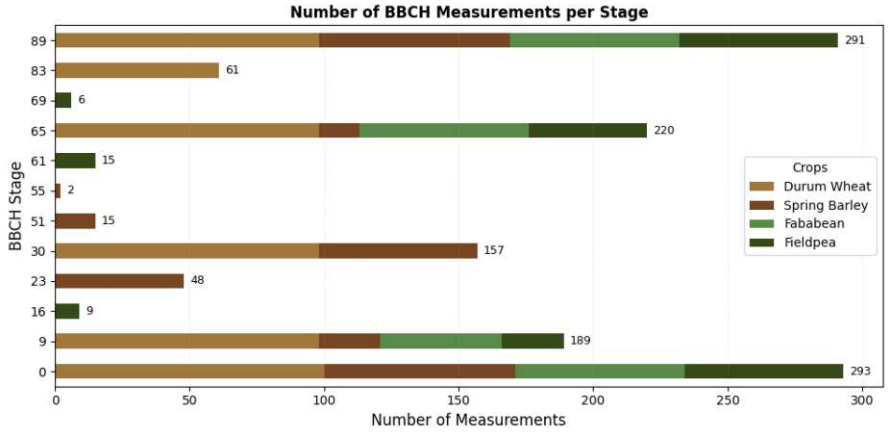


Figure 5.2: Number of BBCH measurements recorded for each crop species at specific developmental milestones. The stacked bar chart distinguishes between durum wheat (light brown), spring barley (dark brown), faba bean (light green), and field pea (dark green).

Consequently, the modelling effort mapped the selected optimal feature space specifically to these targeted key stages. Because RF models natively lack the ability to capture temporal dependencies inherently present in time-series data [DN20], two additional orientation features were introduced: the previously reached phenological stage *BBCHprev* and the relative time elapsed since that stage was reached *BBCHprevT*. Because developmental trajectories differ fundamentally between species, separate RF models were trained for cereals and legumes. To further account for varying lengths of growing periods (e.g., the extended season of durum wheat compared to spring barley, as previously visualised in Figure 4.2), absolute days after planting were transformed into a relative *Season_Progress* percentage before prediction.

Rather than evaluating this model using R^2 or errors in the numerical BBCH values, performance was assessed by the absolute difference in days between the observed and predicted dates of the key developmental stages [NKM25]. Similar to the approach of Pereira et al. [PDLF+22], different feature groups were first compared, including the full feature set, spectral bands only, and vegetation indices only. The vegetation-index models were more stable across the validation splits, so backward feature selection was continued with this group. The stage-specific results are reported in Section 6.1.

5.1.5 Time-Series and End-of-Season Data Prediction

Following phenology prediction, models derived from the pipeline mentioned above (Section 5.1.3) were developed for the continuous time-series variables: WTOP, CNUP, PLHGT, and LAI. As with the phenology model, RF natively lacks sequential awareness. Thus, alongside *BBCHprev* and *BBCHprevT*, the time-series models were also supplied with the estimated next stage (*BBCHnext*) and the relative days anticipated until it is

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

reached ($BBCHnextT$).

To capture the underlying temporal dynamics, three modelling designs were compared using WTOP as the benchmark target: a standard point-in-time RF regressor, an incremental model estimating daily changes inspired by Vicente-Guijalba et al. [VML15], and an interpolative design. The standard and incremental models achieved validation nRMSEs of 33.4% and 27.8%, respectively. The interpolative model, which additionally incorporated the closest preceding and succeeding field observations and their temporal distances, reduced the validation nRMSE to 16.1%. This comparison motivated its selection for trajectory reconstruction. Figure 5.3 illustrates the trajectories produced by the three designs.

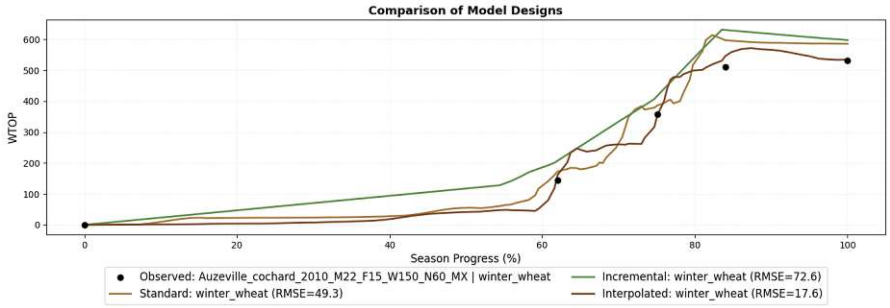


Figure 5.3: Comparison of different modelling designs for time-series forecasting. The solid lines denote the predicted trajectories generated by the models (light brown for standard, green for incremental, and dark brown for interpolated), while the black dots indicate the actual observed field values. The interpolated sequence adheres most closely to the observations, whereas the incremental approach yields the smoothest overall curve.

The feature-importance profiles were also inspected to determine how the designs produced their estimates. The standard model relied primarily on seasonal progress, the incremental model on the NIR band, and the interpolative model on the preceding and succeeding target values, supported by seasonal progress and integrated vegetation indices.

For the interpolative WTOP model, the training and validation nRMSEs were 14.9% and 16.1%, respectively, indicating no signs of overfitting. On this basis, the same interpolation framework was applied to CNUP, PLHGT, and LAI. Detailed performance metrics are reported in Table 6.2.

Following trajectory reconstruction, a secondary predictive RF model was trained on the dense, gap-filled trajectories. This model was used exclusively to generate synthetic time-series for plots containing two or fewer field observations. These plots were excluded from interpolation because their temporal anchors were insufficient to constrain a trajectory. The predictive models retained the temporal, spectral, and management features described above, but did not use preceding or succeeding target observations. Their performance was evaluated on held-out observations using the same metrics and grouped validation procedure as the other RF models. The distributions of observed values, interpolated

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

values, and fully synthetic predictions were additionally compared using Kernel Density Estimation (KDE) for each crop–variable combination.

Finally, predictive RF models were trained for the EOS variables WGRN and NGRN. Their inputs combined the integrated vegetation indices, management variables, and the terminal values of the reconstructed time-series variables. Model performance and feature importances were evaluated using the same validation and interpretation procedures as for the time-series models. The corresponding results are presented in Table 6.3, Figure 6.4, and Tables A.2–A.4.

5.2 Parameterisation of the InterCrop Model

The fundamental objective of the model parameterisation is to identify cultivar-specific crop parameters that minimise the discrepancy between observational predictions and the corresponding *InterCrop* simulation outputs, whilst ensuring that the identified parameter values remain biologically plausible. This can be formalised as a global optimisation problem: find $x^* \in \Omega$ such that $f(x^*) \leq f(x)$ for all $x \in \Omega$, where $f : \Omega \to \mathbb{R}$ is a custom loss function, $\Omega = [0, 1]^d$ denotes a normalised search space, and d represents the number of parameters to be calibrated. Standard Bayesian optimisation (BO) approaches this problem by constructing a global surrogate model over the entire search space, which becomes computationally challenging at high dimensionality as discussed in Section 2.3. To address these limitations, Eriksson et al. [EPG+19] proposed Trust Region Bayesian Optimisation (TuRBO), a method that decomposes the global problem into multiple local optimisation tasks managed through a multi-armed bandit framework.

The optimisation and subsequent sensitivity analysis were restricted to the durum wheat–faba bean combination. This restriction kept the computational and interpretative workload within the scope of the thesis and concentrated the evaluation on the crop combination with the stronger data basis. Repeating both analyses for spring barley–field pea may be considered in future work as discussed in Section 7.4.

For the same reason, the budget was spent on comparing three objective configurations rather than on repeating any of them. Each configuration was executed once, at a fixed seed, so the convergence figures reported in Chapter 6 describe individual search trajectories and no uncertainty intervals were derived for the estimated parameters. The procedure below is therefore designed to make the effect of the reconstructed targets and of the domain-knowledge constraints visible, not to establish the stability of the resulting parameter values.

5.2.1 Trust Region Bayesian Optimisation (TuRBO)

At the core of TuRBO lies the concept of local surrogate modelling within restricted regions of the search space, termed trust regions. To represent global functions of this complexity linear and quadratic approximations are typically insufficient. However, within sufficiently small, localised regions they can provide accurate models. TuRBO combines

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Gaussian process (GP) surrogates with trust regions to leverage both the robustness to noise and principled uncertainty quantification inherent to Bayesian optimisation, whilst maintaining computational efficiency through local approximation. Each trust region is constructed as a hyperrectangle centred at the best solution discovered to date. In the noise-free setting, this best point is taken as the location of the best observed value. When noise is present, it is selected as the observation with the smallest posterior mean under the surrogate model.

The base side length of the trust region, denoted L, is initialised to a value L_{init} and then rescaled for each dimension according to the lengthscale parameters of the GP model (λ_i), whilst preserving a constant total volume of L^d. This adaptive rescaling ensures that the trust region expands and contracts appropriately in directions where the function exhibits steeper or shallower behaviour. Note that, if L encompasses the entire normalised search space, the method reduces to standard global Bayesian optimisation. The trust region expands when the optimiser achieves progress—that is, when improved solutions are identified within the current region—and shrinks when the optimisation appears stagnant, thereby balancing local exploitation with the need to eventually explore neglected areas.

To promote global exploration whilst maintaining the computational benefits of local modelling, TuRBO maintains m trust regions simultaneously, each with its own independent GP model and base side length $L_\ell \leq L_{\max}$ for $\ell \in \{1, \ldots, m\}$. This approach naturally frames the global problem as an exploitation–exploration trade-off. Rather than relying on inefficient random restarts, TuRBO employs a multi-armed bandit mechanism to allocate the evaluation budget, treating each trust region independently. This allows the algorithm to reason probabilistically about which local optimisation run is most promising, thereby concentrating evaluations on regions demonstrating progress whilst avoiding premature convergence to suboptimal local minima.

Within each iteration, TuRBO must select a batch of q candidate points from across all trust regions and subsequently update the local optimisation problems. Thompson sampling provides a principled, theoretically motivated solution to this dual problem. To select the i-th candidate, the algorithm draws an independent sample from the posterior distribution of the GP within each trust region:

$$f_\ell^{(i)} \sim \mathcal{GP}_\ell^{(t)}\left(\mu_\ell(x), k_\ell(x, x')\right) \tag{5.8}$$

where $\mathcal{GP}_\ell^{(t)}$ is the GP posterior for trust region ℓ at iteration t. The candidate is then selected by minimising the sampled function across all trust regions:

$$\left(\ell_i^{(t)}, x_i^{(t)}\right) \in \underset{\substack{\ell \in \{1,\ldots,m\} \\ x \in \mathrm{TR}_\ell}}{\operatorname{argmin}} f_\ell^{(i)}(x) \tag{5.9}$$

This process is repeated for each of the q candidates in the batch. By sampling independently from each trust region's posterior and aggregating samples through minimisation, Thompson sampling handles exploitation within a promising region and inter-regional candidate selection, ensuring that the algorithm concentrates effort on the most promising

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

regions whilst maintaining systematic exploration. The implementation used in this thesis operates a single trust region. With 200 evaluations distributed over 21 parameters, dividing the budget among several regions would leave each surrogate with too few observations to be informative. The restart behaviour is retained, in that a region which has shrunk below its minimum side length is reinitialised from a fresh design. The search is thereby confined to one basin at a time, which is taken up again in Section 7.3. The configuration used for all optimisation runs reported in this thesis, including the evaluation budget and the size of the initial design, is listed in Table A.8. Settings not stated there were left at the defaults of the reference implementation.

5.2.2 Constrained Loss Function

The loss function $f(x)$ was custom-designed to integrate domain knowledge and enforce biological plausibility throughout the optimisation process, following the approach of Mitra et al. [MDPH18]. The function comprises a fitting error component and a suite of biologically motivated constraints.

The fitting error is constructed to ensure accurate temporal and quantitative alignment between simulations and observations. For phenological stages (BBCH), the error metric is the root mean square difference in days between the simulated and the reference date at which stages 9, 30, 65, and 89 are first reached, weighted towards flowering and normalised by a reference deviation of 20 days. This temporal alignment is essential because adequate simulation of variable magnitudes alone is insufficient, when the model cannot capture the correct timing of crop development. For all other variables (WTOP, CNUP, PLHGT, and LAI), the normalised root mean square error (nRMSE) is employed. End-of-season variables such as grain biomass (WGRN) and grain nitrogen (NGRN) are evaluated at crop maturity.

Beyond the fitting error, a set of biologically motivated constraints is computed to guide the optimisation and ensure parameter realism. These constraints are implemented using two complementary strategies. *Soft constraints* add penalties to the loss function that grow with the square of the relative constraint violation, guiding the optimisation toward biologically plausible regions whilst allowing exploration. *Hard constraints*, in contrast, terminate the current iteration immediately when violated, which would spare the corresponding model evaluations. Within the search bounds used here no candidate can violate them, so the three constraints described below act through soft penalties alone.

The first constraint concerns the phenological structure of the crop. The sum of all parameters representing biological days (those accumulating heat units or physiological time) is constrained to remain close to a baseline value. This permits the optimisation to explore in-season timing adjustments without artificially expanding or contracting the overall growing season, which would produce misleading results. The sum may vary by 10% in either direction around that baseline; beyond this band a soft penalty is applied, steering the optimisation toward plausible phenological configurations.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

A second constraint concerns the relation between the minimum and maximum grain nitrogen concentration parameters (GNCmin and GNCmax). A candidate in which GNCmin exceeds GNCmax is biologically impossible and terminates the iteration immediately. This case cannot arise under the bounds used here, since the ±20% intervals of the two parameters are disjoint for both crops (Table A.5). The operative part is the soft penalty, which is applied once the gap between the two values falls below half of the gap between their baselines.

Finally, a third constraint verifies that the maximum plant height parameter (heightMx) is actually achieved within the simulations. Whether this target height is reached depends critically on the interaction between the growth parameter rgrHgt and the heightMx value itself. If simulations fail to attain the specified maximum height, these parameters are mis-specified in a way that standard error metrics would not capture. When this occurs, a soft penalty is added to the loss function, guiding the optimisation algorithm away from parameter combinations that fail to produce the expected growth dynamics.

Through this penalty architecture, domain knowledge is embedded into the optimisation process without removing regions from the search. The penalties preserve the flexibility needed for genuine biological exploration whilst continuously guiding the algorithm toward realistic parameter combinations. The variable weights, the normalisation constants, the admissible bands, and the penalty factors that make up the loss are listed in Table A.7.

5.2.3 Cultivar Grouping and Baseline Parameter Derivation

Before the optimisation could start, the individual cultivars had to be assigned to groups that share a common parameter set. Grouping was based on three traits: maturity type, plant height, and grain protein content. The height distributions per cultivar are shown in Figure 5.4. For most crops, the cultivars were similar enough across all three traits to be treated as a single group. The exception was faba bean, where cultivar Irena was consistently shorter than the others and also matured earlier. These two differences, which occurred in parallel, justified splitting faba bean into two separate groups. Grain protein content was largely comparable across cultivars; the only noticeable spread appeared among the durum wheat cultivars, but as it followed no clear pattern, they were kept together. As a side note, spring barley and field pea showed a high degree of agreement in all three traits. Therefore, subdivision would not be necessary here either.

With the groups defined, a reference value for the maximum plant height parameter (`heightMx`) was taken directly from the observations as the median measured height within each cultivar group (Table 5.2).

The height growth rate parameter (`rgrHgt`) required separate treatment. Unlike the other parameters, it has no counterpart in the underlying SSM-iCrop model, as it was added specifically for the *InterCrop* extension. It needs to be set large enough for the simulated height to be able to reach `heightMx` under standard conditions (Table 4.3), but not excessively high to avoid unrealistic and early growth. A baseline was therefore established through a dedicated calibration loop spanning 18 scenarios, combining three

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

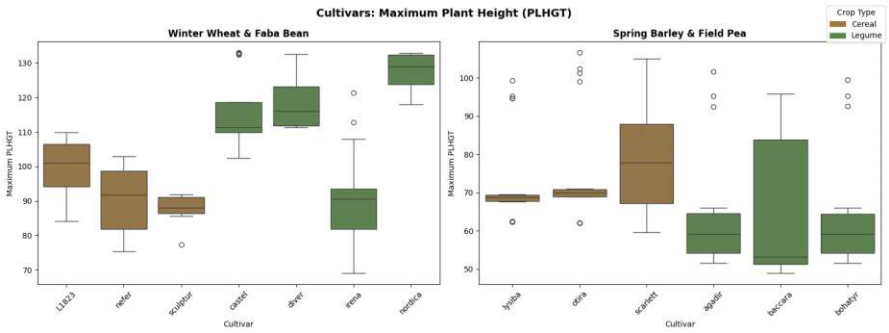


Figure 5.4: Boxplots of measured plant height for each cultivar, used to inform the cultivar grouping. Faba bean cultivar Irena stands out as markedly shorter than the remaining faba bean cultivars, motivating its separation into a distinct group.

Table 5.2: Reference maximum plant height (`heightMx`) derived as the median observed plant height within each cultivar group.

Cultivar group	**`heightMx` (cm)**
`faba bean (medium, early)`	93.0
`faba bean (tall, medium)`	118.5
`durum wheat (tall, medium)`	104.4

weather types (dry, medium, and wet, as used in the sensitivity analysis in Section 5.3), the three soil settings (Table 5.3), and two planting dates in early and late October. The remaining conditions were held fixed at mixed planting with densities of $15\,\mathrm{kg\,ha^{-1}}$ for the legume and $150\,\mathrm{kg\,ha^{-1}}$ for the cereal, and nitrogen application of $120\,\mathrm{kg\,N\,ha^{-1}}$.

For each candidate value of `rgrHgt`, the simulations recorded how often the prescribed `heightMx` was reached within a 5% tolerance. A value that hit the target in more than 70% of scenarios was accepted as a baseline. When the target was missed more frequently, `rgrHgt` was raised in proportion to the shortfall and when every scenario reached `heightMx`, it was lowered slightly to avoid an unnecessarily steep growth rate. The baseline values found this way were then validated against the full 30-year weather record for Tulln, Austria. Figure 5.5 shows this validation for the late-maturing cultivar type under the medium soil setting (SL120; 120 mm PASW). During the subsequent optimisation, `heightMx` was held fixed at its derived median value, and only `rgrHgt` was calibrated starting from this baseline.

Each calibrated parameter was bounded to an interval of ±20% around its baseline, taken from the manually calibrated values of the model creators [MHPN26] or, for `rgrHgt`, from the value derived above. Table A.5 lists the resulting search space for all three calibration units.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

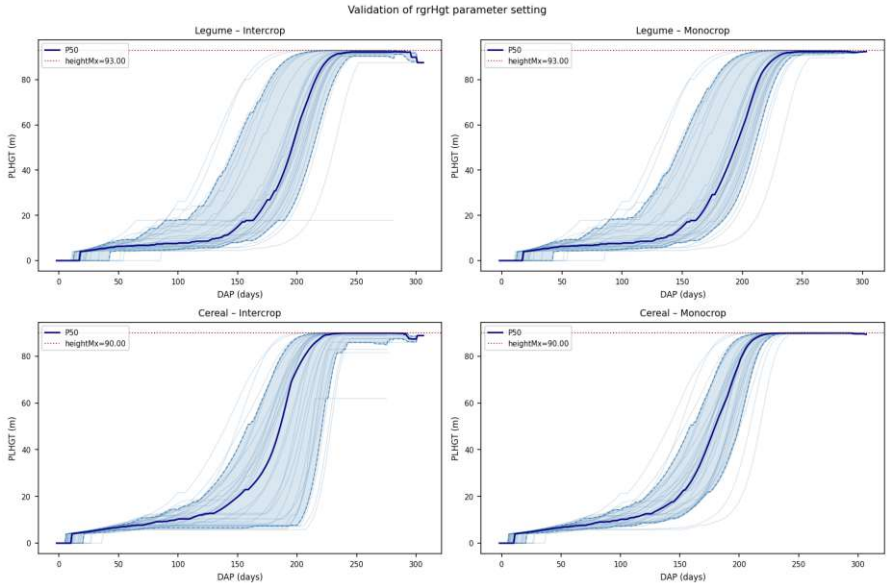


Figure 5.5: Validation of the derived `rgrHgt` baseline against the full 30-year weather record, shown for the late-maturing cultivar type under the medium soil setting (SL120; 120 mm PASW). Each panel plots the simulated plant height trajectories over the season, with the bold line marking the median across all 30 years and the shaded band spanning the 10th to 90th percentile; the dashed line indicates the prescribed maximum height (`heightMx`). Panels are arranged by component (top: faba bean, bottom: durum wheat) and cropping system (left: intercrop, right: monocrop).

5.2.4 Implementation Architecture

The optimisation framework was implemented as a wrapper class for the *InterCrop* Java model. It manages reading template input files, batch-wise filling them with the plot-specific data, executing the *InterCrop* Java model, parsing the output files, computing loss values, and finally iteratively updating the parameters for the next optimisation step. This design allows the underlying Java model to remain unchanged whilst the optimisation logic operates transparently and fully automatically at a higher level of abstraction.

In detail, the optimisation workflow unfolds across several distinct phases. Initially, the framework identifies all cultivar groups present within the predictions and locates the corresponding parameter templates. It then constructs a combined search space across all active groups, with unique parameter names prefixed by their respective group identifiers. This multi-group parameterisation enables simultaneous calibration of different crop varieties within intercropping systems.

The first iteration (Iteration 0) serves as the baseline and employs the manually calibrated parameter values of the model creators [MHPN26]. These expert-selected template parameters are read from the workbook and used to generate batch input files for each

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

location without modification. To manage computational load, plots are organised into batches of up to 20 entries per file. The *InterCrop* Java executable is then invoked on the temporary folder, outputs are parsed, and per-group errors are computed and aggregated. This baseline error establishes a reference point against which all subsequent optimisation iterations are compared. Generated files are preserved for later comparison with optimised results.

During the optimisation loop, TuRBO proposes candidate parameter vectors within the normalised search space. For each candidate, the framework performs the following steps: (1) denormalise the vector and map it to a group-parameter dictionary; (2) evaluate hard constraints and, if violated, skip further evaluation; (3) regenerate all input files with the candidate parameters (clearing previous iteration files except the baseline); (4) execute *InterCrop* and parse the outputs; (5) calculate weighted errors per plot; (6) aggregate per-cultivar-group errors; (7) apply soft constraint penalties; and (8) log comprehensive iteration metadata including timings, per-component errors, and diagnostic information.

After the optimisation finishes, the best parameter vector identified by TuRBO is extracted, mapped back to per-group parameter dictionaries, and returned alongside the complete optimisation history and diagnostic statistics. The best-iteration generated files are preserved to facilitate post-hoc analysis and validation.

5.2.5 Convergence Metrics

To compare optimisation runs beyond their final loss, the search behaviour over the course of the iterations is summarised by two metrics. Let f_k denote the best loss found up to and including iteration k, and f_0 the baseline loss from the manually calibrated parameters [MHPN26] (Iteration 0).

The first metric is the relative improvement at step k, expressing how much the best loss has dropped relative to the baseline:

$$\mathrm{RI}_k = \frac{f_0 - f_k}{f_0} \tag{5.10}$$

A value of 0 indicates no gain over the baseline, whereas a value approaching 1 corresponds to a near-complete reduction of the initial error. Reporting RI_k at a fixed budget allows runs to be compared on equal terms regardless of their absolute loss scale.

The second metric is the area under the convergence curve up to step k, which rewards reaching low losses early rather than only at the end of the budget:

$$\mathrm{AUC}_k = \sum_{i=1}^{k} f_i \tag{5.11}$$

Because AUC_k accumulates the loss across all iterations, two runs that finish at the same final value are distinguished by how quickly they descended: a run that converges early

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

accumulates less error and therefore yields a lower AUC_k. Both metrics are reported in Chapter 6 to characterise the convergence behaviour of the optimisation.

The following section describes the sensitivity analysis framework that was applied to the *InterCrop* model to identify influential parameters and understand their effects across different environmental and management scenarios.

5.3 Combined Sensitivity Analysis and Explainability

Traditional sensitivity analysis quantifies the aggregate impact of individual input parameters on model outputs, but does not reveal how these impacts change across different scales of perturbation [RJS+21]. This study therefore employs a framework that integrates classical sensitivity analysis with explainability concepts. By applying the Variogram Analysis of Response Surfaces (VARS) framework [RG16], the sensitivity of the crop parameters is evaluated across a structured set of environmental and management scenarios. The primary interest is how the crop parameters of each species cross-influence the outputs of both their own and the companion plant under varying conditions. This approach not only identifies the most influential parameters but also reveals the scale at which their influence becomes apparent, thereby providing deeper explanations of the model's behaviour.

5.3.1 The VARS Framework

The VARS framework [RG16] is accessible via the `varstool` software package. VARS characterises the response surface of a model by integrating directional variograms of the model output over the entire input space and across a full range of perturbation scales, denoted as h. The directional variogram function $\gamma(h)$ and the integrated variogram function $\Gamma(H)$ are defined as:

$$\gamma(h) = \frac{1}{2} V[Z(\theta + h) - Z(\theta)] \tag{5.12}$$

$$\Gamma(H) = \frac{1}{2} \int_0^H V[Z(\theta + h) - Z(\theta)]\, dh \tag{5.13}$$

where $\theta = \{\theta_i \text{ for } i = 1, \ldots, n\}$ represents a point in the n-dimensional input space, Z is the respective model output, $V[\cdot]$ denotes the variance operator, and H is the range of perturbation scales of interest [RG16].

The perturbation scale h indicates the magnitude of change applied to an input variable, while $\gamma(h)$ describes the resulting variance in the output (as illustrated in Figure 5.6). This relationship is central to combining sensitivity with explainability: for a given parameter, small variations might produce no noticeable difference in the output, whereas larger perturbations could reveal it as a dominant factor. The integrated variogram provides a measure of cumulative variability up to scale H, such as the IVARS-50 metric evaluated at $H = 1$ where the input range is uniformly scaled to 2.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

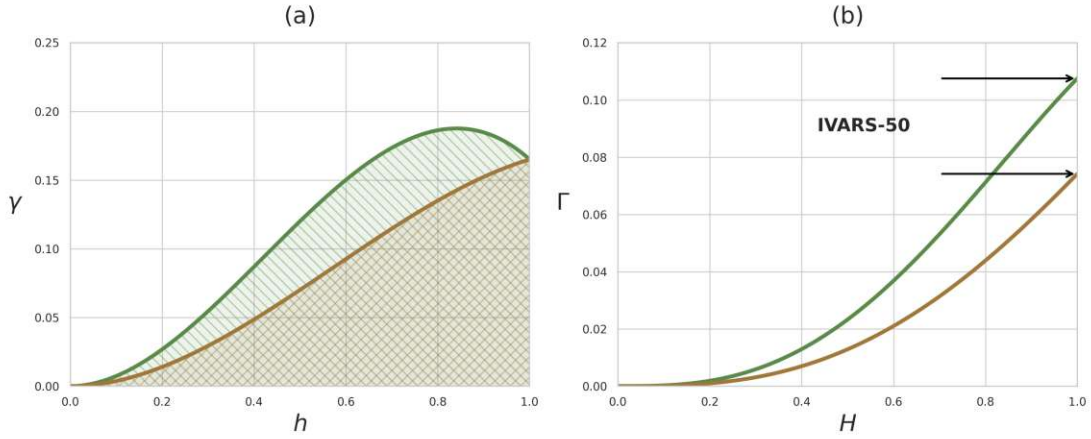


Figure 5.6: An illustrative example of directional and integrated variograms showing IVARS-50 for two inputs. The input range is two, resulting in a range of perturbation scales (h) of one, corresponding to half of the input range. Adopted from [RG16].

To explore the parameter space evenly, the design employs the STAR-VARS sampling strategy [RG16]. Latin Hypercube Sampling generates a set of star centres, each radiating cross-sectional lines through the parameter space to sample the response surface systematically. The algorithm requires specifying the number of star centres and the number of cross-sectional points. For this analysis, 30 star centres and 10 cross-sectional points were selected, which over the 12 crop parameters results in 3,600 model evaluations per scenario.

5.3.2 Parameter Selection and Distribution

The VARS factor set comprises the twelve crop parameters that govern the two intercropping partners, six per species (detailed in Chapter 4, Table 4.3):

- **Cereals (6 parameters):** `MEED_cer`, `GNCmin_cer`, `GNCmax_cer`, `heightMx_cer`, `bdSELBOT_cer`, and `bdANTPM_cer`.
- **Legumes (6 parameters):** `MEED_leg`, `GNCmin_leg`, `GNCmax_leg`, `heightMx_leg`, `bdEMRR1_leg`, and `bdR5R7_leg`.

This set is not identical to the one calibrated in Section 5.2. There, `heightMx` was derived from the observations and held fixed (Section 5.2.3), while `rgrHgt` and `GRTDP` were added to the search space because they determine how quickly the simulated crop approaches the height and rooting-depth limits set by `heightMx` and `MEED`. The sensitivity analysis asks the reverse question, namely how strongly the model responds to those limits themselves, so `heightMx` is perturbed as a factor and the two rate parameters stay at their baseline values.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

All twelve factors follow bounded uniform distributions spanning ±20% around their baseline values, listed in Table A.6, and the same STAR design was reused for every scenario.

The environmental and management drivers are not perturbed as VARS factors, but they span the outer scenario structure. Four drivers were varied at three levels each, namely weather, soil, sowing date, and nitrogen application, producing $3^4 = 81$ scenarios. The three weather settings correspond to meteorological seasons drawn from the historical records (1993–2024) in Tulln, Austria, categorised by total precipitation into Dry (2011), Medium (2006), and Wet (2008). The soil settings represent a light soil (SL60; 60 mm plant-available soil water, PASW), a medium soil (SL120; 120 mm PASW), and a heavy soil (SL240; 240 mm PASW), with their parameters detailed in Table 5.3. Sowing date was set to early, mid, or late October (DOY 278, 288, or 298), and nitrogen application to 0, 60, or 120 kg N ha^{-1}. The full STAR design of the twelve crop parameters is evaluated independently under each of the 81 scenarios, so that the resulting IVARS indices across scenarios are comparable.

Table 5.3: Input parameters for the three soil settings used in the sensitivity analysis.

DLYER	SAT	DUL	LL	ADRY	iniWL	DRAINF	BLD	NORG	FMIN	Nmin
mm			cm^3 cm^{-3}			–	g cm^{-3}	%	–	kg N ha^{-1}
Light soil (SL60; 60 mm PASW)										
300	0.230	0.180	0.120	0.072	0.180	0.6	1.50	0.22	0.040	20.0
300	0.220	0.170	0.120	0.072	0.127	0.6	1.60	0.22	0.009	15.0
300	0.210	0.160	0.120	0.012	0.120	0.5	1.65	0.16	0.005	4.0
300	0.200	0.170	0.120	0.012	0.120	0.5	1.70	0.14	0.001	1.0
Medium soil (SL120; 120 mm PASW)										
300	0.300	0.250	0.140	0.084	0.250	0.5	1.40	0.22	0.040	20.0
300	0.290	0.240	0.140	0.084	0.163	0.5	1.45	0.22	0.009	15.0
300	0.280	0.220	0.130	0.130	0.130	0.4	1.50	0.16	0.005	4.0
300	0.280	0.230	0.130	0.130	0.130	0.4	1.55	0.14	0.001	1.0
Heavy soil (SL240; 240 mm PASW)										
300	0.430	0.380	0.150	0.090	0.380	0.4	1.20	0.22	0.040	20.0
300	0.420	0.370	0.150	0.090	0.187	0.4	1.30	0.22	0.009	15.0
300	0.410	0.360	0.170	0.170	0.170	0.3	1.40	0.16	0.005	4.0
300	0.390	0.340	0.180	0.180	0.180	0.3	1.45	0.14	0.001	1.0

5.3.3 Implementation and Evaluation Metrics

Executing 291,600 model evaluations in total (3,600 star points × 81 scenarios) presented a considerable computational challenge due to the inherent processing load and initialisation overhead of the Java-based *InterCrop* executable. Given that the 81 scenarios are identical for every star point, a single template was created once, and the STAR sampling of the twelve crop parameters was likewise generated only once. For each star point, a single input file was then produced in which only the crop parameters were substituted. The

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

input files were assembled from standard template Excel sheets using the `openpyxl` Python library and passed to the backend via subprocess commands. End to end—including file generation, model execution, output parsing, and the VARS analysis—the full procedure completed in under five hours on a standard consumer laptop.

Upon completion of the simulation runs, the outputs extracted for each intercropping configuration were WGRN and NGRN for both the cereal and legume components. These four output arrays were fed into the offline VARS analyser separately for each of the 81 scenarios, which computed the IVARS sensitivity indices for every crop parameter. IVARS-10, IVARS-30, and IVARS-50 were used to assess sensitivity over perturbations of up to 10%, 30%, and 50% of the parameter range, respectively. Comparing these indices shows whether parameter importance is already apparent under small perturbations or emerges only as a larger part of the parameter range is considered.

The external quality of the sensitivity results was assessed primarily through agronomic plausibility. Where the analysed responses concerned processes shared with monocrop models, their shifts across environmental conditions were compared with the findings reported by Hao et al. [HRW+24] for APSIM and Shen et al. [SZZ+25] for STICS. Agreement with patterns found in these independently developed process-based crop models provides additional evidence that the corresponding *InterCrop* results are plausible. Intercropping-specific responses cannot be evaluated through this comparison because both reference studies concern monocrops. These were instead interpreted against the competition and facilitation mechanisms implemented in *InterCrop*, while the perturbation-scale results were assessed as an additional transparency output of VARS.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 6

Results

6.1 Dataset Enrichment and Synthetic Data Generation

6.1.1 Phenological (BBCH) Prediction

Across the selected BBCH stages, the phenology models achieved an overall mean absolute error of 2.49 days. The standard deviation across the validation splits was 0.26 days, indicating that performance was consistent across the robustness runs. The stage-specific results in Table 6.1 show the largest error for cereal BBCH 30, at 5.59 days. All other stage-specific errors remained below 3.6 days, with the lowest error obtained for legume BBCH 89 at 1.14 days.

Table 6.1: Average absolute difference in days between observed and predicted dates for key BBCH developmental stages. The performance metrics include the mean difference, standard deviation, and the corresponding number of observations (N) for cereals and legumes. The values represent the averages of all validation sets included in the robustness check.

		Avg Diff (Days)		
Crop type	Stage	Mean	Std	N
Cereals	9	2.48	0.14	121
	30	5.59	0.34	157
	65	1.40	0.03	113
	89	1.28	0.07	169
Legumes	9	1.98	0.35	68
	65	3.56	0.86	107
	89	1.14	0.02	122

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6.1.2 Time-Series and End-of-Season Data Prediction

The interpolation models performed most accurately for PLHGT, with a validation nRMSE of 5.7% (Table 6.2). WTOP and CNUP produced similar errors of 16.1% and 16.7%, respectively. LAI was more difficult to reconstruct and reached an nRMSE of 31.1%. The R^2 values followed the same pattern, ranging from 0.97 to 0.99 for WTOP, CNUP, and PLHGT, compared with 0.89 for LAI.

Table 6.2: Validation performance metrics for the time-series interpolation models. RMSE and MAE are expressed in the unit of the respective variable.

Variable	Unit	RMSE	nRMSE (%)	MAE	R^2
WTOP	$\mathrm{g\,m^{-2}}$	52.73	16.13	31.70	0.97
CNUP	$\mathrm{g\,N\,m^{-2}}$	0.98	16.73	0.58	0.97
PLHGT	cm	2.96	5.74	1.91	0.99
LAI	$\mathrm{m^2\,m^{-2}}$	0.38	31.11	0.23	0.89

The second-stage predictive models were less accurate than the interpolation models because they operated without observed target values as temporal anchors. Nevertheless, training the WTOP model on the reconstructed trajectories reduced its validation nRMSE from 33.4% for the original point-in-time model to 28.9% (Table 6.3). Among the remaining variables, PLHGT retained the lowest relative error at 10.4%, whereas CNUP and LAI reached 34.2% and 30.6%, respectively. Figure 6.1 shows three examples of the WTOP trajectories generated for sparse spring-barley plots.

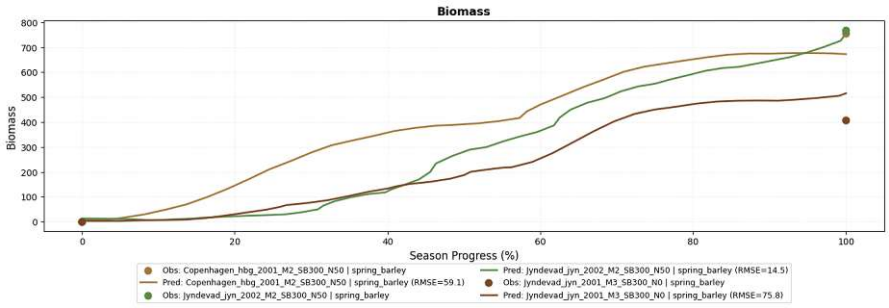


Figure 6.1: Synthetic WTOP trajectories for three spring-barley plots. Lines represent the generated trajectories and dots the observed measurements. The light-brown and green series received 50 $\mathrm{kg\,N\,ha^{-1}}$, whereas the dark-brown series received no applied nitrogen.

Despite restricting the hyperparameter ranges during tuning, the gap between the training and validation sets could not be fully closed in the second predictive phase. This was particularly noticeable for CNUP and LAI, which recorded training nRMSEs of 22.1% and 20.5%, respectively, representing a generalisation gap of around 10 percentage points.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Figure 6.2 shows the corresponding limitations: the LAI model did not reproduce the highest observed values, while the CNUP model tended to underestimate total uptake.

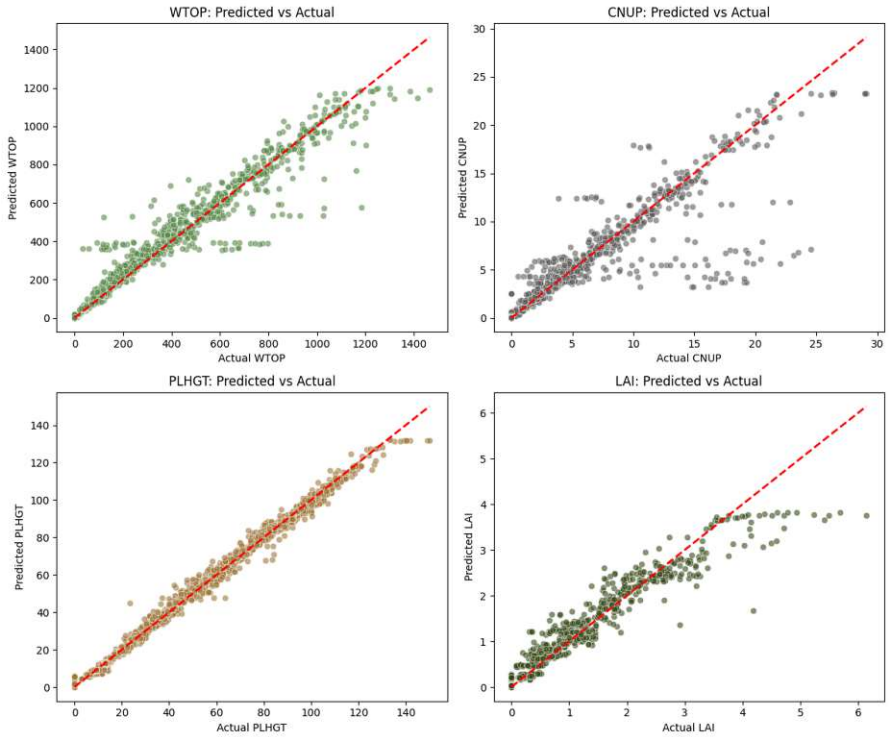


Figure 6.2: Actual versus predicted values for the time-series variables. The red dotted line marks the 1:1 relationship ($y = x$). WTOP predictions (top left, light green) show moderate dispersion. CNUP predictions (top right, grey) tend to underestimate higher observations, while PLHGT predictions (bottom left, brown) lie closest to the 1:1 line. LAI predictions (bottom right, dark green) remain below approximately $4\,\mathrm{m^2\,m^{-2}}$ even where higher values were observed.

The feature-importance profiles differed between the four predictive models. *Season_Progress* dominated PLHGT and CNUP, with importance scores of 0.82 and 0.57, respectively, and was also influential for WTOP together with the integrated NIR band and the integrated MCARI2, CI, and MTVI2 indices. By contrast, *Season_Progress* was absent from the leading LAI features. Integrated CI had the highest importance for LAI at 0.58, followed by *Interrow*, crop type, and integrated or delta values of the Red, NIR, and Green bands. CNUP also drew information from crop or cultivar identity, applied nitrogen, and the integrated Blue band. The complete rankings are provided in Tables A.2 and A.3.

The KDE comparison in Figure 6.3 provides a second assessment of the generated data beyond point-wise prediction errors. For most crop–variable combinations, the interpolated and synthetic distributions followed the observed distributions. The clearest

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

differences occurred for field pea and spring barley, which were underrepresented in the source data, and for LAI across several crops. In particular, the field pea LAI distributions showed little overlap between observations and synthetic predictions.

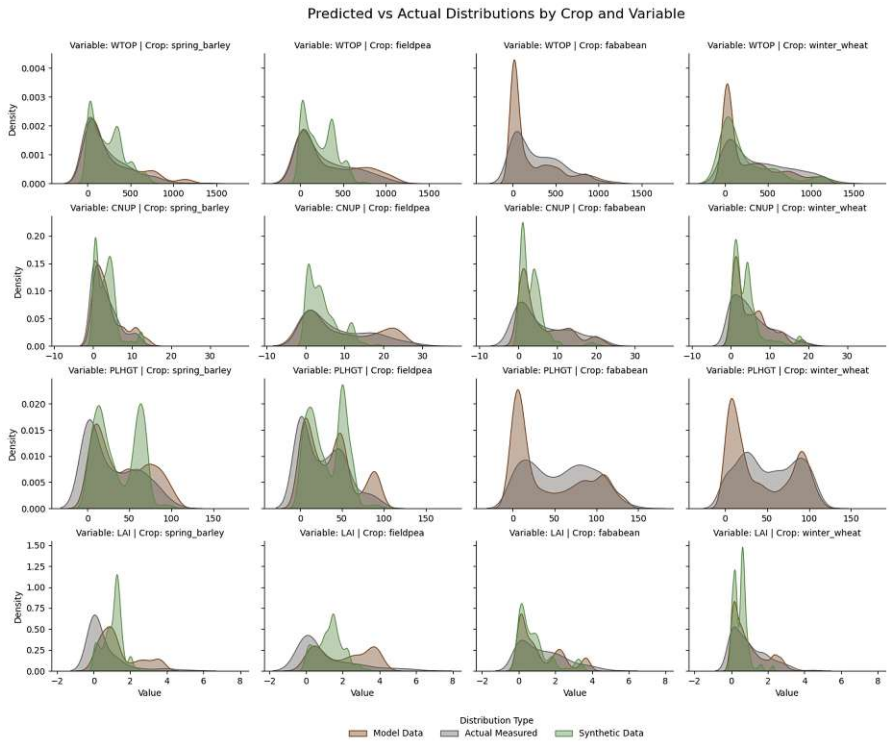


Figure 6.3: Kernel Density Estimation (KDE) of observed values (grey), interpolated values (brown), and synthetic predictions (green). Rows correspond to WTOP, CNUP, PLHGT, and LAI, and columns correspond to crops. The distributions overlap closely for combinations such as durum wheat WTOP, whereas field pea LAI shows little overlap.

The EOS models achieved validation nRMSEs of 22.9% for WGRN and 24.2% for NGRN (Table 6.3). Their training nRMSEs were 17.1% and 17.7%, respectively. The corresponding R^2 values were 0.77 for WGRN and 0.86 for NGRN. Figure 6.4 shows that both models tended to underestimate the highest observed values.

The feature importances indicate that the terminal WTOP value was the principal predictor of WGRN, with an importance score of 0.67, while crop type was the leading predictor of NGRN at 0.42. Integrated Blue contributed to WGRN, and integrated SWIR1 and applied nitrogen contributed to NGRN. The complete feature rankings are reported in Table A.4.

As a simple descriptive test of whether the reconstructed data retained the observed response to nitrogen amount within the data used for parameterisation, separate linear regressions were fitted to the endpoint values of each data group after restricting the analysis to the cultivar groups retained for optimisation. Elasticity expresses the propor-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Table 6.3: Validation performance metrics for the predictive models, including time-series forecasting and end-of-season (EOS) points.

Category	Variable	Unit	RMSE	nRMSE (%)	MAE	R^2
Time-Series	WTOP	$\mathrm{g\,m^{-2}}$	71.92	28.87	39.08	0.94
	CNUP	$\mathrm{g\,N\,m^{-2}}$	1.75	34.22	0.94	0.88
	PLHGT	cm	4.43	10.41	2.56	0.99
	LAI	$\mathrm{m^2\,m^{-2}}$	0.32	30.58	0.17	0.90
End-of-Season	WGRN	$\mathrm{g\,m^{-2}}$	63.81	22.87	46.65	0.77
	NGRN	$\mathrm{g\,N\,m^{-2}}$	1.85	24.18	1.33	0.86

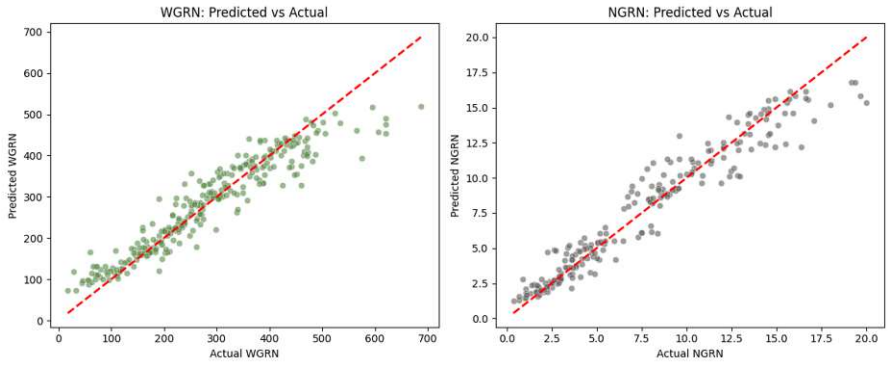


Figure 6.4: Actual versus predicted values for the end-of-season (EOS) variables WGRN (left, green) and NGRN (right, grey). The red dotted line marks the 1:1 relationship ($y = x$). Both models show some underestimation at high observed values.

tional change in the endpoint associated with a proportional change in nitrogen amount. Observed and interpolated values produced similar slopes, correlations, and elasticities for WTOP (elasticity 0.157 and 0.146), PLHGT (−0.029 and −0.021), LAI (0.130 and 0.128), and CNUP (0.109 and 0.104), respectively. The fully predicted values also had a positive elasticity of 0.286 and a Pearson correlation of 0.872 for CNUP, while synthetic LAI had a weak negative elasticity of −0.035 and correlation of −0.126. The synthetic CNUP and LAI groups comprised 58 and 20 cases, respectively, and covered only N0 and N140. These regressions therefore describe the optimisation subset but do not represent the full reconstructed dataset or estimate a continuous nitrogen dose–response relationship.

The final reconstructed dataset comprised 221 plots, with approximately 76 data points per season on average and more than 22,000 data points per variable in total. The following section compares optimisation behaviour when this reconstructed dataset and the raw field observations are used as targets.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

6.2 Model Optimisation Evaluation

Three runs are compared throughout this section. The *Informed* run uses synthetic (model-predicted) target values together with the domain-knowledge penalty in the loss function. Hence, it represents the proposed full pipeline. The *Observed* run replaces those synthetic targets with the raw field measurements and provides a comparison with optimisation on the original observations. The *Unconstrained* run retains the synthetic targets but removes the domain-knowledge penalty, providing a comparison of the search behaviour with and without the constraint terms.

Table 6.4: Convergence comparison of the three optimisation runs at selected iteration counts. For each run the best error achieved so far, the relative improvement over the iteration 0 baseline (Equation 5.10), and the cumulative area under the convergence curve (Equation 5.11) are reported.

Metric	Iteration	Informed	Observed	Unconstrained
Best error	50	0.0899	0.1056	0.0902
	100	0.0889	0.1041	0.0894
	150	0.0887	0.1026	0.0889
	175	0.0886	0.1025	0.0884
	200	0.0885	0.1024	0.0883
Relative improvement (%)	50	9.43	5.82	8.84
	100	10.50	7.12	9.62
	150	10.68	8.52	10.15
	175	10.78	8.62	10.69
	200	10.86	8.67	10.74
AUC sum	50	4.87	5.63	4.90
	100	9.39	10.94	9.45
	150	13.85	16.15	13.95
	175	16.07	18.73	16.18
	200	18.29	21.30	18.40

The Informed run had the lowest best error through iteration 150, after which the Unconstrained run became marginally lower. After 200 iterations, their best errors were nearly identical at 0.0885 and 0.0883, respectively. Relative to their own iteration 0 baselines, the Informed run improved by 10.9% and the Unconstrained run by 10.7%. The Observed run improved from 0.1121 to 0.1024, corresponding to 8.7%. The cumulative AUC at iteration 200 was lowest for the Informed run at 18.29, followed by 18.40 for the Unconstrained run and 21.30 for the Observed run. Figure 6.5 additionally shows the variation in the individual candidate evaluations from which these running best values were derived.

Both synthetic-target runs showed larger relative improvements from their respective

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

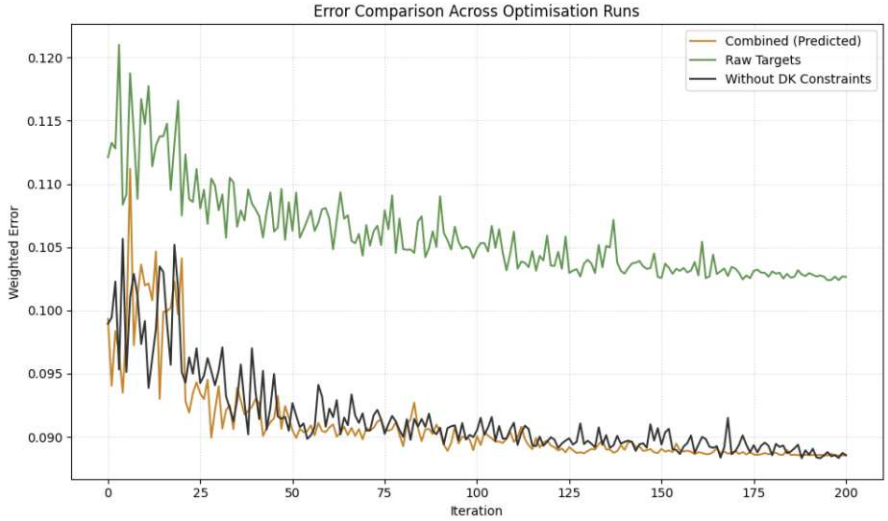


Figure 6.5: Loss values recorded over the course of the optimisation for the three runs: the Informed run (synthetic targets with domain-knowledge constraints), the Observed run (raw field measurements), and the Unconstrained run (synthetic targets without the domain-knowledge penalty). The variation between successive iterations reflects the evaluation of different candidate parameter sets.

baselines than the Observed run at every reported checkpoint. Since the Observed run used a different target dataset and therefore a different loss surface, its absolute loss and AUC values are reported as descriptive comparisons. The Informed and Unconstrained runs used the same synthetic targets and baseline parameter values, but their initial errors were 0.0993 and 0.0989, respectively, because the Informed objective included a small soft penalty for plant height. At iteration 50, the Informed run had a best error of 0.0899 compared with 0.0902, a relative improvement of 9.4% compared with 8.8%, and a cumulative AUC of 4.87 compared with 4.90. Its cumulative AUC remained lower at every checkpoint, while the difference in best error narrowed and reversed after iteration 150.

Because the soft penalties enter the Informed objective additively and are never negative, its reported loss is an upper bound on the fitting error that the Unconstrained objective measures directly. The penalties are added after the weighted variable errors have been averaged, so their magnitude is comparable to that of the fitting error itself rather than negligible against it. The lower Informed values are therefore not produced by the differing objective, since that difference acts in the opposite direction. The same asymmetry makes the reversal in best error after iteration 150 uninformative, as the penalty-free fitting error behind the Informed value of 0.0885 may lie below the Unconstrained 0.0883. All differences between the two runs are small, with 0.6% separating their AUC values at iteration 200, and no repeated runs are available against which this margin could be judged.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table 6.5: Baseline parameter values [MHPN26] and the best parameters identified by each of the three optimisation runs, grouped by calibration unit. Informed: synthetic targets with domain-knowledge constraints; Observed: raw field measurements; Unconstrained: synthetic targets without the domain-knowledge penalty. All values share the baseline of the first column.

Group	Parameter	Baseline	Informed	Observed	Unconstrained
durum wheat (tall, medium)	bdSELBOT	12.00	13.88	11.92	12.18
	bdANTPM	30.00	30.92	30.43	26.54
	rgrHgt	0.0054	0.0048	0.0047	0.0047
	GNCmin	0.0170	0.0173	0.0169	0.0157
	GNCmax	0.0260	0.0264	0.0215	0.0307
	MEED	1200.00	1192.73	1130.75	1064.20
	GRTDP	30.00	32.90	33.45	27.88
faba bean (medium, early)	bdEMRR1	20.00	23.98	23.65	23.94
	bdR5R7	35.00	35.50	37.77	34.70
	rgrHgt	0.0022	0.0021	0.0020	0.0019
	GNCmin	0.0330	0.0302	0.0335	0.0293
	GNCmax	0.0650	0.0666	0.0623	0.0718
	MEED	1000.00	1105.29	944.13	980.10
	GRTDP	30.00	35.54	32.86	32.57
faba bean (tall, medium)	bdEMRR1	23.00	27.58	27.49	27.53
	bdR5R7	40.00	37.72	33.87	37.41
	rgrHgt	0.0020	0.0019	0.0019	0.0019
	GNCmin	0.0330	0.0264	0.0264	0.0265
	GNCmax	0.0650	0.0639	0.0578	0.0709
	MEED	1000.00	1087.54	1070.43	888.67
	GRTDP	30.00	33.37	34.73	28.81

Table 6.5 reports the shared baseline values and the best parameter set found by each run for the three calibration units. Several movements were common to all runs. The three rgrHgt values decreased, bdEMRR1 increased for both faba bean groups, and GNCmin for the tall, medium faba bean group decreased from 0.0330 to 0.0264–0.0265. Other parameters differed more strongly between objectives. For the same faba bean group, MEED increased to 1087.54 in the Informed run and 1070.43 in the Observed run but decreased to 888.67 in the Unconstrained run. Its GNCmax likewise decreased to 0.0639 and 0.0578 in the Informed and Observed runs but increased to 0.0709 without constraints. The runs therefore shared some directional adjustments but reached distinct parameter combinations. For the Informed and Unconstrained runs this divergence occurred at practically the same final error, and the two MEED values of the tall, medium faba bean group span roughly half of the admissible ±20% range. The objective consequently did not identify a unique parameter set within the budget, and the values in Table 6.5 are

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

reported as one member of a group of near-equivalent solutions rather than as a converged estimate.

As a descriptive test, simulations from the parameter set of the Informed run, which represents the full pipeline, were aligned with the observed, interpolated, and, where available, fully predicted values. Mean nRMSE was lower for observed than for interpolated values for WTOP (36.0% versus 51.1%), CNUP (48.4% versus 61.8%), PLHGT (30.9% versus 34.5%), and LAI (68.7% versus 83.7%). The corresponding standard deviations were 19.0 versus 23.4, 21.4 versus 27.3, 14.4 versus 14.3, and 27.4 versus 23.0 percentage points. For the fully predicted groups, mean nRMSE was 78.8% for CNUP ($n = 58$) and 71.4% for LAI ($n = 20$), with standard deviations of 74.4 and 18.9 percentage points. Thus, the final model outputs aligned more closely on average with the field observations than with the interpolated values for all four variables, while alignment with the fully predicted values differed by variable. No inferential test was performed, so these differences are reported descriptively.

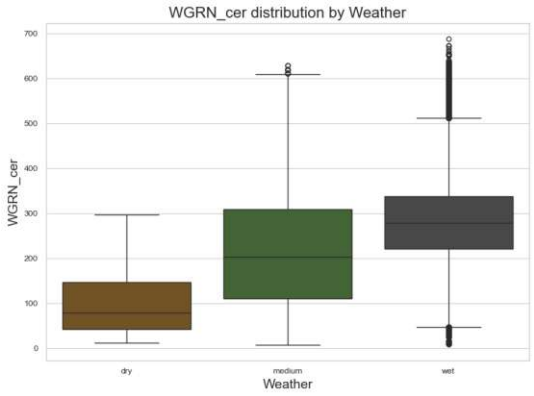
6.3 Sensitivity Analysis Findings

The sensitivity analysis spans 81 scenarios, formed by crossing three levels each of weather (dry, medium, wet), soil setting (SL60, SL120, and SL240), sowing date (5, 15, 25 October) and nitrogen supply (N0, N60, N120). Within every scenario, twelve crop parameters—six for the cereal and six for the legume component—were perturbed, and their influence was quantified for four outputs: NGRN and WGRN of each component, suffixed _cer or _leg.

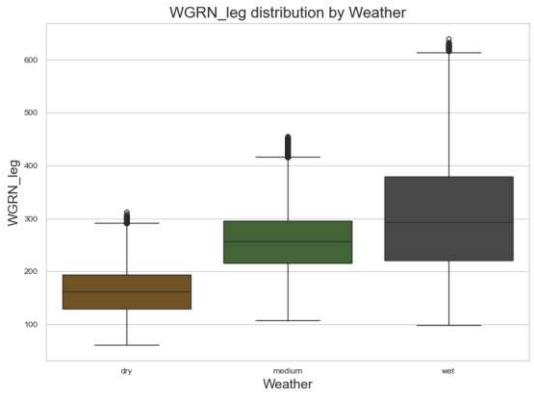
Before turning to the parameter rankings, the distributions of the simulated outputs are compared across weather and nitrogen settings (Figure 6.6). Median WGRN_cer increased from roughly 75 $\mathrm{g\,m^{-2}}$ under dry weather to about 200 $\mathrm{g\,m^{-2}}$ under medium and 280 $\mathrm{g\,m^{-2}}$ under wet conditions. WGRN_leg showed the same direction of change, with medians near 160, 255, and 295 $\mathrm{g\,m^{-2}}$. Median NGRN_cer increased from approximately 2.5 $\mathrm{g\,N\,m^{-2}}$ at N0 to 5.4 at N60 and 6.6 $\mathrm{g\,N\,m^{-2}}$ at N120. In contrast, median NGRN_leg remained close to 9 $\mathrm{g\,N\,m^{-2}}$ across the three nitrogen levels and decreased slightly at the higher levels. Thus, the scenario ensemble produced distinct component-specific responses to the selected weather and nitrogen settings.

Sensitivity is first examined at the whole-model level, taking WGRN_cer as a representative output and comparing its responsiveness across the scenario factors (Figure 6.7). Each panel overlays the three levels of weather, soil, sowing date, or nitrogen amount on the absolute IVARS-50 axes. Across the panels, the largest values generally occur for `heightMx_cer`, `heightMx_leg`, and `bdANTPM_cer`, although their magnitudes vary between settings. Wet weather produces the largest sensitivity values on the two height axes, approaching 800, whereas the values are substantially smaller under dry weather. The sensitivity magnitudes also increase from SL60 to SL240. Differences between sowing dates and nitrogen levels are visible but smaller for most of the displayed parameters. The broad set of influential parameters is therefore similar across these

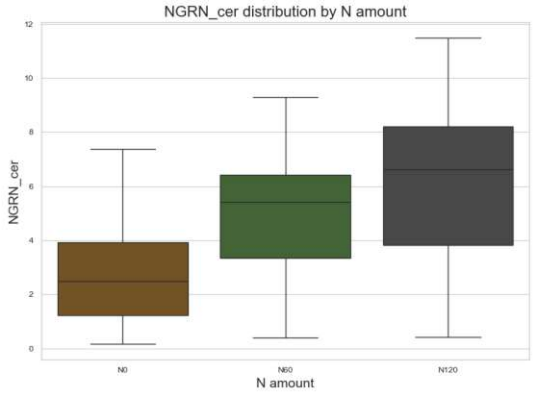
Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.



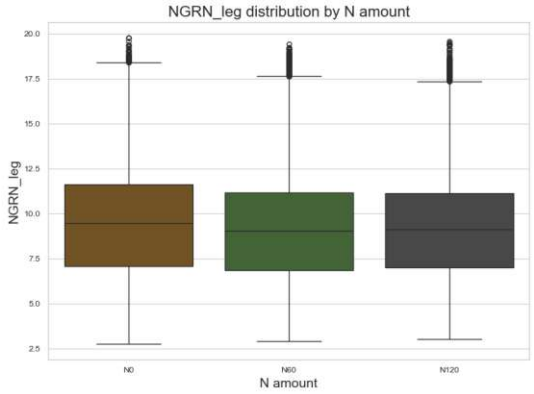
(a) Cereal grain weight by weather.



(b) Legume grain weight by weather.



(c) Cereal grain nitrogen content by nitrogen supply.



(d) Legume grain nitrogen content by nitrogen supply.

Figure 6.6: Distribution of simulated end-of-season outputs across the scenario ensemble. Cereal (left) and legume (right) WGRN are grouped by weather category (top row); cereal and legume NGRN are grouped by applied nitrogen (bottom row). Boxes show the interquartile range and median; whiskers the 1.5 IQR extent.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

aggregated comparisons, while their absolute IVARS-50 values depend on the surrounding scenario.

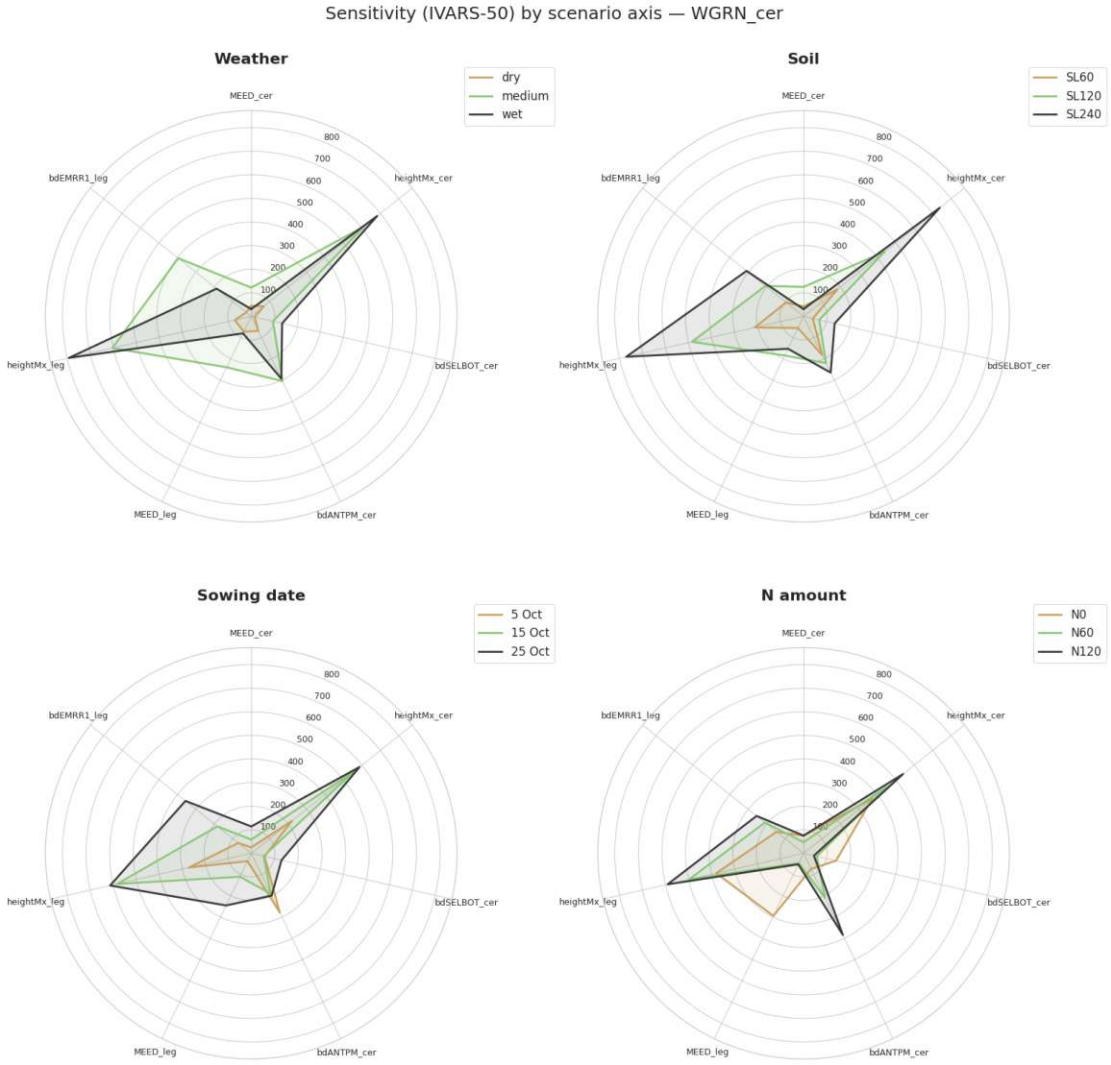


Figure 6.7: WGRN_cer sensitivity across the levels of weather, soil, sowing date, and nitrogen supply. Each panel overlays the three levels of one scenario factor on absolute IVARS-50 axes for the displayed parameters.

Aggregating over the full ensemble identifies the parameters with the highest mean normalised IVARS-50 values for each output (Figure 6.8). For the cereal outputs, `heightMx_cer` reaches approximately 0.55 and `bdANTPM_cer` 0.54–0.59, while `GNCmax_cer` contributes more strongly to NGRN_cer than to WGRN_cer. The legume outputs have their highest values for `bdEMRR1_leg` (0.76 for NGRN_leg and 0.67 for WGRN_leg) and `heightMx_leg` (0.62–0.68), followed by `MEED_leg` (0.34–0.44) and, for WGRN_leg, `GNCmin_leg` (0.43). Cross-component sensitivity is most pronounced for the two height

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

parameters: heightMx_cer has mean normalised values of 0.52–0.58 for the legume outputs, and heightMx_leg reaches 0.62–0.63 for the cereal outputs. Other cross-component values are generally smaller, although several legume parameters retain non-zero sensitivity for the cereal outputs. Among the main effects considered here, plant height therefore shows the clearest cross-component sensitivity.

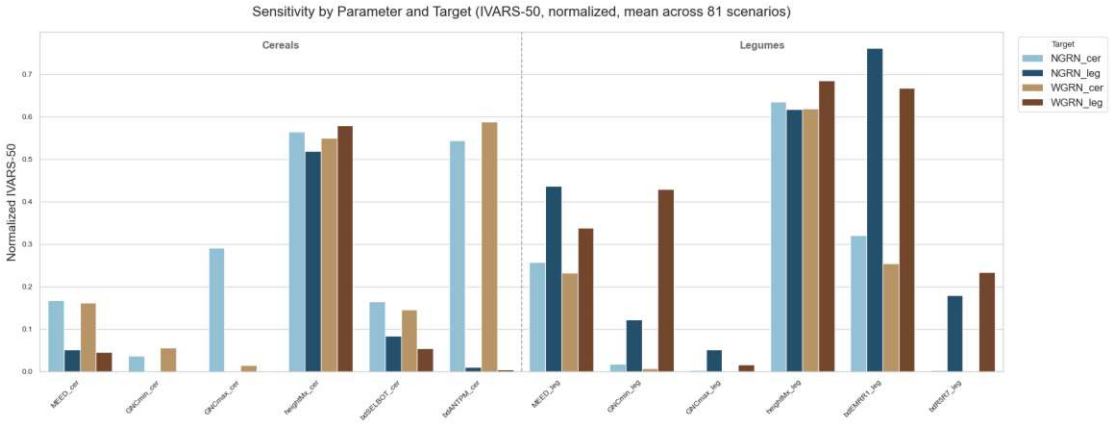


Figure 6.8: Mean normalised IVARS-50 sensitivity of each crop parameter, averaged across the 81 scenarios and grouped by component (cereal parameters on the left, legume parameters on the right). The four bars per parameter correspond to the cereal and legume NGRN and WGRN outputs.

The scenario-resolved heatmaps show changes in the within-scenario rankings that are not visible in the ensemble averages. NGRN_cer is examined across the soil–nitrogen plane under wet weather and a 15 October sowing (Figure 6.9). For SL60, bdANTPM_cer is the highest-ranked parameter at all three nitrogen levels, with GNCmax_cer also reaching row-normalised values of 0.75–0.97. For SL240, heightMx_cer becomes the highest-ranked parameter at N0 and N60, while the normalised value of bdANTPM_cer ranges from 0.01 to 0.35. At SL120, heightMx_leg reaches the row maximum in the displayed scenarios. Within this selected weather and sowing-date setting, the leading parameter for NGRN_cer therefore changes across the soil settings and nitrogen levels.

A second change in ranking appears for WGRN_cer across the weather–sowing-date plane, with nitrogen fixed at N60 and the medium soil setting (SL120) (Figure 6.10). Under dry conditions, bdANTPM_cer is the row maximum and MEED_cer has normalised values of 0.39–0.88. Under wet conditions, heightMx_leg becomes the row maximum for all three sowing dates and heightMx_cer remains between 0.60 and 0.89, while MEED_cer falls to approximately 0.01–0.03. The medium-weather rows lie between these patterns, with heightMx_cer reaching the row maximum for the 15 October sowing. In the displayed scenarios, the relative sensitivity of heightMx_leg thus increases from approximately 0.32 under dry weather to the highest within-row value under wet weather.

Extending this comparison across all soil settings shows a weather-dependent soil pattern in the sensitivity of grain biomass to MEED (Figure 6.11). For WGRN_cer, the dry-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

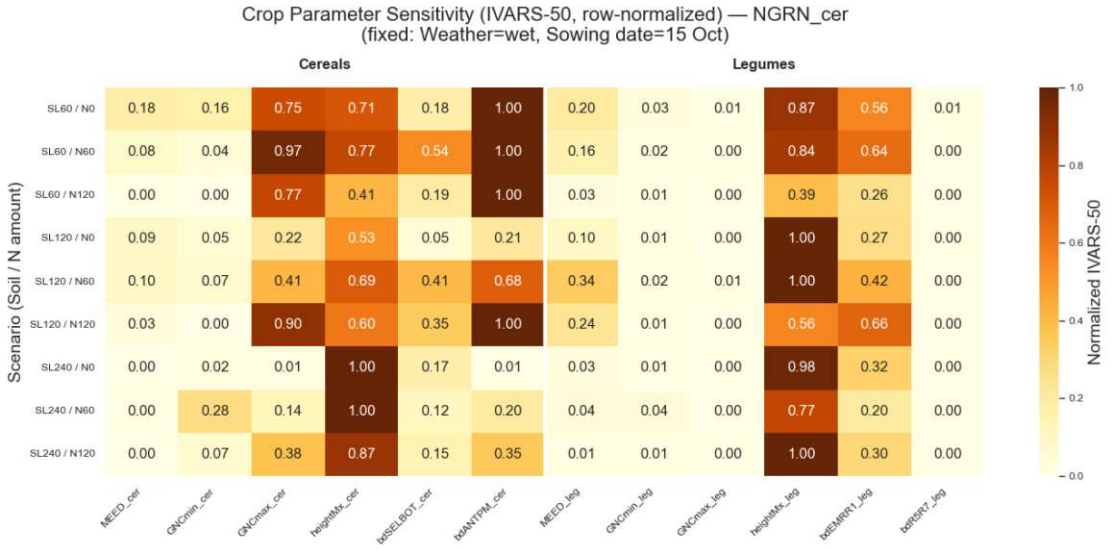


Figure 6.9: Scenario-resolved sensitivity of NGRN_cer across the nine soil–nitrogen combinations, with weather fixed to wet and sowing date to 15 October. Each row is normalised to its own maximum, so cell values range from 0 to 1.

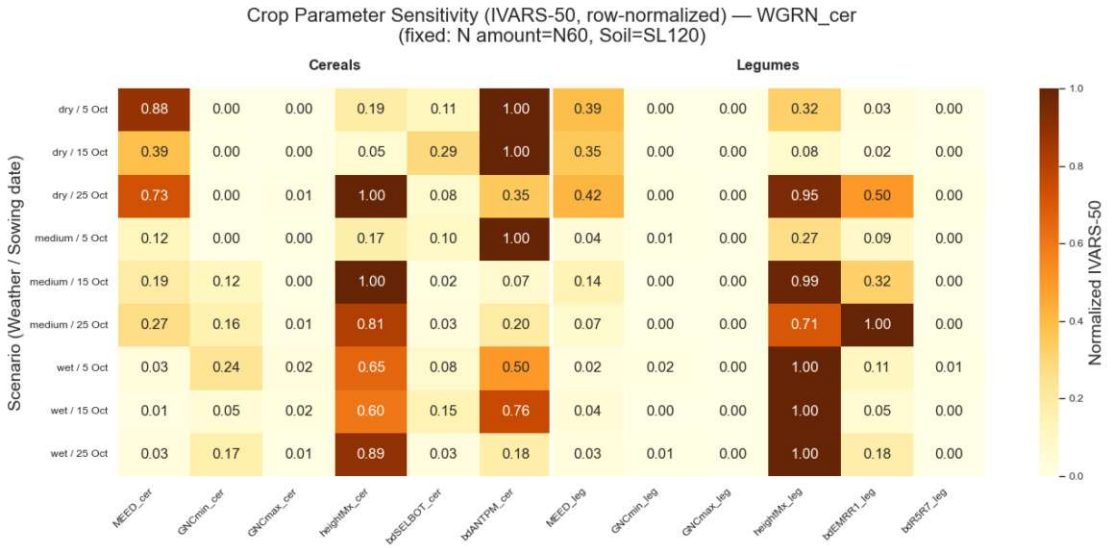


Figure 6.10: Scenario-resolved sensitivity of WGRN_cer across the nine weather–sowing-date combinations, with nitrogen fixed at N60 and the soil setting fixed to SL120. Each row is normalised to its own maximum as in Figure 6.9.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

minus-wet difference in the MEED_cer share of total IVARS-50 was 8.63 percentage points at SL60, 19.62 at SL120, and 1.09 at SL240. The corresponding differences for MEED_leg and WGRN_leg were 16.62, 18.49, and 4.79 percentage points. The range among soil settings was consequently widest under dry weather for both components: 20.21 percentage points for the cereal and 19.34 for the legume, compared with 5.61 and 4.23 under medium weather and 2.39 and 5.64 under wet weather. Thus, drought amplified the soil dependence of the relative MEED sensitivity, with a larger contrast at SL60 and SL120 than at SL240 and the maximum share occurring at SL120 for both components.

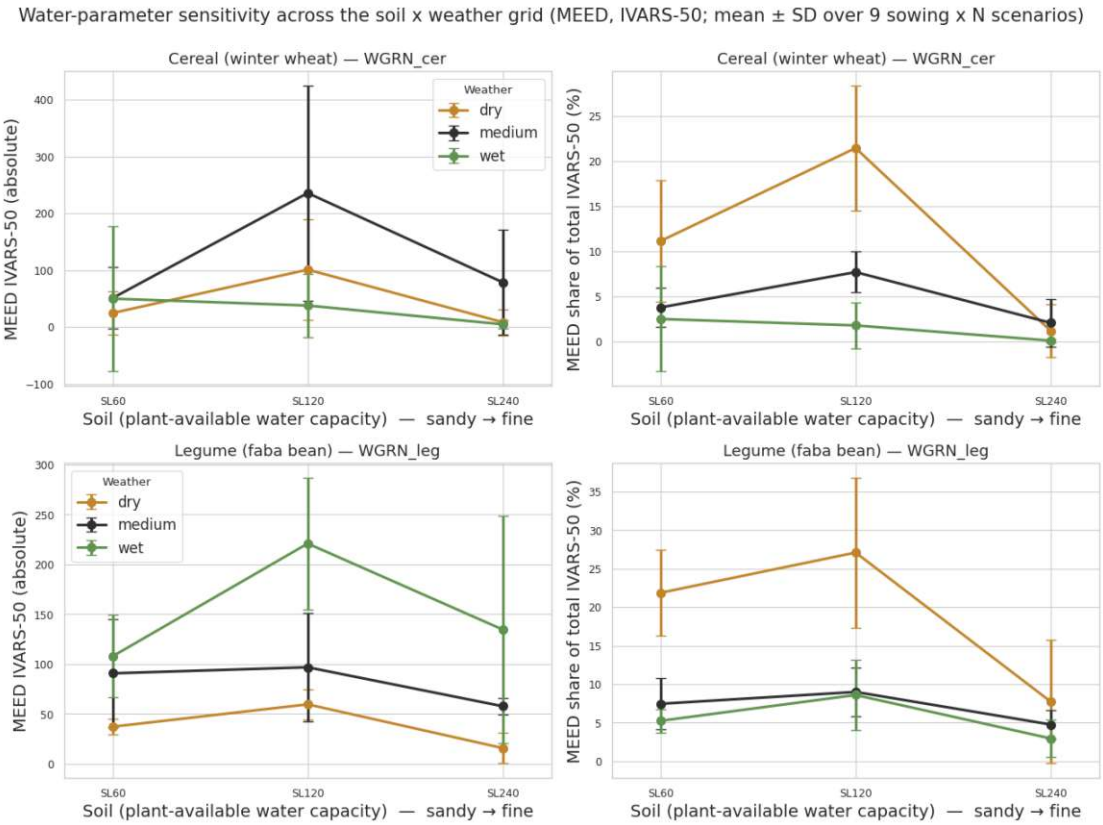


Figure 6.11: Sensitivity of cereal and legume grain biomass to the component-specific MEED parameter across soil and weather settings. The left panels show mean absolute IVARS-50 and the right panels the mean MEED share of total IVARS-50; error bars denote the standard deviation across the nine sowing-date and nitrogen combinations.

The simulated grain-biomass distributions show a less uniform soil–weather pattern (Figure 6.12). For the cereal, the relative separation between the dry and wet medians was largest at SL60 (85%), followed by SL120 (73%) and SL240 (62%). For the legume, it was largest at SL120 (46%), followed by SL240 (43%) and SL60 (37%). The relative range among the SL60 and SL240 medians was largest under dry weather for the cereal (69%,

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

compared with 39% under medium and 22% under wet weather), although its absolute range was slightly larger under medium weather. For the legume, the largest relative and absolute soil range occurred under wet weather (43% and 166.8 g m^{-2}, compared with 37% and 81.4 g m^{-2} under dry weather). The stronger drought-associated soil contrast in MEED sensitivity was therefore not reproduced uniformly in the simulated grain biomass of both components.

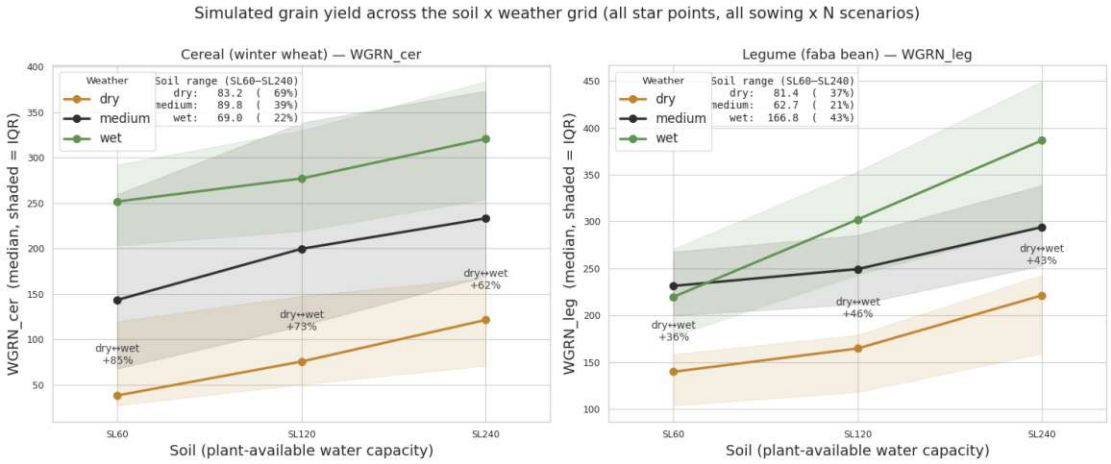


Figure 6.12: Simulated cereal and legume grain biomass across soil and weather settings. Points show medians and shaded bands the interquartile ranges across all STAR points and sowing-date–nitrogen combinations. Annotations give the relative dry–wet separation within each soil setting.

Finally, WGRN_cer sensitivity is compared across the IVARS-10, IVARS-30, and IVARS-50 perturbation scales (Figure 6.13). Each thin line traces one scenario across the twelve parameters, while the bold lines represent the mean profile for each scale. The same group of parameters, including heightMx_cer, bdANTPM_cer, and heightMx_leg, remains prominent across the mean profiles. Their relative emphasis is not identical, however: heightMx_cer becomes more pronounced at the larger scale, and smaller changes are visible for several other parameters. The broad group of influential parameters is therefore retained across scales, while the detailed ranking and relative sensitivity depend partly on the perturbation magnitude.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

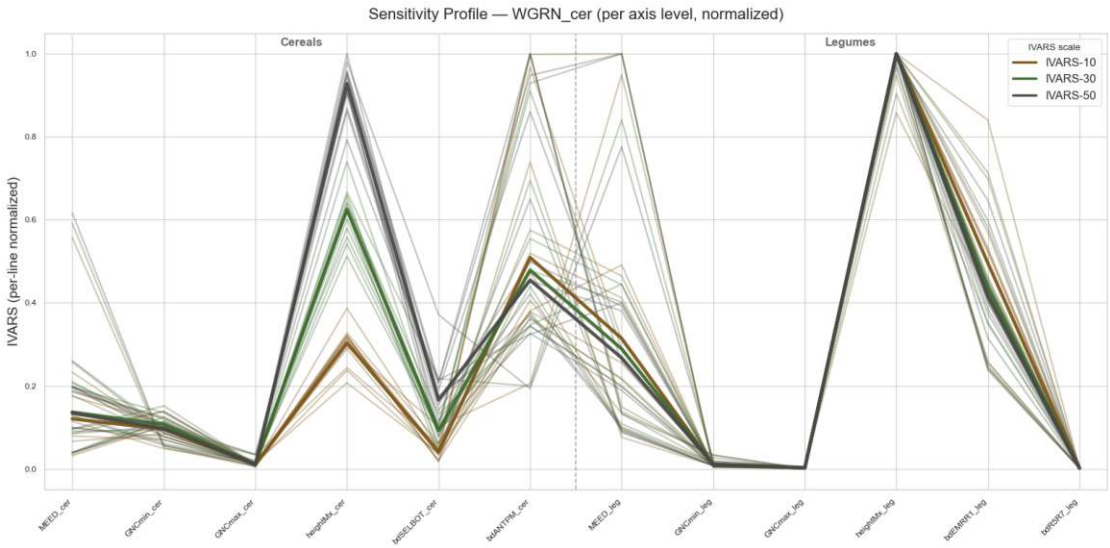


Figure 6.13: Parallel-coordinates plot of row-normalised WGRN_cer sensitivity across the twelve crop parameters. Thin lines show individual scenarios and bold lines the mean profiles for the IVARS-10, IVARS-30, and IVARS-50 scales.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

CHAPTER 7

Discussion and Conclusion

7.1 Answering the Research Questions

The three research questions are answered at two complementary levels. The literature review establishes how the challenges can be approached by identifying suitable methods, while the exemplary study examines whether the selected methods work when applied to *InterCrop* and the available intercropping dataset. The latter provides evidence for a representative but deliberately limited case, not a general validation across process-based intercropping models. The following discussion therefore separates methodological conclusions from observations that depend on this particular dataset and experimental design. The resulting limitations are considered in Section 7.3.

7.1.1 RQ1: Knowledge-Guided Preprocessing and Synthetic Data Generation

The first research question asks how qualitative knowledge constraints can support preprocessing and synthetic data generation under extreme data scarcity, and whether the resulting data are sufficient for model parameterisation. The literature review indicates that the central requirement is to encode the biological structure of the growing season instead of relying on statistical recombination alone. GAN-based approaches require substantially larger datasets, while SMOTE-type methods interpolate within the existing feature space and do not introduce an independent observation of crop development. Remote sensing was therefore selected as an additional source of site-level information, and RF regression was used to connect these signals with the sparse field observations. This provides a concrete answer to the methodological part of RQ1: where observations are too sparse to represent the underlying distribution, data generation should be anchored in an external measurement of crop development and constrained by known temporal and biological relations.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

In the implemented pipeline, this principle was applied at several stages rather than through one isolated constraint. Physical bounds removed invalid reflectance values, while the directional Hampel filter distinguished changes according to whether a spectral band or vegetation index was being processed. The subsequent double-sigmoid fit imposed a seasonal trajectory with ordered green-up and senescence phases, positive rates of change, and different baseline–peak relations for standard and inverse spectral patterns. Phenological ordering was retained in the RF models through the preceding and succeeding BBCH stages and their temporal distances. The two-stage reconstruction then used interpolation where field measurements provided temporal anchors and reserved fully predictive models for plots with two or fewer observations. Qualitative knowledge thus acted both as a filter on the admissible data and as structural guidance for filling the gaps that remained.

The results show that this approach generated enough data to execute the intended parameterisation experiment, although reconstruction quality varied by target. From an initial average of 4.12 temporal observations per variable and plot, the pipeline produced trajectories with approximately 76 points per season and more than 22,000 values per variable across 221 plots. The interpolation models reproduced WTOP, CNUP, and PLHGT with validation nRMSE values between 5.7% and 16.7% and R^2 values between 0.97 and 0.99. LAI was less reliable, with an nRMSE of 31.1%. The fully predictive models were also less accurate than the anchored interpolation models, particularly for CNUP and LAI, and the distributional comparison exposed clear weaknesses for underrepresented field pea and spring barley.

The endpoint analysis provided additional evidence that interpolation retained the management signal present in the observations. The elasticities with respect to nitrogen amount were similar between raw and interpolated endpoints for WTOP (0.157 and 0.146), PLHGT (-0.029 and -0.021), LAI (0.130 and 0.128), and CNUP (0.109 and 0.104). Their slopes and correlations were likewise similar. This agreement is plausible because the interpolation model used surrounding observations as target-derived features, strongly anchoring its endpoints to the measured trajectories. The fully predicted values were less consistent. Synthetic CNUP retained the positive direction but had a substantially larger elasticity of 0.286, while synthetic LAI showed a weak negative response rather than the positive response found in the raw and interpolated data. These synthetic comparisons are fragile because only 20 cases were available at N0 and N140, with no cases at N60 or N80. The high synthetic CNUP correlation of 0.872 may therefore reflect separation between two endpoint groups rather than a learned continuous nitrogen response.

These regressions remain descriptive rather than controlled estimates of management sensitivity. Nevertheless, the close raw–interpolated agreement supports the conclusion drawn from the literature review that, under severe data scarcity, existing observations should be enhanced rather than complete time series generated without target measurements. Purely predictive generation relies on relationships learned from a small dataset and can therefore produce biologically implausible trajectories, whereas interpolation integrates additional knowledge through observed anchor points that constrain the timing

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

and magnitude of crop development. The reconstructed dataset should consequently not be interpreted as a uniformly accurate substitute for field observations. It is better understood as a dense, knowledge-guided enhancement whose reliability depends on the variable, crop, and availability of temporal anchors.

Within the exemplary study, the reconstructed data were sufficient to provide a usable optimisation target. At the fixed budget of 200 iterations, the Informed run improved by 10.9% relative to its own baseline, compared with 8.7% for the Observed run, and lower losses were reached earlier throughout the run. This describes stronger progress relative to each run's own baseline, but it does not show that the reconstructed targets yielded more accurate parameters. The target datasets define different loss surfaces, and their absolute final losses and AUC values are therefore not directly comparable. For none of the runs can full convergence within the 200-iteration budget be guaranteed.

RQ1 therefore requires a differentiated answer. Within this case study, qualitative constraints could guide robust preprocessing and the enhancement of existing observations through directional cleaning rules, bounded seasonal curve shapes, phenological ordering, stage-aware features, and anchored interpolation. This part can be answered positively: it generated sufficient data to parameterise *InterCrop* and preserved the observed endpoint response to nitrogen closely for all four variables. By contrast, the study does not support a positive answer for the generation of complete time series without field anchor points. The original dataset was too small for the fully predictive models to reconstruct all variables and crop groups reliably. More training data might make this approach feasible, but under the conditions examined here, additional integrated knowledge in the form of target measurements or other biologically informative anchor points remains necessary. The comparatively weak reconstruction of LAI and underrepresented crops further shows where new observations are more valuable than further synthetic expansion.

7.1.2 RQ2: Hybrid Optimisation for Parameter Estimation

The second research question concerns how data-driven, heuristic, and domain-knowledge components can jointly improve parameter estimation in an underdetermined process-based model. The literature review provides the rationale for this combination. Bayesian optimisation is suitable when model evaluations are expensive, but a global surrogate becomes increasingly difficult to fit as dimensionality grows. Trust-region methods address this problem by concentrating local surrogate models on promising parts of the parameter space while retaining the ability to restart elsewhere. At the same time, optimisation against sparse observations can overfit parameters or favour biologically implausible simulations. The reviewed constrained approaches show how qualitative knowledge can be translated into penalties and inequalities, so that fit to the data is not the only criterion guiding the search. Hybridisation is therefore useful because the three elements address different parts of the problem rather than acting as interchangeable optimisation techniques.

This division of roles was retained in the exemplary study. The data-driven compo-

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

nent connected the parameterisation to the available cultivars and sites. For example, `heightMx` was derived from the median observed height of each cultivar group, while `rgrHgt` remained adjustable so that the model could reproduce the corresponding height development. The heuristic component was provided by TuRBO, which used local surrogate models, an adaptive trust region, and Thompson sampling to search the combined parameter space within a limited evaluation budget. Wrapping the model execution, loss calculation, and parameter update in one workflow also made the procedure reproducible and adaptable to different cultivar groups. Domain knowledge reduced the effective search problem by fixing parameters that could be estimated directly, bounding those that remained, and adding soft constraints for total phenological duration, the relation between minimum and maximum grain nitrogen concentration, and attainment of maximum plant height. These penalties retained flexibility around uncertain biological expectations rather than ruling candidates out.

Because the Informed and Unconstrained configurations fitted the same reconstructed targets, a similar final data fit is expected rather than evidence for an advantage of either configuration. Their reported objectives are not strictly identical, since the Informed objective also contains the soft penalties, as reflected by its slightly higher baseline error despite shared target data and parameter values. The relevant difference is how the runs approached their endpoints. Since the penalties are additive and never negative, the Informed loss is an upper bound on its fitting error, so the earlier lead and the consistently lower cumulative AUC of that run cannot be artefacts of the additional term. What they do not show is whether the constraints caused them. The margins are small, TuRBO depends on stochastic starting points and search decisions, and only one run per configuration is available, so run-to-run variation cannot be separated from a systematic effect. Faster convergence is therefore not claimed here, and repeated runs across several initialisations would be needed to test the pattern. The contribution of the constraints does not depend on such a convergence advantage. They favour parameter combinations that satisfy the encoded biological expectations without excluding the surrounding region from the search.

Evaluation of the best optimised model against the two target types provides a separate view of model fit. Both were used as optimisation targets, so what follows describes agreement with the calibration data rather than predictive performance. Mean nRMSE was lower for raw than for interpolated targets for all four time-series variables: 36.0% versus 51.1% for WTOP, 48.4% versus 61.8% for CNUP, 30.9% versus 34.5% for PLHGT, and 68.7% versus 83.7% for LAI. These correspond to relative reductions of 30%, 22%, 10%, and 18%, respectively. The consistent ordering shows that, in this evaluation, *InterCrop* agreed more closely on average with the raw observations than with the interpolated trajectories. Interpolation may remove observational noise and thereby improve agreement, but it may also attenuate real peaks, shift timing, or introduce structural bias from the reconstruction model. The results are compatible with the latter effects dominating here, although differences in temporal sampling and endpoint construction may also contribute and the diagnostics do not establish the cause.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

The post-hoc sensitivity analysis provides additional support for the choice of parameter space. Phenological development parameters, `MEED`, grain nitrogen concentration parameters, and plant height all influenced at least one of the analysed outputs, although their importance depended on the crop component and scenario. This indicates that the selected factors represented active controls of model behaviour rather than an arbitrary set of calibration variables. Plant height also showed the clearest cross-component sensitivity, supporting the decision to derive `heightMx` from observations and constrain its relationship with `rgrHgt`. Since the sensitivity analysis was conducted after optimisation, it can retrospectively support the parameter choices, but it did not reduce the search space or guide the optimiser in the experiment reported here.

RQ2 can therefore only be answered for the experiment conducted here. The hybrid configuration combined cultivar-specific observations, trust-region Bayesian optimisation, and biological constraints in one parameterisation procedure. With one run per configuration, it reached low-error regions marginally earlier than the otherwise equivalent Unconstrained run, while both attained nearly identical final errors on their shared targets. This margin is too small and too weakly replicated to indicate a convergence advantage and is reported as a description of the two observed trajectories. The defensible conclusion concerns the role of domain knowledge: hard constraints restricted the search to parameter combinations that satisfied the specified biological relations and avoided costly simulations of invalid candidates, while the soft penalties made deviations from less certain expectations costly. The procedure therefore ensured plausibility with respect to these encoded criteria, independently of whether its observed early advantage would persist under other random initialisations. Full convergence within 200 iterations cannot be guaranteed, no uncertainty intervals were obtained for the final parameters, and distinct objective configurations produced different parameter sets. Neither the accuracy nor transferability of the resulting parameters nor the relative optimisation performance of the three methods can be inferred beyond this experiment without repeated runs and validation on independent observations and further crop combinations.

7.1.3 RQ3: Sensitivity Analysis and Explainability for Model Transparency

The literature review answers the methodological part of RQ3 by identifying VARS as an efficient foundation on which an XAI-oriented analysis can be built. Traditional global SA establishes how strongly variation in each input contributes to variation in a model output over the analysed parameter space. This produces a global importance ranking, but the ranking alone gives little explanation of why a factor is important in one simulation context and less important in another. The XAI contribution in this study is a design principle rather than a second algorithm. That comparison was conducted in Section 2.4, with the aim of establishing what an explanation of a crop model would have to provide. What was carried over is the requirement that an explanation be conditional and scale-resolved, and this requirement determines how the analysis is organised, so that global importance can be examined conditionally. The inner VARS analysis measures

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

parameter sensitivity within one defined scenario, and the outer loop compares the resulting response surfaces across weather, soil, sowing date, and nitrogen supply. A parameter can therefore be described not only as influential, but as becoming more or less influential when a particular environmental or management condition changes. The IVARS-10, IVARS-30, and IVARS-50 indices add a second explanatory dimension by showing whether its importance is already expressed under small perturbations or becomes pronounced only over a wider part of its admissible range. Together, conditional scenarios and perturbation scales move the interpretation from a single statement that a parameter matters towards an account of where and at what scale it matters. Since IVARS quantifies output variability rather than the direction of an effect, these explanations do not by themselves establish whether increasing a parameter raises or lowers the output, nor do they establish a causal mechanism.

The exemplary study demonstrates this distinction for *InterCrop*. The mean IVARS-50 results identified plant height and crop-development parameters as the main controls across the scenario ensemble, but the scenario-resolved analysis showed that this ranking was not fixed. For cereal grain biomass at N60 and SL120, `bdANTPM_cer` was the leading parameter under dry weather, and the normalised sensitivity of `MEED_cer` ranged from 0.39 to 0.88. Under wet weather, both height parameters remained influential, while `MEED_cer` fell to approximately 0.01–0.03. Since `MEED` represents the maximum effective depth of water extraction by roots, its greater importance in the displayed dry scenarios agrees with Hao et al.'s finding that water-acquisition parameters, including maximum root depth, become more influential when water is limiting [HRW+24]. The change from development- and rooting-related sensitivity under dry weather towards height-related sensitivity under wet weather is also consistent with the general conclusion of Hao et al. [HRW+24] and Shen et al. [SZZ+25] that dominant controls shift with the limiting environmental conditions. These agreements provide additional evidence for the general agronomic plausibility of the results shared with monocrop models.

Cereal outputs were most sensitive to `heightMx_cer` and the anthesis-to-maturity duration `bdANTPM_cer`, with `GNCmax_cer` contributing more strongly to cereal grain nitrogen content than to grain biomass. Legume outputs were dominated by the emergence-to-bloom duration `bdEMRR1_leg` and `heightMx_leg`, followed by `MEED_leg` and, for grain biomass, `GNCmin_leg`. This pattern suggests that the timing of reproductive development and the parameters defining canopy development are central controls in the simulated system, while root water extraction and grain nitrogen limits affect particular component–output combinations. That reading follows the definitions of the parameters, but the IVARS values alone do not reveal the intermediate process chain. Time-resolved outputs would be needed to test whether the development parameters act through the duration of biomass and nitrogen accumulation, and whether the height parameters act through canopy competition.

The absolute IVARS-50 comparison also shows that sensitivity on the two height axes was greatest under wet weather and generally increased from SL60 to SL240, whereas aggregated differences among sowing dates and nitrogen levels were smaller for most

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

displayed parameters. For the general crop response, the simultaneous decline in the relative importance of MEED_cer under wet weather is consistent with the shift away from water-acquisition controls reported by Hao et al. [HRW+24] and Shen et al. [SZZ+25] when water is no longer limiting. Within the intercrop, wetter conditions and soil settings with greater PASW may additionally permit enough canopy growth for height-mediated light competition to become more strongly expressed. The sensitivity indices do not separate this intercropping pathway from the general release of water limitation, so targeted comparisons varying water storage, rooting depth, and canopy parameters separately are still required.

The analysis across the complete soil–weather grid strengthens but also qualifies this agreement with the reference studies. Under dry weather, the range among soil settings in the MEED share of total WGRN sensitivity reached 20.21 percentage points for the cereal and 19.34 for the legume, substantially more than under medium or wet weather. Soil conditions therefore affected the relative importance of root water extraction most strongly under drought. The dry–wet contrast was also larger in the light soil with 60 mm PASW than in the heavy soil with 240 mm PASW for both crop components, which is consistent with stronger water-related sensitivity in the light-soil setting. However, the largest amplification occurred in the medium soil with 120 mm PASW rather than at SL60, so the sensitivity pattern does not show a simple monotonic increase towards the lightest soil. For simulated grain biomass the cereal had its largest relative dry–wet separation at SL60, whereas the legume had its largest separation at SL120 and its largest soil-associated range occurred under wet weather. The results thus support a drought-dependent soil influence on MEED sensitivity and partly support a stronger light-soil response, but the latter does not extend uniformly across metrics and crop components.

The soil–nitrogen comparison provides a second example of this conditional behaviour, although it does not reproduce a simple shift towards nitrogen-related parameters under low nitrogen supply. Under wet weather, the leading parameter for cereal grain nitrogen content changed from bdANTPM_cer in SL60 to heightMx_cer in selected SL240 scenarios and heightMx_leg in SL120. GNCmax_cer, which directly defines an upper bound on grain nitrogen concentration, was also strongly associated with NGRN_cer in SL60, reaching normalised sensitivities of 0.75–0.97. This agrees at a general level with the importance of nitrogen-related controls for nitrogen-dependent crop responses reported by Hao et al. [HRW+24] and Shen et al. [SZZ+25]. The reported rankings do not, however, show nitrogen-related crop parameters consistently taking over at N0. The results therefore support context-dependent sensitivity, but not the more specific low-nitrogen ranking found in the reference studies.

Beyond these general plausibility checks, the intercropping analysis provides an additional level of transparency. Maximum plant height showed the clearest cross-component sensitivity: heightMx_cer had mean normalised sensitivities of 0.52–0.58 for the legume outputs, while heightMx_leg reached 0.62–0.63 for the cereal outputs. Several other legume parameters also retained non-zero sensitivity for cereal outputs. These

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

responses are consistent with mechanisms explicitly represented in *InterCrop*. The model partitions radiation according to the LAI and relative height of both components, while water and soil nitrogen are allocated according to the rooting of both species, as described in Section 3.3. The strong cross-component height sensitivity can therefore be connected to implemented light competition, and cross-component responses involving `MEED` or development timing can propagate through competition for water and soil nitrogen. This correspondence increases transparency because the sensitivity results expose the same links that are encoded in the model structure. It does not establish how much each pathway contributes, and targeted simulations are needed to isolate light, water, and nitrogen effects.

The output distributions provide another intercropping-specific result. Cereal grain nitrogen content increased from approximately 2.5 at N0 to 6.6 at N120, whereas legume grain nitrogen content remained near 9 and decreased slightly at the higher nitrogen levels. The relative stability of the legume is consistent with the model allowing legumes to fix atmospheric nitrogen even when no fertiliser is applied. At the same time, fixed nitrogen can facilitate cereal growth through the implemented transfer at the boundary between the crop strips, while both components continue to compete for soil nitrogen. These known mechanisms provide plausible routes for the contrasting component responses and make the result informative about model behaviour. The IVARS analysis does not quantify the separate contribution of fixation, transfer, and competition, and the slight decline in legume grain nitrogen content at higher fertilisation still requires targeted review of nitrogen uptake, fixation, transfer, and biomass allocation outputs.

The perturbation-scale comparison provides a further benefit of the study. Across IVARS-10, IVARS-30, and IVARS-50, `heightMx_cer`, `bdANTPM_cer`, and `heightMx_leg` remained prominent in the mean WGRN_cer profiles. This persistence indicates that the broad group of influential parameters was not created by selecting only one perturbation scale. However, their relative importance was not constant. `HeightMx_cer` became more pronounced as the integrated range increased, and smaller scale-dependent changes appeared for other parameters. The result may indicate that cereal maximum height has a limited effect near some sampled values but produces stronger changes once a sufficiently broad part of its range alters the height relation and light partitioning between components. VARS establishes the scale dependence but not this mechanism, and the current result does not locate a precise threshold. Targeted perturbations around the apparent transition are needed to determine whether the pattern reflects a threshold, gradual nonlinearity, or an interaction with companion-crop structure. This scale resolution is an added contribution to model transparency beyond the scenario-dependent global rankings.

RQ3 can thus be answered at both levels. The literature indicates how sensitivity analysis can be extended towards explainability by resolving global parameter importance according to environmental context and perturbation scale. The exemplary study shows that this approach can reveal such conditional behaviour in *InterCrop*, including a dry-weather increase and stronger soil dependence in the sensitivity of root water-extraction depth, weather-dependent changes in the dominant parameter group, and

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

a direct relationship between grain nitrogen concentration limits and grain nitrogen content sensitivity. Where comparison is possible, agreement with published APSIM and STICS behaviour provides additional evidence for the general agronomic plausibility of these results. The analysis also delivers two benefits beyond this comparison: it exposes cross-component responses associated with implemented competition and facilitation mechanisms, and it shows which sensitivity patterns persist or change across perturbation scales. These additional findings further increase model transparency because they connect output variability to the intercropping structure and reveal nonlinear scale dependence that a single global ranking would conceal. Beyond transparency, knowing which parameters govern which output under which soil and weather conditions is directly usable in practice. It indicates where careful determination of a parameter is worth the measurement effort and where a default value is unlikely to change the simulated outcome, so calibration budgets can be directed at the conditions in which a parameter actually drives the response. `MEED_cer` is the clearest case found here: its normalised sensitivity for cereal grain biomass reached 0.39–0.88 under dry weather but dropped to about 0.01–0.03 under wet weather, and the spread among soil settings was widest under drought. Anyone simulating a dry site on a light soil should therefore invest in a defensible rooting depth, whereas on a well-supplied site the same effort is better spent on plant height and the anthesis-to-maturity duration. These recommendations require further process-level review and targeted simulations before the contribution of individual mechanisms or the presence of thresholds can be established. None of these checks validates the model or its parameters against independent observations. The analysis examines the internal response of the simulator and is not dependent on the partially synthetic calibration targets, yet confidence remains limited to the behaviours tested across the selected parameter ranges and 81 scenarios.

7.2 Conclusion

This study developed and examined a knowledge-guided workflow for parameterising *InterCrop* with limited data. The results show that sparse observations can be enhanced through remote-sensing information, biologically constrained seasonal reconstruction, and anchored interpolation to provide usable optimisation targets. Fully predictive reconstruction was less reliable, particularly for LAI and underrepresented crops. The hybrid optimisation combined these data with trust-region Bayesian optimisation and biological constraints. Within the single experiment, both configurations ended at nearly the same error, and the small differences in their search behaviour do not establish faster convergence, while the constraints ensured that the encoded biological relations were respected. Scenario- and scale-resolved VARS analysis then made the resulting model behaviour more transparent by revealing environmental shifts in parameter importance, cross-component sensitivities, and responses that changed across perturbation scales.

Taken together, the findings demonstrate a coherent way to integrate qualitative knowledge throughout preprocessing, parameter estimation, and model analysis. They support the feasibility of the workflow for the examined durum wheat–faba bean case and provide

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

a basis for evaluating its application to other crop combinations and regions. The contribution is therefore an implemented methodological framework and an exemplary application whose broader reliability must be established through repeated optimisation and independent validation across further environments and crop combinations.

7.3 Limitations of the Study

The data basis imposes several limitations. Remote sensing data were available at site level but not for individual plots, and field pea and spring barley were underrepresented in the observations. Seasonal trajectories were fitted using quality weights rather than an upper-envelope approach. Interpolated and fully predicted synthetic values were also treated identically despite their different levels of observational support. Errors introduced during preprocessing and reconstruction may therefore have propagated into the optimisation.

The optimisation results are similarly restricted by the experimental design. Runs using reconstructed and raw observations are only conditionally comparable because the targets differ in temporal density, smoothness, and noise. The Informed and Unconstrained runs share their targets but not their objective, since only the former adds the soft penalties, so their loss and AUC values do not measure the same quantity either. Several thresholds and hyperparameters were selected from domain knowledge and preliminary experiments rather than systematic tuning, and cultivar groups lacked unambiguous boundaries. Moreover, only one run was conducted per configuration, so variability from stochastic initialisation and search decisions was not quantified. Full convergence within the budget of 200 iterations cannot be guaranteed. The optimiser was also run with a single trust region, so the global exploration that the multi-region variant provides through parallel local searches was traded for a larger share of the budget per surrogate. Together with the absence of uncertainty intervals and of observations withheld from the parameterisation, this means that the reported parameter sets are not established as robust estimates of the underlying cultivar traits. They are the best solutions found under a fixed budget for one target dataset each, and the near-identical final errors at clearly different parameter values indicate that other combinations would fit the same targets almost equally well. What the experiment supports concerns the workflow and its search behaviour, not the transferability of the individual values.

Finally, optimisation and sensitivity analysis were limited to durum wheat–faba bean, and transferability to spring barley–field pea was not evaluated. The sensitivity scenarios covered only three weather seasons and three soil settings. Sensitivity analysis was also conducted post hoc and could therefore not inform parameter selection before optimisation. The link to explainable AI likewise remained methodological: no post-hoc explanation algorithm was applied to *InterCrop*, so the argument that a sensitivity framework can satisfy explainability requirements rests on the comparison in Section 2.4 and on the resulting design of the analysis, not on measured agreement between two attribution methods.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

7.4 Future Research Directions

Future work should extend the present study by testing the transferability of the workflow and by investigating the model responses revealed by the sensitivity analysis in greater detail. An immediate extension of the present work is to apply the complete parameterisation workflow to spring barley–field pea. The observations for this crop combination have already passed through the preprocessing and reconstruction pipeline and are therefore available for optimisation. This extension would provide a first test of whether the selected methods remain applicable to crops with different developmental and competitive characteristics. It would also show which parts of the workflow can be transferred directly and where crop-specific constraints, parameter ranges, or cultivar groupings must be revised.

The optimisation experiments should also be repeated across multiple random seeds. Repeated runs of the Informed, Unconstrained, and Observed configurations would quantify variability arising from initialisation and search decisions. Distributions of convergence measures and estimated parameters could then be compared instead of relying on one trajectory per configuration. This would permit more defensible conclusions about convergence behaviour.

A broader sensitivity analysis should include more seasons and additional crop parameters beyond the twelve factors examined here. Conducting this analysis before optimisation would allow influential parameters to be prioritised, insensitive parameters to be fixed where appropriate, and parameter ranges to be adjusted according to the scales at which model responses emerge. The resulting information could reduce the effective search space and direct measurements towards parameters whose uncertainty most strongly affects the target outputs.

The relationship between sensitivity analysis and explainable AI could also be examined empirically rather than only methodologically. The simulation ensemble generated for the VARS analysis already provides paired parameter and output samples from which a surrogate model could be trained. Applying a post-hoc method such as SHAP to that surrogate and placing its attributions next to the IVARS indices would show whether the two families agree on the dominant factors and, more informatively, where they diverge. Such a comparison would also test whether local explanations recover the scenario dependence that the outer loop makes explicit here.

The observed cross-component sensitivities also need targeted investigation. Controlled simulations should vary parameters related to plant height, development, rooting depth, nitrogen fixation, and nitrogen transfer while recording intermediate outputs for light interception, water uptake, soil nitrogen uptake, fixation, and transfer. Such experiments could separate the contributions of above-ground competition, below-ground competition, and facilitation to the responses of each crop component. Combining these simulations with directional variograms and local response plots would also help determine whether the scale-dependent height response reflects a gradual nonlinearity or a threshold in the intercropping dynamics.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

More generally, *InterCrop* and the parameterisation workflow should be tested with independent observations from different sites, seasons, soils, management regimes, and cultivar combinations. Validation should include intermediate state variables such as biomass, LAI, plant height, and nitrogen uptake alongside end-of-season grain biomass and nitrogen content. This would test whether the estimated parameters reproduce crop development outside the data used for parameterisation and whether the environmental sensitivity patterns identified in this study remain plausible in other contexts.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

APPENDIX A

Appendix

Table A.1: Top ten feature importances for the BBCH Cereals and BBCH Legumes models.

BBCH Cereals		BBCH Legumes	
Feature name	Importance	Feature name	Importance
Season_Progress	0.192	Season_Progress	0.211
BBCHprev	0.155	integrated_CI	0.164
integrated_GNDVI	0.107	integrated_EVI	0.153
integrated_NDVI	0.080	integrated_GNDVI	0.101
integrated_EVI	0.074	integrated_CVI	0.070
integrated_MTVI2	0.066	integrated_MTVI2	0.068
integrated_CI	0.052	integrated_MCARI2	0.067
integrated_MCARI2	0.049	BBCHprev	0.056
integrated_CVI	0.046	integrated_SR	0.039
integrated_SR	0.042	integrated_NDVI	0.037

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table A.2: Top ten feature importances for the WTOP and CNUP predictive models.

WTOP		CNUP	
Feature name	Importance	Feature name	Importance
Season_Progress	0.133	Season_Progress	0.570
integrated_NIR	0.107	Cultivar_Encoded	0.074
integrated_MCARI2	0.107	N_Amount	0.065
integrated_CI	0.087	Crop_Encoded	0.063
integrated_MTVI2	0.085	integrated_Blue	0.057
integrated_NDWI	0.055	N_Fertilisation_Type_Encoded	0.036
delta_GNDVI	0.045	integrated_GNDVI	0.032
BBCHnext	0.044	integrated_NDWI	0.022
delta_NIR	0.041	integrated_MSI	0.017
BBCHprev	0.035	NDWI	0.011

Table A.3: Top ten feature importances for the PLHGT and LAI predictive models.

PLHGT		LAI	
Feature name	Importance	Feature name	Importance
Season_Progress	0.827	integrated_CI	0.575
integrated_NIR	0.055	Interrow	0.058
integrated_CI	0.025	Crop_Encoded	0.049
integrated_MCARI2	0.017	integrated_Red	0.048
integrated_MSI	0.011	delta_NIR	0.038
Crop_Encoded	0.008	NIR	0.037
integrated_NDWI	0.008	integrated_Green	0.034
delta_Blue	0.007	N_Amount	0.029
integrated_Green	0.006	SWIR2	0.029
integrated_SWIR2	0.006	Mixing_Pattern_Encoded	0.022

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table A.4: Top ten feature importances for the WGRN and NGRN predictive models. Only nine features remained for WGRN after backward feature selection.

WGRN		NGRN	
Feature name	Importance	Feature name	Importance
WTOP_predicted	0.671	Crop_Encoded	0.423
Crop_Encoded	0.070	WTOP_predicted	0.179
integrated_Blue	0.054	N_Amount	0.150
PLHGT_predicted	0.039	integrated_SWIR1	0.108
LAI_predicted	0.039	LAI_predicted	0.042
integrated_MSI	0.035	CNUP_predicted	0.030
Cultivar_Encoded	0.034	Cultivar_Encoded	0.015
CNUP_predicted	0.032	PLHGT_predicted	0.007
integrated_NDWI	0.025	integrated_GNDVI	0.007
—	—	integrated_NIR	0.006

Table A.5: Baseline parameter values [MHPN26] for durum wheat and faba bean crop groups. Parameter search spaces are defined as a ±20% range around each respective baseline value.

Durum Wheat		**Faba Bean**			
Parameter	**Tall, Medium**	**Parameter**	**Medium, Early**	**Medium, Tall**	**Unit**
bdSELBOT	9.6 – 14.4	bdEMRR1	16 – 24	18.4 – 27.6	bd
bdANTPM	24 – 36	bdR5R7	28 – 42	32 – 48	bd
rgrHgt	0.00432 – 0.00648	rgrHgt	0.00176 – 0.00264	0.0016 – 0.0024	–
GNCmin	0.01376 – 0.02064	GNCmin	0.0264 – 0.0396	0.0264 – 0.0396	g/g
GNCmax	0.0208 – 0.0312	GNCmax	0.052 – 0.078	0.052 – 0.078	g/g
MEED	960 – 1440	MEED	800 – 1200	800 – 1200	mm
GRTDP	24 – 36	GRTDP	24 – 36	24 – 36	mm/d

Table A.6: Baseline values and factor ranges of the twelve crop parameters perturbed in the sensitivity analysis (Section 5.3). Each factor was drawn from a uniform distribution spanning ±20% around its baseline, rounded to the precision reported here.

Cereal			**Legume**			
Parameter	**Baseline**	**Range**	**Parameter**	**Baseline**	**Range**	**Unit**
bdSELBOT	12	10 – 14	bdEMRR1	23	18 – 28	bd
bdANTPM	30	24 – 36	bdR5R7	40	32 – 48	bd
heightMx	90	72 – 108	heightMx	100	80 – 120	cm
GNCmin	0.0172	0.014 – 0.020	GNCmin	0.033	0.026 – 0.040	g/g
GNCmax	0.026	0.021 – 0.031	GNCmax	0.065	0.052 – 0.078	g/g
MEED	1200	960 – 1440	MEED	1000	800 – 1200	mm

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Table A.7: Numerical settings of the constrained loss function described in Section 5.2. Soft penalties take the form $0.05 \cdot v^2$, where v denotes the relative violation of the respective bound. The termination value in the last row documents the implemented mechanism; it was not triggered by any candidate within the search bounds used here.

Setting	**Value**
Variable weights	
`WGRN` / `NGRN`	0.350 / 0.300
`BBCH`	0.150
`WTOP` / `CNUP`	0.075 / 0.075
`LAI` / `PLHGT`	0.025 / 0.025
Error normalisation	
`WGRN`, `NGRN`	relative deviation of the final value
`BBCH`	RMSE in days divided by 20 d
`BBCH` stage weights (9/30/65/89)	2 / 2 / 3 / 1
`WTOP`, `CNUP`, `LAI`, `PLHGT`	RMSE divided by the reference mean
Aggregation	weighted sum divided by the number of variables available for the plot
Soft penalties	
Penalty weight	0.05
Biological days, durum wheat (tall, medium)	42 bd, admissible 37.8–46.2
Biological days, faba bean (medium, early)	55 bd, admissible 49.5–60.5
Biological days, faba bean (tall, medium)	63 bd, admissible 56.7–69.3
Minimum GNC gap, durum wheat	0.0044 g/g
Minimum GNC gap, faba bean	0.0160 g/g
Height target	$0.95 \cdot$ `heightMx`
Loss returned on a hard violation	10^4 plus the accumulated soft penalty

Table A.8: Optimiser configuration used for the parameterisation runs described in Section 5.2. Settings not listed here were left at the defaults of the reference implementation.

Setting	**Value**
Search dimension d	21 (7 parameters $\times$ 3 calibration units)
Evaluation budget	200
Initial design points	20
Batch size	1
Surrogate model	Gaussian process, Matérn-5/2 kernel
Lengthscales	automatic relevance determination
Candidate selection	Thompson sampling
Random seed	42

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Overview of Generative AI Tools Used

During the creation of this thesis the following AI tools were used as assistants:

- **Google Gemini** was used to brainstorm ideas and to break down methodological concepts into smaller steps.
- **Perplexity AI** supported the literature search for the related-work chapter and the analysis of academic papers.
- **GitHub Copilot** was used during implementation to assist with code development and syntax checking. It was also used to reformulate text passages, support spelling and language checks, and assist with writing and checking LaTeX code.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

List of Figures

4.1 Geographic distribution of the 37 field trial sites across Europe. Circle size is proportional to the number of trials at each location. Adapted from [GMB+24]. 26

4.2 Boxplots of growing period length (days from sowing to harvest) for each cultivar. Left: durum wheat and faba bean; right: spring barley and field pea. Cultivars with markedly different growing period distributions from those used in the manual parameterisation were excluded from the dataset. . . . 28

4.3 Distribution of the number of observations per plot for each of the seven selected variables in the filtered dataset. Boxes show the interquartile range; annotations report the mean number of observations and the total count of plots with at least one observation. 29

5.1 Fitted double sigmoid curve for a selected location representing the smoothing pipeline. Green dots are observed satellite values, and their size is relative to their quality weight. Red crosses indicate outliers identified by the Hampel filter. The blue line represents the fitted curve, and the violet squares are derived synthetic values at the dates of observed field measurements. . . . 36

5.2 Number of BBCH measurements recorded for each crop species at specific developmental milestones. The stacked bar chart distinguishes between durum wheat (light brown), spring barley (dark brown), faba bean (light green), and field pea (dark green). 39

5.3 Comparison of different modelling designs for time-series forecasting. The solid lines denote the predicted trajectories generated by the models (light brown for standard, green for incremental, and dark brown for interpolated), while the black dots indicate the actual observed field values. The interpolated sequence adheres most closely to the observations, whereas the incremental approach yields the smoothest overall curve. 40

5.4 Boxplots of measured plant height for each cultivar, used to inform the cultivar grouping. Faba bean cultivar Irena stands out as markedly shorter than the remaining faba bean cultivars, motivating its separation into a distinct group. 45

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

5.5 Validation of the derived `rgrHgt` baseline against the full 30-year weather record, shown for the late-maturing cultivar type under the medium soil setting (SL120; 120 mm PASW). Each panel plots the simulated plant height trajectories over the season, with the bold line marking the median across all 30 years and the shaded band spanning the 10th to 90th percentile; the dashed line indicates the prescribed maximum height (`heightMx`). Panels are arranged by component (top: faba bean, bottom: durum wheat) and cropping system (left: intercrop, right: monocrop). 46

5.6 An illustrative example of directional and integrated variograms showing IVARS-50 for two inputs. The input range is two, resulting in a range of perturbation scales (h) of one, corresponding to half of the input range. Adopted from [RG16]. 49

6.1 Synthetic WTOP trajectories for three spring-barley plots. Lines represent the generated trajectories and dots the observed measurements. The light-brown and green series received 50 kg N ha^{-1}, whereas the dark-brown series received no applied nitrogen. 54

6.2 Actual versus predicted values for the time-series variables. The red dotted line marks the 1:1 relationship ($y = x$). WTOP predictions (top left, light green) show moderate dispersion. CNUP predictions (top right, grey) tend to underestimate higher observations, while PLHGT predictions (bottom left, brown) lie closest to the 1:1 line. LAI predictions (bottom right, dark green) remain below approximately 4 m^2 m^{-2} even where higher values were observed. 55

6.3 Kernel Density Estimation (KDE) of observed values (grey), interpolated values (brown), and synthetic predictions (green). Rows correspond to WTOP, CNUP, PLHGT, and LAI, and columns correspond to crops. The distributions overlap closely for combinations such as durum wheat WTOP, whereas field pea LAI shows little overlap. 56

6.4 Actual versus predicted values for the end-of-season (EOS) variables WGRN (left, green) and NGRN (right, grey). The red dotted line marks the 1:1 relationship ($y = x$). Both models show some underestimation at high observed values. 57

6.5 Loss values recorded over the course of the optimisation for the three runs: the Informed run (synthetic targets with domain-knowledge constraints), the Observed run (raw field measurements), and the Unconstrained run (synthetic targets without the domain-knowledge penalty). The variation between successive iterations reflects the evaluation of different candidate parameter sets. 59

6.6 Distribution of simulated end-of-season outputs across the scenario ensemble. Cereal (left) and legume (right) WGRN are grouped by weather category (top row); cereal and legume NGRN are grouped by applied nitrogen (bottom row). Boxes show the interquartile range and median; whiskers the 1.5 IQR extent. 62

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6.7 WGRN_cer sensitivity across the levels of weather, soil, sowing date, and nitrogen supply. Each panel overlays the three levels of one scenario factor on absolute IVARS-50 axes for the displayed parameters. 63
6.8 Mean normalised IVARS-50 sensitivity of each crop parameter, averaged across the 81 scenarios and grouped by component (cereal parameters on the left, legume parameters on the right). The four bars per parameter correspond to the cereal and legume NGRN and WGRN outputs. 64
6.9 Scenario-resolved sensitivity of NGRN_cer across the nine soil–nitrogen combinations, with weather fixed to wet and sowing date to 15 October. Each row is normalised to its own maximum, so cell values range from 0 to 1. . 65
6.10 Scenario-resolved sensitivity of WGRN_cer across the nine weather–sowing-date combinations, with nitrogen fixed at N60 and the soil setting fixed to SL120. Each row is normalised to its own maximum as in Figure 6.9. . . . 65
6.11 Sensitivity of cereal and legume grain biomass to the component-specific MEED parameter across soil and weather settings. The left panels show mean absolute IVARS-50 and the right panels the mean MEED share of total IVARS-50; error bars denote the standard deviation across the nine sowing-date and nitrogen combinations. 66
6.12 Simulated cereal and legume grain biomass across soil and weather settings. Points show medians and shaded bands the interquartile ranges across all STAR points and sowing-date–nitrogen combinations. Annotations give the relative dry–wet separation within each soil setting. 67
6.13 Parallel-coordinates plot of row-normalised WGRN_cer sensitivity across the twelve crop parameters. Thin lines show individual scenarios and bold lines the mean profiles for the IVARS-10, IVARS-30, and IVARS-50 scales. . . 68

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

List of Tables

4.1 Number of trials in the full dataset of Gaudio et al. [GMB+24] that reported each of the seven selected plant variables. 27
4.2 Number of unique plots per crop group in the filtered dataset. 28
4.3 Overview of selected *InterCrop* parameters relevant for the parameterisation process. Parameters actively adjusted during the optimisation are indicated in **bold**, whereas the remaining parameters provide context or set upper boundaries. 31
4.4 Summary of the calculated vegetation indices and their corresponding formulas. The cited secondary literature outlines the rationale for their selection, containing references to the foundational primary studies that introduced these formulas. 32

5.1 Spectral band correspondence across Landsat sensor generations (TM, ETM+, and OLI). 34
5.2 Reference maximum plant height (`heightMx`) derived as the median observed plant height within each cultivar group. 45
5.3 Input parameters for the three soil settings used in the sensitivity analysis. 50

6.1 Average absolute difference in days between observed and predicted dates for key BBCH developmental stages. The performance metrics include the mean difference, standard deviation, and the corresponding number of observations (N) for cereals and legumes. The values represent the averages of all validation sets included in the robustness check. 53
6.2 Validation performance metrics for the time-series interpolation models. RMSE and MAE are expressed in the unit of the respective variable. . . . 54
6.3 Validation performance metrics for the predictive models, including time-series forecasting and end-of-season (EOS) points. 57
6.4 Convergence comparison of the three optimisation runs at selected iteration counts. For each run the best error achieved so far, the relative improvement over the iteration 0 baseline (Equation 5.10), and the cumulative area under the convergence curve (Equation 5.11) are reported. 58

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

6.5 Baseline parameter values [MHPN26] and the best parameters identified by each of the three optimisation runs, grouped by calibration unit. Informed: synthetic targets with domain-knowledge constraints; Observed: raw field measurements; Unconstrained: synthetic targets without the domain-knowledge penalty. All values share the baseline of the first column. 60

A.1 Top ten feature importances for the BBCH Cereals and BBCH Legumes models. 81
A.2 Top ten feature importances for the WTOP and CNUP predictive models. 82
A.3 Top ten feature importances for the PLHGT and LAI predictive models. . 82
A.4 Top ten feature importances for the WGRN and NGRN predictive models. Only nine features remained for WGRN after backward feature selection. 83
A.5 Baseline parameter values [MHPN26] for durum wheat and faba bean crop groups. Parameter search spaces are defined as a ±20% range around each respective baseline value. 83
A.6 Baseline values and factor ranges of the twelve crop parameters perturbed in the sensitivity analysis (Section 5.3). Each factor was drawn from a uniform distribution spanning ±20% around its baseline, rounded to the precision reported here. 83
A.7 Numerical settings of the constrained loss function described in Section 5.2. Soft penalties take the form $0.05 \cdot v^2$, where v denotes the relative violation of the respective bound. The termination value in the last row documents the implemented mechanism; it was not triggered by any candidate within the search bounds used here. 84
A.8 Optimiser configuration used for the parameterisation runs described in Section 5.2. Settings not listed here were left at the defaults of the reference implementation. 84

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Bibliography

[AAW+21] F. Akhavizadegan, J. Ansarifar, L. Wang, I. Huber, and S. V. Archontoulis. A time-dependent parameter estimation framework for crop modeling. *Scientific Reports*, 11(1):11437, June 2021.

[ABV24] Y. Akkem, S. K. Biswas, and A. Varanasi. A comprehensive review of synthetic data generation in smart farming by using variational autoencoder and generative adversarial network. *Engineering Applications of Artificial Intelligence*, 131:107881, May 2024.

[ACB17] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 06–11 Aug 2017.

[BAH+06] P. S. A. Beck, C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore. Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sensing of Environment*, 100(3):321–334, February 2006.

[BD16] M. Belgiu and L. Drăguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, April 2016.

[BWVDW+21] H. N. C. Berghuijs, M. Weih, W. Van Der Werf, A. J. Karley, E. Adam, Á. M. Villegas-Fernández, L. P. Kiær, A. C. Newton, C. Scherber, S. Tavoletti, and G. Vico. Calibrating and testing APSIM for wheat-faba bean pure cultures and intercrops across Europe. *Field Crops Research*, 264:108088, May 2021.

[CBHK02] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16:321–357, June 2002.

[CJT+04] J. Chen, P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh. A simple method for reconstructing a high-quality NDVI time-series data

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

set based on the Savitzky–Golay filter. *Remote Sensing of Environment*, 91(3-4):332–344, June 2004.

[DN20] R. A. Davis and M. S. Nielsen. Modeling of time series using random forests: theoretical developments. *Electronic Journal of Statistics*, 14(2):3654–3684, 2020.

[EPG+19] D. Eriksson, M. Pearce, J. Gardner, R. D. Turner, and M. Poloczek. Scalable global optimization via local bayesian optimization. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[FG20] P. Filzmoser and M. Gregorich. Multivariate Outlier Detection in Applied Data Analysis: Global, Local, Compositional and Cellwise Outliers. *Mathematical Geosciences*, 52(8):1049–1066, November 2020.

[GMB+24] N. Gaudio, R. Mahmoud, L. Bedoussac, E. Justes, E.-P. Journet, C. Naudin, H. Hauggaard-Nielsen, E. S. Jensen, E. Pelzer, G. Corre, B. Kammoun, L. Viguier, R. Barillot, A. Couëdel, P. Hinsinger, and P. Casadebaig. A global dataset gathering 37 field experiments involving cereal-legume intercrops and their corresponding sole crops. Dataset, Zenodo, version 2.0.0, November 2024. `https://doi.org/10.5281/zenodo.14160885`.

[GSC+23] A. Gobin, A.-H. M. Sallah, Y. Curnel, C. Delvoye, M. Weiss, J. Wellens, I. Piccard, V. Planchon, B. Tychon, J.-P. Goffart, and P. Defourny. Crop Phenology Modelling Using Proximal and Satellite Sensor Data. *Remote Sensing*, 15(8), April 2023.

[GVBVW+15] P. Grassini, L. G. J. Van Bussel, J. Van Wart, J. Wolf, L. Claessens, H. Yang, H. Boogaard, H. De Groot, M. K. Van Ittersum, and K. G. Cassman. How good is good enough? Data requirements for reliable crop yield simulations and yield-gap analysis. *Field Crops Research*, 177:49–63, June 2015.

[Ham74] F. R. Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393, 1974.

[HLJ25] T. He, M. Li, and D. Jin. Deep learning-based time series prediction for precision field crop protection. *Frontiers in Plant Science*, 16, June 2025.

[HO01] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[HRW+24] S. Hao, D. Ryu, A. W. Western, E. Perry, H. Bogena, and H. J. H. Franssen. Global sensitivity analysis of APSIM-wheat yield predictions to

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

model parameters and inputs. *Ecological Modelling*, 487:110551, January 2024.

[HWWR17] D. He, E. Wang, J. Wang, and M. J. Robertson. Data requirement for effective calibration of process-based crop models. *Agricultural and Forest Meteorology*, 234–235:136–148, March 2017.

[JCD+20] E. S. Jensen, I. R. Chongtham, N. R. Dhamala, C. Rodriguez, N. Carton, and G. Carlsson. Diversifying European agricultural systems by intercropping. *International Journal of Agriculture and Natural Resources*, 47(3):174–186, December 2020.

[JE04] P. Jönsson and L. Eklundh. TIMESAT—a program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, 30(8):833–845, October 2004.

[JKL+18] X. Jin, L. Kumar, Z. Li, H. Feng, X. Xu, G. Yang, and J. Wang. A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, 92:141–152, January 2018.

[KML+15] A. Kross, H. McNairn, D. Lapen, M. Sunohara, and C. Champagne. Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *International Journal of Applied Earth Observation and Geoinformation*, 34:235–248, February 2015.

[KPG+18] P. Kumar, R. Prasad, D. K. Gupta, V. N. Mishra, A. K. Vishwakarma, V. P. Yadav, R. Bala, A. Choudhary, and R. Avtar. Estimation of winter wheat crop growth parameters using time series Sentinel-1A SAR data. *Geocarto International*, 33(9):942–956, September 2018.

[KSS23] J. Księżak, M. Staniak, and J. Stalenga. Restoring the Importance of Cereal-Grain Legume Mixtures in Low-Input Farming Systems. *Agriculture*, 13(2):341, January 2023.

[LZH+19] N. Lu, J. Zhou, Z. Han, D. Li, Q. Cao, X. Yao, Y. Tian, Y. Zhu, W. Cao, and T. Cheng. Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods*, 15(1):17, December 2019.

[MDPH18] E. D. Mitra, R. Dias, R. G. Posner, and W. S. Hlavacek. Using both qualitative and quantitative data in parameter identification for systems biology models. *Nature Communications*, 9(1):3901, September 2018.

[MEB+21] A. M. Manschadi, J. Eitzinger, M. Breisch, W. Fuchs, T. Neubauer, and A. Soltani. Full Parameterisation Matters for the Best Performance of

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Crop Models: Inter-comparison of a Simple and a Detailed Maize Model. *International Journal of Plant Production*, 15(1):61–78, March 2021.

[MHAN25] A. M. Manschadi, J. Heurix, I. B. Ajmera, and T. Neubauer. Conceptualisation and Implementation of the InterCrop Model. Unpublished manuscript, provided by the authors, June 2025.

[MHPN26] A. M. Manschadi, J. Heurix, J. Pfennigbauer, and T. Neubauer. Simulating competition, facilitation, and yield dynamics in cereal-legume intercrops – The SSM-InterCrop model. In iCROPM, editor, *iCROPM2026 Abstract Book*, 2026.

[MPF+22] A. M. Manschadi, M. Palka, W. Fuchs, T. Neubauer, J. Eitzinger, M. Oberforster, and A. Soltani. Performance of the SSM-iCrop model for predicting growth and nitrogen dynamics in winter wheat. *European Journal of Agronomy*, 135:126487, April 2022.

[NAN+22] D. T. Neves, J. Alves, M. G. Naik, A. J. Proença, and F. Prasser. From Missing Data Imputation to Data Generation. *Journal of Computational Science*, 61:101640, May 2022.

[NKM25] T. Neubauer, M. Koelbing, and A. M. Manschadi. Scalable Prediction of Crop BBCH Stages Using Sentinel-2 Plant Phenology Index and Machine Learning. Preprint, available at SSRN, July 2025. `https://doi.org/10.2139/ssrn.5360947`.

[NTND21] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab. Machine Learning for Anomaly Detection: A Systematic Review. *IEEE Access*, 9:78658–78700, 2021.

[PDLF+22] F. R. da S. Pereira, J. P. De Lima, R. G. Freitas, A. A. Dos Reis, L. R. do Amaral, G. K. D. A. Figueiredo, R. A. C. Lamparelli, and P. S. G. Magalhães. Nitrogen variability assessment of pasture fields under an integrated crop-livestock system using UAV, PlanetScope, and Sentinel-2 data. *Computers and Electronics in Agriculture*, 193:106645, February 2022.

[PHC17] Z. Pan, Y. Hu, and B. Cao. Construction of smooth daily remote sensing time series data: A higher spatiotemporal resolution perspective. *Open Geospatial Data, Software and Standards*, 2(1):25, December 2017.

[PJO22] A. Petrovskaia, R. Jana, and I. Oseledets. A Single Image Deep Learning Approach to Restoration of Corrupted Landsat-7 Satellite Images. *Sensors (Basel, Switzerland)*, 22(23):9273, November 2022.

[PRD+25] B. Panigrahi, S. Razavi, L. E. Doig, B. Cordell, H. V. Gupta, and K. Liber. On Robustness of the Explanatory Power of Machine Learning

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

Models: Insights From a New Explainable AI Approach Using Sensitivity Analysis. *Water Resources Research*, 61(3):e2024WR037398, 2025.

[PRD+26] B. Panigrahi, S. Razavi, L. E. Doig, H. Gupta, and K. Liber. On the complementary roles of XVARS and SHAP in explainable AI. *Environmental Modelling & Software*, 203, August 2026.

[QZ24] H. Qu and Z. Zhang. A Time Series Data Augmentation Method based on SMOTE. In *2024 36th Chinese Control and Decision Conference (CCDC)*, pages 5336–5341, Xi'an, China, May 2024. IEEE.

[RC93] P. J. Rousseeuw and C. Croux. Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, 88(424):1273–1283, December 1993.

[RG16] S. Razavi and H. V. Gupta. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. *Water Resources Research*, 52(1):440–455, January 2016.

[RJS+21] S. Razavi, A. Jakeman, A. Saltelli, C. Prieur, B. Iooss, E. Borgonovo, E. Plischke, S. Lo Piano, T. Iwanaga, W. Becker, S. Tarantola, J. H. A. Guillaume, J. Jakeman, H. Gupta, N. Melillo, G. Rabitti, V. Chabridon, Q. Duan, X. Sun, S. Smith, R. Sheikholeslami, N. Hosseini, M. Asadzadeh, A. Puy, S. Kucherenko, and H. R. Maier. The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, 137:104954, March 2021.

[SAA+21] M. R. Smith, E. A. Acquesta, A. L. Ames, A. N. Carey, C. R. Cuellar, R. V. Field, T. M. Maxfield, S. A. Mitchell, E. S. Morris, B. C. Moss, M. N. Nyre-Yu, A. R. Rushdi, M. C. Stites, C. S. Smutz, and X. Z. Zhou. Sage intrusion detection system: Sensitivity analysis guided explainability for machine learning. Technical Report SAND2021-11358, Sandia National Laboratories, September 2021. OSTI ID: 1820253.

[SAB+25] T. F. Stone, J. Alford, P. H. Bečvářová, M. A. M. Eisa, A. H. El-Naggar, M. J. Carpio Espinosa, M. Frąc, J. Álvaro-Fuentes, J. C. García-Gil, K. Krabbe, J. T. Martins, S. I. Pathan, C. Plaza, J. Rasmussen, C. Righini, and M. H. Thorsøe. Food system strategies to increase grain legume-cereal intercropping in Europe. *Agroecology and Sustainable Food Systems*, 49(4):518–542, April 2025.

[SG64] A. Savitzky and M. J. E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8):1627–1639, July 1964.

[SMS13] A. Soltani, V. Maddah, and T. R. Sinclair. SSM-WHEAT: A SIMULATION MODEL FOR WHEAT DEVELOPMENT, GROWTH AND

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

YIELD. *INTERNATIONAL JOURNAL OF PLANT PRODUCTION*, 7(4):711–740, January 2013.

[SPTW18] S. J. Seidel, T. Palosuo, P. Thorburn, and D. Wallach. Towards improved calibration of crop models – Where are we now and where should we go? *European Journal of Agronomy*, 94:25–35, March 2018.

[SRLD24] M. L. Santoni, E. Raponi, R. De Leone, and C. Doerr. Comparison of High-Dimensional Bayesian Optimization Algorithms on BBOB. *ACM Transactions on Evolutionary Learning and Optimization*, 4(3):1–33, September 2024.

[SZZ+25] S. Shen, A. Zheng, D. Zhang, M. T. Harrison, J. E. Olesen, E. E. Rezaei, K. Liu, P. Zhang, W. Li, J. Zou, Z. Wang, Z. Song, W. Yin, H. Chen, and X. Yin. Global sensitivity analysis of STICS model in simulating winter wheat under various nitrogen and water management. *Agricultural Water Management*, 321:109919, December 2025.

[TGMH18] J. Torres, J. Gonçalves, B. Marcos, and J. Honrado. Indicator-based assessment of post-fire recovery dynamics using satellite NDVI time-series. *Ecological Indicators*, 89:199–212, 2018.

[U.S21] U.S. Geological Survey. How do I use a scale factor with Landsat Level-2 science products?, 2021. Accessed 2026-04-12.

[VMG+23] R. Vezy, S. Munz, N. Gaudio, M. Launay, P. Lecharpentier, D. Ripoche, and E. Justes. Modeling soil-plant functioning of intercrops using comprehensive and generic formalisms implemented in the STICS model. *Agronomy for Sustainable Development*, 43(5):61, October 2023.

[VML15] F. Vicente-Guijalba, T. Martinez-Marin, and J. M. Lopez-Sanchez. Dynamical Approach for Real-Time Monitoring of Agricultural Crops. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3278–3293, June 2015.

[WHM+17] C. Wei, J. Huang, L. Mansaray, Z. Li, W. Liu, and J. Han. Estimation and Mapping of Winter Oilseed Rape LAI from High Spatial Resolution Satellite Data Based on a Hybrid Method. *Remote Sensing*, 9(5):488, May 2017.

[XSCV19] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni. Modeling tabular data using conditional GAN. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.
TU Bibliothek Your knowledge hub WIEN

[YJvdS18] J. Yoon, J. Jordon, and M. van der Schaar. GAIN: Missing data imputation using generative adversarial nets. In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5689–5698. PMLR, 10–15 Jul 2018.

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.