Operationalizing digital twins in agriculture with machine learning

Christos Pylianidis

Propositions

1. The value of data is determined by their utility.
   (this thesis)
2. Technology integration tools like digital twins are essential for solving problems of increasing complexity.
   (this thesis)
3. The pursuit of model interpretability hurts decision support.
4. In research, the phrase ‘I don’t know’ leads to more productive meetings.
5. Gym subscriptions by themselves have a placebo effect on muscle gain.
6. Buzzword is a buzzword.

Propositions belonging to this thesis, entitled

Operationalizing digital twins in agriculture with machine learning

Christos Pylianidis
Wageningen, 29 August 2023

Operationalizing digital twins in agriculture with machine learning

Christos Pylianidis

Thesis committee

Promotors
Prof. Dr Ioannis N. Athanasiadis
Personal chair, Wageningen Data Competence Center
Wageningen University & Research

Prof. Dr Bedir Tekinerdogan
Professor of Information Technology
Wageningen University & Research

Co-promotor
Dr Sjoukje A. Osinga
Assistant professor at Information Technology
Wageningen University & Research

Other members
Prof. Dr Kirsten de Beurs, Wageningen University & Research
Prof. Dr Barak Fishbain, Technion, Haifa, Israel
Dr Christophe Pradal, CIRAD, Montpellier, France
Dr Geerten Hengevend, Wageningen University & Research

This research was conducted under the auspices of the Wageningen School of Social Sciences (WASS).

Operationalizing digital twins in agriculture with machine learning

Christos Pylianidis

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 29 August 2023
at 4 p.m. in Omnia.

Christos Pylianidis
Operationalizing digital twins in agriculture with machine learning,
166 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2023)
With references, with summary in English

ISBN: 978-94-6447-657-6
DOI: 10.18174/591164

Contents

Abbreviations

Acronym	Definition
AI	Artificial Intelligence
APSIM	Agricultural Production Systems sIMulator
CPU	Central processing unit
DSS	Decision Support System
GAN	Generative Adversarial Network
IoT	Internet of Things
ML	Machine Learning
MAE	Mean Absolute Error
NRR	Nitrogen Response Rate
RMSE	Root Mean Squared Error
VAE	Variational Autoencoder

Chapter 1

Introduction

1.1 Taking decisions for agricultural systems

Technological advancement has led to a wealth of data for agricultural systems. Large amounts of multi-modal data are generated from different sources and agricultural practitioners need assistance to make them actionable. This role is filled by decision support systems (DSS). Agricultural DSS are "human-computer systems which utilize data from various sources, aiming at providing agricultural practitioners with a list of advice for supporting their decision-making under different circumstances" [1]. DSS can aid with ongoing operations by analyzing the existing conditions in a system (e.g. which patches in a field have the most fertilizer based on satellite images [2]) or by providing support for circumstances that may occur in the future (e.g. exploring the outcome of operations for conflicting objectives like profitability, and societal and environmental effects [3]).

Agricultural DSS have been successfully deployed on multiple levels. On the field level, they have helped farmers to allocate irrigation water more effectively leading to improved nutrient balance [3], and to optimize crop treatment practices to enhance their productivity [4]. On a higher organizational level, they have been employed by water management authorities to balance water use between the field and district levels [3]. On the policy setting level, agricultural policymakers have adopted DSS to take more informed decisions for the animal and crop sectors considering socioeconomic development and greenhouse emissions [5].

Despite the accomplishments of DSS, concerns arise about their application [1]. One such a concern is that DSS rely on static models and processes to provide support. DSS do not adapt to their environment by continuously assimilating new data or recalibrating their embedded models [3]. As a result, assumptions that held on the development of those models might break when the models are transferred to the practitioner's local conditions.

Another concern is that DSS are task specific [1]. They are designed to perform narrow sets of operations around a main task and that is the reason for requiring multiple DSS to manage agricultural activity. This leads to having isolated DSS that ignore the interplay of fundamental factors [3] affecting agricultural operations (e.g. soil variability, crop diseases) and lack a holistic view of the target system.

A final consideration is the inability of DSS to provide direct instructions to practitioners as well as to act on their environment [1]. The reason for this behavior is that DSS are created only with data analysis and visualization in mind, or to provide an overview of actions and their end result, leaving the final decision to the practitioners. Requiring human intervention, DSS impede decision-making processes by hindering automation, and may leave opportunities to go unexploited due to operators choosing suboptimal actions. These flaws have opened discussions about new types of DSS that have proved their value in other disciplines but have not yet found their way to agriculture.

1.2 The next generation of decision support tools

A type of DSS that has become popular are digital twins. Digital twins are used to solve a variety of problems across several disciplines. Each discipline is using digital twins in a different way. As a result, the definition of digital twins varies according to the discipline of application. In this work, we are going to adopt a working definition proposed by [6] which considers a digital twin to be "a dynamic virtual representation of a physical object or system, usually across multiple stages of its lifecycle, that uses real-world data, simulation, or machine learning (ML) models combined with data analysis to enable understanding, learning, and reasoning. Digital twins can be used to answer what-if questions and should be able to present insights in an intuitive way". Key elements in this definition are the words dynamic and representation. Digital twins are dynamic because they evolve over time depending on changes in their environment. Also, digital twins are able to represent objects or systems by individually binding to them and adapting to their local conditions.

Digital twins integrate a variety of operations. For this reason, they are comprised of multiple components. Typical components and their purposes include:

- sensors, responsible for acquiring the current state of the physical object/system
- simulators, predicting futures states based on domain understanding
- learning algorithms, adapting to local conditions
- analytic engines, providing insight into the past, present, and future states
- actuators, being automatically activated when thresholds are reached
- user-interfaces, visualizing analytics and providing information on actuator activation

Digital twins have gained traction in a variety of industries. They are prevalent in manufacturing [7], where they have been used to optimize production floor procedures [8], as well as to create safer working environments for human workers [9]. Likewise, they are commonly found in aviation for predictive maintenance [10] and estimation of the structural life of aircraft [11]. In healthcare, researchers are even aiming for individualized nutrition with digital twins based on the combination of genetics, metabolism, and microbiome [12]. In agriculture, digital twins are not common yet, but recently they have been gaining momentum.

1.3 State of digital twins in agriculture

Back in 2019, when this work started, it was unclear if digital twins were already adopted in agriculture. Deployed digital twin applications were rare, the relevant literature was limited, and the articles describing digital twin applications were lacking implementation details. Reasons behind this delay included a lack of communication for the added value of digital twins, the risk involved in trusting technology applications for complex multidisciplinary problems in high uncertainty environments (especially when these applications automate actions based on analytics), as well as the cost and difficulty of developing them [13].

Nowadays, the concept of digital twins in agriculture is considered attractive. Researchers are investigating ways to materialize the benefits of digital twins promised by other disciplines and there are ongoing discussions on ways to utilize them. Ideas emerge for digital twins in aquaponics to advance urban farming [14], attempts are made to improve the virtual 3D representation of leaves for digital twins in plant breeding [15], conceptual frameworks are created to design digital twins [16], prototypes of digital twins of arable and dairy farms [17] that monitor the nitrogen cycle are researched, and digital twins of tomato crops in greenhouses [17] that predict the effect of cultivation treatments in harvest and financial yield are developed.

1.4 How attractive are digital twins for agricultural applications?

In our view, the attractiveness of digital twins lies in the way they consolidate the physical and virtual worlds. This consolidation is a process of instantiating virtual replicas of physical systems based on a blueprint. The digital twin blueprint. Each replica (digital twin) adapts to its corresponding individual system and is kept in synchronization with the physical world. Here, we use the term 'adapt' to refer to the ability of digital twins to learn the physical system's local idiosyncrasies, as well as the ability to be transferable to different conditions like different crops, locations, and fertilization treatments.

Adaptation is important for digital twins because it allows them to offer individualized curation of complex systems, giving insight into processes happening in each individual system and providing tailored recommendations. Adaptation happens continuously and automatically on digital twins based on incoming data and analytics. This method of adapting to individual systems improves upon current agricultural practices where e.g. fertilizers are applied based on rules of thumb, pesticide doses are calculated based on generic guidelines that only consider the type of pest, and DSS are used to simulate pasture growth across different locations with models that are not calibrated each time.

Automation through integration is another important aspect of digital twins. By integrating a variety of operations like monitoring, analytics, automatic adaptation based on analytics, and acting through actuators, digital twins manage to work with less human intervention than current practices. In this way, automation also removes bottlenecks in decision-making procedures. Digital twins can take decisions faster because they operate continuously, exchanging information between their components in real-time or at regular intervals. In contrast, e.g. agricultural practitioners who need to estimate yield based on fertilizer application have to contact an institute/company that knows how to do it, the institute/company have to find somebody to give them the relevant data, another person has to analyze them, and finally the results have to be communicated back to the practitioners creating in this way a chain of slow synchronous interaction.

1.5 Elements of digital twin adaptation

In other disciplines, researchers have embedded process-based and ML models [18] into the learning components of digital twins to make them adapt to different domains. These models consume large amounts of data during the development of the digital twin blueprint to be preliminarily calibrated and have generic domain knowledge instilled about the physical target system. Additional data are required during digital twin instantiation to provide a first view of the prevailing conditions on each target system and start the adaptation. Also, further data are needed throughout the digital twin's lifespan to regularly recalibrate/update it to reflect changes in the target system environment.

Process-based and ML models consume combinations of historical and observation data. Having data that satisfy the model data requirements is crucial for the initial model calibration/training, as well as for the later adaptation stages. The same applies when the models are asked to make simulations/predictions during deployment, since they expect their inputs to be complete. It is often the case that models are asked to estimate the future value of a quantity. In those cases, they may require the future values of their inputs which should still be available. Also, data should be on the same resolution as to what process-based models expect on their input, and (usually) on the same resolution to what ML were trained on.

1.6 Challenges for digital twin adaptation in agriculture

1.6.1 Data challenges

Agricultural data are produced in large volumes. At the same time, these data cannot be used to create digital twins because there is a discrepancy between the available data, the data requirements of process-based and ML models, and the decisions that are necessary to be made. There are several reasons behind this discrepancy. An important reason is that agricultural experiments take a long time to complete, sometimes even yielding a single data point. Therefore, not enough historical data exist because not many years have passed since experiments started being recorded, or the technology existed to measure environmental quantities.

Also, there is a lack of planning for the future in agricultural experiments which leads to the creation of seemingly large datasets that are obsolete for any other application other than the initial conceptualization. For example, datasets of decades of pasture growing exist, but without including cases where no fertilizer was applied, because at that time this was something that did not interest the corresponding stakeholders. As a result, creating now a digital twin that accounts for pasture fertilization is difficult due to the lack of such data.

Another reason is that agricultural data collection practices lack standardization. These practices vary based on the preferences of each practitioner or the needs of each farm. Different practitioners gather data in different ways (e.g. sampling frequency). As a result, there may seem to be a lot of data but they might be in different resolutions from what existing models expect rendering them unusable.

Data collection equipment also lacks standardization. Depending on the manufacturer, sensors may or may not collect metadata, have different error rates, and inject different levels of noise. This leads to dropping large parts of data during preprocessing to preserve homogeneity. Another reason is that agricultural data are stored in data silos without labels or metadata. Consequently, when they are retrieved for analysis, much of them might be unusable due to being incomprehensible.

Finally, making estimations for future quantities, months ahead before an event occurs, is an integral part of agricultural practices to optimize production, decrease costs and increase profits. These estimations require future values of weather quantities or biophysical factors. Simulators exist to produce these quantities but currently they are capable of producing semi-accurate forecasts only days ahead in the future while also requiring careful calibration. Thus, the further in advance an estimation is needed the more difficult it is for existing models to work.

1.6.2 Model challenges

These data-related issues cause complications with existing process-based and ML agricultural models if we try to deploy them in the same way as they do in other disciplines. Issues can occur in the following ways:

1. Lack of historical/observation data. In this case, a process-based model cannot be calibrated to a new location where data do not exist, and ML models are not a viable

solution since they usually require a lot of data to be developed.

2. Available data have a different resolution than what the models expect. Process-based models usually make simulations by using daily environmental measurements. However, the available data are often on a weekly/bi-weekly/monthly basis, constituting these models inoperative. Similarly, ML models commonly require the input data to be on the same resolution as the data on which they were trained on, otherwise, they again cannot operate.

3. Future states of variables. In some cases, we need to estimate the future value of a quantity. It may be the case that other variables are required, like future weather conditions, which may not be available. Process-based models which require the complete time-series data until that time in the future are not able to work on these occasions.

In the above cases, process-based and ML models cannot function and thus a digital twin would become non-operational, as it would not be able to provide individualized decision support, or to be transferred to the corresponding 'problematic' conditions. These three situations are regularly encountered in agricultural applications. Conclusively, the available models are not suitable in their current state to create digital twins as the available data, the models, and the way that we plan to use them do not fit. Therefore, it would seem appropriate to alter the problem definition from 'how to combine our existing components to create digital twins in agriculture' to 'how to operationalize digital twins in agriculture'.

1.7 In this thesis

The objective of this thesis is to investigate how to operationalize digital twins. We examine this objective from two perspectives. The first is by investigating how to enable decision support for digital twins when the available data are not sufficient to operate process-based or ML models. The second is by considering how to make digital twins transferable to diverse conditions.

To achieve our goal, we use a case study of pasture nitrogen response rate prediction. Nitrogen is the nutrient that pasture draws from the soil in the greatest quantities [19] and thus it becomes a growth-limiting factor [20]. Agricultural practitioners apply nitrogen-containing fertilizer to increase pasture growth rates [21]. However, the relationship between yield increase and fertilizer amount is not linear. Excessive amounts of fertilizer can harm yield, the fertilizer itself has an adverse effect on soil [22] and freshwater [23], and it has a certain (economic) cost. Subsequently, agricultural practitioners need tools to predict the effect of different amounts of fertilizer on their fields' yield, allowing them to manage the trade-off of the aforementioned factors. For the case study, we used a synthetic dataset generated with the Agricultural Production Systems sIMulator (APSIM) [24] and provided by AgResearch Ltd, New Zealand.

The rest of the content is organized as follows:

- In chapter 2, we investigated the adoption of digital twins in agriculture exploring the dimensions of technology readiness level, service provided, physical twins and benefits. We then contemplated on their potential added value and proposed a general roadmap for further adoption.

- In chapter 3, we proposed a general method to make digital twins operational in cases where we do not have enough observations to run ML models, or future weather data to run process-based models, and when the available data are in different resolutions from what the process-based models expect.

- In chapter 4, we examined whether the proposed method of chapter 3 is algorithm independent (independent of the prediction algorithm used). We compared the performance of algorithms that learn the latent space of the input data and checked if their predictions were above a domain-specific error threshold.

- In chapter 5, we investigated how to make digital twins operational in diverse conditions. We developed and evaluated ML metamodels containing data from different amounts of locations, which either contained data from the target location or not, for nitrogen response rate prediction.

- In chapter 6, we considered operationality in diverse conditions in the case where we have data from the target location, but they are sparse, and we examine if we can transfer knowledge (different soils, weather, fertilization rates) from other locations by training metamodels.

- In chapter 7, we reflect on our findings, experiences working on this project, and future prospects.

Chapter 2

Introducing digital twins to agriculture

Published as: Christos Pylianidis, Sjoukje Osinga, Ioannis N. Athanasiadis. Introducing digital twins to agriculture, Computers and Electronics in Agriculture, vol 184, pp 105942, 2021

Abstract

Digital twins are being adopted by increasingly more industries, transforming them and bringing new opportunities. Digital twins provide previously unheard levels of control over physical entities and help to manage complex systems by integrating an array of technologies. Recently, agriculture has seen several technological advancements, but it is still unclear if this community is making an effort to adopt digital twins in its operations. In this work, we employ a mixed-method approach to investigate the added-value of digital twins for agriculture. We examine the extent of digital twin adoption in agriculture, shed light on the concept and the benefits it brings, and provide an application-based roadmap for a more extended adoption. We report a literature review of digital twins in agriculture, identifying use cases, and comparing them with use cases in other disciplines. We compare reported benefits, service categories, and technology readiness levels to assess the level of digital twin adoption in agriculture. We distill the digital twin characteristics that can provide added-value to agriculture from the examined digital twin applications in agriculture and in other disciplines. Then, inspired by digital twin applications in other disciplines, we propose a roadmap for digital twins in agriculture, consisting of examples of growing complexity. We conclude this paper by identifying the distinctive characteristics of agricultural digital twins.

2.1 Introduction

Digital twins (DT) are being increasingly adopted by several disciplines, including the manufacturing [25], automotive [26] and energy [27] sectors, for addressing multidisciplinary problems. DT are digital replicas of actual physical systems (living or not), interweaving solutions of complex systems analysis, decision support and technology integration. DT have gained prominence, partially due to the uptake of Internet of Things technologies, that allow for the monitoring of physical twins at high spatial resolutions, almost in real-time, through miniature devices, producing ever-increasing data streams. DT have been useful for converging the physical and virtual spaces [28], guaranteeing information continuity through the system lifecycle [29], system development and validation through simulation [30], and preventing undesirable system states [31].

The DT concept was coined by M. Grieves in a white paper [32], as a unification of virtual and physical assets in product lifecycle management. Since then, several disciplines have adopted DT, each providing their own definition as there is no generally accepted definition of DT. A working definition for this study considers DT as "*a dynamic virtual representation of a physical object or system, usually across multiple stages of its lifecycle, that uses real-world data, simulation, or ML models combined with data analysis to enable understanding, learning, and reasoning. DT can be used to answer what-if questions and should be able to present insights in an intuitive way*" [33].

The benefits of DT applications include reduced production times and costs, hiding the complexity of integrating heterogeneous technologies, creating safer working environments and establishing more environmentally sustainable operations. DT are utilized by several leading companies and organizations, including Siemens [34], General Electric, NASA, US Airforce [35], Oracle, ANSYS, SAP, and Altair [36]. Furthermore, the recent availability of commercial software tools to develop DT, like Predix [1] and Simcenter 3d [2] [34], is an evidence in itself of increased interest in DT applications.

Information and communication technologies can be leveraged to design and implement the next generation of data, models, and decision support tools for agricultural production systems [37]. Today, technologies like artificial intelligence [38], big data [39] and Internet of Things [40] find their way in practice, and start to converge. Benefits of this convergence have been

[1] Predix is a software platform that facilitates data collection, processing and analytics for industrial applications. The product description can be found in `https://www.predix.io/`.

[2] Simcenter 3d is a software environment that integrates 3d modeling, simulation and data management. It includes modules to capture the dynamics of fluids, composites, acoustics and others. The product description can be found in `https://www.plm.automation.siemens.com/global/en/products/simcenter/simcenter-3d.html`.

demonstrated in DT applications in other disciplines. However, DT are hardly utilized in agricultural applications, and their added value has not yet been discussed extensively. As a result, questions emerge regarding the benefits of DT for agriculture, the characteristics that differentiate them from current practices, and their design and implementation.

The purpose of this work is to investigate the potential added-value of DT in agriculture. To achieve this goal, we will first research the extent to which DT have already been explicitly adopted in agricultural applications, and investigate their reported benefits. Second, we examine the similarities between DT applications in agriculture and other disciplines, to identify opportunities of potential added-value for agricultural DT. Our research questions are formulated as:

- RQ1: To what extent have digital twins been applied in agriculture?
- RQ2: What is a potential application-based roadmap for the adoption of digital twins in agriculture?

To address these questions, we employed a mixed-method approach as exploratory research suggested DT have not been extensively used in agriculture. Thus, a literature review alone would not suffice due to the limited number of reported cases in the literature. Our approach consists of a literature review of existing DT in agriculture, and a survey of case studies in other domains, the latter added to compare with the DT adoption level in agriculture and investigate potential future applications. We searched for DT **use cases** in agriculture, as well as in other disciplines to see how they employ DT. Note that we did not focus on identifying specific DT applications, rather we aimed at generalizing them into abstract, representative use cases. For the use cases identified, we explored the dimensions of maturity, service types and benefits offered. Our methodology is described in detail in Section 2.2 and the results are presented in Section 2.3. In Section 2.4, we discuss our findings concerning the current state of DT in agriculture, the added-value of DT, and we potential areas for future research. Section 2.6 concludes this work.

2.2 Methodology

To answer '*RQ1: To what extent have digital twins been applied in agriculture?*', we identified existing DT use cases in agriculture and extracted attributes which helped us assess how advanced these applications were. To identify use cases, we performed a literature review for DT in agriculture and extracted indicators of maturity, service type and benefits. To capture the development stage of the applications (e.g., idea, lab, production) we chose maturity. Limited use cases of production level DT is an indicator of less widespread use of DT. On the other hand, increased research and deployed applications indicate that DT are still finding their way into agriculture. To describe the purpose of DT on an operational level, we extracted the service type attribute. These services indicate the broader set of operations that DT perform. From the service type, we can understand the complexity of the DT operations, with higher complexity meaning potentially higher added value for the application domain. Also, the service category of DT in agriculture was compared with the service categories found in other disciplines to examine how advanced agricultural DT operations are. Next, to show what is the added-value of DT based on existing applications, we extracted the benefits attribute. Less materialized benefits from the applications indicate limitations for adoption. Below we describe step-by-step how the literature review was performed.

First, we searched in scientific databases and subsequently extended our search to grey literature. We included grey literature because a pre-literature search showed that the peer-reviewed corpus covering DT in agriculture is rather limited. By including grey literature, we also cover work in progress and commercial applications that have not been published in scientific literature.

Second, we checked the corpus for relevance. In scientific publications, we read the abstracts to verify that the topic was about agriculture with references to DT. For the grey literature, we scanned the entire articles to see whether they connect DT to agriculture.

Third, we read all the selected articles and extracted use cases of DT applications. References to similar DT applications between multiple articles were considered only once to avoid redundance. We identified each use case with a number, summarized it in a single paragraph describing its functionality, and extracted the reported benefits.

Fourth, we identified the services offered by each DT use case. We used the service classification initially proposed in [28], and subsequently aggregated in [41]. The categories we used for classifying the use cases are presented in Table 2.1. We categorized the use cases in this way to identify the complexity of operations that DT performed as operation complexity is an indicator of the advancement of DT in agriculture. Also, this categorization helped us compare the types of operation offered by DT in agriculture and other disciplines, and determine any potential gaps to further assess their adoption in agriculture.

Fifth, we categorized the use cases based on their technology readiness level (TRL) to examine whether they are in experimental stage, or if they have

Service Categories	Definition	Typical components					
		Monitoring	Simulation	User interface	Learning	Actuator	Analytics
Real-time monitoring	Monitor and log the status of a system	x		x			
Energy consumption analysis	Analyze the energy consumption of the physical system and find ways to minimize it	x	x	x			x
System failure analysis and prediction	Analyze the data coming from a system to identify the source of failure or when the system is going to need maintenance	x	x	x	x		x
Optimization /update	Find the optimal parameters for the operation of a system and update it to run with those parameters	x	x	x	x	x	x
Behaviour analysis / user operation guide	Analyze human made operations and provide feedback	x		x			x
Technology integration	Bring together different already deployed technologies under the same umbrella to control and visualize operations more easily	x	x	x	x	x	x
Virtual maintenance	Allow users to virtually test different maintenance strategies to find the least intrusive one		x	x			x

Table 2.1: The digital twin service categories used to classify the use cases identified by the literature review. The column *Typical components* lists the components that are usually needed to implement the corresponding services.

been used in production. We partitioned the European Union's TRL scale [42] into three generic levels shown in Table 2.2, and used them to tag the use cases. The first level represents DT which were still in a conceptual phase, the second consists of DT that had a working prototype even without the complete planned functionality, and the third level covers mature DT deployments in production.

Sixth, we identified the physical twin, i.e. the physical system that was twinned in each use case. We classified them in the following categories: living plants or trees, animals, agricultural products, i.e. harvested fruits; agricultural fields, farms, landscapes, farm buildings, as barns, greenhouses or other agricultural buildings, agricultural machinery, including equipment

Aggregated level	European Union technology readiness levels
concept	**1** Basic principles observed **2** Technology concept formulated
prototype	**3** Experimental proof of concept **4** Technology validated in lab **5** Technology validated in relevant environment **6** Technology demonstrated in relevant environment
deployed	**7** System prototype demonstration in operational environment **8** System complete and qualified **9** Actual system proven in operational environment

Table 2.2: The European union TRL grouped into three general levels. Concept level includes European TRL 1-2, Prototype includes levels 3-6 and Deployed includes levels 7-9.

and tractor appliances, and food supply chains and logistics.

Finally, we summarized in a table all the identified use cases, their respective descriptions and the extracted three dimensions - service categories, TRL, and physical twin - to depict the breadth of the application of DT in agriculture. Fig. 2.1 summarizes the methodology for answering RQ1.

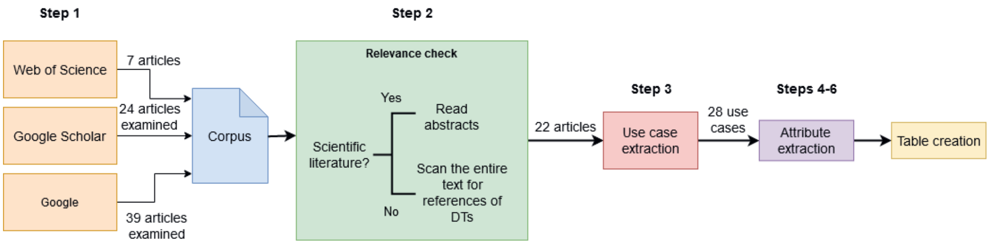


Figure 2.1: The steps followed to search for use cases of digital twins in agriculture.

To answer the second research question, '*RQ2: What is a potential application-based roadmap for the adoption of digital twins in agriculture?*', we searched in literature for use cases aiming to identify the ways in which DT have been successfully applied in other disciplines. Again we aimed at identifying use cases, and extracted indicators of benefits, maturity, discipline, and service type to understand the operations in which DT are most effective and what problems they can solve. First, we searched for peer-reviewed review papers of general DT applications. Second, we scanned the full texts for occurrences of the string 'digital twin', to check if the reviews were related to DT. If DT were just a brief mention and not the main point of the review paper, we considered the reference irrelevant. Third, the remaining articles were examined in alphabetical order based on their title to extract use cases. Repeated mentions of similar use cases were not considered. Fourth, we extracted a short summary of the use cases, the reported benefits that they offered, the discipline, maturity and service categories using the same framework as for research question 1, and the publication and application years. Fifth, we proposed areas of potential application in agriculture, and identified potential benefits based on the use cases in other disciplines. Fig. 2.2 illustrates the methodology for answering RQ2.

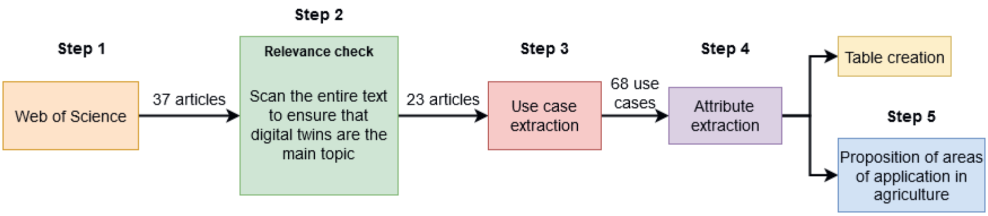


Figure 2.2: The steps we followed to find digital twin use cases in other disciplines so as to answer the second research question.

2.3 Results

2.3.1 Literature review of digital twins in agriculture

For the literature review of DT in agriculture, we first searched in Web of Science [43] using the query `"digital twin*" AND (agri* OR crop* OR farm* OR aqua* OR animal*)`. This query returned results which contain DT and derivatives of `agri`, `crop`, `farm`, `aqua`, or `animal`, to capture cases of DT in subfields of agriculture. The query returned seven results. After the relevance scan the results were reduced to four [44–47]. We then extended the search to Google Scholar [48] using the query `"digital twin" agriculture`. The query returned 947 results. We examined them until five consecutive results were irrelevant (24 results examined), and checked for duplicate results from the previous search in Web of Science, thus reducing the number of results to nine [49–57]. Extending to the Google search engine [58], we used the query `"digital twin" agriculture` which returned 143.000 results. We examined them until five consecutive results were irrelevant or referring to previously found applications (38 results examined). We then checked the extracted results for duplicates from our searches in Web of Science and Google Scholar, eventually reducing the results to nine [59–67]. In total our search yielded 22 sources for DT applications in agriculture. From this result-set, we identified 28 use cases. Following the methodology described in Section 2.2, we summarized each use case and extracted data about the expected benefits, TRL, physical twin, and service category. The results are reported in Table 2.3.

Use case No.	A digital twin:	Benefits	Physical twin	Technology readiness level	Service category: Real-time monitoring	Service category: System failure analysis	Service category: Optimization update	Service category: Technology integration tool	Service category: Energy consumption analysis	Citation
1	of a cow having access to historical and real-time data, able to predict the probability of developing mastitis as a function of various management and treatment decisions	a data organization system for each entity individually which is also queryable and identifies the best response to each query	animal	concept	x	x				[44]
2	of picked mango fruit that captures its temperature variability and biochemical response throughout the cold chain, to evaluate quality losses along the cold chain like firmness and vitamin content	insight into the remaining quality attributes of picked mango fruit	agricultural product	concept	x	x				[45]
3	of a field using data coming from ISOBUS sensors, other field related data, human expertise and machine learning to provide better field prognostics and act faster in the presence of predicted deviations	continuous detailed crop and soil information that allows for faster actions when anomalies occur	agricultural field, farm, landscape	concept	x	x				[46]
4	to emulate the use of unmanned ground vehicles in fields. It accepts the actual landscape of a field as input by utilizing digital elevation models retrieved from Open Street Maps. It recreates the 3D model of the field along with possible additions like trees and static objects. It contains a predefined selection of commercially available unmanned ground vehicles which a farmer can test on the virtual field to find the most efficient for their case	economic and environmental benefits for the farmers since they can choose the optimal machine for their specific field	agricultural machinery	prototype			x	x		[47]

Use case No.	A digital twin:	Benefits	Physical twin	Technology readiness level	Service category: Real-time monitoring	Service category: System failure analysis	Service category: Optimization update	Service category: Technology integration tool	Service category: Energy consumption analysis	Citation
5	of a self-contained aquaponics production unit. The purpose of this digital twin is to balance the fish stock and plants in the unit by monitoring them and controlling the unit automatically. The digital twin uses temperature, light intensity, water flow, pH and dissolved salts sensed data. The virtual unit performs simulations of fish feed, fish weight gain, pH, nitrates and plant growth as what if scenarios to find optimizations on the behavior of the whole system. It does this for production maximization, waste minimization, water conservation, meet quality standards and other production goals.	production maximization, waste minimization, water conservation, quality standards	agricultural building / agricultural machinery	prototype	x	x	x		x	[49]
6	of a pig farm to monitor pig health status and prevent diseases. The digital twin operates by deciding which of the sensed data are useful, performing simulation to find the optimal working conditions of the farm, a control system gets the results of the simulations to apply them to the physical system.The digital twin consists of a layer handling the connectivity between the sensors and their configuration, and a layer analyzing the given conditions in the farm, performing simulations, data handling and visualization. The analysis includes machine / deep learning methods, the results of the simulations are used to control the farm and are presented in an intuitive interface.	improved animal welfare, disease cost reduction	animal / agricultural building	concept	x	x				[68]

2

Use case No.	A digital twin:	Benefits	Physical twin	Technology readiness level	Service category: Real-time monitoring	System failure analysis	Optimization update	Technology integration tool	Energy consumption analysis	Citation
7	of a harvested potato to gain insight into harvester damage to potatoes. During harvesting shocks have the greatest economic impact and potential damage to potatoes. The digital twin of the potato is a plastic object with the weight and size of a real potato, equipped with sensors to detect impacts and rotations. The data is analysed in real time on the harvester and presented to the machine user.	less damage to potatoes and higher profits for the farmers	agricultural product	prototype	x					[51]
8	of a tree and its surrounding in an orchard. The authors created a system that can create a digital twin for every tree in an orchard by using spinning 3D cameras. These cameras monitor the condition of every plant in 3D by capturing indicators that show their health, structure, and fruit quality among others. These digital twins allow the continuous monitoring of orchard production systems to predict stress, disease and crop losses, and develop a self-learning system. This self-learning system can be queried automatically to analyse varying scenarios based on environmental and management parameters.	discovery of higher orchard density layouts, detection of plant degrading indicators	living plant or tree	prototype	x		x			[52]
9	of any agricultural entity, using holographic devices, augmenting the world with camera-based imaging, placing 2D or 3D content in the real world, simulating them, and creating logs and maintenance events	enables the users to see the complete picture of a system leading in improved decision making	agricultural field, farm, landscape	concept	x	x	x	x	x	[53]

Use case No.	A digital twin:	Benefits	Physical twin	Technology readiness level	Service category: Real-time monitoring	System failure analysis	Optimization update	Technology integration tool	Energy consumption analysis	Citation
10	of the cultivated landscape for supporting planners in designing agricultural road networks. The digital twin finds the road network segments with high relevance for agricultural transportation helping planners to modernize these segments according to the agricultural needs. The digital twin creates an information model (described as a UML class) by coupling spatiotemporal information of the cultivated landscape with complex analytical methods.	optimal agricultural road planning, a landscape representation model of high quality that can be reused for other causes	food supply chain and logistics	prototype		x	x	x		[54]
11	of a cow that makes predictions for heat, estrus and health according to its behaviour. It is working based on data from a pedometer attached to the cow as well as company provided location services that accurately detect the cow's movement	animal health analysis and prevention of diseases	animal	deployed	x	x				[55]
12	of feed silos for livestock to monitor their status. It works by placing an IoT device on top of the silos and a cloud platform that allows the stakeholders to access the silo's status through various apps. When the silo stock reaches a certain threshold an alarm is send to the stakeholders phones. It also provides the ability to organize the stock replenishment with a simple action	supply replenishment optimization, cost saving by reducing labour and transport costs, reduction of CO2 from transport emissions by 25%	food supply chain and logistics	deployed	x		x			[55]
13	that allows users to identify pest and diseases in plants. It is based on a mobile app with an on-the-field and on-the-fly systems for fast identification. The user takes photos of the plant and describes the problem, those two constitute the digital twin of the plant. Based on this digital twin a community of experts supports with their opinions to help identify the disease.	fast identification of pest and diseases in plants based on expert opinions	living plant or tree	deployed	x					[55]

Use case No.	A digital twin:	Benefits	Physical twin	Technology readiness level	Service category: Real-time monitoring	Service category: System failure analysis	Service category: Optimization update	Service category: Technology integration tool	Service category: Energy consumption analysis	Citation
14	of a field and its machinery. It provides online visualization of the current position of any machine in the field along with historical movement data. It allows the real-time monitoring of machines and their energy consumption and evaluation of the economic efficiency of the crop management treatment.	insight into how the use of machinery for cropping affects a farmer's economics	agricultural field, farm, machinery / agricultural machinery	deployed	x		x		x	[55]
15	of olive trees to monitor olive fly occurrence. The digital twin is accompanied by an application which uses automated real-time imaging to capture images of pest traps that are then transferred to the digital twin twin. Olive growers can monitor the crop status remotely through the application.	allows for timely reactions to save the crop and product quality, reduce pesticide and labour cost	living plant or tree	deployed	x					[55]
16	of bee colonies. The digital twin is created based on a GPS tracking system along with sensors for humidity, exterior & interior apiary temperature, brood temperature and weight. It provides real-time continuous apiary monitoring that enables beekeepers to remotely control them and make management decision that interact with the bees as little as possible. It allows the beekeepers to manage the food storage reserves, to identify disease and pest infections, to inspect if queenless and swarming states exist, it provides an anti-theft mechanism, and insight into the colony status and hygiene.	maintain healthy bee colony population, prevent pests, nectar flow monitoring	animal / agricultural building	deployed	x	x	x			[55]

Use case No.	A digital twin:	Benefits	Physical twin	Technology readiness level	Service category: Real-time monitoring	Service category: System failure analysis	Service category: Optimization update	Service category: Technology integration tool	Service category: Energy consumption analysis	Citation
17	of a smart farm. The digital twin is built around small services and connects them together. These services provide information of particular systems such as the irrigation and seeding systems. The services use sensed data from soil probes, weather stations, irrigation systems and equipment analyzing and storing them using cloud services. The digital twin then uses these data for visualizations and for decision making actions which are then applied to the physical system through programmable logic controllers (PLAs).	sustainable development, insight into farm operations	agricultural field, farm, landscape	concept	x		x			[56]
18	of the globe's agricultural systems using the WebGIS framework as an organizing principle that connects local, site-specific data generators to a regional and global view of agriculture using technologies like AI, IoT, drones, robots and Big Data, to aid in the development of site-specific conservation and management practices.	supports agricultural industry and government policy makers, increases incomes and global sustainability of agricultural systems	agricultural field, farm, landscape	concept	x		x	x	x	[57]
19	of a vertical farm. The virtual and physical components are interconnected through sensors embedded in the materials of the farm structure that monitor temperature, humidity, luminosity and CO2. Embedding the sensors to the materials allows the digital twin to weight the data closest to the point of interest and establish an ideal value and variations for it. If the measured value is not in the expected range the digital twin controls actuators like air conditioning, air extraction, lighting and misting system. The data gathered by the sensors are analysed in the cloud and provide recommendations to producers to improve their production process.	structure sustainability, decision support, more profitable vertical farming	agricultural building	prototype	x	x	x		x	[59]

2

Use case No.	A digital twin:	Benefits	Physical twin	Technology readiness level	Service category: Real-time monitoring	System failure analysis	Optimization update	Technology integration tool	Energy consumption analysis	Citation
20	of the world's agricultural resources. The digital twin will give instant access to critical data on the world's farmland. It will allow to share insights, materials and connection with the food supply chain.	data democratization, more equitable farming economy, more food at lower cost	agricultural field, farm, landscape	prototype			x	x		[60]
21	of a greenhouse which aids in decision making. It monitors the status of the greenhouse's fans, windows, sprayers and shading net as well as environmental factors like CO_2, temperature, pH and solar radiation to analyze them and simulate different scenarios of decisions. It then visualizes the results to help the user take the optimal decision.	decision support	agricultural building	concept	x		x			[61]
22	of a parch of land. It makes multiple high resolution simulations of the cultivation procedure in parallel, taking into account information from many different sources. It does that continuously to account for several inputs, actions and stresses.	best response identification	agricultural field, farm, landscape	concept	x		x			[62]
23	of an indoor garden that calculates the ideal conditions for plants to grow. It uses the data gathered by a gardening robot such as humidity and nutrient content of the soil as well as simulations to determine what the robot has to do to ensure that each plant gets exactly the right quantity of nutrients and water it needs for ideal growth. The data gathered, the algorithms and the digital twin itself are saved in the cloud.	ideal plant growing conditions	agricultural building	prototype	x		x			[63]
24	for aquaculture combining human intelligence and artificial intelligence to help fishermen develop accurate digital decision-making process for production management	productivity increase, cost reductions	agricultural building	prototype	x		x	x		[64]

Use case No.	A digital twin:	Benefits	Physical twin	Technology readiness level	Service category: Real-time monitoring	Service category: System failure analysis	Service category: Optimization update	Service category: Technology integration tool	Service category: Energy consumption analysis	Citation
25	for livestock that uses a computer vision system installed on the dairy farm along with deep learning to monitor animal behavior and farm operations. It monitors the cows 24/7, sends notifications to the farmer's phone about event in the farm, and provides interpretations using analytic techniques to point which operations can be optimized.	constant cattle monitoring, shows where there is room for improvement, improved milk production and animal well-being	animal / agricultural building	deployed	x		x	x		[65]
26	of tractors . Tractors are fitted with IoT sensors that monitor how they operate in real-time and proactively prevent malfunctions.	prevent equipment malfunctions, improve asset uptime	agricultural machinery	deployed	x		x	x		[66]
27	of tomato crops. The digital twin consists of a 3D simulation model fed with real-time sensor data from a greenhouse. The interactions of the crop variety, the environmental factors and crop management are simulated in the virtual model.	prediction refinement to make better choices for the real crop	living plant or tree	concept	x		x			[67]
28	of an arable or dairy farm to monitor the existing nitrogen cycle. The digital twin consists of crop growth models, soil models and business management systems which are linked with a variety of data like company, weather and sensor data.	improved decisions on whether to supplement nitrogen or to prevent losses	agricultural field, farm, landscape	concept	x	x	x	x		[67]

Table 2.3: The use cases of agricultural DT. Use cases are referred as "uc" and their corresponding numbers in the text. The numbering of the use cases continues for the use cases in other disciplines.

Our search yielded 14 scientific articles. Nine of which were published in journals and five in conference proceedings. Additionally, we identified eight website articles from the grey literature search. Publication outlets, titles and year of publication are summarized in Table 2.4. We observe that the first published references to DT in agriculture date back to 2017, and most of our sources are from 2019 onward.

Citation	Use case No.	Source	Article type	Year	Title
[44]	1	Web of Science	journal	2018	Getting value from artificial intelligence in agriculture
[45]	2	Web of Science	journal	2019	Multiphysics modeling of convective cooling of non-spherical, multi-material fruit to unveil its quality evolution throughout the cold chain
[46]	3	Web of Science	journal	2019	ISO 11783-compatible industrial sensor and control systems and related research: A review
[47]	4	Web of Science	journal	2019	AgROS: A Robot Operating System Based Emulation Tool for Agricultural Robotics
[49]	5	Google Scholar	journal	2019	Digital Twin Technology for Aquaponics: Towards Optimizing Food Production with Dynamic Data Driven Application Systems
[68]	6	Google Scholar	conference	2018	Smart Livestock Farms Using Digital Twin: Feasibility Study
[51]	7	Google Scholar	conference	2019	Business Models for Industrial Smart Services - The Example of a Digital Twin for a Product-Service-System for Potato Harvesting
[52]	8	Google Scholar	journal	2020	Digital Twin for the Future of Orchard Production Systems
[53]	9	Google Scholar	journal	2019	Enabling technologies and tools for digital twin
[54]	10	Google Scholar	journal	2019	Planning Agricultural Core Road Networks Based on a Digital Twin of the Cultivated Landscape
[55]	[11-16]	Google Scholar	conference	2017	Digital twins in farm management: illustrations from the FIWARE accelerators SmartAgriFood and Fractals
[56]	17	Google Scholar	conference	2019	A digital twin for smart farming

[57]	18	Google Scholar	journal	2019	Big Data Analysis for sustainable Agriculture on a Geospatial Cloud Framework
[59]	19	Google	conference	2018	Towards Sustainable Digital Twins for Vertical Farming
[60]	20	Google	website	2018	#twinning: Farming's digital doubles will help feed a growing population using less resources
[61]	21	Google	website	2019	Digital Twin Solutions for Smart Farming
[62]	22	Google	website	2019	Agility in Digital Farming
[63]	23	Google	website	2019	In the digital indoor garden
[64]	24	Google	website	2019	"Digital Twin Solutions for Smart Farming", the III Development AI+HI Total Solution, Awarded R&D 100.
[65]	25	Google	website	2020	Use Cases: Digital Twin in Livestock Farming
[66]	26	Google	website	2018	Digital Twin Excellence: Two Shining Examples
[67]	27, 28	Google	website	2020	WUR is working on Digital Twins for tomatoes, food and farming

Table 2.4: The source, article type and publication year of the use cases for the literature review in agriculture.

The use cases reside in different sub-fields of agriculture: In dairy farming we found DT for the detection of mastitis in cows. Related to apiculture, we found a DT of bee colonies aiming to control their welfare and honey production. In plant production, a DT of tomato crops in a greenhouse aimed to control the growing environment. In agricultural machinery, a DT of tractors was used to emulate their performance prior to purchasing. Other DT included orchards, pig farms and aquaponics production units. We noticed that DT of animals and fields, farms and landscapes are reported with less technical detail. In contrast, DT of agricultural machinery and food supply chains and logistics were often described with more details about their design and operation.

The reported benefits varied, depending on the physical twins. For twins of living systems, like plants and animals, the benefits included early disease identification, production optimization and identification of factors that could degrade their welfare. For agricultural products the benefits were cost savings and improved product quality. Support in crop management decisions allowing for faster action was reported for agricultural fields and farms. Twins of agricultural buildings reported benefits related to growing conditions management and production increase. Lastly, DT in agricultural supply chains and logistics reported benefits included cost savings and more environmentally friendly operations (Table 2.3).

Physical twins include both non-living subjects, like farm buildings such as farm bins or livestock barns, and living subjects, such as arable farms or individual animals. Most of the DT were found for physical twins of agricultural fields, farms, landscapes and buildings. Fewer were found for living plants and animals or agricultural products and the food supply chain. Fig. 2.3 illustrates the types of physical twins identified together with the maturity of the use cases as TRL level.

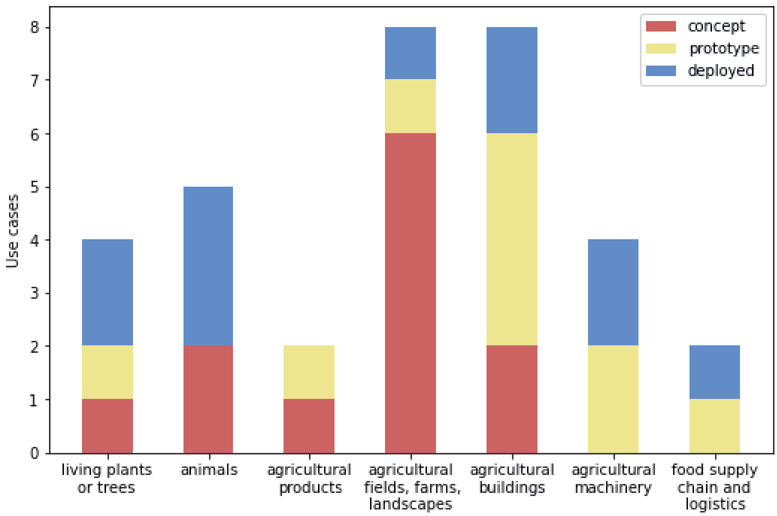


Figure 2.3: Classification of physical twins in agriculture. The colors indicate the maturity level of the DT.

Regarding the TRL, most DT identified in this study were on the conceptual level. In Fig. 2.3 we observe that DT of agricultural fields, farms and landscapes are mostly on the concept level. We also notice that DT in the food supply chain and agricultural machinery have surpassed the concept level stage. Besides, none of the identified agricultural products DT have reached the deployment stage.

Regarding the service categories, the identified use cases of agricultural DT perform energy consumption analysis, real-time monitoring, system failure analysis and prediction, optimization/update, and technology integration. The majority of the DT perform monitoring and optimization operations (Fig. 2.4). We do not observe any pattern of the TRL levels across the service categories.

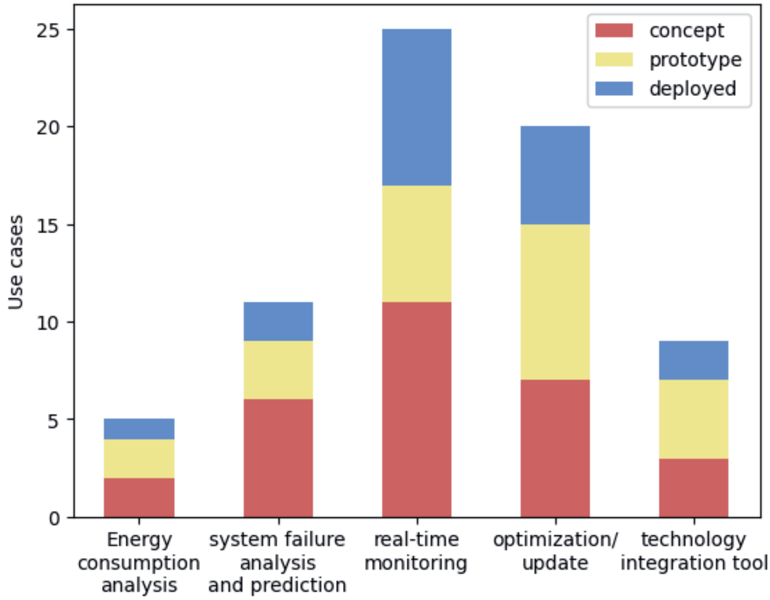


Figure 2.4: Service categories of DT use cases in agriculture. The colors show the maturity level based on TRL.

2.3.2 Digital twins in other disciplines

For the examination of DT in other disciplines we searched in Web of Science using the query `TS="digital twin*" AND ALL=review`. This query returned results that had DT mentioned in the title, abstract, or keywords and had the word *review* mentioned somewhere in the text. Instead of filtering the type of results to reviews only, we chose to search for the word review because some review papers are not always not explicitly tagged as such in Web of Science, or sometimes they miss the word review from their title. The query returned 37 results. After scanning the articles for relevance, the results were reduced to 23 [41, 69–90]. Following the methodology of Section 2.2, we identified 68 use cases, and extracted a short summary, benefits, maturity level, discipline, service categories, year of publication and year of application for each case, reported in Table 2.5.

Use case No.	A digital twin:	Benefits	Technology readiness level	Discipline	Service category: Real-time monitoring	Service category: System failure analysis and prediction	Service category: Optimization update	Service category: Technology integration tool	Service category: Energy consumption analysis	Service category: Virtual maintenance	Citation	Publication year	Application Year
29	as an installer base management system to manage machines	assist in data structuring and management of machines	prototype	manufacturing	x	x	x				[41]	2019	2019
30	for the organization of the production line	handle flexibility of production system	prototype	manufacturing	x		x				[41]	2019	2018
31	for machine reconditioning	machine reconditioning	prototype	manufacturing	x	x					[41]	2019	2018
32	performing machine optimization in the design phase	-	prototype	manufacturing			x				[41]	2020	2018
33	monitoring the interaction of humans and machines to prevent accidents	human safety in workplace	prototype	manufacturing			x				[41]	2019	2019
34	for workplace redesign	improved working conditions, improved productivity	prototype	industry			x				[41]	2019	2019
35	of a building providing ways to make it more energy efficient	building cost and energy consumption estimation, discovery of technical issues that may arise	prototype	construction					x		[69]	2018	2018
36	simulating different scenarios of a biology model to verify its credibility	provide a traceable route to model credibility and acceptance	concept	biology				x			[91]	2017	2017
37	of a vehicle providing historical information and recreating past states and estimating future states	monitor current state, recreate past and future	prototype	automotive industry	x	x					[71]	2019	2017
38	for the optimal organization of a shop floor	improved resource management	concept	manufacturing			x				[72]	2019	2017

Use case No.	A digital twin:	Benefits	Technology readiness level	Discipline	Service category: Real-time monitoring	Service category: System failure analysis and prediction	Service category: Optimization update	Service category: Technology integration tool	Service category: Energy consumption analysis	Service category: Virtual maintenance	Citation	Publication year	Application Year
39	of the production process performing simulations to find the optimal parameters of the production process	trace process performance, find potential improvement	concept	manufacturing			x				[72]	2019	2017
40	for product service systems calculating the optimal parameters for the system	autonomous interaction and further optimization of the components	concept	business			x				[72]	2019	2018
41	for the monitoring of a shop floor and the identification of anomalies	improve visibility of real-time operations, faster and more accurate identification of anomalies	deployed	manufacturing	x	x					[73]	2019	2017
42	calculating optimal assembly schedules	optimal assembly schedules	concept	manufacturing	x		x				[73]	2019	2018
43	for the product design stage	-	prototype	manufacturing			x				[73]	2019	2017
44	for 3D printing monoliths	insight into heat and mass transfer during solidification	concept	biomolecular engineering	x						[74]	2018	2017
45	of a ship that allows for the assessment of the vessel before its construction	virtually assess the safety and performance of vessels	concept	shipping			x			x	[75]	2018	2017
46	of a country that helps bring together different aspects of management and take optimal decisions	-	concept	management	x		x				[75]	2018	2017

Use case No.	A digital twin:	Benefits	Technology readiness level	Discipline	Service category: Real-time monitoring	Service category: System failure analysis and prediction	Service category: Optimization update	Service category: Technology integration tool	Service category: Energy consumption analysis	Service category: Virtual maintenance	Citation	Publication year	Application Year
47	of a mobile network that finds the optimal parameters to reduce energy consumption	lower power consumption	prototype	telecommunications					x		[76]	2019	2019
48	of a manufactured product detecting surface anomalies	improved detection surface anomalies	concept	manufacturing	x						[77]	2019	2019
49	of a smart product monitoring its status	monitoring through product lifecycle, aid in decision making for maintenance and end of product life	concept	manufacturing	x	x	x				[78]	2019	2019
50	to represent an array of contributors and verify the reliability of their operations	reliability verification	concept	manufacturing	x						[78]	2019	2019
51	for the product design phase that helps tot optimize product parameters and identify potential flaws	optimize design scheme, forecast and verify product functions	concept	manufacturing		x					[79]	2018	2017
52	for the monitoring of a manufactured product for the assessment of its performance and identification of flaws	predict lifetime, performance, faults	concept	manufacturing	x	x					[79]	2018	2017
53	to proactively maintain aircraft structure and virtually diagnose problems	reduce cost, improve reliability	prototype	aerospace engineering	x	x					[79]	2018	2012

Use case No.	A digital twin:	Benefits	Technology readiness level	Discipline	Service category: Real-time monitoring	System failure analysis and prediction	Optimization update	Technology integration tool	Energy consumption analysis	Virtual maintenance	Citation	Publication year	Application Year
54	for the monitoring of the aircraft structure and the prediction of its service life	facilitate the management of aircraft service life	deployed	aerospace engineering	x	x					[80]	2019	2011
55	for monitoring the degradation of machine equipment	monitor machine operation and predict surface roughness	prototype	manufacturing	x	x					[80]	2019	2017
56	for the organization of the production line	more reliable, flexible, predictable production process	concept	manufacturing			x				[80]	2019	2016
57	analyzing aircraft wing structural damage	predict fatigue cracks in wings	prototype	aerospace engineering	x	x					[80]	2019	2015
58	bringing many aspects of the manufacturing process under the same umbrella for optimization	smooth interactions among human, machine, product	concept	manufacturing			x				[80]	2019	2017
59	optimizing parameters for a magnet insertion process	production optimization	prototype	manufacturing			x				[80]	2019	2017
60	for geometry assurance of manufactured products	-	concept	manufacturing	x	x					[80]	2019	2017
61	to reduce material waste and prolong machine lifetime	reduce material waste, prolong machine lifetime	concept	manufacturing	x	x					[80]	2019	2017
62	monitoring the operational state of wings		prototype	aerospace engineering	x	x					[80]	2019	2017
63	predicting the time of failure for aircraft tires	improved prediction of probability of failure of tire	prototype	aerospace engineering	x	x					[80]	2019	2017

Use case No.	A digital twin:	Benefits	Technology readiness level	Discipline	Service category: Real-time monitoring	Service category: System failure analysis and prediction	Service category: Optimization update	Service category: Technology integration tool	Service category: Energy consumption analysis	Service category: Virtual maintenance	Citation	Publication year	Application Year
64	for predicting manufacturing parameters	more accurate predictions than classic models	deployed	additive manufacturing			x				[80]	2019	2017
65	for a driver assistance system	reduce complexity, increase flexibility	concept	automotive industry			x				[80]	2019	2017
66	a system to control multiple digital twins of wind turbines	-	concept	renewable energy	x						[80]	2019	2018
67	to control the cooling of a power system	-	prototype	renewable energy	x		x				[80]	2019	2018
68	simulating different scenarios for the construction of a power system	automation, data visualization, decision making	deployed	power systems	x		x				[80]	2019	2017
69	for the monitoring and energy efficient use of the pipes of a wastewater treatment plant	energy saving, forecast faults	deployed	wastewater plant	x				x		[80]	2019	2018
70	optimizing the operations of a wind farm	operation efficiency increase by 20%	prototype	renewable energy			x				[80]	2019	-
71	of a locomotive form its design to its end of life for monitoring and optimization of its parameters	timely operation optimization	deployed	rail industry	x	x	x				[80]	2019	2016
72	of a hospital used for bed planning and work allocation	-	deployed	healthcare			x				[80]	2019	-
73	for maintaining oil/gas facilities in remote areas	improve reliability of oil facility	deployed	oil industry	x	x					[80]	2019	2018

Use case No.	A digital twin:	Benefits	Technology readiness level	Discipline	Service category: Real-time monitoring	System failure analysis and prediction	Optimization update	Technology integration tool	Energy consumption analysis	Virtual maintenance	Citation	Publication year	Application Year
74	for the optimization of an aircraft assembly line	optimize operation efficiency	concept	manufacturing			x				[80]	2019	2017
75	producing prognostics for subsea equipment	cost effectiveness, security of equipment	concept	subsea cable						x	[80]	2019	2017
76	to analyze engine speed, oil pressure and other parameters to prevent vehicle breakdowns	prevent breakdowns, more efficient engine development	prototype	automotive industry	x	x	x				[80]	2019	-
77	to monitor equipment status	enhance operation resilience and flexibility	concept	manufacturing	x	x					[81]	2020	2020
78	of human workers to monitor their health and working conditions and provide productivity optimizations	worker health, production performance	concept	manufacturing			x				[81]	2020	2020
79	of a factory	creation of self organizing factories, with complete operational visibility, flexibility	concept	manufacturing			x				[81]	2020	2020
80	providing insight into production network operations	unprecedented visibility into operation performance, predict future needs in the network	concept	manufacturing			x				[81]	2020	2020
81	to monitor a gas turbine, detect anomalies and perform what-if scenarios	real-time monitoring, in-depth analysis of deviation, scenario assessment	concept	chemical engineering	x	x	x				[82]	2019	2019
82	for material fabrication	growth of material knowledge	concept	material science				x			[83]	2019	2019

Use case No.	A digital twin:	Benefits	Technology readiness level	Discipline	Service category: Real-time monitoring	Service category: System failure analysis and prediction	Service category: Optimization update	Service category: Technology integration tool	Service category: Energy consumption analysis	Service category: Virtual maintenance	Citation	Publication year	Application Year
83	to monitor pig health and prevent diseases	reduced animal disease costs	concept	agriculture	x	x					[46]	2019	2018
84	using ISOBUS sensors to provide better field prognostics	continuous detailed crop and soil information	concept	agriculture	x	x					[46]	2019	2018
85	of a neuromusculoskeletal system to estimate optimal muscle activation patterns	estimate optimal muscle activation patterns	concept	neurorobotics			x				[85]	2019	2019
86	to easily reconfigure production lines	-	concept	manufacturing	x		x				[86]	2018	2017
87	as a test bench for a benching beam	-	prototype	manufacturing			x				[86]	2018	2018
88	to monitor the production line	trace process performance, find potential improvement	prototype	manufacturing			x				[86]	2018	2017
89	to easily reconfigure production lines	trace process performance, find potential improvement	prototype	manufacturing			x				[86]	2018	2018
90	of a vertical milling machine monitoring its health status	avoid sudden downtime	prototype	manufacturing	x	x					[86]	2018	2017
91	of an aluminium smelter	energy reduction, insight into performance and deviations from target	prototype	manufacturing			x		x		[87]	2019	2019
92	for the optimization of the production line	optimize production line functionality	concept	manufacturing			x				[88]	2018	2018

Use case No.	A digital twin:	Benefits	Technology readiness level	Discipline	Service category: Real-time monitoring	System failure analysis and prediction	Optimization update	Technology integration tool	Energy consumption analysis	Virtual maintenance	Citation	Publication year	Application Year
93	for the monitoring and evaluation of a product during its lifetime	status of product through its lifecycle for consumers. Evaluation of product for companies	concept	manufacturing	x						[88]	2018	2018
94	of a city to manage it based on lifestyle and prevent disasters	forecast issues related to lifestyle and disasters	concept	management			x				[89]	2017	2017
95	of tourism destinations to assess different parameters	assess positive and negative impact of tourism, also in sustainability	concept	management	x						[89]	2017	2017
96	serving detailed information about the whole production process	reduced downtime, reduced maintenance costs, machine setup time reduction	prototype	manufacturing			x				[90]	2019	2019

Table 2.5: The use cases of DT in all disciplines. Use cases are referred as "uc" and their corresponding numbers in the text. The numbering of the use cases continues from the use cases in agriculture.

We observed that DT in other disciplines performed energy consumption analysis, real-time monitoring, system failure analysis and prediction, optimization/update, technology integration and virtual maintenance. Most of them performed monitoring and system failure analysis operations (Fig. 2.5). The TRL varied by the year. The earliest documented DT application (2011) was that of an aircraft, which was used in production. From 2011 to 2016, new use cases were scarce. After 2016, many DT applications emerged at the concept and prototype levels, as well as some deployed ones. Applications in the concept stage were more frequent than the ones at the prototype and deployed stages. The reported benefits included cost reductions, energy savings, reduced equipment downtime, quantification of system reliability and safer working environments for personnel.

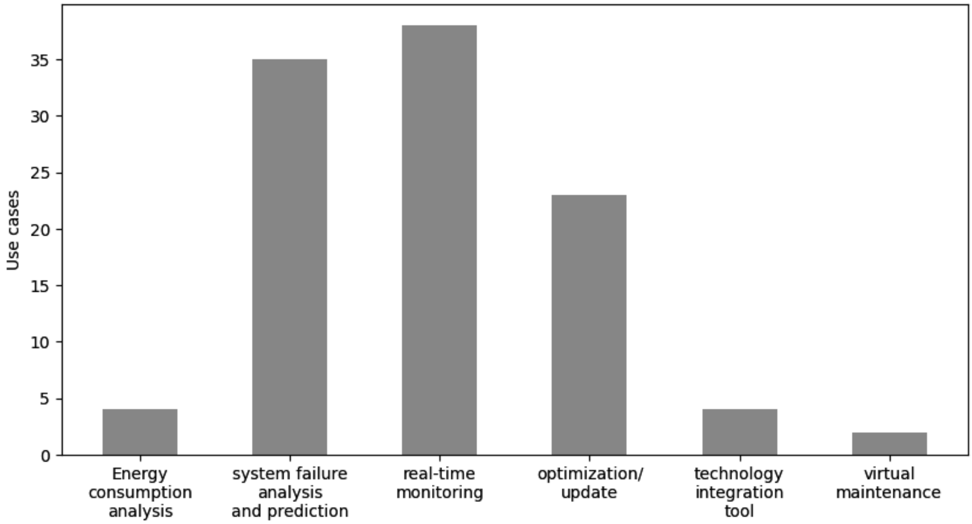


Figure 2.5: The DT service categories for DT in other disciplines. The majority of the proposed DT perform monitoring, optimization and system failure analysis operations.

2.3.3 Threats to validity

The results of the literature review for DT in agriculture showed that there are only a few DT use cases reported in scientific literature. Moreover, 13 *(Table 2.3, uc. 11-16, 20-26)* out of 28 DT use cases were used in the commercial sector and 7 *(Table 2.3, uc. 20-26)* out of those 13 were documented only in non-scientific literature. This may imply that the industry is ahead of academia in the development of DT.

Also, we limited our search to Google Scholar and Google to applications up to 5 consecutive irrelevant or duplicate results. More DT could potentially be found if we examined more results or additional sources.

Another factor that the literature review of this work does not consider is the existence of agricultural applications which are not defined as DT in literature. There are potentially applications that are used as DT but for unknown reasons they were not tagged as such and as a result they were not included in our results.

Besides, in our literature review we included conceptual level DT applications, which means that they are not established applications but work in progress.

2.4 Discussion

2.4.1 Current state of DT in agriculture

In this section, we investigate the state of DT in agriculture by comparing it to the state of DT in other disciplines. The results of the literature review in agriculture show that the available literature is limited. Considering the year of publication, DT have been discussed in other disciplines since 2011 (Table 2.5, *uc. 54*), while in agriculture the first references occurred in 2017. Our interpretation for this delay to investigate DT, is that agricultural researchers are more risk-averse than in other disciplines. A reason may be that in agricultural applications, firms are often small and medium farms. Such farms can bear less risk than bigger companies in other industries, who can afford to experiment and innovate, and thus pioneered DT. Also, DT in other domains are mostly concerned with non-living physical twins, as complex industrial and manufacturing applications. In agriculture, even the non-living physical twins, as those of agricultural buildings, still indirectly interact with plants or animals. The direct or indirect interactions with living systems introduce more challenges for DT in agriculture.

We identified only two overlapping use cases between our searches for agricultural DT, and DT in other disciplines. Use cases (*uc. 83, 84*) correspond to (*uc. 6, 3*). We expected a larger number of overlapping use cases, especially as our search of DT in other disciplines did not exclude agriculture-related use cases. This may be an indication that agricultural DT have not been adopted extensively, as they are not selected as representative use cases in DT reviews.

The benefits of the applications mentioned in the agricultural use cases include cost reductions (*uc. 6*), more detailed information (*uc. 3*), catastrophe prevention (*uc. 15*), positive economic impacts (*uc. 7*), aid in decision making (*uc. 4*) and more efficient management operations (*uc. 12*). Looking at the benefits of DT in other disciplines, we observe that they have a broader range. They also include safer human-machine interaction (*uc. 58*), building cost and energy efficiency estimation (*uc. 35*), and insights into complex multidisciplinary systems (*uc. 94*). DT in agriculture have not yet reached the point to demonstrate similar benefits.

Regarding the TRL, we were initially surprised to see that all levels are approximately equally represented. This large number of field-deployed or production-level DT could indicate a high adoption level in agriculture. However, upon closer inspection, we noticed that 6 out of 8 deployed DT were extracted from a single article [55], reporting on the results of the FIWARE Accelerator Programme [92], whose purpose was to create applications using the FIWARE platform[3]. Apart from the DT deployed by the FIWARE program, we observe that there has been little progress in advancing DT beyond the concept and prototype levels to the production level, where they can be used in real-world conditions. A reason for this may be that in other disciplines there are greater financial incentives, and larger research capacity to try out new technologies, or report their findings at earlier stages. Also, some applications on the conceptual level were described abstractly without any detailed technical design reporting, i.e. *uc. 1, 3, 9.* To our knowledge, Wageningen University and Research has recently introduced an investment theme on Digital Twins, developing twins of tomato crops and arable and dairy farms, but they are still on a conceptual stage (*uc. 27, 28*).

Another interesting finding from Fig. 2.3 was that the supply chain and logistics and agricultural machinery twins were the only ones that did not have any use cases on the conceptual level. While this could be circumstantial, it may also indicate that agricultural DT targeting these sectors are more mature than others. As DT of agricultural supply chains and logistics build upon relatively similar deployments in other supply chains and manufacturing, this could explain their relatively higher level of maturity. However, we did not check thoroughly to what extent DT of agricultural supply chains are concerned with perishables. This argument also pinpoints a significant challenge of DT in agriculture: Most agricultural operations have to do with living subjects, like animals and plants or perishable products, and creating DT for such systems is harder than for non-living human-made systems.

Another reason why most DT are on concept and prototype level might be that agriculture is a slow adopter of technology, partly due to the growing complexity of information technology [57]. To successfully develop DT, the community must become familiar with a variety of related technologies including Internet of Things (IoT), ML and big data. Most of these technologies are still considered new fields of experimentation in agriculture [93], and

[3]A framework of open source components to develop applications for the Internet of Things.

once the community gains confidence around them and adopt best practices for their application, we are likely to see more DT emerging in prototype and deployed levels.

Considering the service categories, most of the agricultural DT offer monitoring and optimization services. Other service categories reported were related to energy consumption analysis, and a few of the DT acted as technology integration tools. In other disciplines, we also came across the *virtual maintenance* category which was absent in agricultural DT. A reason for this gap could be that implementing an advanced technology like DT with more complex operations can be expensive [57], at least in the early experimental phase of its adoption. Applications of DT performing virtual maintenance could be useful for determining the optimal repair/maintenance strategy of agricultural machinery before laying hands on it, similar to repairing subsea equipment in (*uc. 75*).

Regarding the variety of the applications, from Fig. 2.3 we observe that a variety of applications like livestock farming (*uc. 6*), cropping (*uc. 4*) and apiculture (*uc. 16*) are encompassed. Yet, we believe that there is more room for DT to grow in each subfield. In our view, one of the reasons for not having a wider range of applications is the added complexity of the systems that DT pursuit to digitize, especially as this domain is lagging in digitization. Many agricultural systems are living systems, comprising of complex processes, which are harder to model than DT of products or human-made systems. This is in agreement with our findings related to DT in healthcare, another domain that also has to do with living physical twins: Only two use cases were identified related to healthcare (*uc. 22, 46*). Challenges related to living physical twins include capturing underlying processes that are still not well-understood, and accurately monitoring certain processes, for example nitrogen leaching in crop systems. In agricultural systems, it is also common that certain processes are not digitized because there are no financial incentives for doing so.

Another aspect affecting the adoption of DT in agriculture is that the community has to build trust in the interplay of the DT components for its correctness. This trust is essential to create DT that can accurately represent the inner workings of a system, propose maintenance strategies and alternative ways of management. Yet, building this trust in agriculture is difficult, because many decisions affect living systems where, unlike in other disciplines, consequences can be hard to reverse.

The lack of data culture also slows the adoption of DT in agriculture. DT require large amounts of data to operate, and the expected benefits are not eminent in small-scale deployments. In this respect, the lack of a data culture [94] and compartmentalization of agricultural systems understanding inhibits DT development and decreases potential for adoption. As a last note, integrating DT components and updating them in real-time can be daunting. For a community that is highly interdisciplinary and less information technology-oriented [95], this is a major turnoff.

2.4.2 The added-value of digital twins

This review identified few applications of DT in agriculture, with several of them being only superficially described in the corresponding articles. This suggests that DT benefits have not been clearly communicated to the agricultural community yet. Consequently, the community has not yet had the chance to investigate how they could utilize them and include them in their current practices. In this section, we pinpoint in the form of characteristics the benefits that DT can bring to agriculture. The characteristics can be seen in Fig. 2.6.

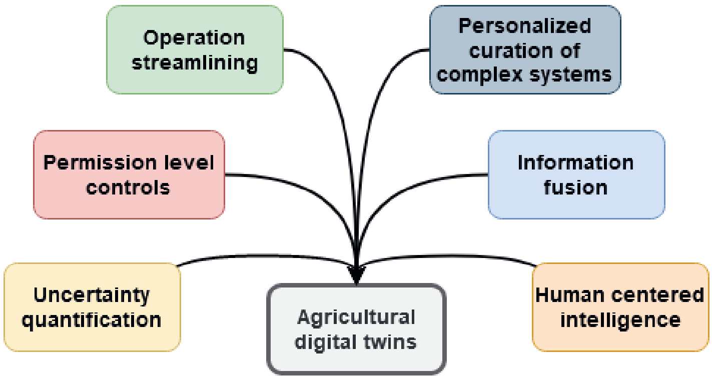


Figure 2.6: The characteristics of DT that can benefit agricultural applications.

The vision behind DT is to offer **personalized curation of complex systems**. This means that DT can account for system idiosyncrasies, that are often too complex to be accounted for in a generic model. DT adapt to local conditions in each individual physical twin, by fusing data and learning from them. DT are customized to mimic the individual characteristics of each system instance and deployment, and expose the system under different perspectives like system health, operation effectiveness, and profitability.

Streamlining of operations is another characteristic of DT. They offer an automated pipeline of operations like data acquisition from sensors, performing simulations, creating reports and controlling actuators. These operations are executed continuously, without requiring the attention, time and expertise of the users. DT bring together operations that previously were offered by a range of tools, hide their complexity, save time and remove context switching obstacles for the users. In this way, DT democratize technology and make it available to a wider range of stakeholders.

A key aspect of DT is **information fusion**, as they integrate and enrich information originating from several heterogeneous sources. DT observe physical twins from different perspectives by using multiple sources of data and assessing possible outcomes of actions. Information fusion combined with the continuous nature of operations depicts the complete picture of the past and current state of the system, and allows to estimate future states.

Uncertainty quantification is another characteristic of DT. DT can take into account the cumulative effect of the involved uncertainties since they observe systems from different angles. This information can then be customized and communicated to the stakeholders according to their expertise.

DT often embed **permission level controls**. The type of reports and controlling mechanisms can vary, based on the user of the application. This makes it possible to create different levels of transparency, depending on the sensitivity of the handled data and the importance of the operations taking place.

Finally, DT may demonstrate **human-centered intelligence** to control mechanisms for aspects that were neglected in the past, like human-machine interaction for safer working environments.

2.5 The future of digital twins in agriculture

The added value of DT has not yet materialized in agricultural applications. DT could be used pervasively, on different spatial and temporal scales and with varying levels of complexity, depending on their components and the desired functionality. We expect that the future of DT will evolve from simpler cases, exhibiting fewer components, to more sophisticated ones. We propose a roadmap for the development of DT in agriculture, starting from simple DT applications, with fewer components and simpler functionality, gradually adding components and functionality, to demonstrate the full potential of DT.

On a fundamental level, a DT will include monitoring, user interface and analytic components. These components are the first step towards empowering a DT to monitor and analyze agricultural systems and offer a continuous stream of operations. An example DT with these components could be deployed to monitor the microclimate of a greenhouse and provide insights for its management. In this case, the DT would monitor environmental indicators like solar radiation, humidity and CO_2, analyze them according to user-defined thresholds and report its findings, similar to the use case (*uc. 21*).

A slightly enhanced DT could include actuator components to control fans and windows in a greenhouse. The monitoring and control operations would be performed continuously, notifying different stakeholders with information that is relevant to them. For instance, in the case of consecutive stormy days, the DT would notify the farmer that it closed the windows because the temperature dropped, and notify the supply chain stakeholders that the production will be delayed because the plants cannot grow fast enough with the current weather. Also, the DT will report which indicators surpassed certain thresholds, thus taking specific actions using its actuators, and consequently assuring the stakeholders of its correct operation. Similar twins could be deployed to food silos (*uc. 12*) to keep track of their stock and autonomously organize their proactive replenishment, notifying the supply chain stakeholders and farmers respectively, and to livestock farms to keep track of environmental indicators that are known to affect animal welfare (*uc. 25*).

Further enhancing DT with simulation components is necessary for them to support decision-making based on past and future predicted states of the physical twin. A dairy farm DT could use simulation to forecast the occurrence of mastitis due to intensive milking for each individual cow. Utilizing this DT, a farmer could evaluate multiple milking scenarios and choose the one that strains the cow the least (*uc. 1*). Data analysis and simulation would happen in local or guaranteed cloud infrastructure to ensure data privacy. More advanced, simulations could investigate factors that have already lead to the appearance of mastitis, and result in improved breeding decisions. On an agricultural farm, DT of fields could use simulation to approximate the behavior of equipment in local conditions (*uc. 4*). Utilizing such a DT, a farmer could test a harvester, before purchasing it, on her local field with different weather scenarios to measure fuel consumption and plant damage.

Incorporating a learning component brings agricultural DT to the next level. A learning component may allow DT to assist in management operations for systems where the underlying mechanisms are unclear. In the case of a livestock farm, a DT with learning capabilities would be able to find patterns in real-time and in historical environmental data that could facilitate the onset and spread of diseases like swine fever. This would help stakeholders to take proactive measures to prevent not only the spread but also the appearance of diseases (*uc. 6*). Additionally, the DT would identify the most important variables shaping these patterns, estimate related risks, and clearly communicate the involved uncertainties, by presenting probability metrics for example.

Towards Digital Earth [96], a large-scale DT of an agricultural landscape, may consist of multiple DT of individual farms, each with several learning components. Such a DT will be able to consider the inter-field dynamics regarding water flow, fertilizer dispersion and nutrient leaching. It would provide variable fertilizer rates, based on site-specific intelligence, for example what amount can be absorbed by each field without being dispersed to other fields, and how much each field should be irrigated considering groundwater levels, and the availability of irrigation infrastructure. This would happen by learning from historical data about how the amount of fertilizer and irrigation affected the crop yield and depleted the nutrients of each field in the past. Ultimately, the DT would constantly improve itself in defining the acceptable fertilizer amounts and irrigation through continuous learning, also learning from the past decisions of the individual farmer. Besides, capitalizing on this information would lead to the creation of better cropping patterns, using different constraints like weather, profitability and field nutrient replenishment rate.

Further improving agricultural twins with a human-machine interface component would allow the establishment of safer working environments. A DT of a harvester with a human-machine interface component could trace the position of the workers and their actions to ensure that the machine is distant enough to avoid injuries (*uc. 33*). Also, a DT of grain bins could detect human presence inside the bin with cameras, and stop the procedures that cause grain movement to prevent entrapment. This is crucial as a large number of injuries occur every year with agricultural equipment due to the lack of safety measures [97].

Overall, DT can be applied to several agricultural subfields like plant and animal breeding, aquaponics, vertical farming, cropping systems and livestock farming. Adopting DT can start with simple setups, that can be gradually enhanced with more components to make them more intelligent and autonomous.

2.5.1 Considerations regarding the application of DT in agriculture

The application of DT in agriculture also involves potential pitfalls. As mentioned in [44], controlling physical twins through their virtual counterparts may lead to a lack of attention to the real-world systems. In agriculture, such neglect could cause irreversible damage, as DT are applied to living physical twins, among other things.

There are also cases where DT are not yet feasible, due to the large amount of resources they require to be developed, and the high complexity of the physical twins [98]. This could be the case of some agricultural system interactions that cannot be accurately quantified yet. There are also concerns about the technology skills required to create DT [99]. DT development requires specialized knowledge from several technology domains, which can be a serious threat in an already multidisciplinary domain like agriculture.

Synchronization between the physical and virtual twins is another target that is difficult to achieve [100]. In agriculture, human-made systems like agricultural equipment could be easier to synchronize with the virtual system, unlike natural systems such as animals or land parcels.

Also, the integration of DT components can be difficult [13]. In agriculture, this could be the case for combining the simulation and monitoring components for crops, as they rely on different infrastructures, software and end-users.

Last but not least, the widespread success of DT in agricultural applications does not only depend on technology, skills, or data infrastructures and availability but the involved business aspect. As with any new technology that is to be introduced in a farm, DT need to demonstrate their added value and the return on investment.

2.6 Conclusion

Returning to our first research question, we found that there are already a few applications of DT in agriculture. However, they are in primary stages and are not designed thoroughly enough to offer the benefits that other disciplines enjoy. Exceptions included some deployed applications that were part of a European Union-funded program. We believe that there is still a long way to go before the agricultural community can fully seize the benefits of DT. Agricultural researchers and stakeholders should make an effort to stay up-to-date with technological advancements and seek to find links between agricultural problems, and problems that are solved with DT in other disciplines.

Regarding the second research question, we proposed a roadmap of applications, starting from DT with simpler functionality, incrementally adding components to gradually demonstrate the benefits that are already present in other disciplines. As for the twins themselves, we foresee that there will be some confusion in the coming years about what a DT is and when a technology can be considered a DT. Research has been done to classify technologies based on how close they are to becoming DT [25], but it is still difficult to identify when a system can be called a DT. For the needs of most agricultural applications, we suggest that a DT should have at least the monitoring, interface and analytic components.

We identified two distinctive characteristics of DT in agriculture while reviewing the use cases and proposing our application roadmap. The first difference is that many agricultural DT involve directly or indirectly living systems and perishable products. While DT are ideal to provide insights into such complex systems and incorporate non-deterministic processes, their integration with the physical twin can be difficult. This is further amplified due to the idiosyncrasies of living physical twins. The second difference lies in the spatio-temporal dimension of their operation. DT in other disciplines range between the size of an airplane to that of a factory. Agricultural DT range from individual plants and animals to twins of land parcels, farms, or regions. As such, one may need to consider effects across these scales. On the temporal dimension, agricultural DT differ due to the slower response rates of their physical twins. Agricultural processes like the growing of plants tend to evolve relatively slow, so at least initially there is no need for high-frequency interactions between physical and digital twins. These two characteristics of agricultural DT need to be considered when developing DT inspired by DT in other disciplines.

As a final note, given the potential for the adoption and the benefits of applying DT in agriculture, we strongly believe that they have the prospect to bring a technological breakthrough in the near future.

Chapter 3

Simulation-assisted machine learning for operational digital twins

Published as: Christos Pylianidis, Val Snow, Hiske Overweg, Sjoukje Osinga, John Kean, Ioannis N. Athanasiadis. Simulation-assisted machine learning for operational digital twins, Environmental Modelling & Software, vol 148, pp 105274, 2022

Abstract

In the environmental sciences, there are ongoing efforts to combine multiple models to assist the analysis of complex systems. Combining process-based models, which have encoded domain knowledge, with machine learning models, which can flexibly adapt to input data, can improve modeling capabilities. However, both types of models have input data limitations. We propose a methodology to overcome these issues by using a process-based model to generate data, aggregating them to a lower resolution to mimic real situations, and developing machine learning models using a fraction of the process-based model inputs. We showcase this method with a case study of pasture nitrogen response rate prediction. We train models of different scales and test them in sampled and unsampled location experiments to assess their practicality in terms of accuracy and generalization. The resulting models provide accurate predictions and generalize well, showing the usefulness of the proposed method for tactical decision support.

3.1 Introduction

Digital twins are established in several industries, including manufacturing [101], healthcare [102], automotive [26]. Their ability to replicate physical systems and provide decision support through data fusion, simulation, and technology integration makes them attractive to apply in complex multidisciplinary problem-solving. Recently, digital twins have drawn the attention of the environmental sciences community. Researchers are exploring digital twins in hydrology [103], agriculture [104], smart farming [16], livestock farming [105], remote sensing [106] and earth sciences [107]. Recently, the European Union has announced plans for a high-resolution Earth digital twin that aims at actionable intelligence from (big) data streams [108, 109]. In the US, the research agenda for intelligent systems in geosciences [110] aims to incorporate extensive knowledge about the physical, geological, chemical, biological, ecological, and anthropomorphic factors that affect the Earth system while leveraging recent advances in data-driven research.

3

Digital twins intertwine data streams from a variety of in-situ or remote sensors with simulation and learning components. These components are then used to estimate future system states and offer an understanding of how complex mechanisms evolve. Digital twins incorporate sensor data streams with process-based models (PBM) or ML models, to provide insights by analyzing what-if scenarios, or provide operational decision support for managing and controlling complex systems. PBMs implement mathematical representations of physical processes and their interactions, and estimate future system states by numerical integration. While PBMs embody system understanding, they require many inputs and tend to be computationally intensive. ML models follow an empirical, data-driven approach in making predictions based on large collections of historical data. ML models are computationally fast in making predictions and robust with noisy data, but typically harder to interpret, and expensive to develop from data.

Digital twins need to be operational in a variety of data availability conditions. Their operation depends on the ability of the underlying models to cope with missing data streams or different resolutions. Problems with limited data arise when digital twins have to make decisions for the not-immediate future and quantities have to be forecasted. Also, their application in locations where data are sparse or non-existent (unsampled locations) can be challenging. Another concern is that transitions between different aggregation levels may be impossible due to the difference in the detail of the data that models expect. Therefore, digital twins need models or techniques to create models, that are able to handle such cases in order to provide operational decision support.

ML models can be versatile to a varying extent and resolution of input data.

However, they generally require large volumes of data for their development, accompanied by labels that are not easily available in environmental sciences. Techniques like few-shot learning [111] seem promising to learn from small datasets, but still novel research is needed to develop ML approaches that incorporate prior knowledge about environmental processes [112] and use it to effectively supplement the available data [110]. A path forward could be to employ synthetically generated datasets from simulations that mimic real conditions, which can be effectively used for developing ML models [110].

In this work, we showcase an approach to create ML models which tackles the challenges of data availability and data resolution while providing operational decision support for digital twins. We propose a method which (a) does not need forecasted data to be operational, (b) is applicable to locations where data are not yet available to calibrate PBMs, and (c) is applicable in cases where the available data do not have the resolution expected by the PBMs. We then demonstrate its usefulness in the context of a case study. In the case study, we create ML models of different scales to predict pasture nitrogen response rate (NRR) and examine their reliability by assessing their predictive and generalization capacity.

The rest of the paper is organized as follows: in section 3.2 we describe the requirements of PMBs, the proposed method and related work. In section 3.3, we present the case study and the methodology to experimentally evaluate the proposed method. Section 3.4 reports the results of our experiments, followed by a discussion in section 3.5, and the conclusion (in section 3.6).

3.2 Simulation-assisted machine learning

3.2.1 Process-based model data requirements

PBMs typically require several high-resolution data streams as inputs to simulations [113, 114]. Data availability becomes a problem with PBMs when applying models in new locations, where no or little data have been collected yet. In such cases, input data need to be estimated or collected, which can be a lengthy and expensive process. Also, when input data are available, they are needed in a prolonged temporal horizon of interest. For example, daily weather forecasting may be necessary for in-season crop model predictions [115]. Without such detailed forecasts of inputs, PBMs can make estimations only up to the present day. They may extend their reach to the near future if quantitative short-term weather forecasts are available. Otherwise, PBMs are used with historical data to estimate probability or risk distributions based on simulations, e.g. as in [116], and often together with data assimilation techniques to integrate them with sensor observations of system states [117].

Another factor affecting the operational use of PBMs is data resolution. Usually, sensor input is not available at the resolution required by the models. For example, input data streams may be available on a weekly basis, while models require daily inputs [118]. Data availability and resolution are two factors that can prohibit the use of existing PBMs in digital twins. A depiction of the data requirements of PBMs can be seen in Fig. 3.1a.

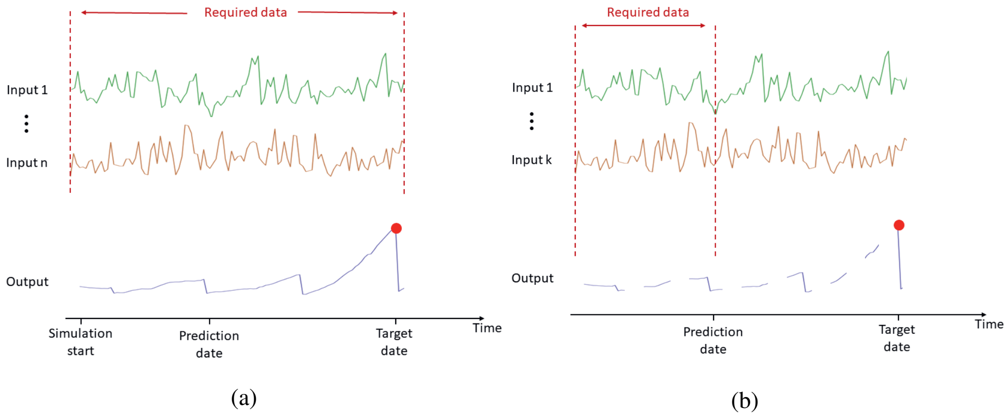


Figure 3.1: Data requirements of PBMs (a) and our approach (b). The PBM needs n inputs that span through the entire duration of the simulation to produce an output at the end of the simulation. The model of our approach requires a subset k, $k < n$, of the PBM inputs. The required data are limited to what has been observed prior to the prediction date (i.e. the date on which the prediction is required to be made). The red circles represent the outputs (predictions) of the model.

3.2.2 Requirements for operational decision making

In order to have digital twins for operational decision making, we need models which are able to operate when less data are available. Specifically, we identified three requirements. First, we need models which can make predictions for the future with data only until the prediction date, without requiring the future values of variables. Second, these models should be accurate in locations where historical data are available (sampled locations[1]) but also in locations where data have not been collected in the past (unsampled locations). Third, the models should be able to work in cases where high-resolution data are not available e.g. due to lower frequency sampling rates or when less input data streams are available in unsampled locations. The data requirements of such models can be seen in Fig. 3.1b.

[1]Throughout the manuscript we use the term *location(s)*, but without loss of generality this can be considered as *situation(s)*, when considering non-spatially explicit systems.

3.2.3 Proposed method

To satisfy the requirements for operational decision making, we can train ML models on PBM input/output data (so they are also metamodels, see paragraph 3.2.4), discard data we do not need, and then aggregate on lower resolutions. Having a PBM, a target variable and historical data to make simulations, we propose the following steps from an application-based perspective:

1. Define the decision horizon, i.e. how far in the future predictions are going to be made. Based on this boundary, we know how much data we need to retain, as any data after the prediction date are going to be discarded.

2. Choose an aggregation level for the retained data (wherever applicable), with lower resolution than the original data. This will allow the ML model to make predictions even when high-resolution data are not available.

3. Generate data. To generate data we need to define a hyperspace of input combinations for the model. We can choose a full factorial design [119] to contain all the possible combinations of the input variables, or decide to retain only the physically consistent combinations.

4. If possible, discard inputs/output datastreams of the PBM. The fewer inputs the better, because in this way the data requirements of the model are reduced. This decision can be made based on domain knowledge or feature selection procedures.

5. Finally, develop one or several ML models using the data resulting from the above steps.

Evaluation is an important factor to verify that the created models are useful for operational/tactical decision making. A practical way to estimate the predictive capacity of the models is to compare their errors with a threshold based on domain expertise. Also, the models should be tested for their generalization capacity. A way to do this is to consider both sampled and unsampled locations for testing experiments, where data from some locations are excluded from the model training sets, and examine model performance in the excluded locations. Another evaluation aspect is to determine the appropriate training data size of the models. The more variability a model has seen in its training data the more accurate prediction and generalization capacity it should have. In the case where more data do not increase prediction performance it could mean that they do not add any variability and hence we do not need to generate much data in the future. In our case, data quantity is controlled by the amount of data that we generate with the PBM. Therefore, an evaluation step could be to test models of different scales by including different amounts of locations, years, or other parameters.

3.2.4 Related work

Efforts to overcome the inherent shortcomings of PBMs for operational decision-making have been focused on combining PBMs with ML through the concept of metamodeling. Metamodels (also called surrogate or hybrid models) refer to models which mimic the behavior of other models [120]. ML metamodels have been used in agricultural and environmental sciences to cope with a variety of problems. To instill domain knowledge to ML models the authors of [121] train a neural network on PBM output using a custom loss function to predict the water temperature in lakes. To reduce the long execution times of PBMS, metamodels have been employed to predict maize yield and compare the results with those of the PBMs and ML models [122, 123]. To accelerate sensitivity analysis, metamodels have been trained on the output of agricultural simulators [124]. Also, hydrological metamodels have been evaluated for their performance in terms of speed and accuracy [125, 126], as well as generalization capacity (domain adaptation) in unsampled areas [127]. Likewise, to extrapolate at regional and national levels, metamodels have been deployed in environmental management [113]. Lastly, to work in situations where PBM inputs are not available, the authors of [128] create metamodels to predict pre-season maize yield for decision support.

The aforementioned studies focus on each of the advantages of metamodeling individually, whether it is domain knowledge imputation, faster computation times, improved generalization capacity over PBMs, or working with less data. Also, most of these studies make an effort to create models that predict the variable of interest at any time of its evolution, similar to what PBMs do, i.e. by predicting state variables for each simulation step. In this work, we introduce a generic method to exploit these advantages, as well as to deal with data resolution problems which were not explicitly mentioned in those studies, and also we do it for a specific point in time in the future of the target variable.

3.3 Methods

3.3.1 Overview

To assess the method described in 3.2.3 we performed a case study of grass pasture NRR prediction in different locations (see Fig. 3.2) of New Zealand.

The application of nitrogen along with environmental factors such as temperature and time of year greatly affects pasture growth [129], so it is important to know the nitrogen response rate.

To examine the reliability of our models we performed a sampled and an unsampled location experiment. In the sampled location experiment, we assessed the predictive capacity of the models in cases where data from the testing locations are available. In the unsampled location experiment, we examined the generalization capacity of the models in cases where data from the testing locations are unavailable. For both sampled and unsampled location experiments, we iteratively considered each location to be a testing location to be able to better establish our verdicts. To argue about the predictive and generalization capacities we used a case study-specific example where we compared the models' performance with a threshold that makes sense for crop practitioners. Also, we created models of different scales by using various amounts of data for training, and examined how data quantity included in training affects their performance.

3.3.2 Case study

The target of our prediction was the expected two-month nitrogen response rate (NRR; kg of additional, i.e. compared to not applying any fertilizer) of pasture dry matter grown in the two months after fertilizer application per kg of N fertilizer applied. As in most countries, pastures in New Zealand suffer a chronic deficiency of nitrogen [130, 131] and farmers apply nitrogen-containing fertilizers to increase pasture growth rates [132, 133]. Nitrogen fertilizer can be applied regularly (e.g. after each grazing event) or more tactically to manipulate the supply of pasture available to feed stock. As fertilizer costs increase, environmental concerns about leaching of nitrogen increase and/or the prices received for meat and milk decrease, farmers become more interested in understanding when best to apply nitrogen fertilizer to obtain the best NRR. Current NRR estimators are based on rules-of-thumb that consider the month of year, soil temperature, soil nitrogen, or pasture growth rate [134–136].

There are PBMs that can estimate NRR based on site (soil properties, pasture type) and the prevailing conditions (weather) but they have limited usefulness as operational estimators of NRR, because the weather for the two months after a proposed current or future application date are not known, and such data are required to run the model. Also, while there are some NRR data available from experiments, they are sparse and not sufficient to train ML models.

3.3.3 Data generation

We used APSIM v7.10 r4191 [24, 137] to generate the training and testing data. Pasture growth was simulated with the AgPasture module [138] which has already been demonstrated to be a reasonable estimator of pasture growth in New Zealand [139, 140]. The range of input conditions covered eight contrasting locations in New Zealand (Fig. 3.2) and are given in Table 3.1.

Pasture NRR is known to be influenced by soil water and nitrogen availability, temperature, and solar radiation. The combinations of input conditions were designed to provide coverage across these variables, along with 40 years of historical weather data from the New Zealand Virtual Climate Station Network [118, 141], which gave a rich source of variation in weather after fertilizer application.

A hyperspace of parameters was created using the full factorial of the input conditions and put into APSIM. The total number of generated simulations was 1,658,880. After removing the control simulations, (see Table. 3.1) 1,382,400 remained.

Figure 3.2: The eight locations included in the generated dataset.

3.3.4 Data preprocessing

The data generated by APSIM were processed to form a regression problem where the target variable was the NRR and the inputs were the weather, treatment options regarding the fertilizer and irrigation, and biophysical variables. First, the NRR was calculated at two months after fertilization for each non-control simulation. Second, from the generated daily data only the samples in

Table 3.1: The simulation parameters of APSIM. The factorial of those parameters was used to create a hyperspace of input combinations.

Simulation parameters	
Location	daily weather from eight sites spanning the country
Soil water	42, 67, 110 and 177 mm of plant-available water stored to 600 mm deep
Soil fertility	carbon concentration in the top 75 mm of 2, 4, and 6%
Irrigation	irrigated with a centre-pivot or dryland
Fertilizer year	years 1979-2018
Fertilizer month	January-December
Fertilizer day	5th, 15th and 25th of the month
Fertilizer rate	0 (control), 20, 40, 60, 80 and 100 kg N /ha

a window of 28 days before fertilization were retained. This window was selected because in the experience of the authors, pasture 'loses memory' of past conditions relatively quickly provided it is not under- or over-grazed. Weather data after the fertilization were also not considered as such data would be unavailable under operational conditions. Third, the generated data were split into 80/20% training/test sets based on years to avoid information leakage during the later stages of preprocessing. The training and test sets included the year ranges 1979-2010 and 2011-2018, respectively. Fourth, the weather and biophysical variables were aggregated using their weekly mean values. Finally, only a subset of the variables was preserved. This subset included weather variables, simulation parameters (soil water, soil fertility, irrigation, fertilizer month, fertilizer rate), and biophysical variables produced by APSIM (above ground pasture mass, net increase in herbage above-ground dry matter, potential growth if there was no water and no N limitation, soil water stored from 0 to 300mm, soil temperature at 300mm, soil temperature at 50mm, herbage nitrogen concentration in dry matter). These variables were preserved because they were considered to be likely drivers and also known prior to fertilization (to ensure operational usefulness), based on expert knowledge of the authors.

3.3.5 Model scale

Different models were created using different amounts of data. We considered models on three scales: local, regional, and national, each including a different number of locations. The criterion for selecting the locations differed, based on whether the experiment was performed in sampled or unsampled locations.

In the sampled location experiment, the locations were selected based on a climate matching process. The degree of climatic similarity between sites was assessed using the CLIMEX "Match Climates" algorithm [142]. This algorithm produces a composite match index (CMI, from 0 to 1) which indicates the similarity between two locations in weekly average maximum and minimum temperatures, total annual rainfall, seasonal pattern of rainfall, relative humidity and modelled soil moisture. The required climate data were obtained for the nearest 0.05° location from NIWA's Virtual Climate Station Network [141] for the period 1979 to 2010, i.e. using data only from the training set. The results were expressed as a matrix of pairwise CMIs between all sites. In this experiment, the local model included data from the sampled location, the regional model from the sampled location and the best two matches for this location, and the national model data from all the locations.

In the unsampled location experiment, the locations included in each model were selected based on minimum haversine distances from the testing locations. The reason for not using climate matching with CLIMEX was the assumption that data from the unsampled locations were not available, and as a result climate matching could not be performed. The local model included data from the nearest neighbor of the unsampled location, the regional from the three nearest neighbors, and the national from all the locations except the unsampled one. See Fig. 3.3 for a visualization of training models of different size, and Table A.2 in the appendix for the locations included in each model.

3

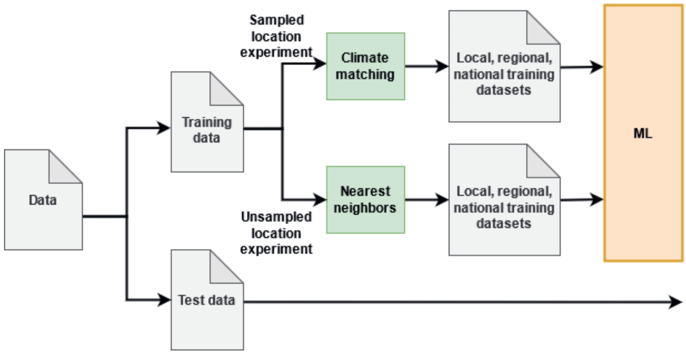


Figure 3.3: The splitting of the processed data to create models for testing in sampled and unsampled location experiments.

3.3.6 Machine learning pipeline

The models were developed with the Random Forest algorithm. Random Forest was selected based on the results of preliminary exploration (see Table A.3 in the appendix). Feature selection was not performed since we had only a few features, which were all considered explanatory. Training data were standardized for each location and experiment, and test set data were standardized with the corresponding scaler. Categorical variables like irrigation (*on/off*) were converted to ordinal. Hyperparameter tuning was performed using Bayesian optimization with 25 iterations and the 5-fold cross-validation score as a metric for each iteration. The tuned parameters can be seen in Table 3.2.

Table 3.2: Tuned parameters of Random Forest and their ranges.

Random Forest parameters	
n_estimators	50-800
max_depth	3-12
min_samples_split	30-500
min_samples_leaf	30-500
max_features	0.33

3.3.7 Evaluation

The predictive capacity of the models was evaluated using the root mean squared error and the residuals of the models on a monthly and yearly basis. A threshold of 5 kgDM/ha/KgN[2] (NRR) was selected, based on expert knowledge, to investigate if the models were accurate enough from a practical perspective. To test the generalization of the models, RMSE and residuals were also examined against the threshold of 5 in unsampled locations.

3.3.8 Experimental setup

The data preprocessing stage was carried out utilizing the *Apache Spark* framework in standalone mode. The ML models were developed using the *scikit-learn* library in *Python*. The experiments took place in a computing node with an *Intel Xeon E5-2630 v4* CPU and 120GB of RAM.

3.3.9 Software availability

The code used for the case study of this paper can be found on GitHub [3].

3.4 Results

In the following sections we present RMSE values and residual plots for sampled and unsampled locations. The errors of the models fluctuated depending

[2] kg of dry matter/ha/kg of nitrogen
[3] https://github.com/BigDataWUR/simulation-assisted-ML

on model scale, location, month and year of application, and whether the location was considered to be sampled/unsampled. None of the models proved to be universally better on all the locations or in both the sampled/unsampled testing experiments. However, some of them showed higher performance and generalization capacity than others in certain cases.

3.4.1 Sampled locations experiment

For the sampled location experiment, regional models had lower RMSEs than the local and national in 4 out of 8 locations, but the error differences between the models were smaller than 0.03. National models had the second-best performance. RMSEs for each model and location can be seen in Table 3.3.

Prediction residuals are illustrated in Fig. 3.4. We observe that errors were mostly below the operational threshold of 5 kgDM/ha/KgN. Exceptions were the months January and February which showed errors close to 5 in some cases. On a closer inspection, we observed large fluctuations based on whether there was irrigation or not (Fig. A.1). In the non-irrigated case, we noticed that for January, February and December the residuals were larger than our threshold of 5 kgDM/ha/KgN. For the other months the performance was well below our threshold. In the irrigated case, the residuals took considerably smaller values.

On a yearly basis (Fig. 3.5), the candles of the residuals were below 2.5, except for Ruakura in 2016 and some years in Lincoln which were higher than 2.5 but still lower than 5. Separating the irrigated and non-irrigated cases, we found that the irrigated cases had residuals consistently lower than our threshold. For the non-irrigated cases (Fig. A.3) we observed that the years 2015, 2016 had larger residuals in several locations.

3.4.2 Unsampled location experiment

In the unsampled location experiment (Fig. 3.4), we observed that the performance of the models generally decreased compared to the sampled experiment. This decrease was more evident in Lincoln and Kokatahi while in the rest of the locations the differences are minor. The regional models outperformed the national and local models in 4 locations (Fig. 3.3). The performance of the regional models was close to that of the national models in many cases. The only location where a local model outperformed the other two was in Mahana.

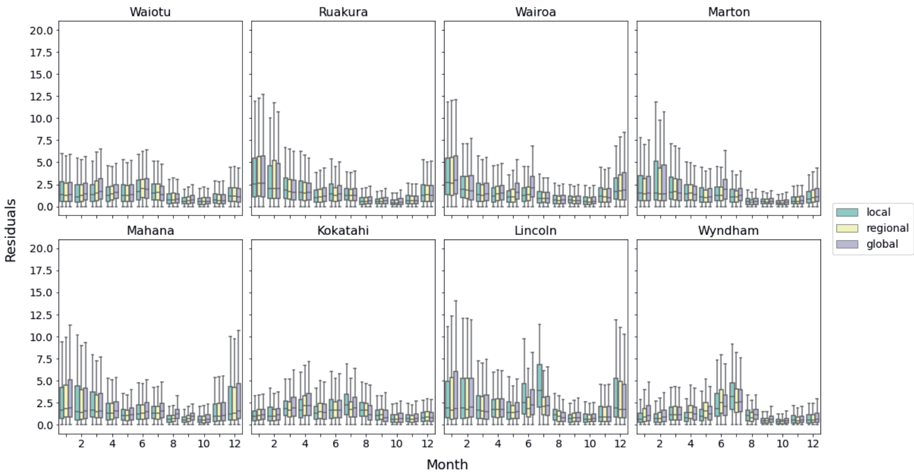
From the residual plots on a monthly basis (Fig. A.2) we observed considerable variation in the residuals between the irrigated and non-irrigated cases. Also, we noticed that the interquartile ranges had been increased compared to

Table 3.3: The test set RMSEs of the models in the sampled/unsampled location experiments.

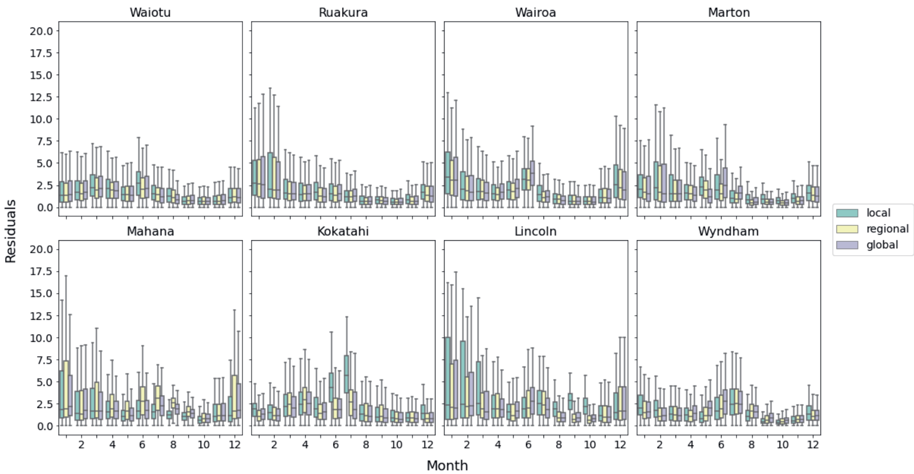
Experiment	Model size	Waiotu	Ruakura	Wairoa	Marton	Mahana	Kokatahi	Lincoln	Wyndham
Sampled locations	Local	2.42	2.84	2.7	2.92	3.16	1.97	3.86	2.13
	Regional	2.32	2.83	2.65	2.63	3.08	1.99	3.47	2.13
	National	2.3	2.78	2.76	2.72	3.26	2.09	3.56	2.1
Unsampled locations	Local	2.53	2.94	3.15	2.93	3.26	3.15	4.57	2.44
	Regional	2.4	2.82	2.86	2.66	3.97	2.4	3.88	2.28
	National	2.39	2.8	2.96	2.83	3.46	2.27	3.92	2.34

the sampled locations, especially for the local models, and were higher than 5 in many occasions, with the largest errors happening in Lincoln.

From the residual plots on a yearly basis (Fig. A.4), we observed that the interquartile ranges had been increased compared to the sampled location experiment. Again, the years 2014-2016 had the widest interquartile ranges, with those of the Lincoln local model displaying the largest errors. Except for those years, we could say that the performance of each model is stable across the years, for each location.

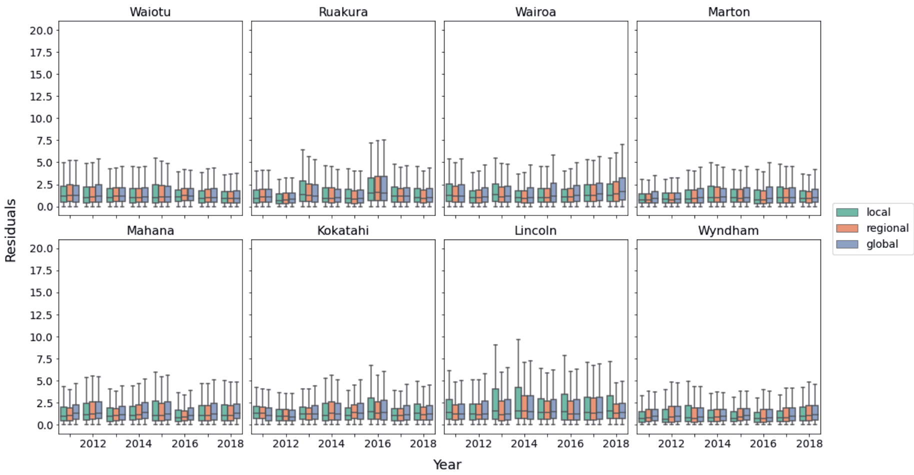


(a)

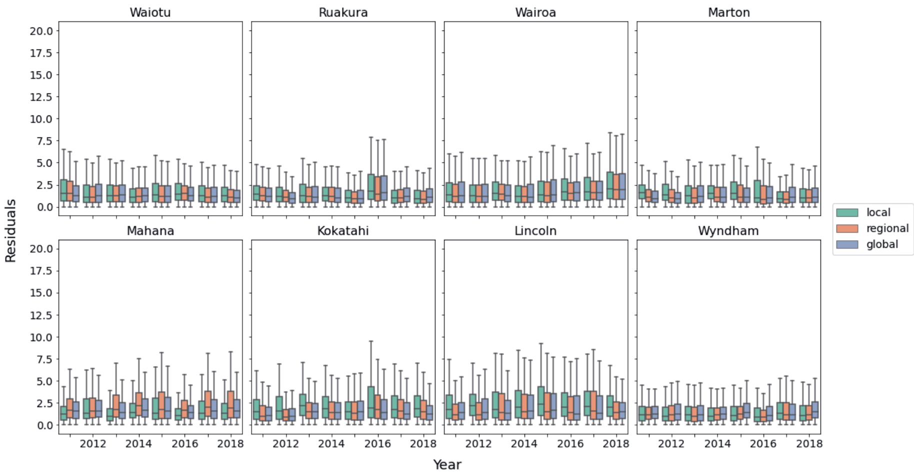


(b)

Figure 3.4: **Monthly** test set residuals of models for **sampled** (a) and **unsampled** (b) location experiments.



(a)



(b)

Figure 3.5: **Yearly** test set residuals of models for **sampled** and (a) and **unsampled** location experiments (b).

3.5 Discussion

In our experiments, the models captured in most cases sufficient variation from the data to achieve RMSEs lower than the threshold of 5 kgDM/ha/KgN. This means that they could be potentially used in practical applications where weather data after fertilization are missing, or data are on a lower resolution than those that APSIM expects. These results persisted in the unsampled location experiment, thus providing evidence that the models are operational in locations where data do not exist to calibrate PBMs, as well as locations not included in the training set of the models. In the following sections, we interpret the results of the local, regional and national models, and discuss the models as a product of the proposed model development methodology.

3.5.1 Predictive capacity (sampled locations)

When separately analyzing the irrigated/non-irrigated cases, we observed that the lack of irrigation hindered the predictive capacity of the models. The reason for this impediment is that when no irrigation is provided, the weather conditions become the driving factor of the NRR, because the grass relies solely on rain to grow. Therefore, as several uncertainty factors pile up (weather volatility, NRR sensitivity to weather, predictions two months in the future without knowing the weather), the results are expected to deteriorate, but they are not indicative of the general model performance. The deterioration was sharper during the spring/summer months November, December, January and February because irregular rainfall is most critical in these seasons. Also, the performance degradation was not the same in all the locations since some locations have more favorable weather conditions than others.

Comparing the models, we observed small differences in model performance. At first glance, this seems counter-intuitive since bigger models were trained on supersets of the smaller model data. This means that they have the same information to learn from, and thus they should perform at least equally well. However, this is not the case since the smaller models seem to benefit more from additional data from locations with similar climates than from having more data from locations with less similar climates. Regarding the national models, they have somewhat higher RMSEs than the other two models because they include data from all the locations, which makes them harder to adapt to local conditions.

The models showed good performance through each month of the year for the irrigated case. Interquartile ranges were mostly below 5 which means that 50% of the values lie within this range. In different locations we see different months having the largest residuals. This has to do with the variation in their microclimates, since rainfall and temperatures can be disparate. Residuals went as high as 7.5 in Lincoln, which is characterized by low precipitation amounts as can be seen in Fig. A.5.

Also, we observed that the errors of the models are consistent throughout the years. There is some variation for 2016 and 2018 in Kokatahi and Wairoa, mainly due to local weather conditions and extreme events which the models were unable to capture. This aligns with our expectation, since extreme events are rare (so there are only a few in the dataset) and also because their presence may be imbalanced between the training/test sets. Yet in most cases the model shows adequate predictive capacity, even eight years after the last year that was included in the training set.

From the perspective of model operationalization, the models proved that they can complement the PBMs to provide predictions of adequate accuracy, and overcome the problem of data availability to a certain degree. This degree depends on the level of uncertainty involved in the predictions and the ML pipeline used to build the models. In a digital twin, these models could provide the first line against working with limited data. Several models could be included with different tasks. For example, a model providing predictions for the irrigated case of a specific location with a specific soil type, one trained on extreme weather, one on non-irrigated cases and so on. These ensembles of models could potentially capture a large degree of variation while waiting for more data to become available.

3.5.2 Model generalization (unsampled locations)

In the unsampled location experiment, the differences between the models became more evident, as the local models' performance deteriorated more than the others (Fig. A.2). The reason behind this difference is that the local models had data from only one location, which was not the location where the testing happened. On the other hand, the bigger models were favored in this experiment since they included data from multiple locations and could extrapolate better. This phenomenon is more noticeable in the non-irrigated cases (Fig. A.2a) where the local model shows high deviations from the simulated NRRs. Having said that, with the exception of January, February and December, the RMSEs were below 5 for all models. Those three months included temperatures higher than 25°C (Fig.A.6) which can be harmful to the grass, and when combined with the non-irrigated case the uncertainty for the future increases.

On the monthly residual plots of the unsampled location experiment (Fig. A.2a) we saw a more detailed picture of model performance with respect to size. Many times the residuals surpassed our threshold, especially those of the local and regional models. From these cases we can deduce that the national models are superior to the local and regional ones. The cases where the national models had increased interquartile ranges happened on the same months and locations as in the sampled location experiment (e.g. January-February in Ruakura, February-March in Marton). The latter observation means that the increased ranges are not a matter of hindered generalization among locations, but of an inability to capture variability in those climates due to the features included in the models.

From the residuals on a yearly basis we observed that the errors are mostly consistent across the years in each location. The local models showed the highest fluctuations throughout the years (like in Ruakura and Lincoln). The regional models had the second-highest discrepancies throughout the years (like in Mahana, Lincoln). The national models were the most stable ones. This behavior can be attributed to the amount of data included in each model, because the more data from different locations a model includes the more divergent weather conditions it has seen. This means that it can generalize better in the weather conditions of the years to come. Also, it is interesting to see that models can generalize in unsampled locations many years (8) later since the last year included in the training sets.

From an operational perspective, the models showed a capacity to generalize in previously unseen conditions. A recommendation we would make when starting modeling in unsampled locations would be to begin with a national model rather than a model from the single nearest/similar location. In digital twins where existing models cannot be applied due to lack of calibration data or insufficient observation training data, these models can provide a first impression of variables of interest in the future, even though there are still limitations. Again, the model performance could be improved by training for more specific scenarios and using more advanced ML techniques.

3.5.3 Future Work

This line of research could be improved further by generating data from multiple PBMs, and by trying different aggregation levels to find a balance between performance and working with low-resolution data. Also, it would be beneficial to evaluate model performance against ground truth data, which were not available for this case study. Regarding the case study, the data preprocessing and ML procedures could be adapted to better fit the domain of the application by using custom features, performing training/test splits which better balance underrepresented phenomena between the sets, or using stratified sampling to select which simulations are going to be included in each set. More elaborated ML model architectures could further improve performance metrics.

3.6 Conclusions

In this work, we introduced a method to develop operational digital twins by creating models which overcome the problems of data availability and data resolution. We showcased this method using a grass pasture nitrogen response rate case study.

Experimental results verified that this method is able to produce digital twins to offer tactical advice in highly non-linear situations where local conditions and treatment options affect the outcome of the predictions.

The ability of the models to provide accurate predictions in different locations, for both sampled and unsampled experiments, indicates that they can adequately capture the variability encoded in process-based models. The developed models were able to capture the target variable, even without having the complete weather and biophysical time series. This practically allows to develop operational digital twins in cases of limited data availability. Also, model predictions were made on field-level using weekly data instead of daily data that a process-based model would require. As a result, digital twins using these models are capable of operating in situations where process-based models cannot. These advantages, combined with the fact that we did not need to forecast any future weather values to get those results, differentiate this method from the creation of metamodels which just summarize process-based models, and demonstrate that simulation-assisted machine learning is able to offer advice in practical conditions.

Chapter 4

Learning latent representations for operational nitrogen response rate prediction

Appeared as: Christos Pylianidis, Ioannis N. Athanasiadis. Learning latent representations for operational nitrogen response rate prediction, AI for Earth Sciences workshop at International Conference on Learning Representations 2022, 2022

Abstract

Learning latent representations has aided operational decision-making in several disciplines. Its advantages include uncovering hidden interactions in data and automating procedures which were performed manually in the past. Representation learning is also being adopted by earth and environmental sciences. However, there are still subfields that depend on manual feature engineering based on expert knowledge and the use of algorithms which do not utilize the latent space. Relying on those techniques can inhibit operational decision-making since they impose data constraints and inhibit automation. In this work, we adopt a case study for nitrogen response rate prediction and examine if representation learning can be used for operational use. We compare a Multilayer Perceptron, an Autoencoder, and a dual-head Autoencoder with a reference Random Forest model for nitrogen response rate prediction. To bring the predictions closer to an operational setting we assume absence of future weather data, and we are evaluating the models using error metrics and a domain-derived error threshold. The results show that learning latent representations can provide operational nitrogen response rate predictions by offering performance equal and sometimes better than the reference model.

4.1 Introduction

Latent representation learning has been adopted in several disciplines to extract and handle hidden interactions between the input variables allowing for more informed decisions. In geosciences, representation learning algorithms emerge [143] that perform visual analogies in the latent space, similar to how Word2vec can be leveraged to learn how words appear in similar contexts. In medicine, latent representation learning is used [144] to work with incomplete multi-modality data to learn independent representations for the prediction of Alzheimer's appearance. In biology, latent representations are used to model unmeasured quantities like pain and stress [145]. Representation learning has also found its way to the earth and environmental sciences. Examples include learning better representations of 2D coordinates [146], and extracting unknown basin characteristics [147]. However, it has been observed [148] that this is not the case for several subfields, where practitioners prefer to use features based on expert knowledge and already proven algorithms that do not explore the latent space. This creates missed opportunities to examine whether improved predictive performance can be achieved, or new interactions to be found, or even to automate prediction pipelines. A representative case of such a missed opportunity is with estimating nitrogen application for fertilization purposes.

Nitrogen is the nutrient that crops and pasture draw from the soil in the greatest quantities [19] and thus it becomes a growth-limiting factor [20]. Nitrogen deficiency has been associated with low yields [21], and pastures in several countries suffer from it [130, 131]. Farmers apply nitrogen-containing fertilizer to increase pasture growth rates but environmental concerns rise as nitrogen has been linked to soil[22], freshwater and atmosphere pollution[23]. Subsequently, agricultural practitioners are asked to control nitrogen application with precise doses based on NRR[1]. To control nitrogen application, research is being directed towards modern systems like digital twins [149] which can aid decision support through automation and data integration. However, digital twins require components, such as process-based and ML models, that are able to predict NRR across several months in the future to be considered operational. Process-based models that can calculate NRR exist but they are of limited use as the weather months after nitrogen application is unknown yet required to run the model. Also, while NRR observations exist from experiments, they are sparse and not enough to train ML models.

In a recent study [150], we presented a methodology to tackle these data-related problems by training ML models based on process-based model out-

[1]Amount of extra kg of yield for every kg of nitrogen applied ($kg_{yield}/ha/kg_{nitrogen}$)

put. We predicted pasture NRR two months ahead of the prediction date, assuming an absence of intermediate weather data. However, we performed common practices of environmental sciences like selecting features solely based on expert knowledge, averaging weather variables, and feeding all those to Random Forest (RF) [151]. Hence, the latent space of data was not explored, and it was left unchecked if we could achieve similar performance with higher resolution data by learning the latent space. That would be important to examine since methods that learn the latent space have shown to perform equally or better than approaches that do not, as they may capture interactions that are not yet understood. Also, it would promote automation in systems like digital twins by removing the step of manual feature extraction. In this work, we are going to treat this study as a stepping stone, as it proved that we can have accurate NRR predictions in limited data settings, in a situation where an ML model and a process-based model alone were not operational.

Here, we perform a systematic comparison of different architectures to learn the latent space of a synthetic dataset for NRR prediction. We adopt the case study and data provided by [150] and we use RF as a reference for comparing the performance of the architectures. We learn the latent representations of the inputs/outputs of a process-based model and predict NRR with a Multilayer Perceptron (MLP), an autoencoder (AE), and a dual-head autoencoder (DAE). We perform multiple runs for each architecture as well as RF to verify the robustness of each model. We then evaluate the results using error metrics as well as a domain-derived error threshold.

4.2 Materials and Methods

4.2.1 Case study & Data generation

The case study was concerned with finding the pasture NRR for two sites (Fig. B.7) in New Zealand. The prediction target was the NRR of pasture dry matter grown in the two months after fertilizer application. Data generation was performed with APSIM [24]. The simulation parameters of APSIM covered conditions that are known to affect pasture growth. The full factorial [119] of those parameters was created and put to APSIM. The range of each parameter can be seen in Table B.4.

4.2.2 Data preprocessing

The generated data were processed to form a regression problem. The target variable was the NRR and the input variables were the weather, fertilizer amount, fertilization month, irrigation and a subset of biophysical variables produced by APSIM. From the generated daily data, only the data within the first 28 days prior to fertilization were preserved because pasture is supposed to 'lose memory' of past conditions after that time frame. **Weather data after that these 28 days were also discarded as they would be unavailable in operational conditions**. The remaining data were split into 67.5% training, 12.5% validation, and 20% test sets, based on years, to avoid information leakage during later processing stages. The validation set included the years [1979, 1987, 1999, 2007], the training set years [1979-2010] excluding the validation years, and the test set [2011-2018].

4.2.3 Architectures

4.2.3.1 Multilayer Perceptron

An MLP was put in the comparison to examine how its latent space learning capabilities compared with learning compressed representations of an AE. The loss was given by equation 4.1. The network topology can be seen in Fig. 4.1. Training parameters can be found in Appendix 7.4.

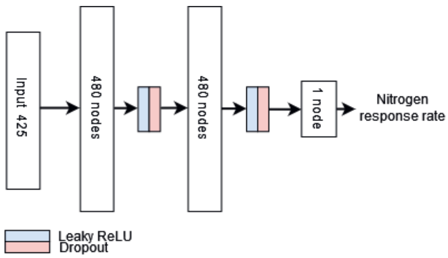


Figure 4.1: The topology of the MLP.

4.2.3.2 Autoencoder

An AE was selected to create a compressed representation of the input variables. The AE included skip connections similarly to [152] from the encoder to the decoder to lessen degradation [153]. The reconstruction loss was given by equation 4.2. After training the AE, the decoder was removed and replaced by an MLP. Training was performed again for the MLP (with loss given by equation 4.1, and a frozen encoder) to learn to predict NRR. The autoencoder topology can be seen inside the dashed line of Fig. 4.2.

$$L_{nrr} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (4.1) \qquad L_{rec} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{425}(y_{ij} - \hat{y_{ij}})^2 \quad (4.2)$$

where $N = batch\ size.$

4.2.3.3 Dual-head autoencoder

The encoder and decoder parts were the same as of the 'simple' AE. The addition was that the compressed representation was then directed to an MLP which carried out the NRR prediction task. The network topology can be seen in Fig. 4.2. Both the AE and the MLP were trained simultaneously, with the total loss being the summation of the equations 4.1, 4.2.

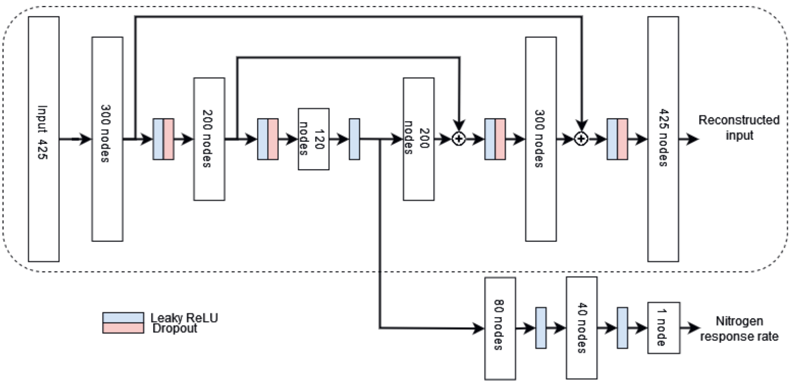


Figure 4.2: The topologies of the AE (inside the dashed border), and the DAE (altogether).

4.2.4 Evaluation

The performance of the different architectures was compared using the *mean absolute error* (MAE), the variance learned from the latent representations using R^2, and the standard deviation of the predictions. Also, the predictive capacity of the models was assessed using a domain-derived error threshold of 5 $kg_{yield}/ha/kg_{nitrogen}$. Prediction residuals systematically above that threshold constituted a model incapable for operational use. Each architecture, as well as RF, were ran 5 times with different seeds to verify the robustness of the results.

4.3 Results

In Table 4.1, we see the error metrics for each architecture and location aggregated over the runs. RF has the lowest error and highest explained variance for both locations. AE has the largest error and lowest R^2. DAE has the lowest errors among the architectures with just a slight edge over MLP. Regarding the standard deviations of the predictions, AE has the lowest deviation and DAE the highest.

Table 4.1: Error metrics for each architecture and RF aggregated over the runs. σ refers to the standard deviation of the predictions of the runs.

	RF			**MLP**			**AE**			**DAE**		
	MAE	R^2	σ	MAE	R^2	σ	MAE	R^2	σ	MAE	R^2	σ
Waiotu	1.55	0.68	3.53	1.85	0.62	3.63	2.26	0.45	3.34	1.72	0.65	3.62
Mahana	1.87	0.61	4.16	2.19	0.53	4.57	2.72	0.38	4.02	2.07	0.5	4.9

In Fig. 4.3, we can see how the residuals of the different architectures compared to RF across months. The residuals were aggregated over years and the five runs. For the first location, Waiotu, we observe that all candle bodies are below the domain-derived threshold that we set, with some upper whiskers overcoming the threshold. DAE seems to be the best performing architecture, since it has the shortest body of the three and also lower medians. Also, DAE appears to have slightly lower errors than RF in several cases. AE appears to have the largest errors, with large candles and extended upper whiskers. For Mahana, most candles are below our threshold but with larger bodies than Waiotu and taller upper whiskers. For January and December AE is above and close to the threshold respectively, generally having the highest errors. MLP and DAE seem to outperform RF for January, February, and December. Again, DAE has the best performance of the three architectures with candles being lower than the rest and lower medians.

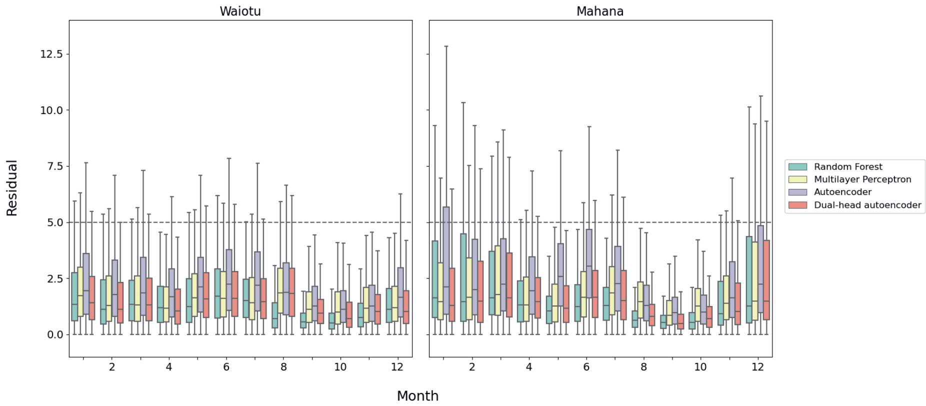


Figure 4.3: Residuals for each architecture and RF aggregated over years and runs. The horizontal dashed line indicates the domain-derived threshold. The body of the candles represents 50% of the values, and the bottom and top whiskers 25% each. The horizontal lines inside the candles show the median.

4.4 Discussion

From a performance-oriented perspective, we could deduce that RF is the best model by looking at the error metrics. However, we cannot judge how much better it is from MLP and DAE or how well the different architectures learned because their errors and standard deviations were similar. A more clear case is that of AE, which underperforms the rest of the models considerably. The standard deviation of its predictions might be the smallest but this may be due to learning a small part of the lower dimensional manifold, created on the output of the encoder, and thus not being able to offer varied predictions.

Examining the residuals of the architectures, we observe that they provide predictions mostly within our domain-derived threshold. The multiple runs and yearly aggregation demonstrate the stability[2] of the models, showcasing their robustness. This conveys that the models were able to extract latent representations which allow them to be potentially used in an operational setting. AE appeared to be the weakest model since its candles were generally larger, exceeding our threshold in Mahana. The two stage training (first autoencoder, then replacing then decoder with an MLP) may have caused it to weigh more on learning how to reconstruct its inputs rather than how NRR is connected with them. On the other hand, the MLP was able to extract more meaningful representations for NRR predictions something evident from the fact that in

[2]Variation between the months exists due to seasonality. December to February is summer in New Zealand with conditions that increase uncertainty for pasture growth and thus NRR errors.

many months it was on par and sometimes better than RF. Similarly, DAE performed equal or better than RF in most months for both locations. This may imply that the latent space that MLP and DAE learned covered aspects which were not represented in the manually derived expert features of RF. Also, the performance gap between AE and DAE showed that optimizing simultaneously for two tasks when one task depends on the other can make the network learn better representations in the context of this study.

An aspect potentially affecting the results of the architectures is how well input features can be represented in the latent space learned by the models. APSIM has a binary input variable to control the existence of irrigation which materially changes NRR. This variable is the only signal outside of APSIM that indicates this type of change. In our architectures there are several layers and this signal may be difficult to be preserved and projected in the latent space. On the contrary, for algorithms like RF this signal is not lost and can easily change how predictions are made. This may be a reason for not having higher performance with the different tested architectures and something to be accounted for when learning latent representations from environmental data.

4.5 Conclusion and Future work

In this study, we assessed the ability of three neural network architectures to learn the latent space of process-based model output for operational decision support. We compared the results with those of RF which was already proved operational in another study. The results were promising since all architectures were able to learn representations that captured enough variation to be considered operational. The MLP and DAE outperformed RF in certain cases, showing that they can uncover latent factors from the input space which accounted for more variability than manually selected features based on domain knowledge. This is an important step towards providing operational decision support in modern systems like digital twins, avoiding feature engineering in certain cases and automating prediction pipelines.

In the future, we would like to experiment with more synthetic datasets to examine if we can generalize our findings to other case studies. Also, we would like to validate the models using observation data to further verify how operational the created models are. Another important aspect would be to experiment with architectures that provide explicit interpretability of the latent space and examine how this space compares with expert-derived features.

Chapter 5

Location-Specific vs Location-Agnostic Machine Learning Metamodels for Predicting Pasture Nitrogen Response Rate

Published as: Christos Pylianidis, Ioannis N. Athanasiadis. Location-Specific vs Location-Agnostic Machine Learning Metamodels for Predicting Pasture Nitrogen Response Rate, International Conference on Pattern Recognition International Workshops and Challenges, vol 12666, pp 45-54, 2021

Abstract

In this work we compare the performance of a location-specific and a location-agnostic machine learning metamodel for crop nitrogen response rate prediction. We conduct a case study for grass-only pasture in several locations in New Zealand. We generate a large dataset of APSIM simulation outputs and train machine learning models based on that data. Initially, we examine how the models perform at the location where the location-specific model was trained. We then perform the *Mann–Whitney U test* to see if the difference in the predictions of the two models (i.e. location-specific and location-agnostic) is significant. We expand this procedure to other locations to investigate the generalization capability of the models. We find that there is no statistically significant difference in the predictions of the two models. This is both interesting and useful because the location-agnostic model generalizes better than the location-specific model which means that it can be applied to virgin sites with similar confidence to experienced sites.

5.1 Introduction

Environmental data are growing in an unprecedented way [154]. Many domains of Environmental Research utilize those data and combine them with Machine Learning (ML) techniques [155] to enable understanding. However, there are domains like grassland-based primary production systems where certain areas (e.g. pasture production, nitrogen leaching) have limited, low quality data, making them poor candidates for ML applications. In such areas, dynamic models are deployed to seek causality and make predictions based on first principles but sometimes they need data that is not available.

ML has been used in a complementary way with dynamic models to summarize them and capture their embedded knowledge. The resulting ML models are also known as metamodels, surrogate models or emulators. The knowledge summarization is achieved by training ML models using the output of dynamic model simulations. Advantages of this technique include the reduction in need of observation data [156], the use of fewer inputs [113] and faster computation times [157] for large scale systems than the dynamic models. The paradigm of summarizing dynamic models is applied in several disciplines from physics [158] to hydrology [125].

Dynamic model summarization has also been studied in agriculture [159]. Several studies have examined the application of ML surrogate models for sensitivity analysis [124], the performance of different ML algorithms for crop model summarization [128] and the amount of data needed for accurate predictions [128]. In these works, the authors trained ML models in generated datasets to examine how well the models can generalize, using either one or all the available locations, and not testing in other locations. However, the generalization capability of a model over multiple locations does not mean that it performs better than a model specifically trained for that location. Since there are cases where the interest lies in absolute performance or generalizability of the summarization model it would be compelling to investigate how location-specific and location-agnostic models compare in those aspects.

The purpose of this work is to investigate the performance difference of location-specific and location-agnostic ML metamodels using a case study approach. To achieve this goal, we first generate a large dataset across several locations using a crop simulation framework. Second, we aggregate the generated data and train a ML model using all the available locations, and a second ML model using only one location. Next, we test the ML models on a dataset comprised of samples of the latter location. We compare the results using statistical metrics, and examine if they are statistically different using the *Mann–Whitney U test* [160] which has been used for comparing ML models in other works [161]. Finally, we investigate the trade-off between model performance and generalizability by testing the models in the rest of the locations of our dataset.

5.2 Materials and methods

5.2.1 Case study, data description

A case study was performed to predict the grass-only pasture NRR in different locations in New Zealand. The application of nitrogen along with environmental factors such as temperature and time of year greatly affects pasture growth [129] so it is important to know the NRR. Our dataset consisted of grass pasture growth simulations performed with the APSIM modeling and simulation framework [24]. A hyperspace of parameters was created and put to the simulator. The simulation parameters for APSIM included daily historical weather data from eight locations in New Zealand and management treatment options which can be seen in Table 5.1. The cross-product of those parameters was used to create a hyperspace of input combinations for APSIM. The total number of simulations was 1,658,880 which should have yielded 1,382,400 NRRs. However, the input combinations included application of fertilizer at times when pasture growth was near zero because of dry soil conditions or cold temperatures. These were excluded from the analysis as the calculated N response rate was known to be unreliable. In total there were 1,036,800 response rates available for further analysis. Our target was to predict the 3-month NRR – the additional pasture dry matter grown in the three months after fertilizer application over that from a non-fertilizer control divided by the kg of nitrogen in the fertilizer applied. The outputs of APSIM consisted of the NRR, biophysical variables related to fertilizer concentration in grass and moisture in soil.

Table 5.1: The simulation parameters of APSIM. The cross-product of those parameters was used to create a hyperspace of input combinations.

	Simulation parameters
Location	weather from eight sites spanning the country
Soil water	42 or 77 mm of plant-available water stored to 600 mm deep
Soil fertility	carbon concentration in the top 75 mm of 2, 4, or 6%
Irrigation	irrigated with a centre-pivot or dryland
Fertilizer year	all years from 1979 to 2018
Fertilizer month	all months of the year
Fertilizer day	5th, 15th and 25th of the month
Fertilizer rate	0 (control), 20, 40, 60, 80 and 100 kg N /ha

5.2.2 Data preprocessing

The generated data were preprocessed to formulate a regression problem where the target variable was the NRR and the inputs were the weather, some treatment options regarding the fertilizer and irrigation, and some biophysical variables. The generated data were aggregated from a daily to a simulation basis, to imbue memory to the data. First, the data were split into training and test sets to avoid information leakage during the latter stages of processing. The split happened based on the year, taking one year to the test set every five years and the rest to the training set. The resulting percentage of training and test samples was 80/20%. Second, from the generated daily data only the samples in a window of 28 days before fertilization were kept. This range was selected because grass pasture is known to not be affected by past conditions further than this window provided it is not under- or over-grazed. Also, weather data after the first fertilization was not considered because preliminary work has shown that it is not needed to achieve meaningful results. Third, only the variables related to the weather, simulation parameters, NRR and to some of the biophysical variables were preserved which were considered to be likely drivers, based on expert knowledge of the NRR. Fourth, the weather and biophysical variables were aggregated using their weekly mean values. Finally, the aforementioned steps were repeated once to form an aggregated dataset containing all the locations, and once for each of the eight locations contained in our dataset. The output of those steps was an aggregated dataset (training set) for the location-specific model, an aggregated dataset (training set) for the location-agnostic model, and an aggregated dataset (test set) for each location.

5.2.3 Machine learning pipeline

The aggregated datasets were then passed to the ML stage. In this stage, the training and test data were standardized using the same data transformer to keep the same mean for both transformations. To clarify further, each test set was using the scaler of the location-agnostic model and the location-specific model so that each model can have a version of the test set according to the mean of its training set. Categorical variables were converted to ordinal by substituting them with numbers. Then, hyperparameter optimization was performed to the Random Forest algorithm using gridsearch with 5-fold cross-validation. The gridsearch parameters were *n_estimators* {200, 300, 400, 500}, *max_depth* {3, 5, 7, 11}, *min_samples_split* {2, 3, 4, 8, 16}, *min_samples_leaf* {1, 2, 4, 8, 16} and *max_features* {0.33, sqrt, None}. The out-of-bag score was used for the building of the Random Forest trees. No feature selection was performed because the number of features was small (64) compared to the size of the training datasets ($1,044,060$ and $130,095$ samples for the multiple and single locations correspondingly). After training, the optimized models of the location-agnostic and location-specific models were tested using the test set of location *Waiotu* where the location-specific model was trained. The pipeline of the ML stage is shown in Fig. 5.1.

5.2.4 Evaluation

The performance of the location-specific and location-agnostic models was first evaluated by comparing error metrics (MAE, RMSE, R^2) of their results on the test set. Then, the *Mann–Whitney U test* was performed on the models' results on the test set to see if the differences were significant. The *Mann–Whitney U test* examines if the distributions of the populations of two groups are equal and it was preferred among other statistical tests because first it is non-parametric, second it assumes that the pairs in the samples do not come from the same populations and third that the observations are ordinal, all of which fit our problem. Consequently, error metrics and the *Mann–Whitney U test* were calculated for the rest of the locations to test the models' generalizability.

5.2.5 Implementation

The data preprocessing stage was developed utilizing the Apache Spark framework. The ML models were developed using the *scikit-learn* library in Python. The experiments took place in a Databricks node consisting of 96 cores and 384GB of RAM to speed up procedures through parallelization.

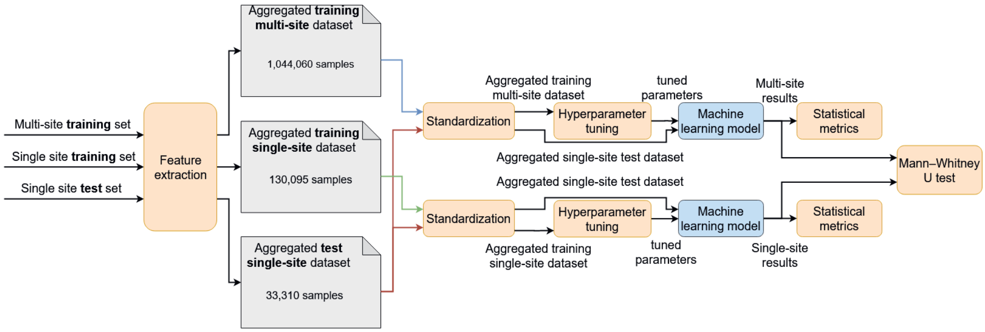


Figure 5.1: The pipeline for the training and testing of the models on the location where the location-specific model was trained. At the end there is also the evaluation stage. The process starts by taking the training and test datasets from the preprocessing stage. It has to be explicitly noted that hyperparameter tuning was performed only on the training set. More specifically the test set was the same for both models but it was standardized for each model individually to preserve the same mean which was used for each training set.

5.3 Results

The hyperparameter tuning procedure selected the following parameters for both models: *n_estimators* 400, *max_depth* 11, *min_samples_split* 2, *min_samples_leaf* 1, *max_features* 0.33. The results of the ML models on the training and test sets are shown in Figure 5.2, along with the distributions of the simulation and the model predictions. We observe that the angle between the identity and regression lines on the test set is smaller for the location-specific model which means that it fits better the location-specific test data. The data points on the test set of the location-agnostic model are more dispersed. Also, we notice that the distributions of the location-specific and location-agnostic model predictions on the test set appear to be similar. The mean and variance of the distributions appear to be close as it can be seen in Table 5.2.

Regarding the error metrics, in Table 5.3 we observe the MAE, RMSE and coefficient of determination (R^2) for both models on the test set of each location. For the location where the location-specific model was trained (Waiotu), we observe that the location-specific model performs better than the location-agnostic model. For the rest of the locations, the location-agnostic model outperforms the location-specific one.

In Table 5.3 we also observe the results of the *Mann–Whitney U test* for each location. For the location where the location-specific model was trained (*Waiotu*) we see that there is no statistically significant difference between the models. The same applies to the location *Ruakura*.

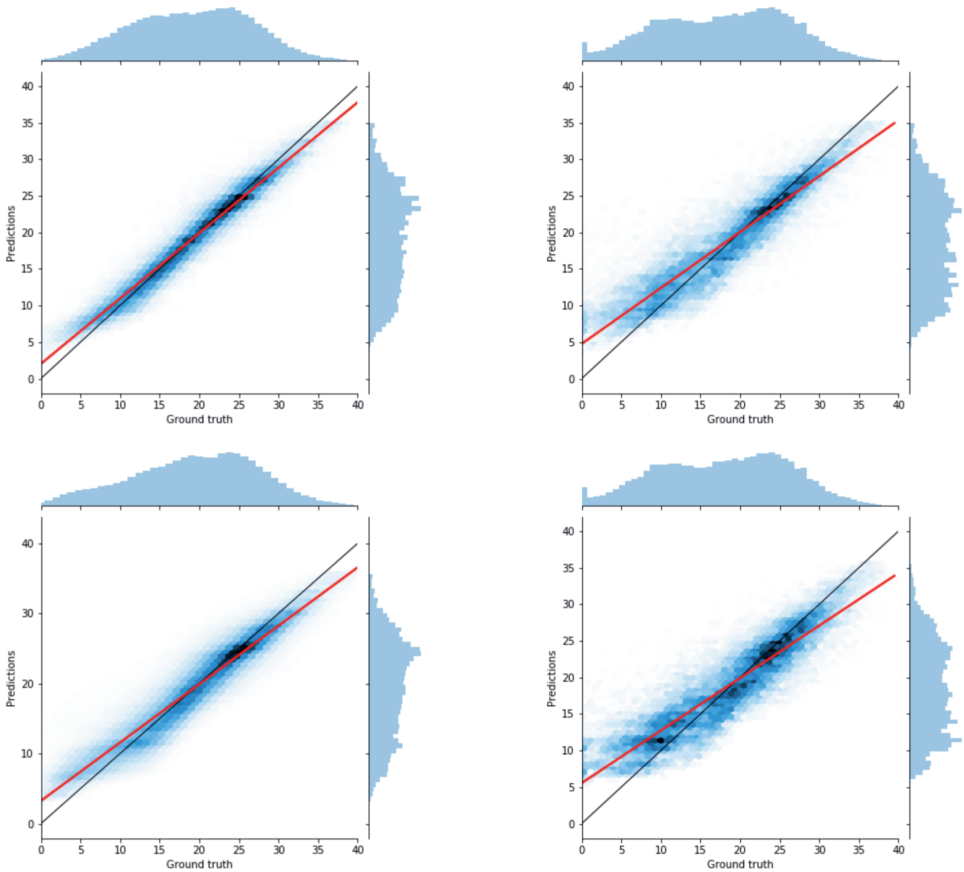


Figure 5.2: The results of the **location-specific model (top row)** and **location-agnostic model (bottom row)** for the **training (left)** and **test (right)** sets. The test set is common for both models and contains data from the location where the location-specific model was trained (*Waiotu*). On the vertical axes are the predictions of the model and on the horizontal axes the simulated values. On top and right of the plots are the distributions of the simulated and predicted values correspondingly. The black lines are the identity lines. The red lines are the regression of Prediction on Ground truth. Darker spots indicate that more predictions fall on the same area.

Table 5.2: The distribution characteristics of the two models for the test set predictions on the location where the location-specific model was trained.

	Location-specific	Location-agnostic
Mean	18.47	18.44
Variance	44.73	40.99
Skewness	0.11	0.18
Kurtosis	-0.90	-0.82

Table 5.3: The error metrics of the location-specific and site agnostic models on the different locations. On the first row are the locations existing in our dataset. *Waiotu* is the location where the site specific model was trained. On the second row are the MAE, RMSE and R^2, for each model and location. The blue and red colors indicate the models with the highest and lowest performance correspondingly, for each location and error metric. On the third row, statistically significant difference on *Mann–Whitney U test* between the predictions of the two models is denoted with as asterisk.

		Waiotu	**Ruakura**	**Wairoa**	**Marton**	**Mahana**	**Kokatahi**	**Lincoln**	**Wyndham**
Location-specific	MAE	2.37	2.72	2.92	3.27	3.44	4.36	4.96	5.62
	RMSE	3.19	3.62	4.03	4.41	4.2	5.81	6.63	7.29
	R^2	0.85	0.78	0.68	0.66	0.66	0.5	0.41	0.38
Location-agnostic	MAE	2.71	2.13	2.71	2.06	2.29	2.56	2.88	2.31
	RMSE	3.55	2.95	3.91	2.83	3.04	3.33	4.08	3.06
	R^2	0.81	0.85	0.7	0.86	0.82	0.83	0.78	0.89
Mann–Whitney U test				*	*	*	*	*	*

5.4 Discussion

The results showed slightly better error metrics for the location-specific model over the location-agnostic model for *Waiotu*. The reason may be that the location-specific model learns the local conditions better since they are only from this location and fewer than those included in the training of the site-agnostic model. For the rest of the locations, the location-agnostic model performs better because it was trained with more data, which also included these locations and as a result, it can generalize better. An interesting finding is that the errors of the location-specific model increase as we move further away from *Waiotu*, as shown in Fig. 5.3. The locations can be seen in Fig. 5.4. This finding indicates that the further away a prediction is made from the training location, the higher the error will be for a location-specific model. On the other hand, the location-agnostic model is not affected since it was trained in a larger dataset which included data from those locations.

Another finding was that there was no statistical difference between the predictions of the two models for *Waiotu*. The location-specific model may perform better but it seems that the gain is marginal and is lost when moving to other locations. The second location with no statistical difference between the models' predictions is *Ruakura*. We assume that this happens because *Ruakura* and *Waiotu* are close to each other and as a result, environmental factors do not vary substantially between those locations.

We deduct that there seems to be a trade-off between accuracy and generalization performance. The location-specific model is trained on a smaller dataset and overfits the data. As a result it performs better for *Waiotu* but the

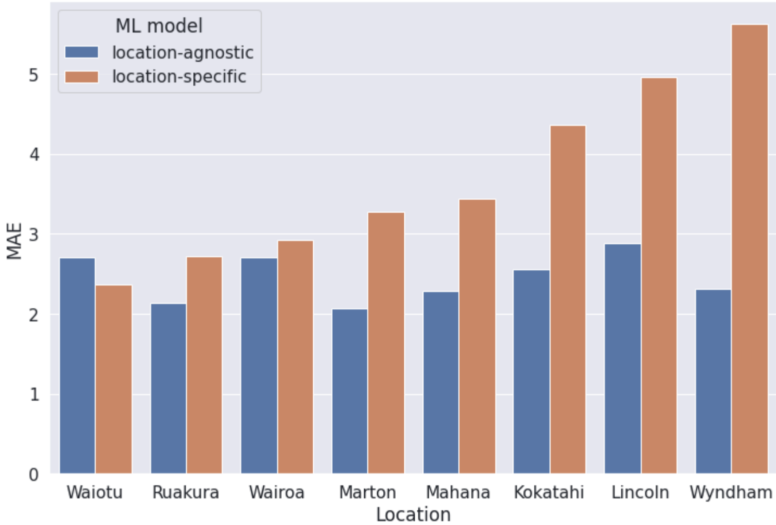


Figure 5.3: MAE of the location-specific and agnostic models for all the locations in our dataset. On the vertical axis is the error and on the horizontal the locations. The orange and blue colors indicate the results of the location-specific and agnostic models respectively.

Figure 5.4: The locations in New Zealand which were included in our dataset. On the top right is *Waiotu* which was used to train the location-specific model. Right next to *Waiotu* is *Ruakura*. The rest of the locations are further away.

location-agnostic model generalizes better. In our opinion, the decision for which model to deploy depends on the use. We emphasize though that the performance difference in this case study is not dramatic for *Waiotu*. On the other hand, the generalization performance is evident especially as we move further away from the location where the location-specific model was trained.

5.5 Limitations

A limitation of our study regarding the performance comparison of the ML models is that the location-agnostic model was trained using data from all the locations. As a result we did not test how the models would perform in a location that would be new to both of them.

Another limitation is that the performance of both models was affected by the way we partitioned years into the training/test split. That is due to seasonality in the generated data, which was not taken into account when performing the split.

5.6 Conclusion & future work

In this work, we examined the performance difference between a location-specific and a location-agnostic metamodel using error metrics and the *Mann–Whitney U test*. We tested the models in different locations including the location where the location-specific model was trained. We found that the location-specific model performs better for the location where it was trained, although not in a statistically significant way. Also, the error metrics in other locations showed that the location-agnostic model generalizes better.

Future work could include the setup of the methodology in a way to test location-specific models for all the available locations to examine if the results will be the same. Also, a location could be left out of both training sets to allow testing in a new location for both models. Besides, different machine learning algorithms could be deployed and tuned even further. The performance of the models could also be improved by adding complex features and features based on agronomic knowledge.

Chapter 6

Domain adaptation with transfer learning for pasture digital twins

Submitted for publication to an international journal: Christos Pylianidis, Michiel G.J. Kallenberg, Ioannis N. Athanasiadis.

Abstract

Domain adaptation is important for agricultural applications because the underlying systems have their own individual characteristics. Applying the same treatment practices (e.g. fertilization) to different systems may not have the desired effect due to those characteristics. Domain adaptation is also an inherent aspect of digital twins. In this work, we examine the potential of transfer learning to be used for domain adaptation in pasture digital twins. We use a synthetic dataset of grass pasture simulations to pretrain and fine-tune machine learning metamodels for nitrogen response rate prediction. We investigate the outcome in locations with diverse climates, and examine the effect on the results of including more weather data and more agricultural management practices during the pretraining phase. We find that transfer learning seems promising to make the models adapt to new conditions. Moreover, our experiments show that adding more weather data on the pretraining phase has a small effect on fined-tuned model performance compared to adding more management practices but more work is needed to further study this behavior.

6.1 Introduction

DSS are widely used in agriculture to convert data to practical knowledge [1, 162]. A paradigm of DSS that has recently found its way to agriculture is that of digital twins [104]. Digital twins are expected to merge the physical and virtual worlds by providing a holistic view of physical systems, through data integration, adaptation to local conditions, and continuous monitoring. They have started gaining traction with data architectures and applications for greenhouses [163, 164], conceptual frameworks for designing and developing them [16], and case studies in aquaponics [14].

A factor differentiating digital twins from existing systems is their ability to adapt to local conditions [165]. Following the digital twin paradigm, in contrast to generic models which apply global rules across all systems, we can create a blueprint that contains a high-level view of how a system works, and then instantiate it to systems with diverse conditions and let it adjust to them. In agriculture, adaptation to local conditions (or domain adaptation) is important because systems are affected by multiple local factors, and characterized by high uncertainty, due to nature's variability. Decisions have to account for the variability in weather conditions, types of soil, and agricultural management (i.e. fertilization, irrigation, crop protection actions). Examples of failing to adapt include wrong estimations of yield [166], failure to detect plant drought stress [167], and expensive equipment that does not work the way it is supposed to be [168].

A challenge to applying domain adaptation techniques in agricultural digital twins lies in data-related issues. These issues occur because the process-based and ML comprising the digital twins have difficulties operating with missing data, or data do not conform with their input requirements. ML models usually require large amounts of data to be trained, along with labels that are not readily available in agriculture. Also, it is beneficial for them to have data that cover large variability of the original domain but usually the majority of the agricultural field observations are concentrated in a few locations with similar weather and the same agricultural practices. On the other hand, process-based models require their inputs to be complete. This can be a problem when those inputs are from future states of variables (e.g. weather, biophysical factors) and requires bringing in additional tools to estimate them.

A workaround to data-related challenges is to use surrogate models, often also called metamodels. Metamodels mimic the behavior of other (typically more complex) models [120]. ML metamodels combine the advantages of ML models (learning patterns from data, operating with noisy data) and process-based models (operating based on first principles). A way to develop ML metamodels is to apply ML algorithms to the output of process-based model simulations. In this way, the ML algorithms can use a large corpus of synthetic data, and more importantly extract the embedded domain knowledge contained in them. This technique has been proven to work well for instilling domain knowledge of water lake temperature to models [121], and working with data of different resolutions and absence of future weather values in NRR prediction [150]. However, the effectiveness of metamodels has not been investigated in conjunction with domain adaptation techniques in the context of agricultural digital twins.

Domain adaptation can be achieved with techniques like data assimilation and transfer learning. Data assimilation refers to the practice of calibrating a numerical model based on observations. This technique has been applied for grassland management digital twins [169], and digital twins for adaption to climate change [109]. Transfer learning refers to the utilization of knowledge obtained by training for a task, to solve a different but similar task. To the best of our knowledge, domain adaptation through transfer learning has not been thoroughly discussed in the context of digital twins for agriculture. An application we found was for plant disease identification, where the authors used a pretrained version of ImageNet and then continued training on a dataset containing images of diseased plants [170]. However, in other sectors we find that transfer learning has been considered in several cases for digital twins [171–173]. Consequently, the applicability of transfer learning as a domain adaptation practice has not been extensively examined for agricultural digital twins.

In this work, we explore the potential of transfer learning to be used for domain adaptation in digital twins. To this end, we use a case study of digital twins predicting pasture NRR[1] at farm level. We use a synthetic dataset of grass pasture simulations and develop ML metamodels with transfer learning to investigate their adaption to new conditions. Our main question is:

- Q: How well can we transfer field-level knowledge from one location to another using transfer learning?

To answer this question, we examine it from different angles and form the following sub-questions:

- q1: How domain adaptation with transfer learning is affected by including more variability in agricultural management practices?
- q2: How domain adaptation with transfer learning is affected by including more variability in weather data?
- q3: How well does domain adaptation with transfer learning perform when applied to locations with different climate from the original one?

[1] additional kg/ha of dry matter harvested per kg of nitrogen fertilizer applied

6.2 Methodology

6.2.1 Overview

To assess how well we can transfer field-level knowledge from one farm to another we performed a case study of grass pasture NRR prediction in different locations across New Zealand. We have a dataset of pasture growth simulations based on historical weather data from sites with different climates (Fig. 6.1), soil types, and fertilization treatments. Based on these data we pretrained ML metamodels in an **origin** location and fine-tune them in a **target** location to predict NRR and see how tuning affects model performance in both pretraining and fine-tuning test sets.

To obtain more dependable results, we pretrained in an origin climate and fine-tuned in two target climates that differ from each other. Also, we experimented with the amount of weather data included in the models as well as the number of soil types and fertilization levels. We created different setups and examined their results across several years, and for multiple runs using different seeds.

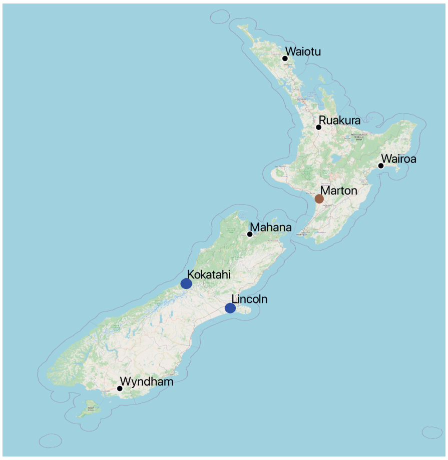


Figure 6.1: The sites contained in our dataset. With the brown color is the site in the origin climate (*Marton*, climate 1), and with the blue the sites in the target climates, (Kokatahi and Lincoln, climate 2 and 3 respectively)

6.2.2 Data generation

The simulations comprising our dataset were generated with APSIM [24] using the AgPasture module [138]. This module has been proven to be an accurate estimator of pasture growth in New Zealand [139, 140]. The simulation parameters covered conditions that are known to affect pasture growth. The full factorial [119] of those parameters was created and given as input to APSIM. The range of the parameters can be seen in Table 6.1.

Table 6.1: The full factorial of the presented parameters was used to generate simulations with APSIM.

Parameter	**Range**
Weather	daily weather from 8 sites
Soil water capacity	42, 67, 110 and 177 mm of plant-available water
Soil fertility	2, 4, and 6% of carbon concentration
Irrigation	irrigated, non-irrigated
Fertilization year	1979-2018
Fertilization month	January-December
Fertilization day	5^{th}, 15^{th} and 25^{th} of the month
Fertilizer amount	0, 20, 40, 60, 80 and 100 kg N / ha

6.2.3 Case study

In our experiments, we considered only the simulations where no irrigation was applied because this scenario is closer to the actual pasture growing conditions in New Zealand. Additionally, we only considered the autumn (March, April, May) and spring (September, October, November) months because these are the months in which agricultural practitioners are most interested in deciding how much fertilizer to apply.

To derive the NRR from the growth simulations, we calculated the additional amount of pasture dry matter harvested in the two months after fertilizer application per kg of nitrogen fertilizer applied.

Regarding the prediction scenario, we assumed to have weather and biophysical data only four weeks prior to the prediction date since pasture is supposed to not have memory beyond that point. Also, from the prediction date until the harvest date (two months later) we assumed that no data were available.

6.2.4 Experimental setup

Throughout the setup we create two types of models. The first type is trained on the data of the original location, and we call it 'origin model'. The second type is fine-tuned with the data of the target location, by using the origin model as a basis, and we call it 'target model'. We train different models using various setups which help us answer the sub-questions q1-q3. To answer q1, we considered two setups where variability comes from the number of agricultural management conditions included in the pretraining datasets:

- 1 type of soil and 2 types of fertilization treatments
- 3 types of soil and 5 types of fertilization treatments

To answer q2, we considered two setups where the digital twin blueprint contains training data from:

- 10 years of historical weather
- 20 years of historical weather

Consequently, for q1 and q2 there are four setups namely:

- Low weather and agromanagement variabilities (s1)
- High weather variability, low agromanagement variability (s2)
- Low weather variability, high agromanagement variability (s3)
- High weather and agromanagement variabilities (s4)

containing varying amounts of training data based on soil type, fertilization treatment, and the number of historical weather years. The details for each setup can be seen in Table C.10. To answer q3, we considered three locations from our dataset with diverse climates. The origin location (location 1) where pretraining takes place, and two target locations. The target locations were selected based on the climate similarity index CCAFS [174], to be dissimilar with the origin location to varying degrees (see Fig. C.8). Also, weather factors that are known to affect pasture growth were considered, namely precipitation and temperature. Location 2 is characterized by more frequent rainfall and lower temperatures than the origin location 1, and location 3 by less frequent rainfall and a wider range of temperatures than location 1. The respective plots can be seen in Fig. C.9.

Finally, we took measures to make the results more dependable. To alleviate the effect of imbalanced sets due to anomalous weather, we examined how transfer learning works across several years by sliding the corresponding training/validation/test sets across five years. Also, to see how robust were the models we trained each one of them five times with different seeds in each setup and sliding year.

6.2.5 Data processing

The APSIM synthetic dataset was further processed to form a regression problem whose inputs were weather and biophysical variables as well as management practices. Initially, the NRR was calculated at two months after fertilization. Then, data was filtered to contain only simulations for the non-irrigated case. After that, only daily weather data in a window of four weeks prior to the prediction date were retained. Weather data between the prediction and target dates were also discarded because such data would be unavailable under operational conditions. Next, simulations with NRR less than 2 were removed as they were attributed to rare extreme weather phenomena which were not relevant to model for this study. From the remaining data only the daily weather variables regarding precipitation, solar radiation, minimum and maximum temperature were preserved. From the biophysical outputs of APSIM only above ground pasture mass, herbage nitrogen concentration in dry matter, net increase in herbage above-ground dry matter, potential growth if there was no water and no nitrogen limitation soil, and temperature at 50cm were preserved because they were considered likely drivers of yield (and known prior to the prediction date) based on expert knowledge. Additionally, from the simulation parameters only soil fertility, soil water capacity, fertilizer amount and fertilization month were retained to be put to the models as inputs. The data were then split into training/validation/test sets according to the experimental setup. Z-score normalization followed, with each test set being standardized with the scaler of the corresponding training set. The fertilization month column was transformed into a sine/cosine representation.

6.2.6 Neural network architecture

The selected architecture was a dual-head autoencoder which proved to be accurate for NRR prediction tasks in another study [150]. The architecture consisted of an autoencoder with LSTM layers whose purpose is to learn to condense the input weather and biophysical time series, and a regression head with linear layers whose task is to predict the NRR (Fig. 6.2). The combined loss is derived by summing the reconstruction loss and the NRR prediction loss.

The hyperparameters of the origin model were selected based on a preliminary study and were the same across all setups and years. For the target model, hyperparameter tuning was performed with gridsearch for each setup, year, and seed. The hyperparameters of the origin models and the search space for the target models can be seen in Table C.6 and C.7.

For the part of tuning the network in different climates, no layer was

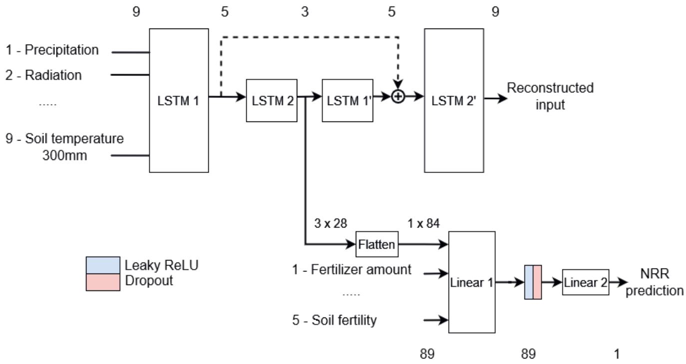


Figure 6.2: The autoencoder architecture used to pretrain and fine-tune the models. The numbers on the top and bottom of the architecture indicate the number of features in the input/output of each component. The inputs to the encoder were nine time-series variables. The compressed representation of those time-series (output of *LSTM 2*) along with five scalars were concatenated and directed to a multi-layer perceptron

frozen.

6.2.7 Evaluation

Pretrained and target models were evaluated on the test set of the origin location as well as the target location. This was done to examine how well they absorbed new information and fast they were forgetting old information. The difference in performance between the origin and target models was measured with R^2. R^2 was reported as an average across the five seeds, for each setup, and each year. Also, the standard deviations of R^2 between the seeds were examined to see how stable the performance is across the runs.

6.3 Results

For both target locations we observe that fine-tuning increased the average R^2 across the runs on the **target location test set** for most setups. For s1, s2 this behavior was consistent in both location 2 (Fig. 6.3) and location 3 (Fig. 6.4). For s3, tuning offered marginal improvements in both locations. In the case of s4, the results varied between the locations, as in location 2 there was no improvement and even degradation in years 2004-2005 (Fig. 6.3d), and in location 3 minor improvements (Fig. 6.4d). The standard deviations of the target models on the target location test sets were within the [0.01, 0.08] range (Fig C.11, C.12).

origin model	1992-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	-0.62	-0.56	-0.85	-0.86	-1.05				
validation			-0.14	-0.34	-1.23	-1.24	-0.84		
origin test (Clim 1)					-0.73	-0.79	-0.79	-0.72	-1.54
target test (Clim 2)					-0.25	-0.13	-0.16	-0.26	-0.34

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.71	0.74	0.7	0.77	0.77				
validation			0.72	0.67	0.7	0.79	0.67		
origin test (Clim 1)					-0.21	-0.4	-0.06	0.11	0.06
target test (Clim 2)					0.65	0.74	0.68	0.72	0.78

(a) Setup s1: Low weather and agromanagement variabilities

origin model	1982-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.01	0.06	0.05	0.06	0.06				
validation			0.09	0.13	0.06	0.01	0.06		
origin test (Clim 1)					0.04	0.04	0.06	0.06	0
target test (Clim 2)					0.18	0.17	0.11	0.15	0.18

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.73	0.77	0.73	0.79	0.79				
validation			0.74	0.7	0.73	0.8	0.68		
origin test (Clim 1)					-0.2	-0.4	-0.06	0.11	0.07
target test (Clim 2)					0.68	0.76	0.71	0.74	0.79

(b) Setup s2: High weather variability, low agromanagement variability

origin model	1992-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.74	0.74	0.72	0.73	0.75				
validation			0.73	0.71	0.77	0.75	0.62		
origin test (Clim 1)					0.78	0.76	0.62	0.46	0.56
target test (Clim 2)					0.79	0.82	0.85	0.82	0.88

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.89	0.88	0.87	0.91	0.86				
validation			0.9	0.85	0.8	0.88	0.86		
origin test (Clim 1)					0.73	0.76	0.59	0.45	0.57
target test (Clim 2)					0.77	0.83	0.87	0.85	0.88

(c) Setup s3: Low weather variability, high agromanagement variability

origin model	1982-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.73	0.74	0.74	0.74	0.76				
validation			0.72	0.71	0.81	0.79	0.68		
origin test (Clim 1)					0.8	0.79	0.68	0.49	0.58
target test (Clim 2)					0.82	0.84	0.83	0.81	0.88

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.9	0.88	0.85	0.91	0.91				
validation			0.89	0.85	0.82	0.88	0.85		
origin test (Clim 1)					0.77	0.78	0.68	0.45	0.56
target test (Clim 2)					0.77	0.84	0.83	0.84	0.87

(d) Setup s4: High weather and agromanagement variabilities

Figure 6.3: R^2 for the setups of the origin models (**climate 1**), and target models in **climate 2**. The results are presented as averages across the 5 seeds for each setup and year. On Fig. 6.3a the brown and blue colors indicate which training, validation and test set correspond to each experiment due to the sliding years. Same colors represent sets of the same experiment. For example, with the brown color the training set of the origin model included years 1992-2001, validation set 2002-2003, and both test sets years 2004-2005. On the experiment with the blue color the training set included years 1993-2002, validation 2003-2004, and both test sets 2005-2006. The leftmost cell of the results is colored (green, yellow, pink) as the corresponding set is colored, and has a width equal to the amount of training years included in it. For the other 4 sliding years, only the last year of each set is shown with grey color

Tuning also increased the average R^2 on the **origin location test set** for s1 (Figs. 6.3a, 6.3b) and s2 (Figs. 6.4a, 6.4b). However, on s3 and s4 the performance remained stable or deteriorated depending on the year. The standard deviations of the target models on the origin location test sets were within the [0.01, 0.3] range. The standard deviation of the target models on the pretraining test sets for s1 and s2 were within the range [0.03, 0.32], and for s3 and s4 [0.02, 0.19].

Another observation is that the R^2 of the origin model on s1 was negative in both locations for all years (Figs. 6.3a, 6.4a). The corresponding standard deviations were also high as shown in Figs. C.11a, C.12a.

A remark is also the high volatility of R^2 depending on the year, of both origin and target models in the origin and target locations test sets. Performance becomes more stable as more weather and agromanagement variability are added (e.g. s1 to s2, s1 to s3) but there were years like 2004-2005 in location 2, 2008-2009 on location 3 where R^2 dropped substantially. The standard deviations (Figs. C.11, C.12) also became lower across the years as more agromanagement variability was added.

origin model	1992-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	-0.62	-0.56	-0.85	-0.86	-1.05				
validation			-0.14	-0.34	-1.23	-1.24	-0.84		
origin test (Clim 1)					-0.73	-0.79	-0.79	-0.72	-1.54
target test (Clim 3)					0	-0.04	-0.28	-0.07	-0.03
target model	**1999-2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
training	0.36	0.46	0.41	0.45	0.43				
validation			0.28	0.32	0.36	0.48	0.42		
origin test (Clim 1)					0.26	0.06	0.07	0.01	-0.41
target test (Clim 3)					0.27	0.34	0.4	0.19	0.21

(a) Setup s1: Low weather and agromanagement variabilities

origin model	1982-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.01	0.06	0.05	0.06	0.06				
validation			0.09	0.13	0.06	0.01	0.06		
origin test (Clim 1)					0.04	0.04	0.06	0.06	0
target test (Clim 3)					0.06	0.07	0.09	0	-0.11
target model	**1999-2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
training	0.37	0.49	0.44	0.48	0.45				
validation			0.3	0.34	0.39	0.51	0.45		
origin test (Clim 1)					0.26	0.05	0.07	0.02	-0.39
target test (Clim 3)					0.29	0.37	0.43	0.2	0.22

(b) Setup s2: High weather variability, low agromanagement variability

origin model	1992-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.74	0.74	0.72	0.73	0.75				
validation			0.73	0.71	0.77	0.75	0.62		
origin test (Clim 1)					0.78	0.76	0.62	0.46	0.56
target test (Clim 3)					0.54	0.56	0.58	0.25	0.22
target model	**1999-2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
training	0.48	0.63	0.73	0.66	0.63				
validation			0.61	0.57	0.6	0.65	0.61		
origin test (Clim 1)					0.78	0.65	0.44	0.34	0.47
target test (Clim 3)					0.54	0.6	0.59	0.24	0.24

(c) Setup s3: Low weather variability, high agromanagement variability

origin model	1982-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.73	0.74	0.74	0.74	0.76				
validation			0.72	0.71	0.81	0.79	0.68		
origin test (Clim 1)					0.8	0.79	0.68	0.49	0.58
target test (Clim 3)					0.51	0.5	0.54	0.25	0.21
target model	**1999-2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
training	0.53	0.64	0.71	0.67	0.65				
validation			0.61	0.57	0.58	0.64	0.59		
origin test (Clim 1)					0.8	0.7	0.48	0.3	0.4
target test (Clim 3)					0.5	0.57	0.57	0.27	0.27

(d) Setup s4: High weather and agromanagement variabilities

Figure 6.4: Average R^2 for the various setups of the origin models (**climate 1**), and target models in **climate 3**. The figures should be read following the pattern of Fig. 6.3a

One more finding is that adding more weather variability while keeping the agromanagement practices unchanged had a negligible (positive) effect on the performance of the target models. This pattern can be observed for both locations when transitioning from s1 to s2 (e.g Fig. 6.3a to 6.3b), and from s3 to s4 throughout the years (e.g Fig. 6.3c to 6.3d). Also, in those scenarios, the standard deviations of the target models on the pretraining test sets did not decrease when extra weather variability was added. On the other hand, increasing the management practices while keeping the same weather variability seemed to increase the R^2 of both models in both test sets. This can be seen when transitioning from s1 to s3 (e.g Fig. 6.4a to 6.4c), and from s2 to s4 (e.g Fig. 6.4b to 6.4d).

6

6.4 Discussion

Starting with some general remarks about model performance, for transfer learning tasks there is usually a model that works well which then undergoes further training. Here, the first impression is that the performance of the origin model on s1 and s2 is inadequate. This is potentially due to the selected architecture and the way training was performed. In those setups the samples were too few (see Fig. C.10), and the architecture had a lot of weights. As a result, the network may not have been able to extract meaningful features in those cases. Also, the performance increase on the pretraining test set after tuning may indicate that extra information is included in the tuning training data, but it could also mean that the worse performance was due to training for too few epochs.

Another remark is that the target models achieve considerably higher R^2 on location 2 than on location 3. This behavior could be attributed to the weather conditions of each location. Location 2 is characterized by more precipitation, reducing in this way the uncertainty of having less water during the period of sixty days that for which we assume that no weather data are available from the prediction to the target date. As a result, the NRR values concentrate on a narrower range, and models have an easier task explaining variance.

Regarding fine-tuning, it seems to make the models able to generalize better in the target locations than the models which have not seen this extra information before. Especially for setups s1 and s2, the results indicate that transfer learning adds value when the available soil, and fertilization management data are limited in quantity. This statement is supported by the consistency of the results which come from several years, and two diverse locations, suggesting that this behavior is not year or location dependent. For the same setups, the decrease in the standard deviation after fine-tuning strengthens the claim that the improved performance is not a coincidence.

On the contrary, when there is sufficient variability in the soil and fertilization management practices, the role of fine-tuning becomes ambiguous. It may seem that transfer learning increases the generalization capacity of the models for most of s3 and s4 cases, even though improvements are marginal. However, these improvements are so small that they get counteracted by the standard deviation of the successive runs. Also, depending on the year (e.g. 2004-2005 for location 2) fine-tuning may be harmful as it decreases R^2 further than the standard deviation of the five runs. Performing more runs with different seeds or testing in different years could potentially yield different results than those observed. Consequently, we cannot assess the merits of fine-tuning in those cases.

Moving on to the effect of adding more weather variability in the origin models, we saw that the differences in performance were small. This pattern was observed for both target locations, and the reasons behind its appearance may vary. We could presume that adding weather variability does not help the models enough to extract information relevant to NRR prediction. This could be the case if in those extra years the weather was very different from the weather of the target locations. Another case would be that since we have a gap of sixty days between the prediction and target dates, and assuming the absence of extreme phenomena, the weather is more loosely connected to the NRR prediction than other factors like soil type and fertilization practices.

A more apparent reason for the effect of adding more weather variability is that we potentially observe the effect of increasingly higher sample sizes. In Fig. C.10 we see the number of samples in each setup. Adding more weather data (s1 to s2, or s3 to s4) doubles the samples included in the pretraining data. However, with the current experimental setup, adding soil types and fertilization treatments (s1 to s3, or s2 to s4) increases the number of samples by a much higher degree. Therefore adding more weather variability to the pretraining sets has little (but positive) effect on the model test sets, which seems small compared to adding more soil types and fertilization treatments because with the latter we have many more samples. The increase of R^2 of the target models on the pretraining test sets seems to support this argument. Adding more weather data to a model from a target location could help explain the variability in that location. However, here we see that it also helps to explain variance in the original location, prompting that this increase is not due to the so different conditions supposedly existing on the new data but just an increased sample size. For this reason, this phenomenon is more evidently expressed at s1 and s2 where sample sizes are lower.

6.5 Limitations

There are cases where it is unclear if the improvement in R^2 on the tuning location test set comes from adding samples with information about local conditions or from just the continuation of training with extra samples. To be able to better deduct those cases the set sizes should be equal between the different setups s1-s4. The challenge there would be to create representative sets for all setups, sliding years, and target locations.

Another limitation is that we used the same neural network architecture for all the setups. This architecture has many weights that need to be calibrated and in setups with fewer samples it may not be appropriate to use. A simpler architecture might have given different results.

With the provided experimental setup we created two types of models, the ’pretrained’ (origin) and ’fine-tuned’ (target). The origin models contained an increasing number of samples from the origin location based on the setup, and the fine-tuned a fixed number of samples from the target location. However, we did not include in the study the results of models trained only on the data from the target location. Preliminary tests with the chosen architecture showed that such models had negative R^2 in all setups and high standard deviations, so they were omitted. A more thorough investigation would include such models with simpler architectures, or different algorithms with features aggregated on a weekly/biweekly basis to decrease the number of parameters that have to be calibrated.

In regards to the data splits, in a more practical application the test set years would be closer to the training set years. With the current setup the training and test sets are 2 or 3 years apart. With such gaps the weather may change substantially leading to non-representative sets. An alternative setup would be to have these years closer and maybe remove the validation set and perform a k-fold cross validation for hyperparameter tuning instead.

6.6 Conclusion

In this work, we examined the application of transfer learning as a way to make field-level pasture digital twins adapt to local conditions. We employed a case study of pasture NRR prediction, and investigated factors that affect the efficiency of the adaptation procedure. Different setups had varying outcomes but generally transfer learning seems to provide a promising way for digital twins to learn the idiosyncrasies of different locations.

Revisiting q1, based on our experiments variability in soil type and fertilization treatment seemed to help the models explain a large fraction of variance in the target locations. Therefore, for field-deployed applications practitioners could try to gather as much data as possible with this kind of variability or generate them. On the other hand, for q2 we found that the addition of extra weather variability had a small impact on model performance. Thus, adding more variability in soil and agricultural management practices should be of higher priority. In both cases, more work is needed to verify the degree to which large sample sizes start to affect the results. Regarding q3, transfer learning appears to work for diverse climates with performance differences depending on the prevailing local conditions. Again, more work is needed to test its efficiency in climates that are even more diverse and characterized by more extreme phenomena.

Finally, to answer our main question, the above are evidence that we can transfer field-level knowledge to a degree that models can explain an adequate portion of variance in the target locations. In this respect, transfer learning has the potential for making digital twins adapt to different conditions by working in different climates, and with different types of variability. Practitioners could create blueprints of digital twins with origin models and then adapt to different locations by instantiating them there preferably with samples that contain varied soil types and fertilization treatments.

Chapter 7

Synthesis

This chapter is separated into three parts. The first part (Section 7.1) contains the findings of this thesis based on the five previous chapters. The second (Section 7.2), contains reflections on the findings, and expands on the experiences acquired regarding the interplay of data, experimental design, and collaboration. The third (Sections 7.3,7.4), discusses the impact of this work and possible future directions.

7.1 Research findings

The purpose of this thesis was to examine ways to operationalize digital twins in two respects. First, by **enabling decision support** in cases where the available data are not sufficient in volume. Second, by making them **transferable** to different conditions. Table 7.1 summarizes the findings of our investigation.

Table 7.1: Contribution of papers to objectives. Cells with '*auxiliary*' indicate work which helped us to better identify the objectives.

Chapter	**Objective**		**Contribution to objectives**
	Enabling decision support	**Transferability**	
2	(auxiliary)	(auxiliary)	Literature review to identify the adoption rate, benefits, and prospects of digital twins in agriculture
3	x		A metamodeling framework to enable decision support by overcoming data issues
4	x		Examination of the framework generality in terms of ML algorithms
5		x	Investigation of transferability with rich data or no data from the target conditions
6		x	Exploring transferability with a few data from the target conditions, through domain adaptation with transfer learning

Chapter 2

In this chapter, we examined the adoption of digital twins in agriculture by searching for reported applications in the literature and comparing those with other disciplines. We found that most agricultural digital twin applications had low technology readiness levels and did not provide advanced services reported in other disciplines. Reported reasons for this delay were the difficulty of synchronizing living systems with their virtual counterparts, issues with data accessibility, maintenance, and standardization, difficulties in trusting complex technological systems, and lack of investment. Also, we proposed a roadmap describing how digital twins could be further developed in agriculture based on recent developments in other disciplines. The literature review aided us in understanding a practical dimension of digital twin adoption, pointing us to the direction of operationalization. The review also made evident that digital twins are not yet widespread for decision support due to data related issues, and because they cannot perform operations that make digital twins special like adaptation to local conditions.

Chapter 3

In this work, we proposed a methodology based on ML metamodels to enable decision support for digital twins in the following cases:

- There are not enough observations available to train ML models
- We have data from the domain where we want to deploy a digital twin, but they are in a different temporal resolution from what process-based models expect (e.g. we might have weekly soil moisture data but our model requires data on a daily resolution)
- Process-based or ML models require data that are not available yet (e.g. future weather)

We tested the methodology using a case study of pasture NRR prediction in New Zealand. Several metamodels containing different amounts of data were created and tested in scenarios where data samples from new locations were either available or not available. The metamodels were evaluated using a domain-specific error threshold, and in most cases, they were able to provide accurate predictions. Embedding such models to digital twins would allow them to make predictions in data limited settings and be operational in cases where other existing tools fall short.

Chapter 4

The methodology of chapter 3 was further evaluated with various neural network architectures to examine whether it is algorithm independent. The architectures were a multilayer perceptron, and two types of autoencoders. The first autoencoder reconstructed weather variables and after training the decoder was replaced with a multilayer perceptron performing our prediction task. The second autoencoder optimized simultaneously for weather reconstruction and NRR prediction. The results showed that the metamodeling methodology is independent to the types of architectures we tested. This finding is important for enabling decision support in digital twins because their embedded models' algorithms are selected based on the available data (e.g. data modality, number of samples), and the type of application (e.g. transfer learning, explainability).

Chapter 5

In this study, we compared the performance of a metamodel developed with training data in New Zealand from only one location ('location-specific' metamodel), with a model trained with data from all the available locations in our dataset ('generic' metamodel). We tested the models iteratively in each location of our dataset and evaluated their accuracy[1] with common error metrics. Also, we compared if the differences in the distributions of their predictions were statistically significant with a statistical test. The results verified that if we do not have data from the location where we want to bring our digital twin, the larger amount of synthetic data with more variability we generate the better. Based on this finding, in case we do not have data from a target location, we can attempt to make our digital twins transferable there by developing ML metamodels on historical data from surrounding or other locations. In this way, we may enable decision support for that location in contrast to other models or tools which are not calibrated for that domain.

Chapter 6

The purpose of this chapter was to examine the potential of transfer learning for domain adaptation in pastoral digital twins. We chose a location in our dataset and created training datasets with different combinations of weather years, types of soil, and agro-management practices. We then used these datasets to train metamodels ('origin' models). Next, we took the metamodels and transferred them to locations with different climate conditions based on a climate similarity index. Training continued in those new locations, and the resulting metamodels ('target' models) were compared with the origin models in the target locations. Domain adaptation with transfer learning appeared to have potential for operating digital twins in diverse domains but there is more work to be done in order to pinpoint the factors that affect it. This finding provided evidence that ML metamodels are promising for domain adaptation in digital twins. As a result, the benefits of the metamodeling approach reported in the previous chapters are also applicable here. More importantly, this study showcased that decision support can be enabled in diverse domains by transferring digital twins through domain adaptation.

[1] Degree of closeness of measurements of a quantity to that quantity's actual value. Not classification accuracy.

7.2 Reflections

7.2.1 Metamodeling as a versatile way to enable decision support

The proposed methodology of chapter 3 is based on the generation of simulations from which we can extract the embedded domain knowledge and remodel it with ML algorithms to fit our task. A few years of historical weather data accompanied by a set of agro-management practices (fertilization or irrigation events) were enough to create a space of inputs based on which a large synthetic dataset was created to satisfy the needs of ML models for sample representativeness, data quantity and variability. This comes as a stark contrast to observation agricultural datasets, which lack representativeness due to practitioners employing the same agro-management practices based on rules of thumb in all their fields. With such non-contrastive data, it is difficult to create digital twins that learn and adapt to individual conditions. Another, implicit, benefit of operating with simulated data generated by a well-tested and broadly accepted process-based model is that we can focus on extracting the domain knowledge embedded in the data rather than be concerned about the effect of noise in them.

Employing synthetic data to enable decision support is an active field of research [175], as it can benefit disciplines where data are hard to acquire. Applications based on this principle are found in other disciplines, such as recognizing human actions based on generated data from 3D models [176], creating autonomous driving systems trained on simulated traffic conditions [177], and detecting fraud in networks with models trained on synthetic logs [178]. These examples exhibit the versatility of this approach and motivate its broader use. Agriculture and environmental sciences are characterized by poor data management practices, and they could find relief by adopting this paradigm for developing models using simulators and deep generative models (e.g. VAEs, GANs).

Regarding the generality of this approach to making predictions in agriculture, another line of research was conducted to predict potato tuber weight with generated data from the TIPSTAR crop growth model. This is still a work in progress but the impression we have from a preliminary examination with a small sized generated dataset is that we could enable decision support with accurate predictions, which strengthens our confidence in this method. In the same way, multiple case studies across different crops may be combined to further assess the usefulness of this method across different domains, process-based models, and ML algorithms.

7.2.2 To validate on observations, or not to validate on observations

Regarding the evaluation of the metamodels, a mix of common error metrics (e.g. R2, RMSE), domain-specific error thresholds, and statistical tests was used. For the studies of chapters 3 and 4, we evaluated the metamodels based on their residuals from what APSIM predicted, and an error threshold communicated with experts from New Zealand. For chapters 5 and 6 we used error metrics to assess accuracy and a statistical test to compare them with each other. Consequently, the metamodels have been trained on synthetic data and have been evaluated against the corresponding process-based model. No observations were involved in the testing phase and as a result we cannot deduce about their performance with data collected from sensors on fields. There is a compelling argument though, that since APSIM is actually used to provide decision support to agricultural practitioners and has been validated with observations, we could presume that approximating the accuracy of APSIM would create adequately accurate ML metamodels for field-deployed applications. Also, considering the accuracy of APSIM as the upper bound for the ML metamodel performance would seem reasonable, but there are cases in literature [179][176] where metamodels showed improved generalization capacity over their training data generators. Therefore, lack of observation data for testing purposes during experimental design, should not inhibit the exploration of synthetic datasets and the benefits they may provide.

7.2.3 Inputs to data generators, variability, and metamodel performance

An integral part of the proposed metamodeling method is to create a space of inputs for the process-based model that is representative of the problem at hand. The input space should cover as much variability as possible from the conditions expected to be found during the inference time of the metamodels. If the input parameters are not varied, the simulated quantities may be constrained to ranges that do not cover enough variability for our modeling task.

This effect became noticeable in chapter 5 where we compared a location-specific model containing training data from only one location, with a generic model which included data from multiple locations. The performance of the location-specific model was deteriorating when tested in climates dissimilar to the climate of its training data location, because it was trained on simulations that did not include enough climate variability. In contrast, the generic model made accurate predictions for all the locations where it was tested because it had seen samples with similar weather patterns. Similarly, in chapter 6 we found that target models containing limited variability on their pretraining data (either climate or agro-management practices) did not perform as well as those which had seen more variable conditions.

To decide the number of parameter sets and the ranges of the parameters, researchers working with synthetic data should make a judgment of how complex their generator is and how its subsystems interact. Usually, agricultural and environmental process-based models combine multiple modules which include cross-scale effects, thus exhibiting complex behavior. Small differences in their initial conditions may significantly affect the outcome of simulations. Therefore, it is important to create input spaces that can capture this variability in the output.

7.2.4 Decision support through ML algorithm independence

As described in chapter 4, the applied metamodeling methodology is independent of the ML algorithm employed to create the models. This flexibility makes it applicable to various problems and domains while also preserving the benefits of overcoming the data issues mentioned in chapter 3. For example, in environmental sciences tabular datasets are common. As a result, many researchers aim to create decision support systems by using algorithms that perform well with such data, like random forest. Besides, other datasets have to do with evolving quantities, like weather attributes, so they are represented as time-series. Algorithms that learn the latent space of input data, such as neural networks, may be more relevant to time-series prediction tasks than random forest. Similarly, researchers can alternate between time-series and tabular data representations, (or even representations suitable for CNNs) depending on the number of samples in the datasets. These representations will result in different performances depending on the algorithm used. All those cases can be accommodated with the proposed methodology and a model selection procedure can assist in selecting the algorithm. Also, applications for providing decision support to different domains or creating more explainable metamodels can be facilitated since some algorithms can accommodate new data points after training, and others are inherently explainable. Consequently, the proposed method to overcome limited numbers of observations, different resolutions, and working without forecasted data, while also enabling these applications is important to create operational digital twin applications.

7.2.5 Transferability with no data from the target location

The first step towards transferability is to enable decision support in new domains without having to perform domain adaptation. Ideally, a digital twin should have a broad sense of the prevailing conditions in a target domain before trying to adapt to it. Towards this direction, in chapter 5 we investigated how metamodels performed in different domains (climates in this case) when they have no data versus rich data from a target domain. We recommended having a blueprint of a digital twin developed on multiple sites around the location of interest. An observation in our experiments was that the further away (in distance) the target location was from the location where the location-specific metamodel was trained, the worse its accuracy was. This finding is related to the similarity of the involved climates and not the distance of the corresponding locations. Therefore, our recommendation still applies, with the reservation that for nearby locations with disparate weather characteristics it may not be enough to make the digital twins have an initial sense of the prevailing conditions in the target area.

Making predictions without data from the target domains is something that the AI (Artificial Intelligence) community has experience doing. Techniques like zero-shot/few-shot learning [180] are applied to differentiate new diseases from previously known ones, to make autonomous driving cars recognize unknown objects (concept cars, signs with graffiti), and to identify new species of animals [181]. There is growing interest in such techniques, since ML applications enter increasingly more fields of research where data are scarce due to the expertise required to annotate them, or the rarity of events. Digital twins could benefit from those methods, especially those which synchronize to ‘open-world‘ systems and are more likely to encounter unforeseen events.

The work described in chapter 5 could also be interpreted as a case of zero-shot learning. That’s because we investigated how to transfer models to new locations from which we do not have any data, in a way similar to how zero-shot methods work. Zero-shot methods achieve making predictions by connecting seen and unseen domains through properties of the dataset samples [182]. For example, a classifier which outputs the species of an animal in an image, can be modified to classify new species by outputting properties like color or number of legs instead of the species. Likewise, the properties we used were soil characteristics, agro-management options, and weather conditions. More elaborate properties could have potentially yielded better results.

7.2.6 One generic model to rule them all

In chapter 5 we saw that the generic metamodel outperformed the location-specific one in all the locations except for the location where the location-specific model was trained. A question arising for practical applications is how many models we need to include in our digital twin in case we want to provide decision support for multiple locations. Should we use a location-specific model for each location or combine everything into a generic model? The answer lies in the application needs. If the application needs models with better generalizability we should opt for generic models. Alternatively, if absolute performance is essential, we could create individual models for each domain. Development time and maintenance of these models should also be considered. Depending on the algorithm that we use to create models we may need to find important variables, balance datasets, perform augmentations, and tune models individually for each of these domains. Similarly, during operational use the incoming data and model predictions should be checked for drifts. This can become tedious when there are a lot of models involved, especially since they will be part of a larger entity (digital twin) where they are intertwined with other monitoring and optimization operations.

7.2.7 Transferability with sparse data from the target location

For chapter 6, we initially experimented with transferring knowledge between locations about pasture yield by considering changes in multiple domains (climate, non-fertile soil to fertile, low fertilization levels to high, etc.) simultaneously. This was a setup that closely resembled how a digital twin would be applied in a practical situation. However, it proved to be a more complex problem than expected as the target models could not achieve a meaningful performance improvement. Thus, we scaled back to changing only the climate and creating setups with incrementally more variability in the pretraining datasets. The number of domains and the variability in them is an important factor to consider when making agricultural digital twins, since learning local conditions is important but trying to adapt to multiple domains simultaneously can be challenging. As a result, a digital twin which considers fewer domains or lower resolutions to limit variability may be more appropriate.

Domain adaptation with transfer learning may also work or not, depending on the combination of how well we model the problem into our algorithms and the quality of the data. For example, we might have a lot of variability in our dataset around nitrogen application. We can set up our algorithm to understand differences in yield coming from nitrogen application, not just providing the nitrogen amount as an input, and thus it might perform well in this case. However, if we try to transfer on the same location for another quantity (e.g. biomass, NDVI), it might not work due to the dataset not being varied enough to capture this relationship, or due to containing extreme cases that affect the quantities of interest disproportionately. Consequently, when making digital twins adapt to new domains, it would be beneficial to have several variables that can help us get the information we want, in case we are unable to estimate some of them.

Pretraining on synthetic datasets and performing domain adaptation with observations is a technique gaining ground in AI applications. Companies [183] use this technique to first simulate traffic, then to train autonomous driving systems based on them, and finally to adapt these systems using 'real' traffic footage. Similarly, when we started working in chapter 6, we attempted to pretrain models with synthetic data from the New Zealand dataset and fine-tune them using observations from pasture trials in the Netherlands. A large part of the effort was devoted to trying to preserve as many observation samples as possible. That was challenging due to differences in agro-management practices, and variables with no exact counterpart (soil fertility, soil water capacity) between the observation and synthetic datasets. We attempted to restructure the problem to better match the designs of the two datasets, to no avail. We also contacted people experienced in running relevant process-based models to assist us in running simulations, but due to time constraints these collaborations did not work. A lesson we learned is that when attempting to pretrain models on synthetic data and fine-tune them on observations, there should be a preliminary examination of the observation data, to ensure that the generated data fit the setup of the observation dataset.

7.2.8 Data as a factor that inhibits digital twin adoption

In chapter 2, we found that one of the reasons for the delayed adoption of digital twins in agriculture was the lack of data culture. Our experiences concur with this finding. To build on our ideas to operationalize digital twins we needed data either in the form of observations or simulations.

We realized that data are difficult to find. If in Wageningen it is challenging to create a registry for relevant datasets and their contents, then what can be expected from other places where agricultural research is not so advanced? As so, it appears to be a disparate relation of important agricultural research happening, in an environment which facilitates the creation of quality datasets, but data end up being 'hidden'.

Another finding was that data are difficult to acquire. Even if you learn about the existence of data relevant to your project it is challenging to get them. The trust of the data owners has to be earned and their interest in your work has to be developed. A collaboration has to be formed where both parties are going to benefit from. This process is not always straightforward since the involved parties may have divergent goals. Expectation management is important to make the data owners aware of the risks entailed for the outcomes of the experiments, so that they do not feel let down when things do not go as expected but they have already shared their data.

Also, we found that data are not well maintained. Datasets suffer from labelling and documentation issues. A considerable amount of time and effort has to be spent understanding what the measured variables represent and how they connect to each other. There are also data organization problems. Usually, data lie in databases in multiple tables. These tables are poorly documented and designed. Involving data owners to understand how data are organized or what columns mean, may not lead to significant progress, either.

Contemplating on these findings we realized that data do exist but not really. With considerable effort data can be found and acquired, but they are mostly unusable for the type of decision support that digital twins are meant to provide.

7.2.9 Thoughts on collaborative data-driven research

Effect of multidisciplinary teams

From the beginning of this project, it became evident that support from multidisciplinary parties would be beneficial to create experimental designs that better address practical problems. Involving these parties allowed us to have a panoramic view of the problems they faced and adjust our designs appropriately. Also, different parties had expertise in different domains and were more effective at explaining the results under different perspectives. For example, in chapter 3 the metamodels were overestimating the NRR in the range of 0 to 5. To find out why that happened we had to combine the knowledge of people who knew about growing pasture, how the simulations were generated, and why this effect appears from a data science perspective. Additionally, we were able to have a better understanding of how accurate our results were due to domain-derived error metrics communicated with these parties.

The devil is in the details

A realization regarding experimental design was from how many angles a problem could be tackled depending on the problem formulation. For example, in chapter 5 we had to decide whether we are going to make predictions using regression or classification. In a practical setting, regression may have had more sense as practitioners could work easier with a number. However, the same problem could be solved with classification by predicting ranges of NRR. Also, we had to decide whether we were going to treat our sensor data as tabular, time-series, or in another format. This decision was taken considering the number of samples available and the number of weights of each neural network architecture. Similarly, in chapter 6 we had to choose for the data splits between what is more correct from a ML perspective versus what is more prevalent in relevant agricultural studies. The criterion here was the time required to carry out the experiments. In another case, in chapter 3 we had to decide whether we were going to pursue good predictions for extreme weather conditions or not. That would be interesting from a research perspective, but in reality, agricultural practitioners already know based on experience in what ranges the yield will be when such events occur. All those choices made sense from different perspectives. Alternatives paths may had produced similar or better outcomes. Yet, at some point decisions had to be made in order to move on and produce an output.

Technical resource availability

Available technical infrastructure was an important factor affecting experimental design. Several times we had to migrate between the research group's servers, Google Colab, AWS, and Azure, to make sure that we have the available computational resources to run the experiments to the extent that we wanted to. This includes a budget to cover the expenses, and a balance between fast iteration versus more organized development practices. Switching between these platforms can sometimes take time since the codebase has to be adjusted and their internals to be learned. Also, moving large datasets between the platforms can be time-consuming.

Another technical challenge is to create datasets consisting of simulations with process-based models. Operating these models may not always be a straightforward process as many of them are incrementally built over time and they rely on old undocumented codebases and combinations of outdated programming languages. The situation is improving over time with efforts like PCSE [184] which is an environment that allows using multiple process-based models under a common interface. An alternative solution is to containerize [185] these models when a working setup has been achieved. The models can then be executed through the containers in any system, without additional setup. Getting a working setup though is a time-consuming process that does not contribute to the actual purpose of the experiments.

Therefore, moving towards a more data-driven way of working and incorporating different tools for data wrangling and analytics, one should not underestimate the time needed to choose these tools or combine them into workflows when creating experimental designs.

7.3 Impact

Research community

From the beginning of this work, we tried to reach out to the agricultural research community as well as the AI audience, through conferences, workshops, and networking, to show the types of problems that agriculture faces when attempting to apply ML solutions. We made the first steps to defining the state of digital twins in agriculture and we created a roadmap for their evolution. Next, we focused on how a digital twin would be useful in practical applications by discovering ways to operationalize it. Emphasis was given to the distinction between integrating technologies to create digital twins and how these integrations could actually become useful. We showed that there are several obstacles that have to be surpassed to develop digital twins to a stage where they can offer the type of decision support that we find in other industries. Additionally, we showed that data-driven techniques are integral for digital twins, but they also have limitations. To this end, we pushed towards a combined use of process-based and ML models. We showed the benefits of such applications, demonstrating in this way that the debate for which type of model is better is non-relevant and in effect damaging to the progress of agricultural technological applications. Our applied research is relevant for agriculture, environmental sciences and maybe even a broader range of AI applications, since it a) exposes the difference between having the available resources (large databases, technological and mathematical tools) and being able to make them actionable, and b) provides ways to constitute the resources actionable.

Society

Regarding ties to society, a large part of the Dragon project (funding source) concerned the dissemination of data-driven techniques to the agricultural community. Within this framework we organized a workshop in a summer school in Novi Sad, Serbia, about the use of tools to handle large amounts of data in agricultural applications. We included a practical use case where participants could experiment with data. The audience had varied backgrounds, from academic researchers to company representatives, and agricultural practitioners who were curious about these technologies. Another initiative was the creation of a MOOC in edX [2] about the application of Big Data technologies in the agri-food domain. The material of the course was designed to be accessible to a wide range of audiences. Students can solve exercises and experiment

[2] https://www.edx.org/course/big-data-for-agri-food-applications

with relevant tools and data in a cloud platform, without needing to download or install anything on their system. The course has already run for a second year in a row, with more than a thousand students enrolled from around the world. Furthermore, throughout this work we collaborated with a variety of agriculture stakeholders to find ways to jointly solve practical problems. An outcome of such a collaboration was that a fertilization company in New Zealand got interested in our metamodeling methodology to operationalize digital twins and provided funding to continue this research. Finally, the code developed during this work is publicly available along with documentation on GitHub [3].

Agricultural community

In this work we pointed to how the lack of data culture inhibits the adoption of digital twins in agriculture. We provided evidence that the lack of data sharing, and data management practices restrict the evolution of agriculture in this direction. We believe that by pinpointing these factors and showcasing that they give birth to additional obstacles (that need additional research and funding to be dealt with), the agricultural community will become more aware of them and take steps toward improving its practices. Even more, because the technology to handle large amounts of (sensitive) data exists to facilitate this process, as well as the practices to do so, which have been evolved to a great extent by other industries.

7.4 Future work

Related to the metamodeling methodology

Throughout this work, we developed and evaluated models based mostly on synthetic data. Consequently, we cannot deduct about the predictive capacity of the models in actual field conditions. A question arising here is to what extent our models are transferable to field conditions. It would be interesting to validate these models using observation datasets that come from fields and see if and how the outcome diverges from our previous results.

On another note, to showcase the proposed metamodeling methodology, we used a case study where we had to make a prediction 60 days in advance without any intermediate data. An alternative research direction would be to examine how large we can make this data gap and still have accurate predictions, or how performance is affected by the size of the gap. The same study could be extended by investigating the performance difference between these

[3] https://github.com/BigDataWUR

experiments and their counterparts, which have the gaps filled with simulated data. Potential questions to be answered are whether there is any performance gain, and if this gain is worth the effort of acquiring/generating data to fill the gaps.

Continuing the line of experimenting with synthetic data, another unexplored approach would be to train metamodels based on the output of multiple process-based models. Different models might contain knowledge fragments of how a system evolves, which combined may give a more complete picture of the system under inspection. The question here would be whether ML algorithms can explore this space of simulations and combine this information to potentially outperform individual process-based models.

Related to transferability

Regarding the transferability of the metamodels, a topic not examined here that could enable greater transferability would be to investigate the factors affecting different kinds of domain adaptation. For example, our domain of adaptation might be the type of crop, climate, soil and so on. In the case of crops, are either weather or soil conditions of the origin and target locations more important? In the case of climate, is it soil or agro-management practices? Similarly, for other domains.

An alternative line of research would be to obtain observations from different domains (weather, soil, crop) and investigate whether it is possible to pretrain on synthetic data and transfer to observations. A difficulty in such a study would be to match the agro-management practices of the synthetic and observation datasets, as well as the prediction designs. For example, the fertilization strategies might be different (single vs multiple applications) or the growing period might differ. Another obstacle would be to match the variables contained in the synthetic data with the observations. The reason is that simulations usually contain richer data than what can be measured in field conditions (e.g. daily plant growth, soil nutrients), and metamodels created based on those extra variables will need to have them also from the observation data. Research questions could include how the performance changes between the pretrained (simulated data) and fined-tuned models (simulated data + observations) on the target domains; in which domains the performance is more sensitive; and what the successful types of transfer are (between climates, soils, crops etcetera).

More general issues

A driving factor of agricultural systems is the weather. An important research direction would be to examine how to better incorporate extreme weather phenomena to synthetic datasets and examine how to create metamodels that perform well in weather extremes. An inhibiting factor for these tasks is that weather timeseries are limited in size, since accurate weather measurement is being performed only for decades, and in most places even less than that. Research questions could include how to balance extreme phenomena in these datasets to allow metamodels to learn better, and how metamodels trained solely on extreme weather data would perform.

Lastly, a line of research could involve the level of individualization that we want our digital twins to have. Individualization could be embedded in the variables/features of our ML algorithms. For example, throughout this work the variables/features used did not pinpoint the location where they come from. There was no year, longitude, latitude, or other identifying factors that could help the models identify where a sample comes from and 'remember' properties of that domain to aid the predictions. This information was excluded in order to check whether the models can connect climate and agro-management practices to yield prediction. Other digital twins may have the objective to achieve absolute predictive accuracy. In such cases, our case studies could be repeated to answer if we can achieve different levels of individualization using more personalized variables/features.

Appendices

Chapter 3: Locations included in each model

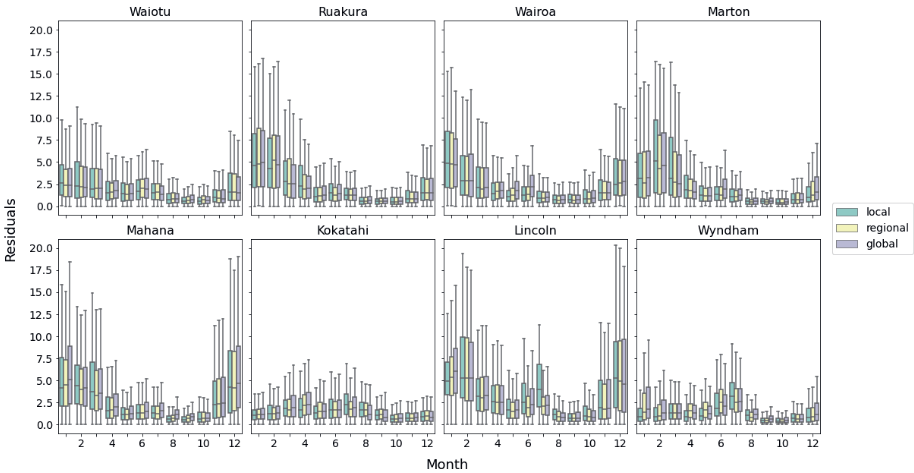
Table A.2: Location data included in each model. The locations of the sampled location experiment were chosen based on climate similarity while the ones of the unsampled location experiment were based on haversine distance.

Target location	**Scenario**					
	Sampled location			**Unsampled location**		
	Local	**Regional**	**Global**	**Local**	**Regional**	**Global**
Waiotu	Waiotu	Waiotu, Wairoa, Ruakura,	all	Ruakura	Ruakura, Wairoa, Marton	all but Waiotu
Ruakura	Ruakura	Ruakura, Marton, Wairoa,	all	Wairoa	Wairoa, Marton, Waiotu	all but Ruakura
Wairoa	Wairoa	Wairoa, Ruakura, Waiotu	all	Marton	Marton, Ruakura, Mahana	all but Wairoa
Marton	Marton	Marton, Mahana, Ruakura	all	Wairoa	Wairoa, Mahana, Ruakura	all but Marton
Mahana	Mahana	Mahana, Marton, Ruakura	all	Marton	Marton, Kokatahi, Lincoln	all but Mahana
Kokatahi	Kokatahi	Kokatahi, Waiotu, Wairoa	all	Lincoln	Lincoln, Mahana, Wyndham	all but Kokatahi
Lincoln	Lincoln	Lincoln, Mahana, Marton	all	Kokatahi	Kokatahi, Mahana, Wyndham	all but Lincoln
Wyndham	Wyndham	Wyndham, Marton, Mahana	all	Lincoln	Lincoln, Kokatahi, Mahana	all but Wyndham

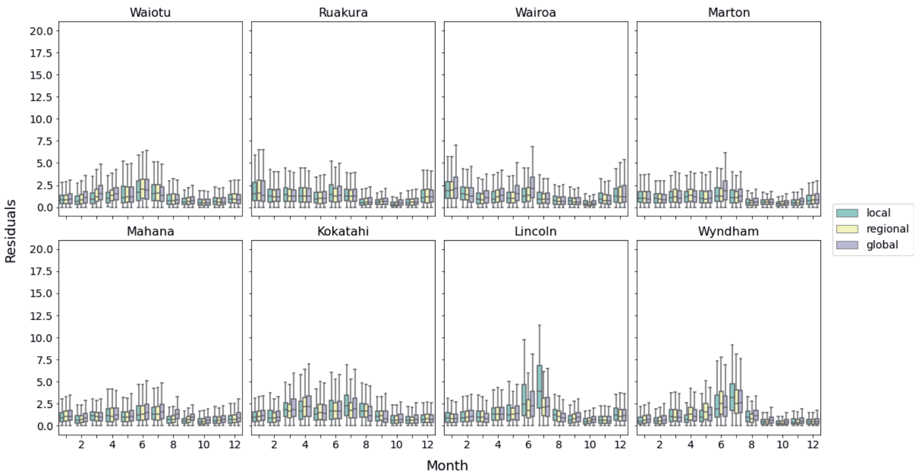
Table A.3: The gridsearch and RMSE results of different machine learning algorithms for training in Ruakura and testing in Waiotu, with the yearly split mentioned in the text, as a preliminary test to choose an algorithm. The gridsearch parameters as denoted as found in *scikit-learn*'s documentation. The parameters in bold are those that gridsearch selected for each algorithm.

	Gridsearch parameters	**RMSE**
Random Forest	n_estimators:[100, **200**] max_depth:[3, 7, **12**] min_samples_split:[**10**, 20] min_samples_leaf:[**10**, 30] max_features:[0.33]	2.51
Gradient Boosting Trees	learning_rate:[0.05, **0.1**, 0.2] n_estimators:[100, **200**] min_samples_split:[**10**, 20] min_samples_leaf:[**10**, 30] max_depth:[3, 7, **12**] max_features:[0.33]	2.52
Linear Support Vector Regression	C:[0.2, **0.5**, 1] epsilon:[**0.05**, 0.1, 0.2] loss:[epsilon_insensitive, **squared_epsilon_insensitive**]	2.68
Elastic Net	alpha: [**0.2**, 0.5, 1] max_iter: [**500**, 1000, 2000] l1_ratio: [**0.2**, 0.5, 0.8]	2.69
Support Vector Regression	kernel:[rbf] C:[0.2, 0.5, **1**] epsilon:[0.05, 0.1, **0.2**]	2.78
Multi-Layer Perceptron	hidden_layer_sizes:[(40,), (40,40), **(60,60)**] activation:[relu] batch_size:[**32**, 64] max_iter:[100] early_stopping:[True] n_iter_no_change:[20]	3.98

Chapter 3: Rainfed vs irrigated plots

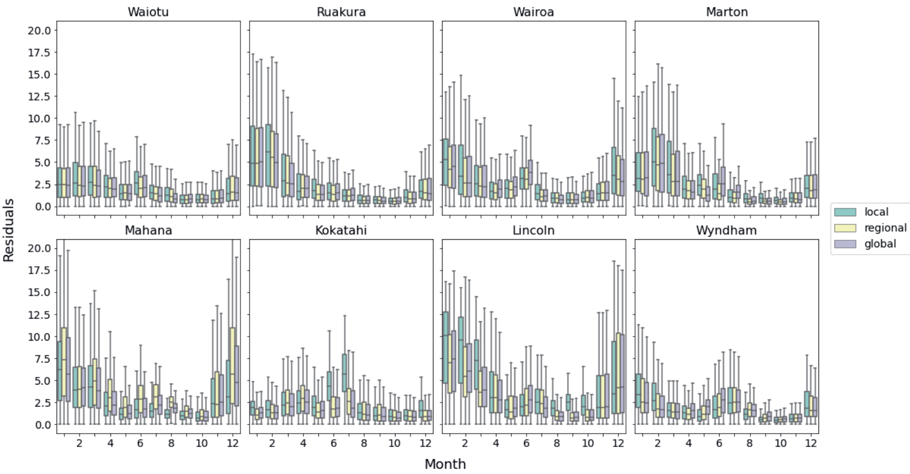


(a)

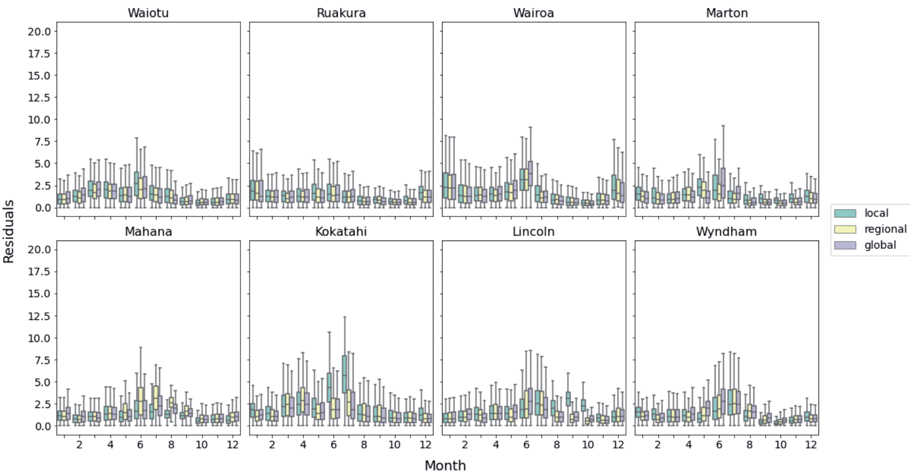


(b)

Figure A.1: **Monthly** test set residuals of models for **sampled** locations in rainfed (a) and irrigated cases (b).

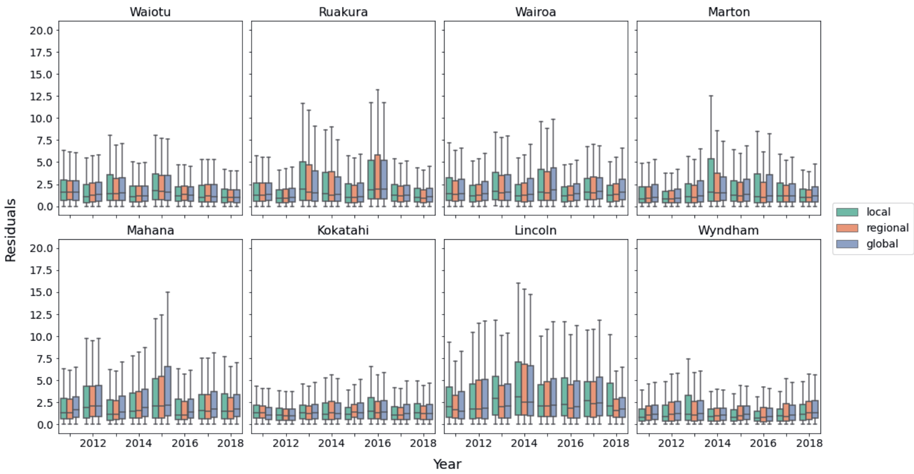


(a)

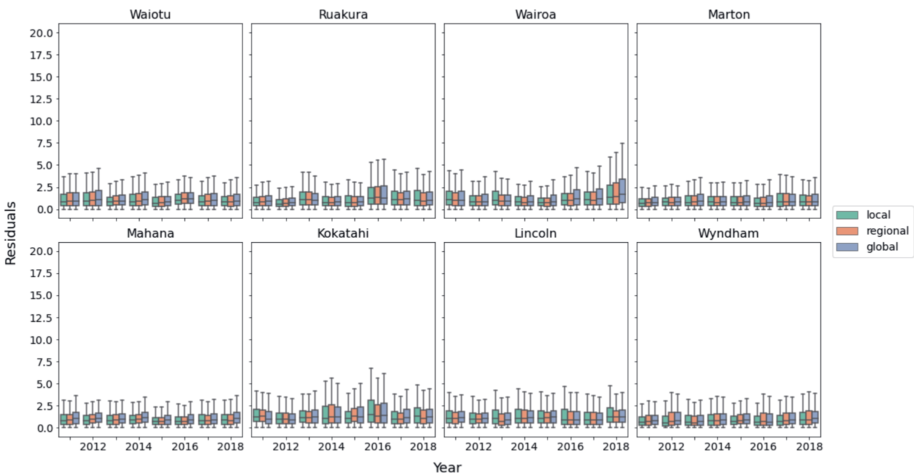


(b)

Figure A.2: **Monthly** test set residuals of models for **unsampled** locations in rainfed (a) and irrigated cases (b).

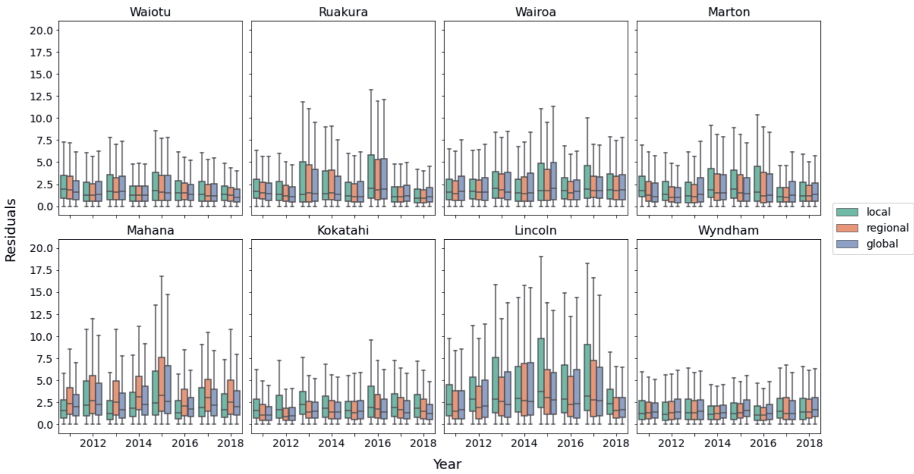


(a)

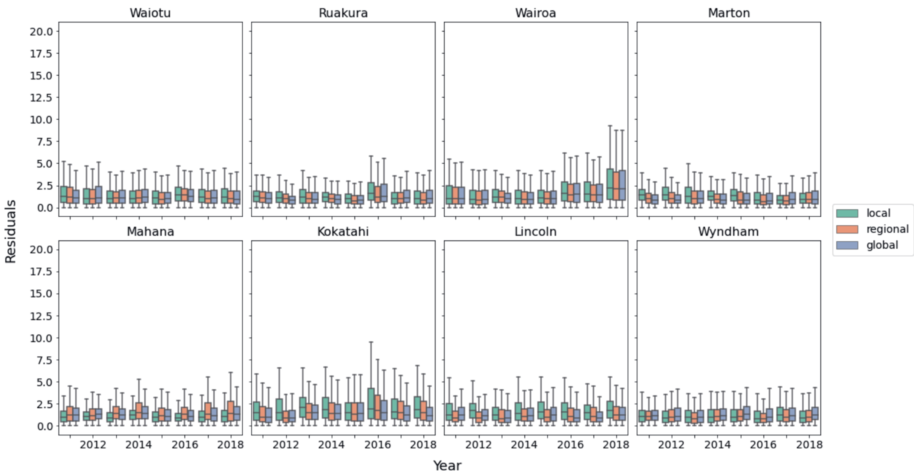


(b)

Figure A.3: **Yearly** test set residuals of models for **sampled** locations in rainfed (a) and irrigated cases (b).



(a)



(b)

Figure A.4: **Yearly** test set residuals of models for **sampled** locations in rainfed (a) and irrigated cases (b).

Chapter 3: Weather plots

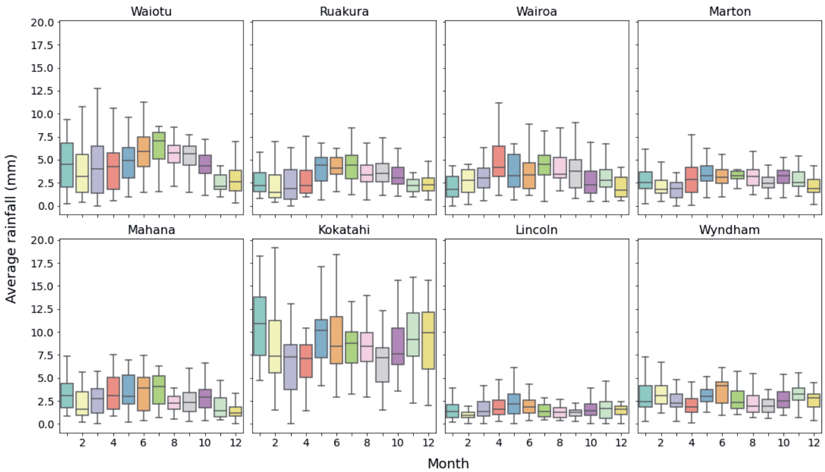


Figure A.5: Average rainfall per month and location for the four weeks that we assume to have data.

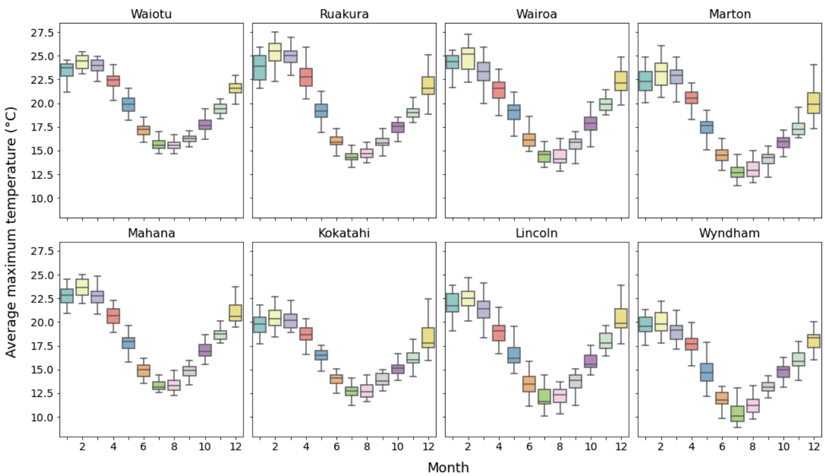


Figure A.6: Average maximum temperature per month and location for the four weeks that we assume to have data.

Chapter 4: Case study sites

Figure B.7: New Zealand sites. Sites on the red circles are the ones included in this work.

Chapter 4: APSIM simulation parameters

Table B.4: APSIM simulation parameters and their ranges. The full factorial of those parameters comprised the input to APSIM.

Parameter	Range
Weather	daily weather from 3 sites
Soil water	42, 67, 110 and 177 mm of plant-available water
Soil fertility	2, 4, and 6% of carbon concentration
Irrigation	irrigated, non-irrigated
Fertilizer year	1979-2018
Fertilizer month	January-December
Fertilizer day	5^{th}, 15^{th} and 25^{th} of the month
Fertilizer amount	0, 20, 40, 60, 80 and 100 kg N / ha

Chapter 4: Tuning and training

The training data were standardized for each location independently. The test and validation data were standardized with the corresponding training scaler, to have the same mean and standard deviation.

The number of layers, nodes in each layer, optimizer parameters, and dropout rate for each architecture were based on the results of a preliminary study. The MLP had two hidden layers with 480 nodes each, optimization with Adam (lr=0.001, weight_decay=0.0001), dropout rate 20%, batch size 64 and 100 epochs. The AE had five hidden layers (300, 200, 120, 200, 300 nodes), optimization with AdamW (lr=0.0003, weight_decay=0.01), dropout rate 10%, batch size 64 and 60 epochs. After training, the decoder was replaced with an MLP with two hidden layers of 180 nodes each and training for 60 epochs. The DAE had the same autoencoder and optimizer as AE, with an addition of an MLP connected to the output of the encoder. The MLP had two hidden layers (80, 40 nodes). The whole network was trained with batch size 64, for 100 epochs.

RF took as input weekly aggregated features which were only a few and were considered explanatory so no feature selection took place. Hyperparameter tuning was performed using Bayesian optimization with 25 iterations and the 5-fold cross-validation score as a metric for each iteration. The tuned parameters can be seen in Table B.5.

Table B.5: The parameters tuned during Bayesian optimization for RF.

Parameters	**Range**
n_estimators	50-800
max_depth	3-12
min_samples_split	30-500
min_samples_leaf	30-500
max_features	0.33

Chapter 6: Climate similarity

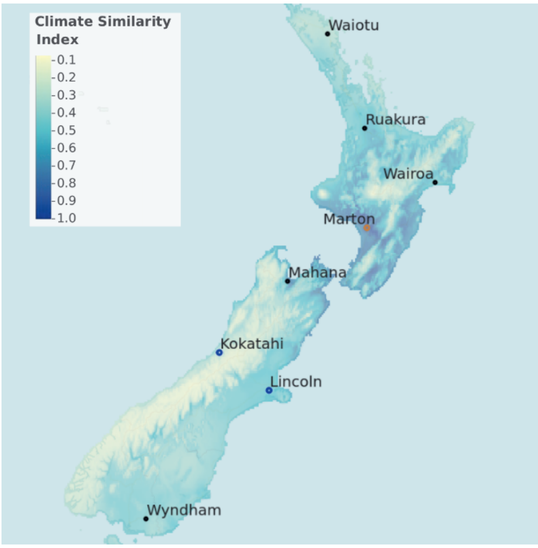
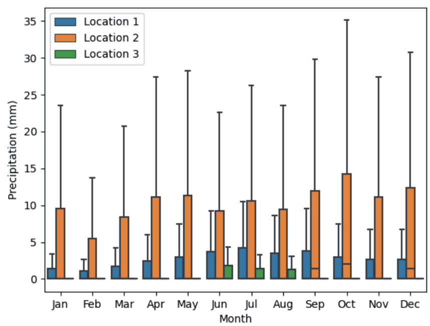
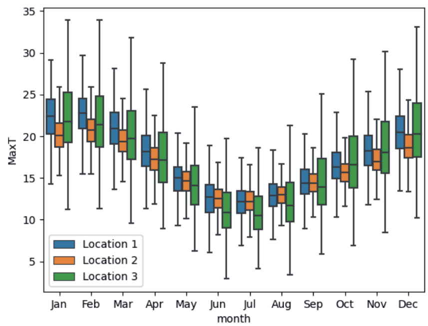


Figure C.8: CCAFS similarity index across New Zealand. The weather parameters for the similarity were precipitation and average temperature. Location 1 (Marton) is colored in brown, and locations 2 (Kokatahi), 3 (Lincoln) in blue. The darker the color on the map, the more similar the climate is to location 1. Location 2 had index value 0.354, and location 3 0.523



(a) Precipitation



(b) Maximum temperature

Figure C.9: Weather parameters known to affect pasture growth for the climates included in this study. The parameters are presented across the months and are aggregated over the years

Chapter 6: Experimental setup simulation parameters and amount of samples

origin model (s1)	Weather Years	Fertilization days	Fertilization months	Fertilizer amounts	Soil Fertilities	Soil water capacities	Total samples
training	10	3	12	2	1	1	720
validation	2	3	12	2	1	1	144
origin test (Clim 1)	2	3	12	2	1	1	144
origin model (s2)							
training	20	3	12	2	1	1	1440
validation	2	3	12	2	1	1	144
origin test (Clim 1)	2	3	12	2	1	1	144
origin model (s3)							
training	10	3	12	5	3	4	21600
validation	2	3	12	5	3	4	4320
origin test (Clim 1)	2	3	12	5	3	4	4320
origin model (s4)							
training	20	3	12	5	3	4	43200
validation	2	3	12	5	3	4	4320
origin test (Clim 1)	2	3	12	5	3	4	4320
target model (s1-s4)							
training	3	3	6	2	1	1	108
validation	2	3	6	2	1	1	72
target test (Clim 2, 3)	2	3	6	2	1	1	72

Figure C.10: Number of parameters and total samples used in each training/validation/test set of each setup

Chapter 6: Model hyperparameters

Table C.6: The fixed hyperparameters of the origin models.

Hyperparameter	Value
learning rate	$4 * 10^{-}5$
batch size	64
epochs	60

Table C.7: The search space for the hyperparameters of the target models.

Hyperparameter	Values
learning rate	$[4 * 10^{-}5, 10^{-}4]$
batch size	[2, 74]
epochs	[7, 15, 30]

Chapter 6: Results - standard deviations

origin model	1992-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.44	0.41	0.56	0.55	0.72				
validation			0.22	0.34	0.76	0.7	0.57		
origin test (Clim 1)					0.52	0.51	0.51	0.46	0.88
target test (Clim 2)					0.3	0.26	0.24	0.29	0.41

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.04	0.04	0.05	0.02	0.03				
validation			0.04	0.01	0.04	0.02	0.04		
origin test (Clim 1)					0.28	0.32	0.14	0.07	0.11
target test (Clim 2)					0.02	0.04	0.04	0.03	0.04

(a) Setup s1: Low weather and agromanagement variabilities

origin model	1982-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.04	0.04	0.05	0.05	0.06				
validation			0.05	0.04	0.05	0.05	0.06		
origin test (Clim 1)					0.04	0.04	0.05	0.05	0.08
target test (Clim 2)					0.08	0.08	0.06	0.07	0.08

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.04	0.04	0.04	0.02	0.02				
validation			0.04	0.01	0.02	0.01	0.03		
origin test (Clim 1)					0.24	0.3	0.15	0.07	0.08
target test (Clim 2)					0.01	0.03	0.03	0.02	0.08

(b) Setup s2: High weather variability, low agromanagement variability

origin model	1992-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.01	0.02	0.01	0.01	0.01				
validation			0.01	0.01	0.02	0.03	0.03		
origin test (Clim 1)					0.02	0.04	0.05	0.05	0.04
target test (Clim 2)					0.04	0.03	0.01	0.01	0.02

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.01	0.03	0.04	0.01	0.05				
validation			0.01	0.01	0.04	0.01	0.01		
origin test (Clim 1)					0.09	0.03	0.04	0.04	0.04
target test (Clim 2)					0.03	0.02	0.03	0.01	0.01

(c) Setup s3: Low weather variability, high agromanagement variability

origin model	1982-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.01	0.01	0	0	0.01				
validation			0.01	0.01	0.02	0.02	0.02		
origin test (Clim 1)					0.02	0.02	0.02	0.04	0.02
target test (Clim 2)					0.04	0.02	0.02	0.01	0.01

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0	0.03	0.02	0.01	0				
validation			0.01	0.01	0.03	0.01	0.01		
origin test (Clim 1)					0.04	0.01	0.02	0.03	0.02
target test (Clim 2)					0.03	0.02	0.02	0.01	0.02

(d) Setup s4: High weather and agromanagement variabilities

Figure C.11: Standard deviations of the various setups for origin models (**climate 1**), and target models in **climate 2**

origin model	1992-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.44	0.41	0.56	0.55	0.72				
validation			0.22	0.34	0.76	0.7	0.57		
origin test (Clim 1)					0.52	0.51	0.51	0.46	0.88
target test (Clim 3)					0.09	0.15	0.26	0.11	0.11

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.06	0.07	0.08	0.06	0.06				
validation			0.05	0.02	0.05	0.07	0.05		
origin test (Clim 1)					0.03	0.12	0.13	0.19	0.3
target test (Clim 3)					0.02	0.04	0.05	0.03	0.04

(a) Setup s1: Low weather and agromanagement variabilities

origin model	1982-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.04	0.04	0.05	0.05	0.06				
validation			0.05	0.04	0.05	0.05	0.06		
origin test (Clim 1)					0.04	0.04	0.05	0.05	0.08
target test (Clim 3)					0.04	0.05	0.05	0.02	0.03

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.06	0.06	0.08	0.06	0.06				
validation			0.05	0.02	0.05	0.07	0.05		
origin test (Clim 1)					0.04	0.13	0.14	0.18	0.24
target test (Clim 3)					0.03	0.04	0.05	0.03	0.04

(b) Setup s2: High weather variability, low agromanagement variability

origin model	1992-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.01	0.02	0.01	0.01	0.01				
validation			0.01	0.01	0.02	0.03	0.03		
origin test (Clim 1)					0.02	0.04	0.05	0.05	0.04
target test (Clim 3)					0.04	0.05	0.02	0.05	0.04

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.04	0.06	0.03	0.02	0.03				
validation			0.02	0.03	0.02	0.03	0.02		
origin test (Clim 1)					0.02	0.19	0.05	0.03	0.08
target test (Clim 3)					0.04	0.05	0.03	0.06	0.05

(c) Setup s3: Low weather variability, high agromanagement variability

origin model	1982-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.01	0.01	0	0	0.01				
validation			0.01	0.01	0.02	0.02	0.02		
origin test (Clim 1)					0.02	0.02	0.02	0.04	0.02
target test (Clim 3)					0.03	0.04	0.03	0.02	0.02

target model	1999-2001	2002	2003	2004	2005	2006	2007	2008	2009
training	0.05	0.04	0.03	0.04	0.02				
validation			0.02	0.02	0.02	0.03	0.03		
origin test (Clim 1)					0.02	0.1	0.07	0.07	0.02
target test (Clim 3)					0.03	0.04	0.04	0.06	0.06

(d) Setup s4: High weather and agromanagement variabilities

Figure C.12: Standard deviations of the various setups for the origin models **climate 1**, and target models in **climate 2**

References

[1] Zhaoyu Zhai, José Fernán Martínez, Victoria Beltran, and Néstor Lucas Martínez. Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture*, 170:105256, 2020. ISSN 0168-1699. doi:https://doi.org/10.1016/j.compag.2020.105256. URL https://www.sciencedirect.com/science/article/pii/S0168169919316497.

[2] Antonio Comparetti and Jose Rafael da Silva. Use of Sentinel-2 Satellite for Spatially Variable Rate Fertiliser Management in a Sicilian Vineyard. *Sustainability*, 14(3), 2022. ISSN 2071-1050. doi:10.3390/su14031688. URL https://www.mdpi.com/2071-1050/14/3/1688.

[3] Iffat Ara, Lydia Turner, Matthew Tom Harrison, Marta Monjardino, Peter deVoil, and Daniel Rodriguez. Application, adoption and opportunities for improving decision support systems in irrigated agriculture: A review. *Agricultural Water Management*, 257:107161, 2021. ISSN 0378-3774. doi:https://doi.org/10.1016/j.agwat.2021.107161. URL https://www.sciencedirect.com/science/article/pii/S0378377421004388.

[4] Elzbieta Czembor, Zygmunt Kaczmarek, Wiesław Pilarczyk, Dariusz Mańkowski, and Jerzy H. Czembor. Simulating Spring Barley Yield under Moderate Input Management System in Poland. *Agriculture*, 12(8):1091, 7 2022. doi:10.3390/agriculture12081091.

[5] Irina Pilvere, Aleksejs Nipers, Agnese Krievina, Ilze Upite, and Daniels Kotovs. LASAM Model: An Important Tool in the Decision Support System for Policymakers and Farmers. *Agriculture (Switzerland)*, 12(5), 5 2022. ISSN 20770472. doi:10.3390/agriculture12050705.

[6] Martin Harris Andy Stanford-Clark, Erwin Frank-Schultz. What are digital twins?, 2019. URL https://developer.ibm.com/articles/what-are-digital-twins/.

[7] Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51(11): 1016–1022, 1 2018. ISSN 24058963. doi:10.1016/j.ifacol.2018.08.474.

[8] W. Wang, Yingfeng Zhang, and R. Y. Zhong. A proactive material handling method for CPS enabled shop-floor. *Robotics and Computer-Integrated Manufacturing*, 61, 2 2020. ISSN 07365845. doi:10.1016/j.rcim.2019.101849.

[9] John O. Oyekan, Windo Hutabarat, Ashutosh Tiwari, Raphael Grech, Min H. Aung, Maria P. Mariani, Laura López-Dávalos, Timothé Ricaud, Sumit Singh, and Charlène Dupuis. The effectiveness of virtual environments in developing collaborative strategies between industrial robots and humans. *Robotics and Computer-Integrated Manufacturing*, 55:41–54, 2 2019. ISSN 07365845. doi:10.1016/j.rcim.2018.07.006.

[10] D. Guivarch, E. Mermoz, Y. Marino, and M. Sartor. Creation of helicopter dynamic systems digital twin using multibody simulations. *CIRP Annals*, 68(1):133–136, 1 2019. ISSN 17260604. doi:10.1016/j.cirp.2019.04.041.

[11] Eric J Tuegel, Anthony R Ingraffea, Thomas G Eason, and S Michael Spottswood. Reengineering Aircraft Structural Life Prediction Using a Digital Twin. *International Journal of Aerospace Engineering*, 2011:154798, 2011. ISSN 1687-5966. doi:10.1155/2011/154798. URL `https://doi.org/10.1155/2011/154798`.

[12] Kalliopi Gkouskou, Ioannis Vlastos, Petros Karkalousos, Dimitrios Chaniotis, Despina Sanoudou, and Aristides G. Eliopoulos. The "virtual Digital Twins" Concept in Precision Nutrition, 11 2020. ISSN 21565376.

[13] R. Kurth, R. Tehel, T. Päßler, M. Putz, K. Wehmeyer, C. Kraft, and H. Schwarze. Forming 4.0: Smart machine components applied as a hybrid plain bearing and a tool clamping system. *Procedia Manufacturing*, 27:65–71, 2019. ISSN 23519789. doi:10.1016/j.promfg.2018.12.045.

[14] A Ghandar, A Ahmed, S Zulfiqar, Z Hua, M Hanai, and G Theodoropoulos. A Decision Support System for Urban Agriculture Using Digital Twin: A Case Study With Aquaponics. *IEEE Access*, 9:35691–35708, 2021. ISSN 2169-3536. doi:10.1109/ACCESS.2021.3061722.

[15] Wei Li, Deli Zhu, and Qing Wang. A single view leaf reconstruction method based on the fusion of ResNet and differentiable render in plantgrowthdigitaltwinsystem. *Computers and Electronics in Agriculture*, 193:106712, 2022. ISSN 0168-1699. doi:https://doi.org/10.1016/j.compag.2022.106712. URL `https://www.sciencedirect.com/science/article/pii/S0168169922000291`.

[16] Cor Verdouw, Bedir Tekinerdogan, Adrie Beulens, and Sjaak Wolfert. Digital twins in smart farming. *Agricultural Systems*, 189, 4 2021. ISSN 0308521X. doi:10.1016/j.agsy.2020.103046.

[17] Willem Jan Knibbe, Lydia Afman, Sjoerd Boersma, Marc-Jeroen Bogaardt, Jochem Evers, Frits van Evert, Jene van der Heide, Idse Hoving, Simon van Mourik, Dick de Ridder, and Allard de Wit. Digital twins in the green life sciences. *NJAS: Impact in Agricultural and Life Sciences*, 94(1):249–279, 2022. doi:10.1080/27685241.2022.2150571. URL `https://doi.org/10.1080/27685241.2022.2150571`.

[18] T G Ritto and F A Rochinha. Digital twin, physics-based model, and machine learning applied to damage detection in structures. *Mechanical Systems and Signal Processing*, 155:107614, 2021. ISSN 0888-3270. doi:https://doi.org/10.1016/j.ymssp.2021.107614. URL `https://www.sciencedirect.com/science/article/pii/S0888327021000091`.

[19] Reza Ramdan Rivai, Takuji Miyamoto, Tatsuya Awano, Rie Takada, Yuki Tobimatsu, Toshiaki Umezawa, and Masaru Kobayashi. Nitrogen deficiency results in changes to cell wall composition of sorghum seedlings. *Scientific Reports*, 11(1):23309, 2021. ISSN 2045-2322. doi:10.1038/s41598-021-02570-y. URL `https://doi.org/10.1038/s41598-021-02570-y`.

[20] Chunyan Zhou, Jing Le, Dengxin Hua, Tingyao He, and Jiandong Mao. Imaging analysis of chlorophyll fluorescence induction for monitoring plant water and nitrogen treatments. *Measurement*, 136:478–486, 2019. ISSN 0263-2241. doi:https://doi.org/10.1016/j.measurement.2018.12.088. URL `https://www.sciencedirect.com/science/article/pii/S026322411831234X`.

[21] Yangjun Zhang, Lubin Tan, Zuofeng Zhu, Lixing Yuan, Daoxin Xie, and Chuanqing Sun. TOND1 confers tolerance to nitrogen deficiency in rice. *The Plant Journal*, 81(3):367–376, 2 2015. ISSN 0960-7412. doi:https://doi.org/10.1111/tpj.12736. URL https://doi.org/10.1111/tpj.12736.

[22] Mei Han, Mamoru Okamoto, Perrin H Beatty, Steven J Rothstein, and Allen G Good. The Genetics of Nitrogen Use Efficiency in Crop Plants. *Annual Review of Genetics*, 49(1):269–289, 11 2015. ISSN 0066-4197. doi:10.1146/annurev-genet-112414-055037. URL https://doi.org/10.1146/annurev-genet-112414-055037.

[23] Xin Zhang, Eric A Davidson, Denise L Mauzerall, Timothy D Searchinger, Patrice Dumas, and Ye Shen. Managing nitrogen for sustainable development. *Nature*, 528(7580):51–59, 2015. ISSN 1476-4687. doi:10.1038/nature15743. URL https://doi.org/10.1038/nature15743.

[24] Dean P. Holzworth, Neil I. Huth, Peter G. deVoil, Eric J. Zurcher, Neville I. Herrmann, Greg McLean, Karine Chenu, Erik J. van Oosterom, Val Snow, Chris Murphy, Andrew D. Moore, Hamish Brown, Jeremy P.M. Whish, Shaun Verrall, Justin Fainges, Lindsay W. Bell, Allan S. Peake, Perry L. Poulton, Zvi Hochman, Peter J. Thorburn, Donald S. Gaydon, Neal P. Dalgliesh, Daniel Rodriguez, Howard Cox, Scott Chapman, Alastair Doherty, Edmar Teixeira, Joanna Sharp, Rogerio Cichota, Iris Vogeler, Frank Y. Li, Enli Wang, Graeme L. Hammer, Michael J. Robertson, John P. Dimes, Anthony M. Whitbread, James Hunt, Harm van Rees, Tim McClelland, Peter S. Carberry, John N.G. Hargreaves, Neil MacLeod, Cam McDonald, Justin Harsdorf, Sara Wedgwood, and Brian A. Keating. APSIM - Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling and Software*, 62:327–350, 12 2014. ISSN 13648152. doi:10.1016/j.envsoft.2014.07.009.

[25] Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihn. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51(11):1016–1022, 2018. ISSN 24058963. doi:10.1016/j.ifacol.2018.08.474. URL https://doi.org/10.1016/j.ifacol.2018.08.474https://linkinghub.elsevier.com/retrieve/pii/S2405896318316021.

[26] F. Caputo, A. Greco, M. Fera, and R. Macchiaroli. Digital twins to enhance the integration of ergonomics in the workplace design. *International Journal of Industrial Ergonomics*, 71:20–31, 5 2019. ISSN 18728219. doi:10.1016/j.ergon.2019.02.001.

[27] Krishnamoorthi Sivalingam, Marco Sepulveda, Mark Spring, and Peter Davies. A Review and Methodology Development for Remaining Useful Life Prediction of Offshore Fixed and Floating Wind turbine Power Converter with Digital Twin Technology Perspective. In *2018 2nd International Conference on Green Energy and Applications (ICGEA)*, pages 197–204. IEEE, 3 2018. ISBN 978-1-5386-5234-3. doi:10.1109/ICGEA.2018.8356292. URL https://ieeexplore.ieee.org/document/8356292/.

[28] Fei Tao, Jiangfeng Cheng, Qinglin Qi, Meng Zhang, He Zhang, and Fangyuan Sui. Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, 94(9-12):3563–3576, 2 2018. ISSN 0268-3768. doi:10.1007/s00170-017-0233-1. URL http://link.springer.com/10.1007/s00170-017-0233-1.

[29] Sebastian Haag and Reiner Anderl. Digital twin – Proof of concept. *Manufacturing Letters*, 15(February):64–66, 1 2018. ISSN 22138463. doi:10.1016/j.mfglet.2018.02.006. URL https://doi.org/10.1016/j.mfglet.2018.02.006https://linkinghub.elsevier.com/retrieve/pii/S2213846318300208.

[30] Stefan Boschert and Roland Rosen. Digital Twin—The Simulation Aspect. In Hehenberger Peter, , and David Bradley, editors, *Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and their Designers*, pages 59–74. Springer International Publishing, Cham, 2016. ISBN 978-3-319-32156-1. doi:10.1007/978-3-319-32156-1_5. URL https://doi.org/10.1007/978-3-319-32156-1_5.

[31] Michael Grieves and John Vickers. Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In Kahlen Franz-Josef, , Shannon Flumerfelt, and and Alves Anabela, editors, *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*, pages 85–113. Springer International Publishing, Cham, 2017. ISBN 978-3-319-38756-7. doi:10.1007/978-3-319-38756-7_4. URL https://doi.org/10.1007/978-3-319-38756-7_4.

[32] Michael Grieves. Digital Twin: Manufacturing Excellence through Virtual Factory Replication. Technical report, Florida Institute of Technology, 2014.

[33] Andy Stanford Clark, Erwin Frank Schultz, and Martin Harris. What are digital twins? Technical report, IBM, 2019. URL https://developer.ibm.com/articles/what-are-digital-twins/.

[34] Elisa Negri, Luca Fumagalli, and Marco Macchi. A Review of the Roles of Digital Twin in CPS-based Production Systems. *Procedia Manufacturing*, 11(June): 939–948, 2017. ISSN 23519789. doi:10.1016/j.promfg.2017.07.198. URL http://dx.doi.org/10.1016/j.promfg.2017.07.198https://linkinghub.elsevier.com/retrieve/pii/S2351978917304067.

[35] T. Mukherjee and T. DebRoy. A digital twin for rapid qualification of 3D printed metallic components. *Applied Materials Today*, 14:59–65, 3 2019. ISSN 23529407. doi:10.1016/j.apmt.2018.11.003.

[36] Qinglin Qi, Fei Tao, Ying Zuo, and Dongming Zhao. Digital Twin Service towards Smart Manufacturing. In *Procedia CIRP*, volume 72, pages 237–242. Elsevier B.V., 2018. doi:10.1016/j.procir.2018.03.103.

[37] Sander J.C. Janssen, Cheryl H. Porter, Andrew D. Moore, Ioannis N. Athanasiadis, Ian Foster, James W. Jones, and John M. Antle. Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology. *Agricultural Systems*, 155:200–212, 7 2017. ISSN 0308521X. doi:10.1016/j.agsy.2016.09.017. URL https://linkinghub.elsevier.com/retrieve/pii/S0308521X16305637.

[38] Diego Inacio Patricio and Rafael Rieder. Computer vision and artificial intelligence in precision agriculture for
}grain crops: A systematic review. *COMPUTERS AND ELECTRONICS IN AGRICULTURE*, 153:69–81, 10 2018. ISSN 0168-1699. doi:10.1016/j.compag.2018.08.001.

[39] Sjaak Wolfert, Lan Ge, Cor Verdouw, and Marc-Jeroen Bogaardt. Big Data in Smart Farming - A review. *AGRICULTURAL SYSTEMS*, 153:69–80, 5 2017. ISSN 0308-521X. doi:10.1016/j.agsy.2017.01.023.

[40] Olakunle Elijah, Tharek Abdul Rahman, Igbafe Orikumhi, Chee Yen Leow, and M H D Nour Hindia. An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges. *IEEE INTERNET OF THINGS JOURNAL*, 5(5):3758–3773, 10 2018. ISSN 2327-4662. doi:10.1109/JIOT.2018.2844296.

[41] Chiara Cimino, Elisa Negri, and Luca Fumagalli. Review of digital twin applications in manufacturing. *Computers in Industry*, 113, 12 2019. ISSN 01663615. doi:10.1016/j.compind.2019.103130.

[42] European Commision. Technology readiness levels (TRL); Extract from Part 19 - Commission Decision C(2014)4995". Technical report, European Commision, 2014. URL https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf.

[43] Web of Science. URL https://clarivate.com/webofsciencegroup/solutions/web-of-science/.

[44] Matthew J. Smith. Getting value from artificial intelligence in agriculture. *Animal Production Science*, 2018. ISSN 1836-0939. doi:10.1071/AN18522. URL http://www.publish.csiro.au/?paper=AN18522.

[45] G. Tagliavini, T. Defraeye, and J. Carmeliet. Multiphysics modeling of convective cooling of non-spherical, multi-material fruit to unveil its quality evolution throughout the cold chain. *Food and Bioproducts Processing*, 117:310–320, 9 2019. ISSN 09603085. doi:10.1016/j.fbp.2019.07.013.

[46] Dimitrios S. Paraforos, Galibjon M. Sharipov, and Hans W. Griepentrog. ISO 11783-compatible industrial sensor and control systems and related research: A review. *Computers and Electronics in Agriculture*, 163(June):104863, 8 2019. ISSN 01681699. doi:10.1016/j.compag.2019.104863. URL https://linkinghub.elsevier.com/retrieve/pii/S0168169919304636.

[47] Naoum Tsolakis, Dimitrios Bechtsis, and Dionysis Bochtis. Agros: A robot operating system based emulation tool for agricultural robotics. *Agronomy*, 9(7), 2019. ISSN 20734395. doi:10.3390/agronomy9070403.

[48] Google Scholar, . URL https://scholar.google.com/.

[49] Gary Tan, Axel Lehmann, Yong Meng Teo, and Wentong Cai, editors. *Methods and Applications for Modeling and Simulation of Complex Systems*, volume 1094 of *Communications in Computer and Information Science*. Springer Singapore, Singapore, 2019. ISBN 978-981-15-1077-9. doi:10.1007/978-981-15-1078-6. URL http://link.springer.com/10.1007/978-981-15-1078-6.

[50] Seng-Kyoun Jo, Dae-Heon Park, Hyeon Park, and Se-Han Kim. Smart Livestock Farms Using Digital Twin: Feasibility Study. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1461–1463. IEEE, 10 2018. ISBN 978-1-5386-5041-7. doi:10.1109/ICTC.2018.8539516. URL https://ieeexplore.ieee.org/document/8539516/.

[51] A. Kampker, V. Stich, P. Jussen, B. Moser, and J. Kuntz. Business models for industrial smart services - the example of a digital twin for a product-service-system for potato harvesting. In *Procedia CIRP*, volume 83, pages 534–540. Elsevier B.V., 2019. doi:10.1016/j.procir.2019.04.114.

[52] Peyman Moghadam, Thomas Lowe, and Everard J. Edwards. Digital Twin for the Future of Orchard Production Systems. *Proceedings*, 36(1):92, 2 2020. doi:10.3390/proceedings2019036092. URL https://www.mdpi.com/2504-3900/36/1/92.

[53] Qinglin Qi, Fei Tao, Tianliang Hu, Nabil Anwer, Ang Liu, Yongli Wei, Lihui Wang, and A. Y.C. Nee. Enabling technologies and tools for digital twin. *Journal of Manufacturing Systems*, 2019. ISSN 02786125. doi:10.1016/j.jmsy.2019.10.001.

[54] Thomas Machl, Andreas Donaubauer, and Thomas H Kolbe. Planning Agricultural Core Road Networks Based on a Digital Twin of the Cultivated Landscape. In *Full Paper Journal of Digital Landscape Architecture*, pages 316–327, Dessau-Roßlau, 2019. ISBN 978-3-87907-663-5. doi:10.14627/537663034.

[55] Jan Willem Kruize. Digital twins in farm management: illustrations from the FIWARE accelerators SmartAgriFood and Fractals. In *7th Asian-Australasian Conference on Precision Agriculture*, 2018. doi:10.5281/zenodo.893662. URL `https://www.researchgate.net/publication/322886729`.

[56] Rafael Gomes Alves, Gilberto Souza, Rodrigo Filev Maia, Anh Lan Ho Tran, Carlos Kamienski, Juha-Pekka Soininen, Plinio Thomaz Aquino-Jr, and Fabio Lima. A digital twin for smart farming. In *IEEE Global Humanitarian Technology Conference*, Seattle, 2019.

[57] Jorge A. Delgado, Nicholas M. Short, Daniel P. Roberts, and Bruce Vandenberg. Big Data Analysis for Sustainable Agriculture on a Geospatial Cloud Framework. *Frontiers in Sustainable Food Systems*, 3:54, 7 2019. doi:10.3389/fsufs.2019.00054. URL `https://www.frontiersin.org/article/10.3389/fsufs.2019.00054`.

[58] Google, . URL `https://www.google.com/`.

[59] José Monteiro, João Barata, Marco Veloso, Luís Veloso, and João Nunes. Towards Sustainable Digital Twins for Vertical Farming. In *Thirteenth International Conference on Digital Information Mangement (ICDIM 2018)*, pages 234–239, Berlin, 2018. ISBN 9781538652442.

[60] IBM Research. #twinning: Farming's digital doubles will help feed a growing population using less resources., 2018. URL `https://www.research.ibm.com/5-in-5/seed/`.

[61] R&D WORLD. Digital Twin Solutions for Smart Farming, 2019. URL `https://www.rdworldonline.com/rd100/digital-twin-solutions-for-smart-farming/`.

[62] Michael R. Collins. Agility in Digital Farming, 2019. URL `https://www.precisionag.com/digital-farming/agility-in-digital-farming/`.

[63] Aenne Barnard. In the digital indoor garden, 2019. URL `https://new.siemens.com/global/en/company/stories/research-technologies/digitaltwin/digital-indoor-garden.html`.

[64] Jing Ming Chiu, Joy Yen, and Ying Shen. "Digital Twin Solutions for Smart Farming", the III Development AI+HI Total Solution, Awarded R&D 100., 2019. URL `https://web.iii.org.tw/Press/NewsDtl.aspx?nsp_sqno=269&fm_sqno=14`.

[65] Bhushan Mokal and Arpit Sharma. Use Cases: Digital Twin in Livestock Farming, 2020. URL `https://www.aranca.com/knowledge-library/articles/ip-research/use-cases-digital-twin-in-livestock-farming`.

[66] Thomas Ohnemus. Digital Twin Excellence: Two Shining Examples, 2020. URL `https://www.digitalistmag.com/iot/2018/06/14/digital-twin-excellence-2-shining-examples-06175901`.

[67] Wageningen University & Research. WUR is working on Digital Twins for tomatoes, food and farming, 2020. URL https://www.wur.nl/en/newsarticle/WUR-is-working-on-Digital-Twins-for-tomatoes-food-and-farming.htm.

[68] Seng-Kyoun Jo, Dae-Heon Park, Hyeon Park, and Se-Han Kim. Smart Livestock Farms Using Digital Twin: Feasibility Study. In *International Conference on ICT Convergence : "ICT Convergence Powered by Smart Intelligence"*, Jeju Island, 2018. ISBN 9781538650417.

[69] Sakdirat Kaewunruen, Panrawee Rungskunroch, and Joshua Welsh. A Digital-Twin Evaluation of Net Zero Energy Building for Existing Buildings. *Sustainability*, 11(1):159, 12 2018. ISSN 2071-1050. doi:10.3390/su11010159. URL http://www.mdpi.com/2071-1050/11/1/159.

[70] Eann A. Patterson and Maurice P. Whelan. A framework to establish credibility of computational models in biology, 10 2017. ISSN 00796107.

[71] Paula Fraga-Lamas and Tiago M. Fernández-Caramés. A Review on Blockchain Technologies for an Advanced and Cyber-Resilient Automotive Industry. *IEEE Access*, 7:17578–17598, 2019. ISSN 21693536. doi:10.1109/ACCESS.2019.2895302.

[72] Pai Zheng, Zuoxu Wang, Chun Hsien Chen, and Li Pheng Khoo. A survey of smart product-service systems: Key aspects, challenges and future perspectives, 10 2019. ISSN 14740346.

[73] Dawn M Tilbury. Annual Review of Control, Robotics, and Autonomous Systems Cyber-Physical Manufacturing Systems. *Annu. Rev. Control Robot. Auton. Syst*, 2:427–443, 2019. doi:10.1146/annurev-control-053018. URL https://doi.org/10.1146/annurev-control-053018-.

[74] Stephen J A Dewitt, Anshuman Sinha, Jayashree Kalyanaraman, Fengyi Zhang, Matthew J Realff, and Ryan P Lively. Critical Comparison of Structured Contactors for Adsorption-Based Gas Separations. *Annual Review of Chemical and Biomolecular Engineering*, 9(1):129–152, 2018. doi:10.1146/annurev-chembioeng. URL https://doi.org/10.1146/annurev-chembioeng-.

[75] Ruth N. Bolton, Janet R. McColl-Kennedy, Lilliemay Cheung, Andrew Gallan, Chiara Orsingher, Lars Witell, and Mohamed Zaki. Customer experience challenges: bringing together digital, physical and social realms. *Journal of Service Management*, 29(5):776–808, 10 2018. ISSN 17575818. doi:10.1108/JOSM-04-2018-0113.

[76] Rui Dong, Changyang She, Wibowo Hardjawana, Yonghui Li, and Branka Vucetic. Deep Learning for Hybrid 5G Services in Mobile Edge Computing Systems: Learn from a Digital Twin. *IEEE Transactions on Wireless Communications*, 18(10):4692–4707, 10 2019. ISSN 15582248. doi:10.1109/TWC.2019.2927312.

[77] Yuval Cohen, Maurizio Faccio, Francesco Pilati, and Xifan Yao. Design and management of digital manufacturing and assembly systems in the Industry 4.0 era, 12 2019. ISSN 14333015.

[78] Tetsuo Tomiyama, Eric Lutters, Rainer Stark, and Michael Abramovici. Development capabilities for smart products. *CIRP Annals*, 68(2):727–750, 1 2019. ISSN 17260604. doi:10.1016/j.cirp.2019.05.010.

[79] Qinglin Qi and Fei Tao. Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. *IEEE Access*, 6:3585–3593, 1 2018. ISSN 21693536. doi:10.1109/ACCESS.2018.2793265.

[80] Fei Tao, He Zhang, Ang Liu, and A. Y.C. Nee. Digital Twin in Industry: State-of-the-Art. *IEEE Transactions on Industrial Informatics*, 15(4):2405–2415, 4 2019. ISSN 15513203. doi:10.1109/TII.2018.2873186.

[81] Yuqian Lu, Chao Liu, Kevin I.Kai Wang, Huiyue Huang, and Xun Xu. Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues, 2 2020. ISSN 07365845.

[82] Venkat Raman and Malik Hassanaly. Emerging trends in numerical simulations of combustion systems. *Proceedings of the Combustion Institute*, 37(2):2073–2089, 2019. ISSN 15407489. doi:10.1016/j.proci.2018.07.121.

[83] William Yi Wang, Jinshan Li, Weimin Liu, and Zi Kui Liu. Integrated computational materials engineering for advanced materials: A brief review. *Computational Materials Science*, 158: 42–48, 2 2019. ISSN 09270256. doi:10.1016/j.commatsci.2018.11.001.

[84] Dimitrios S. Paraforos, Galibjon M. Sharipov, and Hans W. Griepentrog. ISO 11783-compatible industrial sensor and control systems and related research: A review, 8 2019. ISSN 01681699.

[85] Claudio Pizzolato, David J. Saxby, Dinesh Palipana, Laura E. Diamond, Rod S. Barrett, Yang D. Teng, and David G. Lloyd. Neuromusculoskeletal modeling-based prostheses for recovery after spinal cord injury. *Frontiers in Neurorobotics*, 13, 2019. ISSN 16625218. doi:10.3389/fnbot.2019.00097.

[86] Mohammed M. Mabkhot, Abdulrahman M. Al-Ahmari, Bashir Salah, and Hisham Alkhalefah. Requirements of the smart factory system: A survey and perspective, 6 2018. ISSN 20751702.

[87] Amit Gupta and Biswajit Basu. Sustainable Primary Aluminium Production: Technology Status and Future Opportunities. *Transactions of the Indian Institute of Metals*, 72(8):2135–2150, 8 2019. ISSN 09751645. doi:10.1007/s12666-019-01699-9.

[88] Morteza Ghobakhloo. The future of manufacturing industry: a strategic roadmap toward Industry 4.0. *Journal of Manufacturing Technology Management*, 29(6):910–936, 10 2018. ISSN 1741038X. doi:10.1108/JMTM-02-2018-0057.

[89] Dongwook Kim and Sungbum Kim. The role of mobile technology in tourism: Patents, articles, news, and mobile tour app reviews. *Sustainability (Switzerland)*, 9(11), 11 2017. ISSN 20711050. doi:10.3390/su9112082.

[90] Francesco Longo, Letizia Nicoletti, and Antonio Padovano. Ubiquitous knowledge empowers the Smart Factory: The impacts of a Service-oriented Digital Twin on enterprises' performance. *Annual Reviews in Control*, 47:221–236, 1 2019. ISSN 13675788. doi:10.1016/j.arcontrol.2019.01.001.

[91] Eann A. Patterson and Maurice P. Whelan. A framework to establish credibility of computational models in biology, 10 2017. ISSN 00796107.

[92] FIWARE Foundation. FIWARE Accelarator Programme, 2020. URL `https://www.fiware.org/community/fiware-accelerator-programme/`.

[93] Bruno Basso and John Antle. Digital agriculture to design sustainable agricultural systems. *Nature Sustainability*, 3(4):254–256, 2020. ISSN 2398-9629. doi:10.1038/s41893-020-0510-0. URL `https://doi.org/10.1038/s41893-020-0510-0`.

[94] James W. Jones, John M. Antle, Bruno Basso, Kenneth J. Boote, Richard T. Conant, Ian Foster, H. Charles J. Godfray, Mario Herrero, Richard E. Howitt, Sander Janssen, Brian A. Keating, Rafael Munoz-Carpena, Cheryl H. Porter, Cynthia Rosenzweig, and Tim R. Wheeler. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural Systems*, 155:269–288, 7 2017. ISSN 0308521X. doi:10.1016/j.agsy.2016.09.021.

[95] Philip Brown, Adam Daigneault, and Joshua Dawson. Age, values, farming objectives, past management decisions, and future intentions in New Zealand agriculture. *Journal of Environmental Management*, 231:110–120, 2 2019. ISSN 10958630. doi:10.1016/j.jenvman.2018.10.018.

[96] Michael F Goodchild, Huadong Guo, Alessandro Annoni, Ling Bian, Kees de Bie, Frederick Campbell, Max Craglia, Manfred Ehlers, John van Genderen, Davina Jackson, Anthony J Lewis, Martino Pesaresi, Gábor Remetey-Fülöpp, Richard Simpson, Andrew Skidmore, Changlin Wang, and Peter Woodgate. Next-generation Digital Earth. *Proceedings of the National Academy of Sciences of the United States of America*, 109(28):11088–94, 7 2012. ISSN 1091-6490. doi:10.1073/pnas.1202383109. URL `http://www.ncbi.nlm.nih.gov/pubmed/22723346http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3396470`.

[97] Rohan Jadhav, Chandran Achutan, Gleb Haynatzki, Shireen Rajaram, and Risto Rautiainen. Review and Meta-analysis of Emerging Risk Factors for Agricultural Injury. *Journal of Agromedicine*, 21(3):284–297, 7 2016. doi:10.1080/1059924X.2016.1179611. URL `https://doi.org/10.1080/1059924X.2016.1179611`.

[98] Timothy D. West and Mark Blackburn. Is Digital Thread/Digital Twin Affordable? A Systemic Assessment of the Cost of DoD's Latest Manhattan Project. In *Procedia Computer Science*, volume 114, pages 47–56. Elsevier B.V., 2017. doi:10.1016/j.procs.2017.09.003.

[99] Mika Lohtander, Elodie Garcia, Minna Lanz, Jarno Volotinen, Juho Ratava, and Jarno Kaakkunen. Micro Manufacturing Unit - Creating Digital Twin Objects with Common Engineering Software. In *Procedia Manufacturing*, volume 17, pages 468–475. Elsevier B.V., 2018. doi:10.1016/j.promfg.2018.10.071.

[100] Behrang Ashtari Talkhestani, Nasser Jazdi, Wolfgang Schlögl, and Michael Weyrich. A concept in synchronization of virtual production system with real factory based on anchor-point method. In *Procedia CIRP*, volume 67, pages 13–17. Elsevier B.V., 2018. doi:10.1016/j.procir.2017.12.168.

[101] Bin He and Kai Jian Bai. Digital twin-based sustainable intelligent manufacturing: a review. *Advances in Manufacturing*, 9(1), 3 2021. ISSN 21953597. doi:10.1007/s40436-020-00302-5.

[102] Ying Liu, Lin Zhang, Yuan Yang, Longfei Zhou, Lei Ren, Fei Wang, Rong Liu, Zhibo Pang, and M. Jamal Deen. A Novel Cloud-Based Framework for the Elderly Healthcare Services Using Digital Twin. *IEEE Access*, 7:49088–49101, 2019. ISSN 21693536. doi:10.1109/ACCESS.2019.2909828.

[103] Agnethe N Pedersen, Morten Borup, Annette Brink-Kjær, Lasse E Christiansen, and Peter S Mikkelsen. Living and Prototyping Digital Twins for Urban Water Systems: Towards Multi-Purpose Value Creation Using Models and Sensors. *Water*, 13(5), 2021. ISSN 2073-4441. doi:10.3390/w13050592. URL `https://www.mdpi.com/2073-4441/13/5/592`.

[104] Christos Pylianidis, Sjoukje Osinga, and Ioannis N Athanasiadis. Introducing digital twins to agriculture. *Computers and Electronics in Agriculture*, 184:105942, 2021. ISSN 0168-1699. doi:https://doi.org/10.1016/j.compag.2020.105942. URL `https://www.sciencedirect.com/science/article/pii/S0168169920331471`.

[105] Suresh Neethirajan and Bas Kemp. Digital twins in livestock farming, 4 2021. ISSN 20762615.

[106] Stefano Nativi, Paolo Mazzetti, and Max Craglia. Digital ecosystems for developing digital twins of the earth: The destination earth case. *Remote Sensing*, 13(11), 6 2021. ISSN 20724292. doi:10.3390/rs13112119.

[107] Huadong Guo, Stefano Nativi, Dong Liang, Max Craglia, Lizhe Wang, Sven Schade, Christina Corban, Guojin He, Martino Pesaresi, Jianhui Li, Zeeshan Shirazi, Jie Liu, and Alessandro Annoni. Big Earth Data science: an information framework for a sustainable planet. *International Journal of Digital Earth*, 13(7):743–767, 7 2020. ISSN 1753-8947. doi:10.1080/17538947.2020.1743785. URL `https://doi.org/10.1080/17538947.2020.1743785`.

[108] Paul Voosen. Europe builds 'digital twin' of earth to hone climate forecasts, 10 2020. ISSN 10959203.

[109] Peter Bauer, Bjorn Stevens, and Wilco Hazeleger. A digital twin of Earth for the green transition. *Nature Climate Change*, 11(2):80–83, 2021. ISSN 1758-6798. doi:10.1038/s41558-021-00986-y. URL `https://doi.org/10.1038/s41558-021-00986-y`.

[110] Yolanda Gil, Suzanne A Pierce, Hassan Babaie, Arindam Banerjee, Kirk Borne, Gary Bust, Michelle Cheatham, Imme Ebert-Uphoff, Carla Gomes, Mary Hill, John Horel, Leslie Hsu, Jim Kinter, Craig Knoblock, David Krum, Vipin Kumar, Pierre Lermusiaux, Yan Liu, Chris North, Victor Pankratius, Shanan Peters, Beth Plale, Allen Pope, Sai Ravela, Juan Restrepo, Aaron Ridley, Hanan Samet, Shashi Shekhar, Katie Skinner, Padhraic Smyth, Basil Tikoff, Lynn Yarmey, and Jia Zhang. Intelligent Systems for Geosciences: An Essential Research Agenda. *Communications of the ACM*, 62(1):76–84, 12 2018. ISSN 0001-0782. doi:10.1145/3192335. URL `https://doi.org/10.1145/3192335`.

[111] Yang Li and Jiachen Yang. Meta-learning baselines and database for few-shot classification in agriculture. *Computers and Electronics in Agriculture*, 182:106055, 2021. ISSN 0168-1699. doi:https://doi.org/10.1016/j.compag.2021.106055. URL `https://www.sciencedirect.com/science/article/pii/S0168169921000739`.

[112] Anuj Karpatne, Gowtham Atluri, James H. Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2318–2331, 10 2017. ISSN 1041-4347. doi:10.1109/TKDE.2017.2720168. URL `http://ieeexplore.ieee.org/document/7959606/`.

[113] Maharavo Marie Julie Ramanantenasoa, Sophie Génermont, Jean Marc Gilliot, Carole Bedos, and David Makowski. Meta-modeling methods for estimating ammonia volatilization from nitrogen fertilizer and manure applications. *Journal of Environmental Management*, 236:195–205, 4 2019. ISSN 10958630. doi:10.1016/j.jenvman.2019.01.066.

[114] Dimitrios A. Kasampalis, Thomas K. Alexandridis, Chetan Deva, Andrew Challinor, Dimitrios Moshou, and Georgios Zalidis. Contribution of remote sensing on crop models: A review, 2018. ISSN 2313433X.

[115] Kaitlin Togliatti, Sotirios V. Archontoulis, Ranae Dietzel, Laila Puntel, and Andy VanLoocke. How does inclusion of weather forecasting impact in-season crop model predictions? *Field Crops Research*, 214:261–272, 12 2017. ISSN 03784290. doi:10.1016/j.fcr.2017.09.008.

[116] Iris Vogeler, Rogerio Cichota, and Val Snow. Identification and testing of early indicators for N leaching from urine patches. *Journal of Environmental Management*, 130:55–63, 2013. ISSN 03014797. doi:10.1016/j.jenvman.2013.08.047.

[117] W A Dorigo, R Zurita-Milla, A J W de Wit, J Brazile, R Singh, and M E Schaepman. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation*, 9(2):165–193, 2007. ISSN 0303-2434. doi:https://doi.org/10.1016/j.jag.2006.05.003. URL `https://www.sciencedirect.com/science/article/pii/S0303243406000201`.

[118] R Cichota, V O Snow, and A B Tait. A functional evaluation of virtual climate station rainfall data. *New Zealand Journal of Agricultural Research*, 51(3):317–329, 9 2008. ISSN 0028-8233. doi:10.1080/00288230809510463. URL `https://doi.org/10.1080/00288230809510463`.

[119] Jiju Antony. 6 - Full Factorial Designs. In Jiju Antony, editor, *Design of Experiments for Engineers and Scientists (Second Edition)*, pages 63–85. Elsevier, Oxford, 2014. ISBN 978-0-08-099417-8. doi:https://doi.org/10.1016/B978-0-08-099417-8.00006-7. URL `https://www.sciencedirect.com/science/article/pii/B9780080994178000067`.

[120] Robert W Blanning. The construction and implementation of metamodels. *SIMULATION*, 24(6): 177–184, 1975. doi:10.1177/003754977502400606. URL `https://doi.org/10.1177/003754977502400606`.

[121] Anuj Karpatne, William Watkins, Jordan Read, and Vipin Kumar. Physics-guided Neural Networks (PGNN): An Application in Lake Temperature Modeling. *ACM SIGKDD International Conference on Knowledge discovery and data mining*, 10 2017. URL `http://arxiv.org/abs/1710.11431`.

[122] Mohsen Shahhosseini, Guiping Hu, Isaiah Huber, and Sotirios V. Archontoulis. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports*, 11(1), 12 2021. ISSN 20452322. doi:10.1038/s41598-020-80820-1.

[123] Michael J. Roberts, Noah O. Braun, Thomas R. Sinclair, David B. Lobell, and Wolfram Schlenker. Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, 12(9), 9 2017. ISSN 17489326. doi:10.1088/1748-9326/aa7f33.

[124] Daniel W. Gladish, Ross Darnell, Peter J. Thorburn, and Bhakti Haldankar. Emulated Multivariate Global Sensitivity Analysis for Complex Computer Models Applied to Agricultural Simulators. *Journal of Agricultural, Biological, and Environmental Statistics*, 24(1):130–153, 3 2019. ISSN 15372693. doi:10.1007/s13253-018-00346-y.

[125] Zhang Ruixi, Zen Remmy, Xing Jifang, Arsa Dewa Made Sri, Saha Abhishek, and Bressan Stéphane. Hydrological Process Surrogate Modelling and Simulation with Neural Networks. In Hady W. Lauw et al., editors, *Advances in Knowledge Discovery and Data Mining*, pages 449–461, Cham, 2020. Springer International Publishing. ISBN 978-3-030-47436-2. doi:10.1007/978-3-030-47436-2_34.

[126] Nathalie Villa-Vialaneix, Marco Follador, Marco Ratto, and Adrian Leip. A comparison of eight metamodeling techniques for the simulation of N 2O fluxes and N leaching from corn crops. *Environmental Modelling and Software*, 34:51–66, 6 2012. ISSN 13648152. doi:10.1016/j.envsoft.2011.05.003.

[127] Bernard T. Nolan, Christopher T. Green, Paul F. Juckem, Lixia Liao, and James E. Reddy. Metamodeling and mapping of nitrate flux in the unsaturated zone and groundwater, Wisconsin, USA. *Journal of Hydrology*, 559:428–441, 4 2018. ISSN 00221694. doi:10.1016/j.jhydrol.2018.02.029.

[128] Mohsen Shahhosseini, Rafael A. Martinez-Feria, Guiping Hu, and Sotirios V. Archontoulis. Maize yield and nitrate loss prediction with machine learning algorithms. *Environmental Research Letters*, 14(12), 12 2019. ISSN 17489326. doi:10.1088/1748-9326/ab5268.

[129] A G Gillingham, J D Morton, and M H Gray. Pasture responses to phosphorus and nitrogen fertilisers on east coast hill country: 2. Clover and grass production from easy slopes. *New Zealand Journal of Agricultural Research*, 51(2):85–97, 2008. doi:10.1080/00288230809510438.

[130] C A Rotz, F Taube, M P Russelle, J Oenema, M A Sanderson, and M Wachendorf. Whole-farm perspectives of nutrient flows in grassland agriculture. *Crop Science*, 45(6):2139–2159, 11 2005. ISSN 0011-183X. doi:10.2135/cropsci2004.0523. URL http://www.scopus.com/inward/record.url?scp=27644585386&partnerID=8YFLogxKhttp://www.scopus.com/inward/citedby.url?scp=27644585386&partnerID=8YFLogxK.

[131] David Charles Whitehead. *Grassland nitrogen*. CAB international, 1995. ISBN 9780851989150.

[132] D A Clark, J R Caradus, R M Monaghan, P Sharp, and B S Thorrold. Issues and options for future dairy farming in New Zealand. *New Zealand Journal of Agricultural Research*, 50(2): 203–221, 6 2007. ISSN 0028-8233. doi:10.1080/00288230709510291. URL https://doi.org/10.1080/00288230709510291.

[133] Keith G Pembleton, Richard P Rawnsley, and Lucy L Burkitt. Environmental influences on optimum nitrogen fertiliser rates for temperate dairy pastures. *European Journal of Agronomy*, 45: 132–141, 2013. ISSN 1161-0301. doi:https://doi.org/10.1016/j.eja.2012.09.006. URL https://www.sciencedirect.com/science/article/pii/S1161030112001335.

[134] Waikato Regional Council. Nitrogen fertiliser and effluent. Technical report, 5 2015. URL https://www.waikatoregion.govt.nz/assets/PageFiles/30016/factsheets/CNM%20factsheet%20get_most_fertiliser_effluent_nitrogen_8.pdf.

[135] NZ farm source. Tailoring nitrogen applications to optimise response , 4 2021. URL https://store.nzfarmsource.co.nz/enrichplus/pasture-cropping/tailoring-nitrogen-applications-to-optimise-response.

[136] DairyNZ. Farmfact: Seasonal nitrogen use. Technical report, 7 2012. URL https://www.dairynz.co.nz/publications/farmfacts/fertiliser-and-nutrient-management/farmfact-7-11/.

[137] APSIM. APSIM. URL https://www.apsim.info/.

[138] F. Y. Li, V. O. Snow, and D. P. Holzworth. Modelling the seasonal and geographical pattern of pasture production in New Zealand. *New Zealand Journal of Agricultural Research*, 54(4): 331–352, 2011. ISSN 00288233. doi:10.1080/00288233.2011.613403.

[139] R Cichota, V O Snow, and I Vogeler. Modelling nitrogen leaching from overlapping urine patches. *Environmental Modelling & Software*, 41:15–26, 2013. ISSN 1364-8152. doi:https://doi.org/10.1016/j.envsoft.2012.10.011. URL https://www.sciencedirect.com/science/article/pii/S136481521200271X.

[140] Rogerio Cichota, Iris Vogeler, Armin Werner, Kathryn Wigley, and Brittany Paton. Performance of a fertiliser management algorithm to balance yield and nitrogen losses in dairy systems. *Agricultural Systems*, 162:56–65, 2018. ISSN 0308-521X. doi:https://doi.org/10.1016/j.agsy.2018.01.017. URL `https://www.sciencedirect.com/science/article/pii/S0308521X17308715`.

[141] Andrew Tait, Roddy Henderson, Richard Turner, and Xiaogu Zheng. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. *International Journal of Climatology*, 26(14), 11 2006. ISSN 08998418. doi:10.1002/joc.1350.

[142] Darren Kriticos, Gunter Maywald, Tania Yonow, Eric Zurcher, Neville Herrmann, and Robert Sutherst. *CLIMEX. Version 4. Exploring the Effects of Climate on Plants, Animals and Diseases*. 1 2015.

[143] Neal Jean, Sherrie Wang, Anshul Samar, George Azzari, David Lobell, and Stefano Ermon. Tile2Vec: Unsupervised Representation Learning for Spatially Distributed Data. Technical report, 2018. URL `www.aaai.org`.

[144] T Zhou, M Liu, K H. Thung, and D Shen. Latent Representation Learning for Alzheimer's Disease Diagnosis With Incomplete Multi-Modality Neuroimaging and Genetic Data. *IEEE Transactions on Medical Imaging*, 38(10):2411–2422, 2019. ISSN 1558-254X. doi:10.1109/TMI.2019.2913158.

[145] Andreas Kopf and Manfred Claassen. Latent representation learning in biology and translational medicine. *Patterns*, 2(3):100198, 2021. ISSN 2666-3899. doi:https://doi.org/10.1016/j.patter.2021.100198. URL `https://www.sciencedirect.com/science/article/pii/S2666389921000015`.

[146] Gengchen Mai, Yao Xuan, Wenyun Zuo, Krzysztof Janowicz, and Ni Lao. Sphere2Vec: Multi-Scale Representation Learning over a Spherical Surface for Geospatial Predictions. 1 2022. URL `http://arxiv.org/abs/2201.10489`.

[147] Rahul Ghosh, Arvind Renganathan, Ankush Khandelwal, Xiaowei Jia, Xiang Li, John Neiber, Chris Duffy, and Vipin Kumar. Knowledge-guided Self-supervised Learning for estimating River-Basin Characteristics. 9 2021. URL `http://arxiv.org/abs/2109.06429`.

[148] Maxim Neumann, Andre Susano Pinto, Xiaohua Zhai, and Neil Houlsby. In-domain representation learning for remote sensing. 11 2019. URL `http://arxiv.org/abs/1911.06721`.

[149] Abozar Nasirahmadi and Oliver Hensel. Toward the Next Generation of Digitalization in Agriculture Based on Digital Twin Paradigm. *Sensors*, 22(2), 1 2022. ISSN 14248220. doi:10.3390/s22020498.

[150] Christos Pylianidis, Val Snow, Hiske Overweg, Sjoukje Osinga, John Kean, and Ioannis N Athanasiadis. Simulation-assisted machine learning for operational digital twins. *Environmental Modelling & Software*, 148:105274, 2022. ISSN 1364-8152. doi:https://doi.org/10.1016/j.envsoft.2021.105274. URL `https://www.sciencedirect.com/science/article/pii/S1364815221003169`.

[151] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001. ISSN 1573-0565. doi:10.1023/A:1010933404324. URL `https://doi.org/10.1023/A:1010933404324`.

[152] Lianfa Li, Ying Fang, Jun Wu, and Jinfeng Wang. Autoencoder Based Residual Deep Networks for Robust Regression Prediction and Spatiotemporal Estimation, 2018.

[153] K He, X Zhang, S Ren, and J Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. ISBN 1063-6919. doi:10.1109/CVPR.2016.90.

[154] Rob Lokers, Rob Knapen, Sander Janssen, Yke van Randen, and Jacques Jansen. Analysis of Big Data technologies for use in agro-environmental science. *Environmental Modelling & Software*, 84:494–504, 10 2016. ISSN 13648152. doi:10.1016/j.envsoft.2016.07.017.

[155] Aranildo R Lima, Alex J Cannon, and William W Hsieh. Nonlinear regression in environmental sciences using extreme learning machines: A comparative evaluation. *Environmental Modelling & Software*, 73:175–188, 11 2015. ISSN 1364-8152. doi:10.1016/j.envsoft.2015.08.002.

[156] A. T. Albert, A Rhoades, S Ganguly, D Feldman, A. D. Jones, and M Prabhat. Towards Generative Deep Learning Emulators for Fast Hydroclimate Simulations. In *AGU Fall Meeting Abstracts*, volume 2018, pages IN21C–0723, 12 2018.

[157] Theodore Weber, Austin Corotan, Brian Hutchinson, Ben Kravitz, and Robert Link. Technical note: Deep learning for creating surrogate models of precipitation in Earth system models. *Atmospheric Chemistry and Physics*, 20(4):2303–2317, 2 2020. ISSN 16807324. doi:10.5194/acp-20-2303-2020.

[158] José A Garrido Torres, Paul C Jennings, Martin H Hansen, Jacob R Boes, and Thomas Bligaard. Low-Scaling Algorithm for Nudged Elastic Band Calculations Using a Surrogate Machine Learning Model. *Phys. Rev. Lett.*, 122(15):156001, 4 2019. doi:10.1103/PhysRevLett.122.156001.

[159] Padmaja Ramankutty, Megan Ryan, Roger Lawes, Jane Speijers, and Michael Renton. Statistical emulators of a plant growth simulation model. *Climate Research*, 55(3):253–265, 2013. ISSN 16161572. doi:10.3354/cr01138.

[160] H B Mann and D R Whitney. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann. Math. Statist.*, 18(1):50–60, 10 1947. doi:10.1214/aoms/1177730491.

[161] J. N. Goetz, A. Brenning, H. Petschko, and P. Leopold. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers and Geosciences*, 81:1–11, 8 2015. ISSN 00983004. doi:10.1016/j.cageo.2015.04.007.

[162] Michele Rinaldi and Zhenli He. Chapter Six - Decision Support Systems to Manage Irrigation in Agriculture. In Donald L Sparks, editor, *Advances in Agronomy*, volume 123 of *Advances in Agronomy*, pages 229–279. Academic Press, 2014. doi:https://doi.org/10.1016/B978-0-12-420225-2.00006-6. URL `https://www.sciencedirect.com/science/article/pii/B9780124202252000066`.

[163] Daniel Anthony Howard, Zheng Ma, Jesper Mazanti Aaslyng, and Bo Nørregaard Jørgensen. Data Architecture for Digital Twin of Commercial Greenhouse Production. In *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pages 1–7, 2020. doi:10.1109/RIVF48685.2020.9140726.

[164] Natasja Ariesen-Verschuur, Cor Verdouw, and Bedir Tekinerdogan. Digital Twins in greenhouse horticulture: A review. *Computers and Electronics in Agriculture*, 199:107183, 2022. ISSN 0168-1699. doi:https://doi.org/10.1016/j.compag.2022.107183. URL `https://www.sciencedirect.com/science/article/pii/S0168169922005002`.

[165] Gordon S Blair. Digital twins of the natural environment. *Patterns*, 2(10):100359, 2021. ISSN 2666-3899. doi:https://doi.org/10.1016/j.patter.2021.100359. URL https://www.sciencedirect.com/science/article/pii/S266638992100221X.

[166] B Parkes, T P Higginbottom, K Hufkens, F Ceballos, B Kramer, and T Foster. Weather dataset choice introduces uncertainty to estimates of crop yield responses to climate variability and change. *Environmental Research Letters*, 14(12):124089, 12 2019. doi:10.1088/1748-9326/ab5ebb. URL https://dx.doi.org/10.1088/1748-9326/ab5ebb.

[167] P Schmitter, J Steinrücken, C Römer, A Ballvora, J Léon, U Rascher, and L Plümer. Unsupervised domain adaptation for early detection of drought stress in hyperspectral images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 131:65–76, 2017. ISSN 0924-2716. doi:https://doi.org/10.1016/j.isprsjprs.2017.07.003. URL https://www.sciencedirect.com/science/article/pii/S092427161730271X.

[168] Dario Gogoll, Philipp Lottes, Jan Weyler, Nik Petrinic, and Cyrill Stachniss. Unsupervised Domain Adaptation for Transferring Plant Classification Systems to New Field Environments, Crops, and Robots. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2636–2642, 2020. doi:10.1109/IROS45743.2020.9341277.

[169] Warren Purcell, Anja Klipic, and Thomas Neubauer. A Digital Twin for Grassland Management. In *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pages 1–6, 2022. doi:10.1109/ICECET55527.2022.9873446.

[170] Pelin Angin, Mohammad Hossein Anisi, Furkan Göksel, Ceren Gürsoy, and Asaf Büyükgülcü. Agrilora: A digital twin framework for smart agriculture. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 11(4):77–96, 2020. ISSN 20935382. doi:10.22667/JOWUA.2020.12.31.077.

[171] Yan Xu, Yanming Sun, Xiaolong Liu, and Yonghua Zheng. A Digital-Twin-Assisted Fault Diagnosis Using Deep Transfer Learning. *IEEE Access*, 7:19990–19999, 2019. doi:10.1109/ACCESS.2018.2890566.

[172] Kevin Voogd, Jean Pierre Allamaa, Javier Alonso-Mora, and Tong Duy Son. Reinforcement Learning from Simulation to Real World Autonomous Driving using Digital Twin, 2022. URL https://arxiv.org/abs/2211.14874.

[173] Xuan Zhou, Claudio Sbarufatti, Marco Giglio, and Leiting Dong. A Fuzzy-set-based Joint Distribution Adaptation Method for Regression and its Application to Online Damage Quantification for Structural Digital Twin, 2022. URL https://arxiv.org/abs/2211.02656.

[174] Julián Ramirez-Villegas, Charlotte Lau, Kohler A-K, Andy Jarvis, Nigel W Arnell, Tom M Osborne, and Josh Hooker. Climate analogues: Finding tomorrow's agriculture today. 2011.

[175] Sergey I Nikolenko. *Synthetic Data for Deep Learning*. Springer Cham, 1 edition, 2022. doi:https://doi.org/10.1007/978-3-030-75178-4. URL http://www.springer.com/series/7393.

[176] Yo-Whan Kim, Samarth Mishra, Souyoung Jin, Rameswar Panda, Hilde Kuehne, Leonid Karlinsky, Venkatesh Saligrama, Kate Saenko, Aude Oliva, and Rogerio Feris. How Transferable are Video Representations Based on Synthetic Data? In *36th Conference on Neural Information Processing Systems (NeurIPS 2022) Track on Datasets and Benchmarks*. Massachusetts Institute of Technology. Department of Electrical Engineering and Computer Science, 2022.

[177] Francisca Rosique, Pedro J Navarro, Carlos Fernández, and Antonio Padilla. A Systematic Review of Perception System and Simulators for Autonomous Vehicles Research. *Sensors*, 19 (3), 2019. ISSN 1424-8220. doi:10.3390/s19030648. URL `https://www.mdpi.com/1424-8220/19/3/648`.

[178] Chrystel Gaber, Baptiste Hemery, Mohammed Achemlal, Marc Pasquet, and Pascal Urien. Synthetic logs generator for fraud detection in mobile transfer services. In *2013 International Conference on Collaboration Technologies and Systems (CTS)*, pages 174–179, 2013. doi:10.1109/CTS.2013.6567225.

[179] Jingye Han, Liangsheng Shi, Christos Pylianidis, Qi Yang, and Ioannis N Athanasiadis. DeepOryza: A knowledge-guided machine learning model for rice growth simulation. In *2nd AAAI Workshop on AI for Agric3*, Washington DC, 2022. URL `https://openreview.net/forum?id=L9ankU4Ge-v`.

[180] Mark Palatucci, Dean Pomerleau, Geoffrey Hinton, and Tom M Mitchell. Zero-Shot Learning with Semantic Output Codes. In *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, NIPS'09, pages 1410–1418, Red Hook, NY, USA, 2009. Curran Associates Inc. ISBN 9781615679119.

[181] Mahdi Rezaei and Mahsa Shahidi. Zero-shot learning and its applications from autonomous vehicles to COVID-19 diagnosis: A review. *Intelligence-Based Medicine*, 3-4:100005, 2020. ISSN 2666-5212. doi:https://doi.org/10.1016/j.ibmed.2020.100005. URL `https://www.sciencedirect.com/science/article/pii/S2666521220300053`.

[182] Yongqin Xian, Christoph H Lampert, Bernt Schiele, and Zeynep Akata. Zero-Shot Learning—A Comprehensive Evaluation of the Good, the Bad and the Ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9):2251–2265, 2019. doi:10.1109/TPAMI.2018.2857768.

[183] NVIDIA. NVIDIA DRIVE Sim, 2023.

[184] Allard de Wit. Python Crop Simulation Environment - PCSE, 2023.

[185] Mathijs Jeroen Scheepers. Virtualization and Containerization of Application Infrastructure : A Comparison. 2014.

Summary

Agricultural applications produce data at a fast rate and decision support systems are called to extract insights from them. However, current agricultural decision support systems rely on static models, they are task specific, and they lack automation. This hinders a more advanced type of decision support that modern agricultural systems require. At the same time, the paradigm of digital twins is becoming more prominent in other disciplines. Digital twins seem to be able to overcome the aforementioned limitations and offer benefits that have yet to be realized in agriculture. Despite their success in other disciplines, the potential of digital twins has not been actualized in agriculture. In this thesis we investigate ways to operationalize digital twins, first by enabling them to make predictions when the data are insufficient in amount or temporal resolution, and second by considering their ability to be transferred to diverse conditions.

In chapter 1, we argue about the appeal of digital twins for agricultural applications and describe the challenges that inhibit their adoption. In chapter 2, we review the state of digital twins in agriculture through a literature review. We record technology readiness level, benefits, and types of services they provide in other disciplines, and compare them with their applications in agriculture. Based on the benefits and services provided to other disciplines, we propose a roadmap for their advancement in agriculture.

In chapter 3, we focus on enabling digital twins to provide decision support in situations where data are not in sufficient amounts, or they are in different temporal resolutions from what existing machine learning and process-based models expect. We propose a methodology based on metamodeling to allow making predictions in such cases. We then showcase it, with a case study of pasture nitrogen response rate. We train machine learning metamodels on simulated data produced by APSIM and evaluate them in different conditions with a domain specific error threshold.

For the experiments of chapter 3, we used a single machine learning algorithm to make predictions, Random Forest. Different applications have different needs in terms of data quantity, computational resources, interpretability

and others. Consequently, it is important to examine the generality of chapter's 3 methodology. In chapter 4, we revisit the nitrogen response rate case study to assess whether the proposed metamodeling methodology is algorithm independent. We transition to neural networks and compare a multilayer perceptron, and two variations of autoencoders for predicting nitrogen response rate.

In chapter 5, we examine the transferability of digital twins between locations with diverse conditions. We examine how data variability affects the results when no data from the target locations are available. We develop a machine learning metamodel with training data from a single location, and another one with data from multiple locations. The two models are then evaluated in each location of our dataset iteratively using common error metrics and a statistical test to indicate whether differences in the distributions of their predictions are different.

In contrast to chapter 5, chapter 6 regards the case where sparse data are available from the target locations. In this chapter, we examine the transferability of digital twins by performing domain adaptation with transfer learning. The locations between which we transfer the models are chosen based on a climate similarity tool. We investigate how the amount of weather data and variety of agro-management practices included in the training set of the pretrained models affect adaptation.

Acknowledgments

Back in 2019 I was skeptical about how much my hairline would recede during the course of the PhD. My hair were thick and nice, academics in the deep were flabbergasted. Fast forward to 2023, my hairline is still intact. It even got augmented with white details to reflect light at night keeping me safe when I cycle. This wouldn't be possible without you (to still be great, not the white hair).

Ioannis and **Sjoukje**, I was lucky to have you as supervisors. With you I acquired much more than I expected to be in a PhD.

Bedir, you helped me navigate through the wureacracy.

Dilli, you made the daily office routine more interesting. You kept me in shape with intense ping pong matches and rested with great siestas.

Joep, **Joost**, **Martin**, you taught me about the 30cm (~3 hands for imperial system users) frikandel. You also taught me more about ice-skating, haribo, and nitrogen.

Michiel, **Hilmy**, **Ron**, I always had fun passing from your offices. **Jingye**, I enjoyed discussing with you for different topics, you also made me a pro in mass conservation. **Hiske**, it was nice having you here and trying out different games.

Claudia, **Marieke**, it has been relaxing talking with you during lunch breaks and fun playing ping pong variations.

Matthew, **Maike**, **Matteo**, **Soodeh**, **Mar**, **Sergio**, **Laura**, **Gonzalo**, **Sander** it was always a pleasure taking breaks in your offices and talking with you.

Marco, **Ante**, **Carlos**, you are here because I feel generous.

Theofile, **Nhung**, you too.

There were many other people with whom we worked, had fun, and spent time together. Thank you all!

Publications

Peer-reviewed journal publications

Christos Pylianidis, Val Snow, Hiske Overweg, Sjoukje Osinga, John Kean, Ioannis N. Athanasiadis. Simulation-assisted machine learning for operational digital twins, Environmental Modelling & Software, vol 148, pp 105274, 2022

Christos Pylianidis, Sjoukje Osinga, Ioannis N. Athanasiadis. Introducing digital twins to agriculture, Computers and Electronics in Agriculture, vol 184, pp 105942, 2021

Dilli Paudel, Hendrik Boogaard, Allard de Wit, Sander Janssen, Sjoukje Osinga, **Christos Pylianidis**, Ioannis N. Athanasiadis. Machine learning for large-scale crop yield forecasting, Agricultural Systems, vol 187, pp 103016, 2021

Peer-reviewed book chapter

Christos Pylianidis, Ioannis N. Athanasiadis. Location-Specific vs Location-Agnostic Machine Learning Metamodels for Predicting Pasture Nitrogen Response Rate, International Conference on Pattern Recognition International Workshops and Challenges, vol 12666, pp 45-54, 2021

Peer-reviewed conference publications

Jingye Han, Liangsheng Shi, **Christos Pylianidis**, Qi Yang, Ioannis N. Athanasiadis. DeepOryza: A Knowledge guided machine learning model for rice growth simulation. 2nd AAAI Workshop on AI for Agriculture and Food Systems, 2023

Christos Pylianidis, Ioannis N. Athanasiadis. Learning latent representations for operational nitrogen response rate prediction, AI for Earth Sciences workshop at ICLR2022, 2022

Abstract

Bernardo Maestrini, **Christos Pylianidis**, Janne Kool, Keiji Jindo, Dilli Paudel, Ioannis Athanasiadis, Frits K Van Evert. CSSA and SSSA International Annual Meetings, ASA, 2020

Overview of completed training activities

Name of the learning activity	Department/Institute	Year	ECTS*
A) Project related competences			
A1 Managing a research project			
WASS Introduction Course	WASS	2019	1
Writing the research proposal	WUR	2019	6
Scientific writing	Wageningen in'to Languages	2019	1.8
Dragon project deliverable about the status of the project in 2019	Dragon	2019	0.5
'Introducing digital twins for cropping systems'	WUR data science week	2020	1
Presenting with impact	Wageningen in'to Languages	2021	1
'Introducing digital twins to agriculture' *'Machine learning surrogates for crop models'*	International Environmental Modelling and Software Society conference	2021	1
A2 Integrating research in the corresponding discipline			
3rd International Summer School on Deep Learning Warsaw	IRDTA	2019	2
FTE-35306 Machine Learning	WUR	2019	6
Summer school in Serbia for Machine Learning in agriculture (November)	BIOS	2019	0.5
Workshop Good practises for high-performance computing and cloud	SURF	2019	1
Machine Learning Advances Environmental Science workshop	International Conference on Pattern Recognition, Milan, Italy	2020	1
AI for Earth Sciences workshop	International Conference on Learning Representations, La Jolla, USA	2022	1
Model management, model governance and modelling of modelling for water policy and beyond workshop	International Environmental Modelling and Software Society conference, Brussels, Belgium	2022	1
B) General research related competences			
B1 Placing research in a broader scientific context			
Financial markets	Coursera, Yale university	2020	3
Economics of money and banking	Coursera, Columbia university	2021	3
B2 Placing research in a societal context			
Helped to develop a MOOC regarding Big Data in the agri-food domain in Edx platform	Edx	2022	2
C Career related competences/personal development			
C1 Employing transferable skills in different domains/careers			
Teaching assistant in Big Data course	WUR	2019, 2020	3.5
Lecturer in a summer school Serbia as part of the Dragon project	BIOS	2019	0.5
Total			**36.8**

*One credit according to ECTS is on average equivalent to 28 hours of study load

Colophon

This research was financially supported by the European Union Horizon 2020 Research and Innovation programme (Grant #810775, Dragon).

Cover designed by Christos Pylianidis
Thesis layout by Christos Pylianidis
Printed by ProefschriftMaken