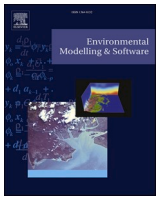
Environmental Modelling and Software 148 (2022) 105274

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Simulation-assisted machine learning for operational digital twins

Christos Pylianidis [a,*], Val Snow [b], Hiske Overweg [a], Sjoukje Osinga [a], John Kean [b], Ioannis N. Athanasiadis [a]

[a] Wageningen University and Research, Wageningen, the Netherlands
[b] AgResearch, Lincoln, New Zealand

ARTICLE INFO

Keywords:
Machine learning
Digital twin
Data availability
Data resolution
APSIM
Metamodel

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.

1. Introduction

Digital twins are established in several industries, including manufacturing (He and Bai, 2021), healthcare (Liu et al., 2019), automotive (Caputo et al., 2019). 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 (Pedersen et al., 2021), agriculture (Pylianidis et al., 2021), smart farming (Verdouw et al., 2021), livestock farming (Neethirajan and Kemp, 2021), remote sensing (Nativi et al., 2021) and earth sciences (Guo et al., 2020). Recently, the European Union has announced plans for a high-resolution Earth digital twin that aims at actionable intelligence from (big) data streams (Bauer et al., 2021; Voosen, 2020). In the US, the research agenda for intelligent systems in geosciences (Gil et al., 2018) 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.

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 machine learning (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

* Corresponding author.
E-mail addresses: christos.pylianidis@wur.nl (C. Pylianidis), val.snow@agresearch.co.nz (V. Snow), hiske.overweg@wur.nl (H. Overweg), sjoukje.osinga@wur.nl (S. Osinga), john.kean@agresearch.co.nz (J. Kean), ioannis.athanasiadis@wur.nl (I.N. Athanasiadis).

https://doi.org/10.1016/j.envsoft.2021.105274
Received 16 July 2021; Received in revised form 3 December 2021; Accepted 7 December 2021
Available online 14 December 2021

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 (Yang and Jiachen, 2021) seem promising to learn from small datasets, but still novel research is needed to develop ML approaches that incorporate prior knowledge about environmental processes (Karpatne et al., 2017a) and use it to effectively supplement the available data (Gil et al., 2019). 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 Gil et al. (2019).

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 rates (NRR) and examine their reliability by assessing their predictive and generalization capacity.

The rest of the paper is organized as follows: in section 2 we describe the requirements of PMBs, the proposed method and related work. In section 3, we present the case study and the methodology to experimentally evaluate the proposed method. Section 4 reports the results of our experiments, followed by a discussion in section 5, and the conclusion (in section 6).

2. Simulation-assisted machine learning

2.1. Process-based model data requirements

PBMs typically require several high-resolution data streams as inputs to simulations (Julie Ramanantenasoa et al., 2019; Kasampalis et al., 2018). 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 (Togliatti et al., 2017). 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 (Vogeler et al., 2013), and often together with data assimilation techniques to integrate them with sensor observations of system states (Dorigo et al., 2007).

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 (Cichota et al., 2008). 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. 1a.

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. 1b.

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 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 (Antony, 2014) 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.

2.4. Related work

Efforts to overcome the inherent shortcomings of PBMs for operational decision-making have been focused on combining PBMs with ML

[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.

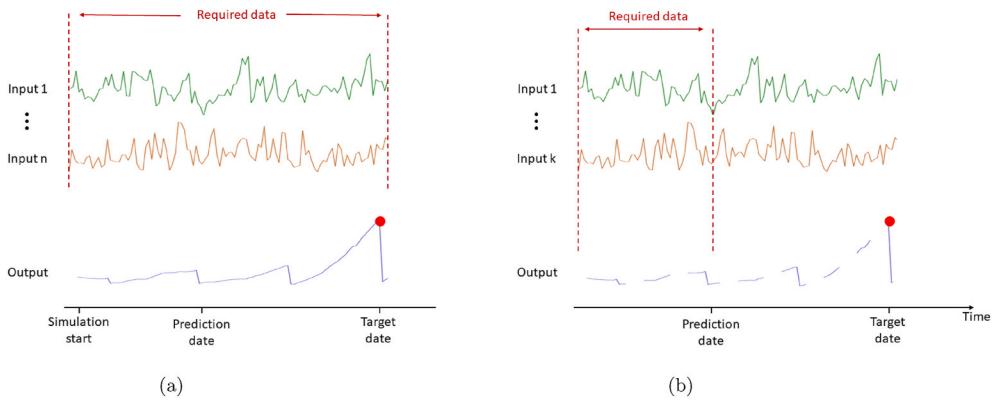


Fig. 1. Data requirements of PBMs (a) and our approach (b). The PBM needs n inputs that span through the entire duration of the simulation to produce an output at the end of the simulation. The model of our approach requires a subset k, $k < n$, of the PBM inputs. The required data are limited to what 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.

through the concept of metamodeling. Metamodels (also called surrogate or hybrid models) refer to models which mimic the behavior of other models (Blanning, 1975). 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 (Karpatne et al., 2017b) 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 (Shahhosseini et al., 2021; Roberts et al., 2017). To accelerate sensitivity analysis, metamodels have been trained on the output of agricultural simulators (Gladish et al., 2019). Also, hydrological metamodels have been evaluated for their performance in terms of speed and accuracy (Zhang et al., 2020; Villa-Vialaneix et al., 2012), as well as generalization capacity (domain adaptation) in unsampled areas (Nolan et al., 2018). Likewise, to extrapolate at regional and national levels, metamodels have been deployed in environmental management (Ramanantenasoa et al., 2019). Lastly, to work in situations where PBM inputs are not available, the authors of (Shahhosseini et al., 2019) 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. Methods

3.1. Overview

To assess the method described in 2.3 we performed a case study of grass pasture NRR prediction in different locations (see Fig. 2) of New Zealand. The application of nitrogen along with environmental factors such as temperature and time of year greatly affects pasture growth (Gillingham et al., 2008), 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.

Fig. 2. The eight locations included in the generated dataset.

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 (Rotz et al., 2005; Whitehead, 1995) and farmers apply nitrogen-containing fertilizers to increase pasture growth rates (Clark et al., 2007; Pembleton et al.,

2013). 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 (Waikato Regional Council, 2015; NZ farm source, 2021; DairyNZ, 2012).

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. Data generation

We used APSIM v7.10 r4191 (APSIM, 2021; Holzworth et al., 2014) to generate the training and testing data. Pasture growth was simulated with the AgPasture module (Li et al., 2011) which has already been demonstrated to be a reasonable estimator of pasture growth in New Zealand (Cichota et al., 2013, 2018). The range of input conditions covered eight contrasting locations in New Zealand (Fig. 2) and are given in Table 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 (Tait et al., 2006; Cichota et al., 2008), 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 1), 1,382,400 remained.

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 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 300 mm, soil temperature at 300 mm, soil temperature at 50 mm, 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.

Table 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

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 (Kriticos et al., 2015). 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 (Tait et al., 2006) 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 for a visualization of training models of different size, and Table A1 in the appendix for the locations included in each model.

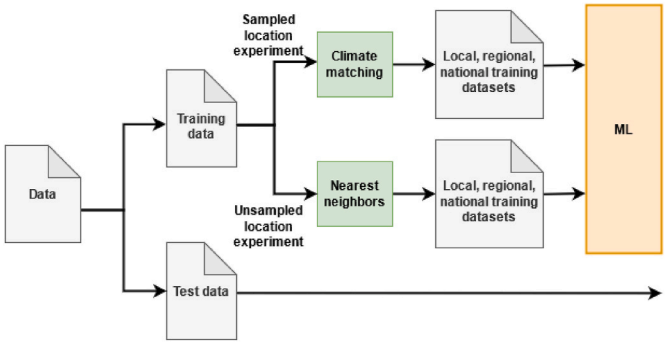


Fig. 3. The splitting of the processed data to create models for testing in sampled and unsampled location experiments.

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 B1 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 2.

3.7. Evaluation

The predictive capacity of the models was evaluated using the root mean squared error (RMSE) 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.

3.8. Experimental setup

The data preprocessing stage was carried out utilizing the *Apache Spark* framework in standalone mode. The machine learning 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 120 GB of RAM.

3.9. Software availability

The code used for the case study of this paper can be found in https://github.com/BigDataWUR/simulation-assisted-ML.

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 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.

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.

Prediction residuals are illustrated in Fig. 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. C1). 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. 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. C3) we observed that the years 2015, 2016 had larger residuals in several locations.

Table 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

[2] kg of dry matter/ha/kg of nitrogen.

4.2. Unsampled location experiment

In the unsampled location experiment (Fig. 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). 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. C2) 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 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. C4), 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.

5. Discussion

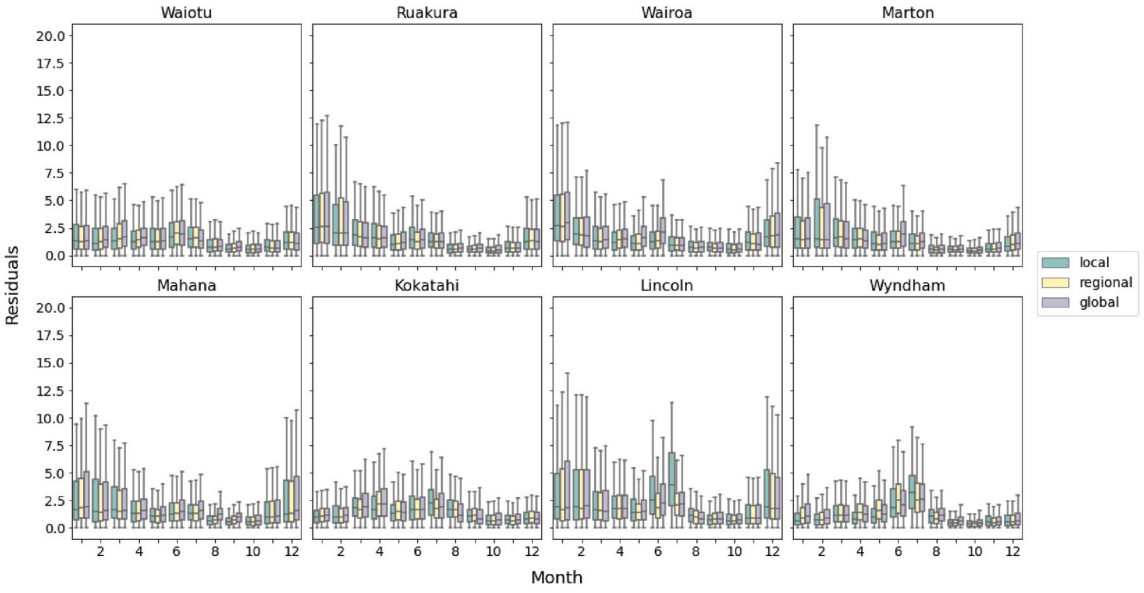
In our experiments, the models captured in most cases sufficient variation from the data to achieve RMSEs lower than the threshold of 5 kgDM/ha/KgN. This means that they could be potentially used in practical applications where weather data after fertilization are missing, or data are on a lower resolution 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.

5.1. Predictive capacity (sampled locations)

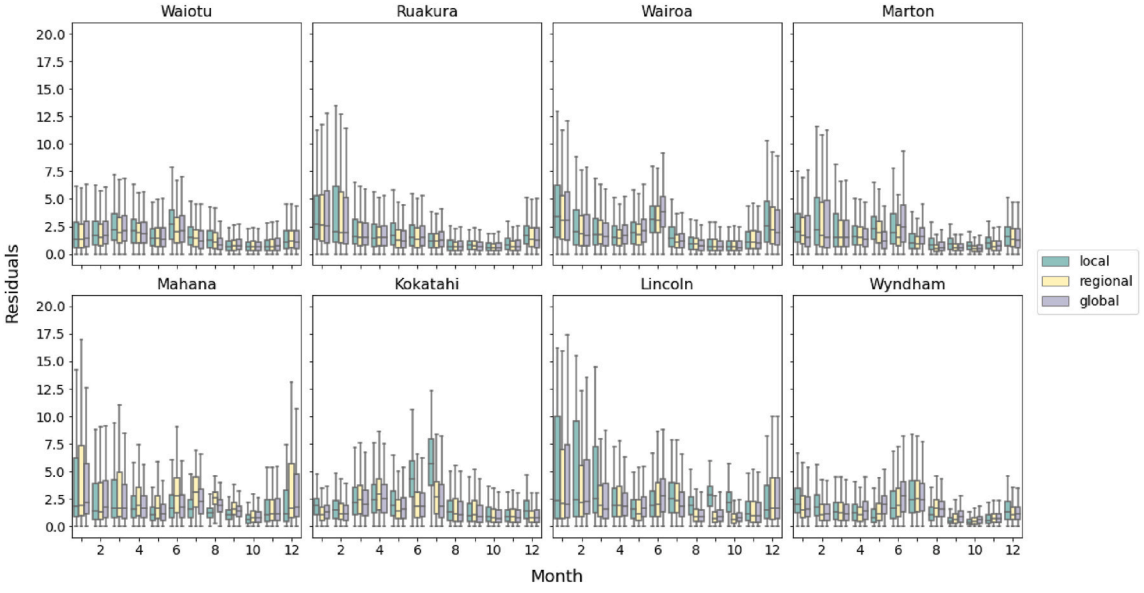
When separately analyzing the irrigated/non-irrigated cases, we observed that the lack of irrigation hindered the predictive capacity of the models. The reason for this impediment is that when no irrigation is provided, the weather conditions become the driving factor of the NRR, because the grass relies 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

Table 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

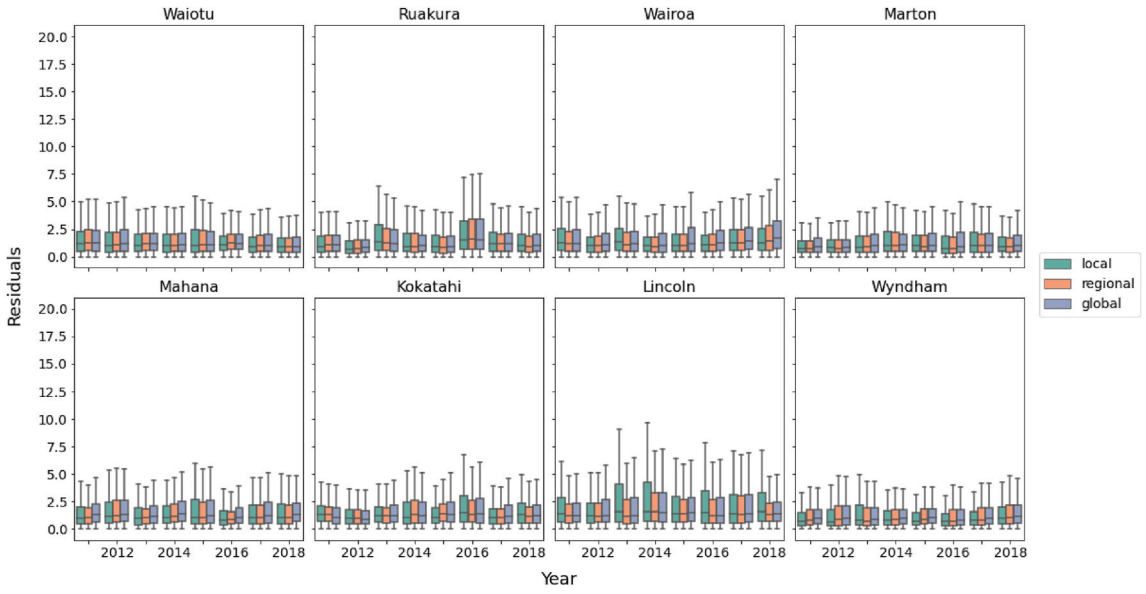


(a)

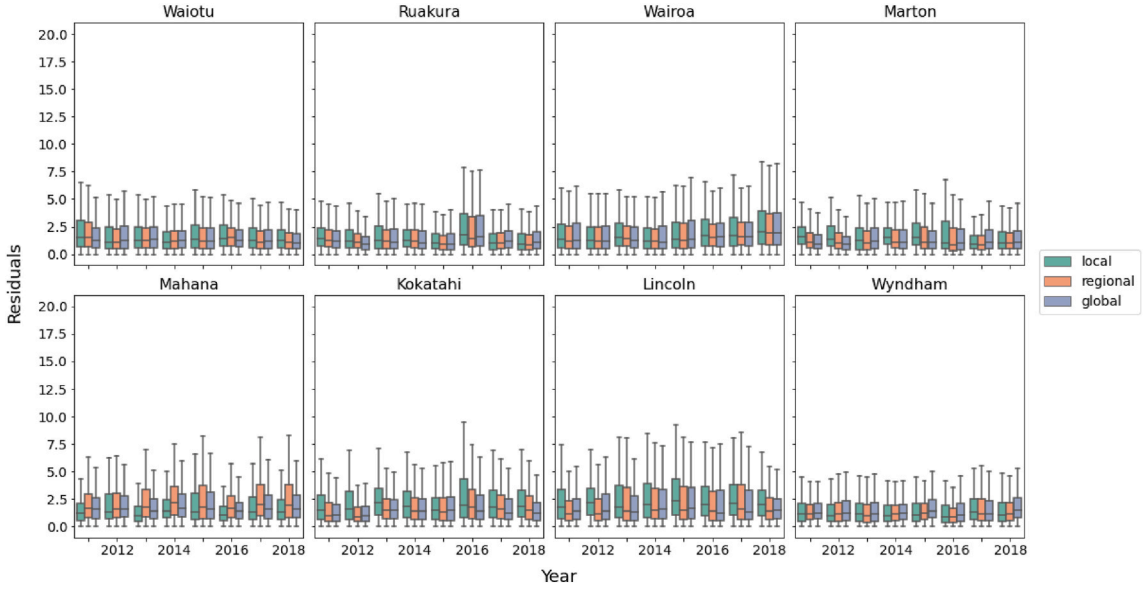


(b)

Fig. 4. Monthly test set residuals of models for **sampled** (a) and **unsampled** (b) location experiments.



(a)



(b)

Fig. 5. Yearly test set residuals of models for **sampled** and (a) and **unsampled** location experiments (b).

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. D1.

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.

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. C2). 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. C2a) 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. D2) 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. C2a) 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.

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.

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.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Given his role as Associate Editor of Environmental Modelling & Software, I.N. Athanasiadis was not involved in the peer-review of this article and had no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to the Editor-in-Chief D.P.Ames.

Acknowledgements

Funding: This work has been partially supported by the European Union Horizon 2020 Research and Innovation program (Grant #810 775, "Dragon"); the Wageningen University and Research Investment Program "Digital Twins" and AgResearch Strategic Science Investment Fund (SSIF) "Emulation of pasture growth response to nitrogen application".

Appendices.

ALocations included in each model

Table A.1
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.

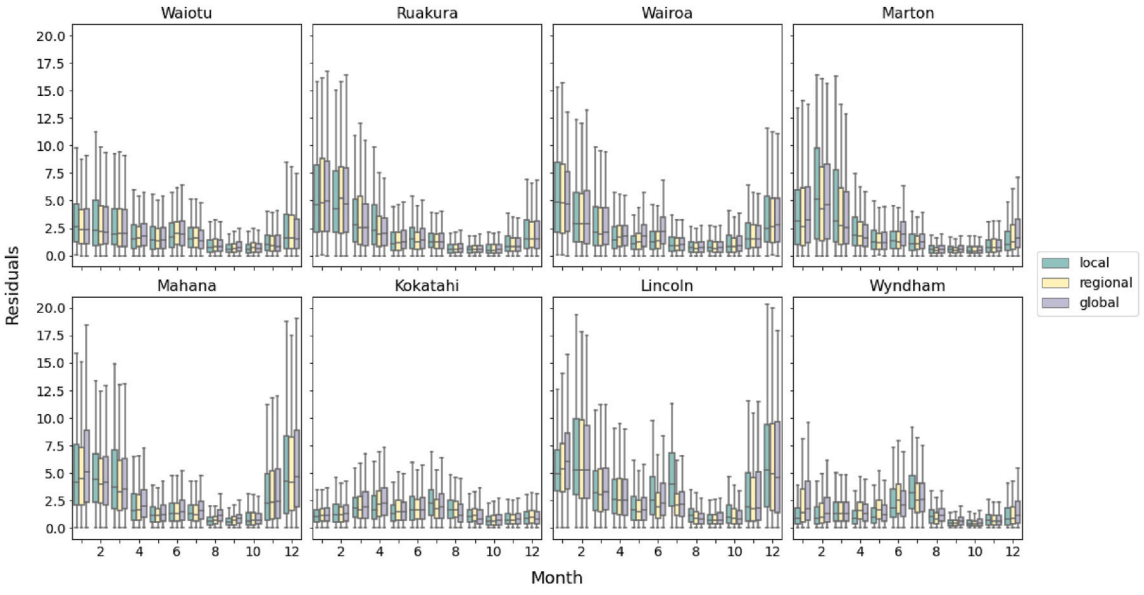
Target location	Scenario					
	Sampled location			Unsampled location		
	Local	Regional	Global	Local	Regional	Global
Waiotu	Waiotu	Waiotu, Wairoa, Ruakura,	all	Ruakura	Ruakura, Wairoa, Marton	all except Waiotu
Ruakura	Ruakura	Ruakura, Marton, Wairoa,	all	Wairoa	Wairoa, Marton, Waiotu	all except Ruakura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

BPreliminary machine learning algorithm comparison

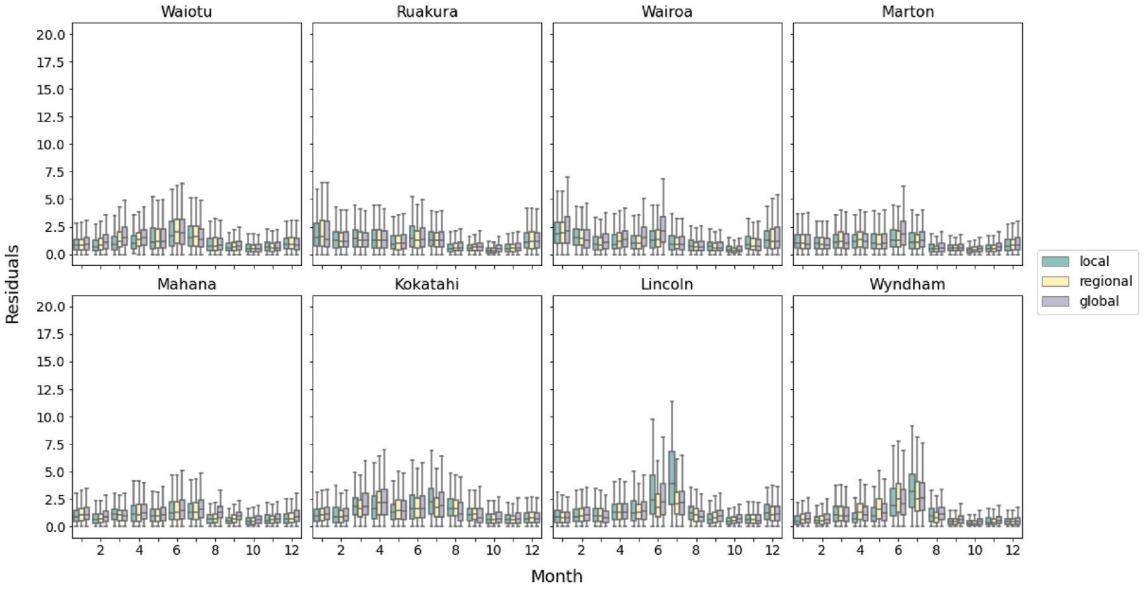
Table B.1
The gridsearch and RMSE results of different machine learning algorithms for training in Ruakura and testing in Waiotu, with the yearly split mentioned in the text, as a preliminary test to choose an algorithm. The gridsearch parameters as denoted as found in *scikit-learn*'s documentation. 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

CRainfed vs irrigated plots

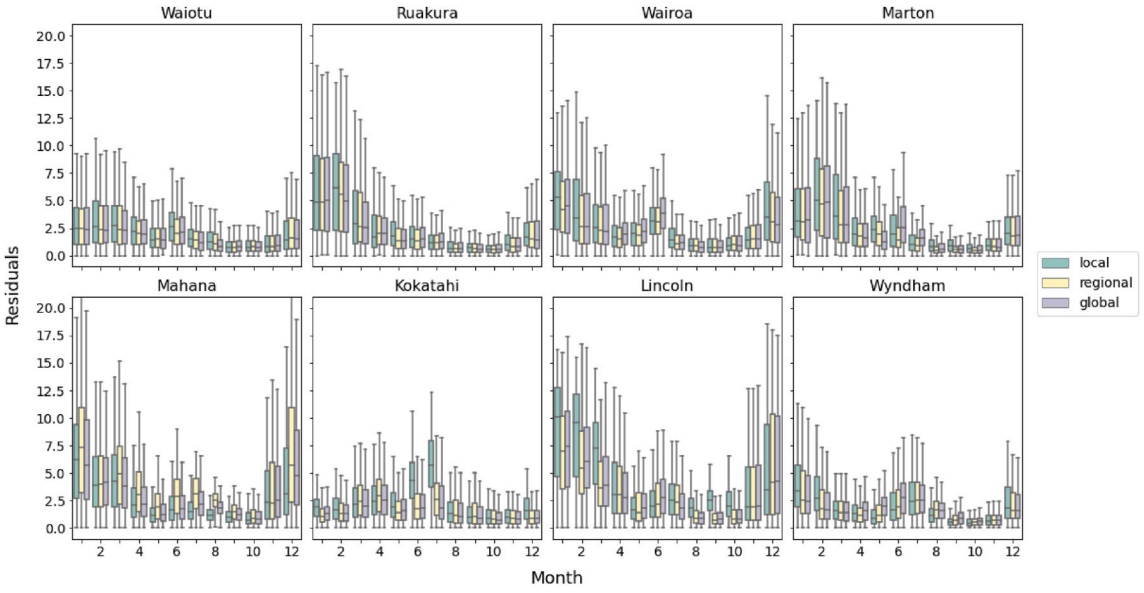


(a)

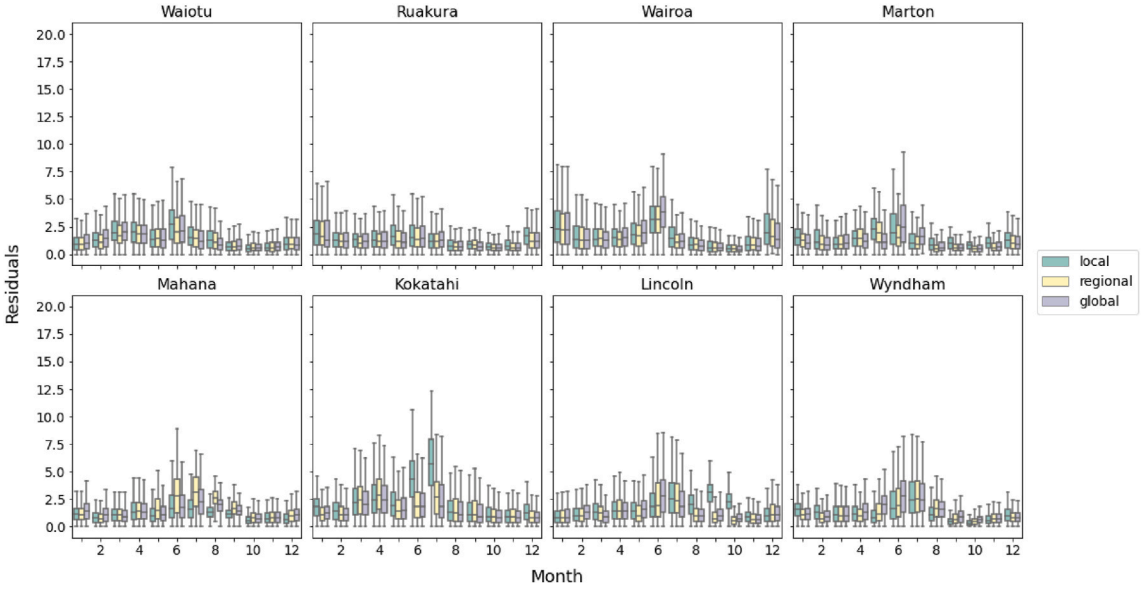


(b)

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

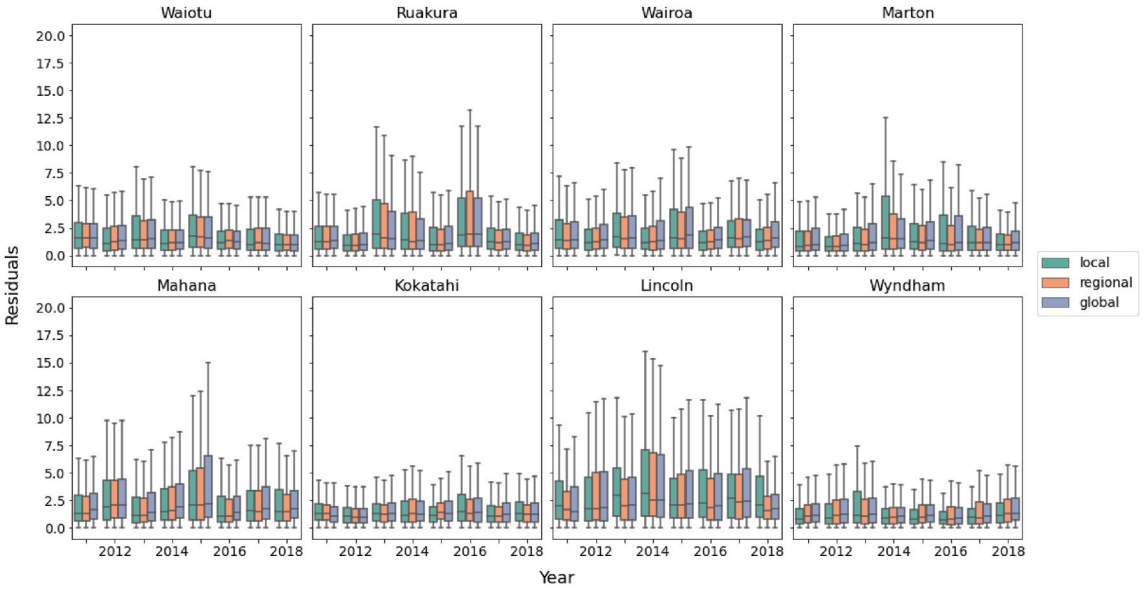


(a)

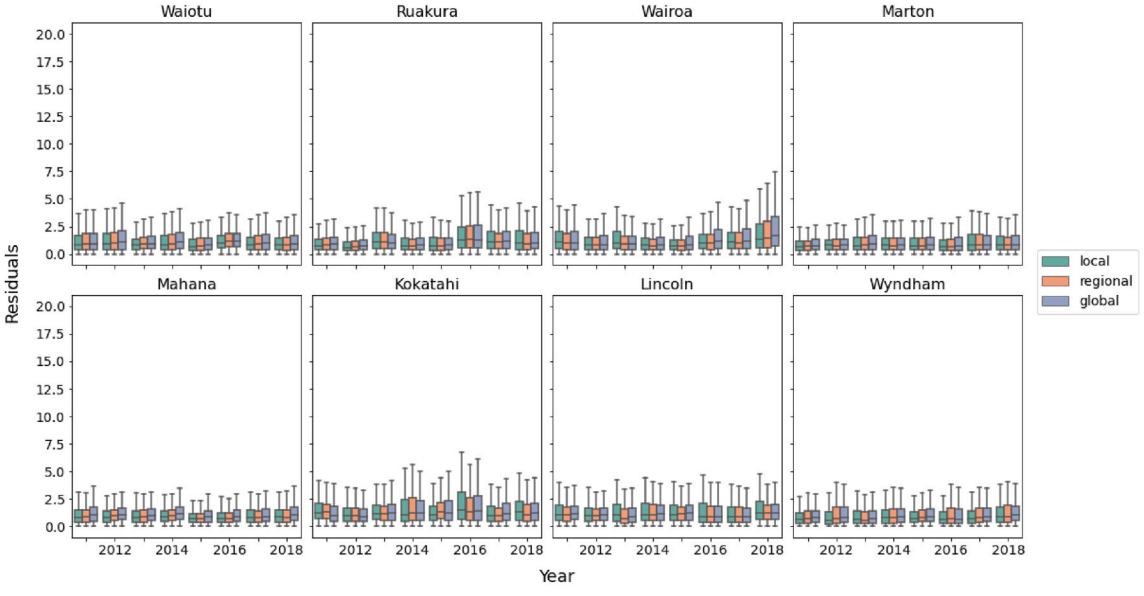


(b)

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

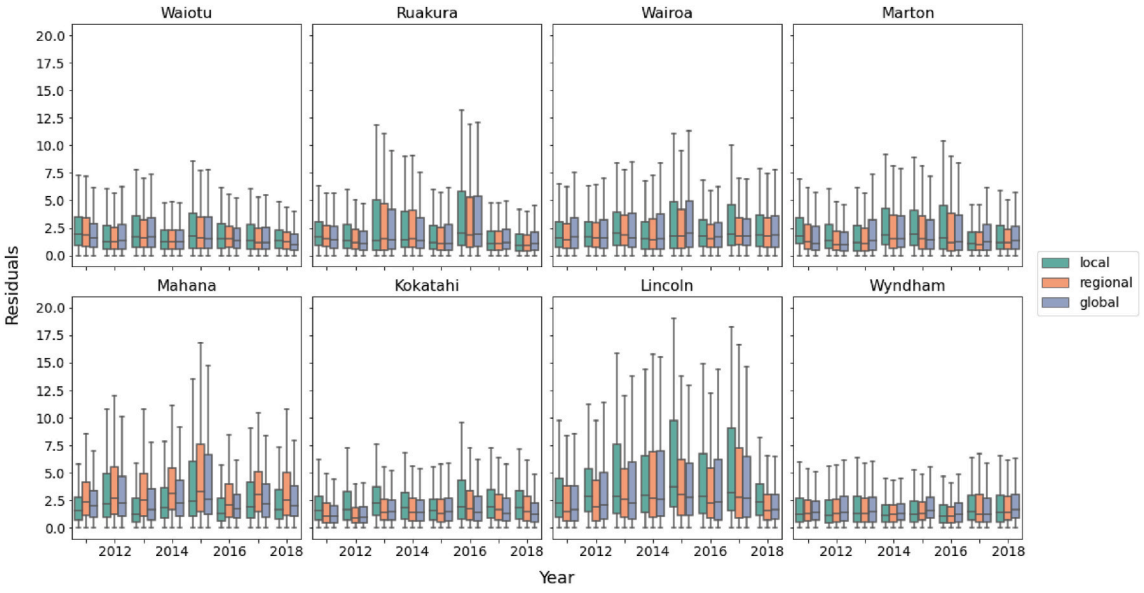


(a)

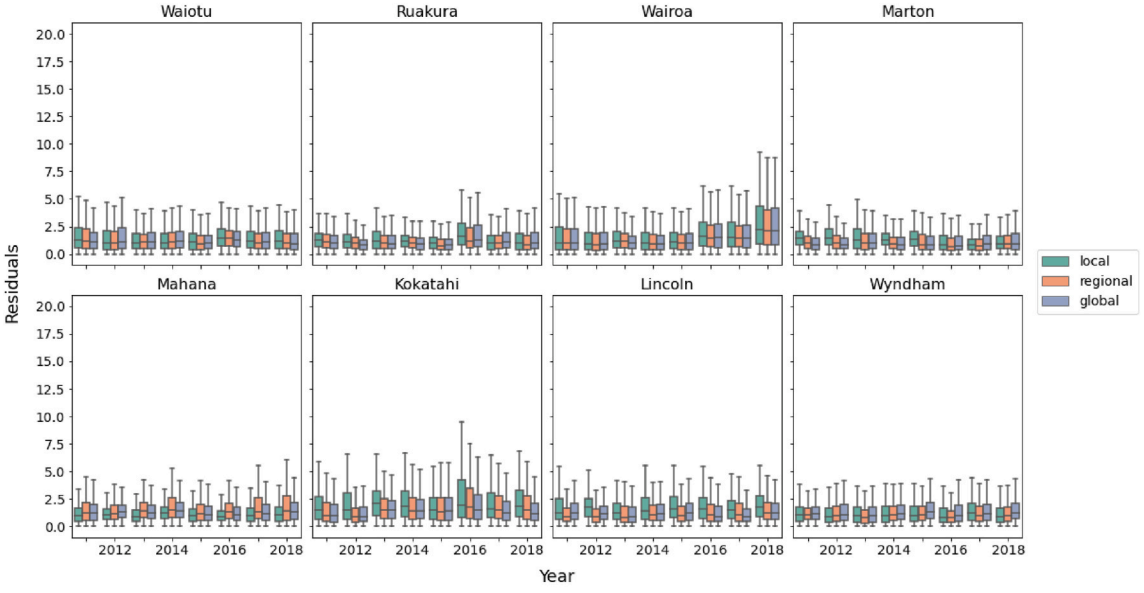


(b)

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



(a)



(b)

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

D. Weather plots

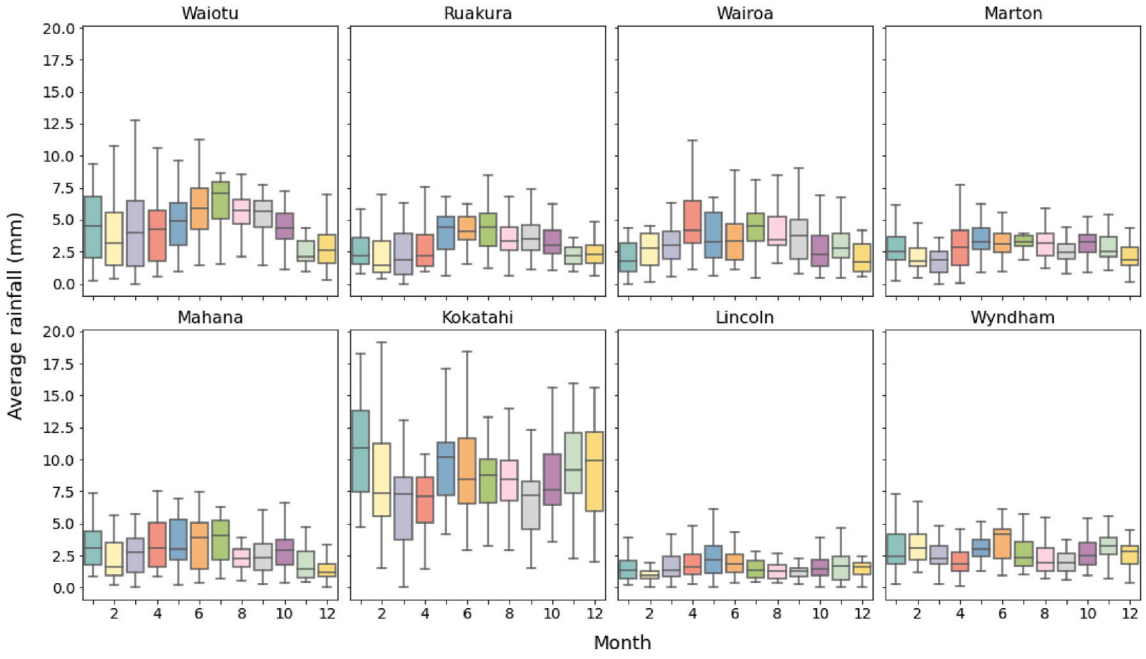


Fig. D.1. Average rainfall per month and location for the four weeks that we assume to have data.

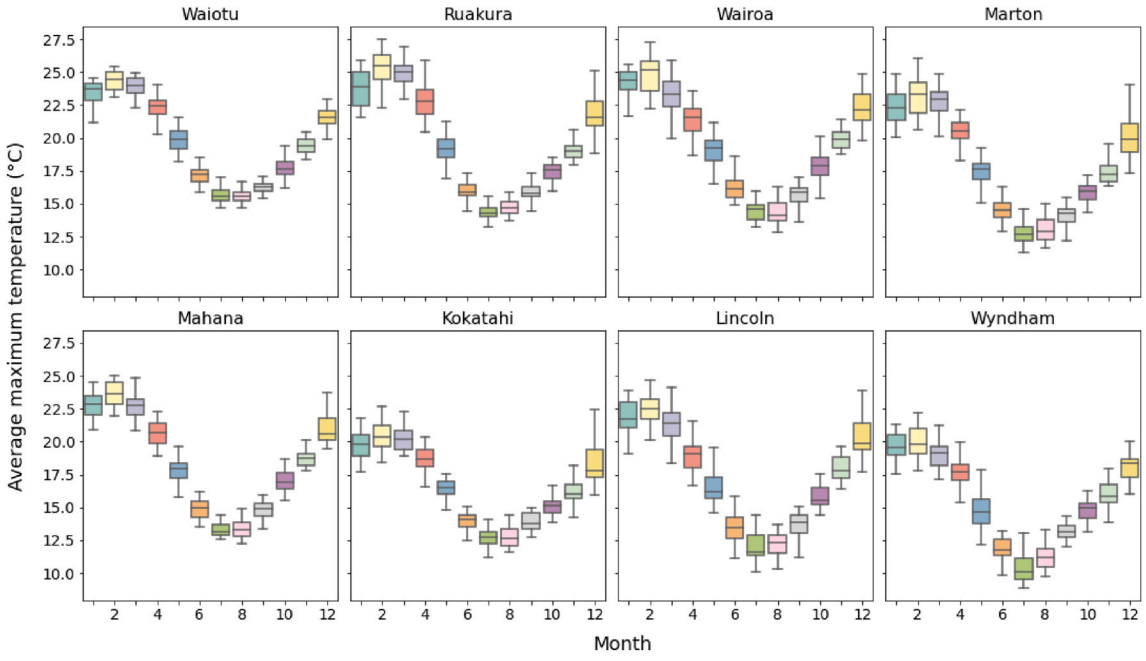


Fig. D.2. Average maximum temperature per month and location for the four weeks that we assume to have data.

References

Antony, Jiju, 2014. 6 - full factorial designs. In: Antony, Jiju (Ed.), Design of Experiments for Engineers and Scientists, second ed. Elsevier, Oxford, ISBN 978-0-08-099417-8, pp. 63–85. https://doi.org/10.1016/B978-0-08-099417-8.00006-7 https://www.sciencedirect.com/science/article/pii/B9780080994178000067.

APSIM. https://www.apsim.info/.

Bauer, Peter, Stevens, Bjorn, Hazeleger, Wilco, 2021. A digital twin of Earth for the green transition. Nat. Clim. Change 11 (2), 80–83. https://doi.org/10.1038/s41558-021-00986-y. ISSN 1758–6798.

Blanning, Robert W., 1975. The construction and implementation of metamodels. Simulation 24 (6), 177–184. https://doi.org/10.1177/003754977502400606.

Caputo, F., Greco, A., Fera, M., Macchiaroli, R., 2019. Digital twins to enhance the integration of ergonomics in the workplace design. Int. J. Ind. Ergon. 71, 20–31. https://doi.org/10.1016/j.ergon.2019.02.001. ISSN 18728219.

Cichota, R., Snow, V.O., Tait, A.B., 2008. A functional evaluation of virtual climate station rainfall data. N. Z. J. Agric. Res. 51 (3), 317–329. https://doi.org/10.1080/00288230809510463. ISSN 0028–8233.

Cichota, R., Snow, V.O., Vogeler, I., 2013. Modelling nitrogen leaching from overlapping urine patches. Environ. Model. Software 41 (15–26). https://doi.org/10.1016/j.envsoft.2012.10.011. ISSN 1364–8152. https://www.sciencedirect.com/science/article/pii/S136481521200271X.

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

Clark, D.A., Caradus, J.R., Monaghan, R.M., Sharp, P., Thorrold, B.S., 2007. Issues and options for future dairy farming in New Zealand. N. Z. J. Agric. Res. 50 (2), 203–221. https://doi.org/10.1080/00288230709510291. ISSN 0028–8233.

DairyNZ, 2012. Farmfact: Seasonal Nitrogen Use. Technical report, 7. https://www.dairynz.co.nz/publications/farmfacts/fertiliser-and-nutrient-management/farmfact-7-11/.

Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. Int. J. Appl. Earth Obs. Geoinf. 9 (2), 165–193. https://doi.org/10.1016/j.jag.2006.05.003. ISSN 0303–2434. https://www.sciencedirect.com/science/article/pii/S0303243406000201.

Gil, Yolanda, Pierce, Suzanne A., Hassan, Babaie, Banerjee, Arindam, Kirk, Borne, Gary, Bust, Cheatham, Michelle, Ebert-Uphoff, Imme, Gomes, Carla, Hill, Mary, Horel, John, Hsu, Leslie, Jim Kinter, Craig, Knoblock, Krum, David, Kumar, Vipin, Lermusiaux, Pierre, Liu, Yan, North, Chris, Pankratius, Victor, Peters, Shanan, Plale, Beth, Allen, Pope, Ravela, Sai, Juan Restrepo, Ridley, Aaron, Samet, Hanan, Shekhar, Shashi, Skinner, Katie, Smyth, Padhraic, Tikoff, Basil, Yarmey, Lynn, Zhang, Jia, 2018. Intelligent systems for geosciences: an essential research agenda. Commun. ACM 62 (1), 76–84. https://doi.org/10.1145/3192335. ISSN 0001–0782.

Gil, Yolanda, Suzanne, A., Pierce, Hassan Babaie, Banerjee, Arindam, Kirk, Borne, Gary, Bust, Cheatham, Michelle, Ebert-Uphoff, Imme, Gomes, Carla, Hill, Mary, Horel, John, Hsu, Leslie, Kinter, Jim, Knoblock, Craig, Krum, David, Kumar, Vipin, Lermusiaux, Pierre, Liu, Yan, North, Chris, Pankratius, Victor, Peters, Shanan, Plale, Beth, Pope, Allen, Ravela, Sai, Restrepo, Juan, Ridley, Aaron, Samet, Hanan, Shekhar, Shashi, 2019. Intelligent systems for geosciences: an essential research agenda. Commun. ACM 62 (1), 76–84. https://doi.org/10.1145/3192335. ISSN 15577317.

Gillingham, A.G., Morton, J.D., Gray, M.H., 2008. Pasture responses to phosphorus and nitrogen fertilisers on east coast hill country: 2. Clover and grass production from easy slopes. N. Z. J. Agric. Res. 51 (2), 85–97. https://doi.org/10.1080/00288230809510438.

Gladish, Daniel W., Darnell, Ross, Thorburn, Peter J., Haldankar, Bhakti, 2019. Emulated multivariate global sensitivity analysis for complex computer models applied to agricultural simulators. J. Agric. Biol. Environ. Stat. 24 (1), 130–153. https://doi.org/10.1007/s13253-018-00346-y. ISSN 15372693.

Guo, Huadong, Nativi, Stefano, Dong, Liang, Craglia, Max, Wang, Lizhe, Sven Schade, Corban, Christina, He, Guojin, Pesaresi, Martino, Li, Jianhui, Zeeshan Shirazi, Liu, Jie, Annoni, Alessandro, 2020. Big Earth Data science: an information framework for a sustainable planet. Int. J. Digit. Earth 13 (7), 743–767. https://doi.org/10.1080/17538947.2020.1743785. ISSN 1753–8947.

He, Bin, Bai, Kai Jian, 2021. Digital twin-based sustainable intelligent manufacturing: a review. Adv. Manuf. 9 (1) https://doi.org/10.1007/s40436-020-00302-5. ISSN 21953597.

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

Julie Ramanantenasoa, Maharavo Marie, Génermont, Sophie, Gilliot, Jean Marc, Bedos, Carole, Makowski, David, 2019. Meta-modeling methods for estimating ammonia volatilization from nitrogen fertilizer and manure applications. J. Environ. Manag. 236, 195–205. https://doi.org/10.1016/j.jenvman.2019.01.066. ISSN 10958630.

Karpatne, Anuj, Atluri, Gowtham, Faghmous, James H., Michael Steinbach, Banerjee, Arindam, Ganguly, Auroop, Shekhar, Shashi, Samatova, Nagiza, Kumar, Vipin, 2017a. Theory-Guided data science: a new paradigm for Scientific discovery from data. IEEE Trans. Knowl. Data Eng. 29 (10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168. ISSN 1041–4347. http://ieeexplore.ieee.org/document/7959606/.

Karpatne, Anuj, Watkins, William, Read, Jordan, Kumar, Vipin, 2017b. Physics-guided neural networks (PGNN): an application in lake temperature modeling. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, vol. 10. http://arxiv.org/abs/1710.11431.

Kasampalis, Dimitrios A., Alexandridis, Thomas K., Deva, Chetan, Challinor, Andrew, Moshou, Dimitrios, Zalidis, Georgios, 2018. Contribution of Remote Sensing on Crop Models: A Review. J. Imag. 4 (4), 52. https://doi.org/10.3390/jimaging4040052.

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

Li, F.Y., Snow, V.O., Holzworth, D.P., 2011. Modelling the seasonal and geographical pattern of pasture production in New Zealand. N. Z. J. Agric. Res. 54 (4), 331–352. https://doi.org/10.1080/00288233.2011.613403. ISSN 288233.

Liu, Ying, Zhang, Lin, Yuan, Yang, Zhou, Longfei, Ren, Lei, Wang, Fei, Liu, Rong, Pang, Zhibo, Jamal Deen, M., 2019. A novel cloud-based framework for the elderly healthcare Services using digital twin. IEEE Access 7, 49088–49101. https://doi.org/10.1109/ACCESS.2019.2909828. ISSN 21693536.

Nativi, Stefano, Mazzetti, Paolo, Craglia, Max, 2021. Digital ecosystems for developing digital twins of the earth: the destination earth case. Rem. Sens. 13 (11) https://doi.org/10.3390/rs13112119. ISSN 20724292.

Neethirajan, Suresh, Kemp, Bas, 2021. Digital Twins in Livestock Farming. Animals 11 (4). https://doi.org/10.3390/ani11041008.

Nolan, Bernard T., Green, Christopher T., Juckem, Paul F., Liao, Lixia, Reddy, James E., 2018. Metamodeling and mapping of nitrate flux in the unsaturated zone and groundwater, Wisconsin, USA. J. Hydrol. 559, 428–441. https://doi.org/10.1016/j.jhydrol.2018.02.029. ISSN 221694.

NZ farm source, 2021. Tailoring Nitrogen Applications to Optimise Response, 4. https://store.nzfarmsource.co.nz/enrichplus/pasture-cropping/tailoring-nitrogen-applications-to-optimise-response.

Pedersen, Agnethe N., Borup, Morten, Brink-Kjær, Annette, Christiansen, Lasse E., Mikkelsen, Peter S., 2021. Living and prototyping digital twins for urban water systems: towards multi-purpose value creation using models and sensors. Water 13 (5). https://doi.org/10.3390/w13050592. ISSN 2073–4441. https://www.mdpi.com/2073-4441/13/5/592.

Pembleton, Keith G., Rawnsley, Richard P., Burkitt, Lucy L., 2013. Environmental influences on optimum nitrogen fertiliser rates for temperate dairy pastures. Eur. J. Agron. 45, 132–141. https://doi.org/10.1016/j.eja.2012.09.006. ISSN 1161–0301. https://www.sciencedirect.com/science/article/pii/S1161030112001335.

Pylianidis, Christos, Osinga, Sjoukje, Athanasiadis, Ioannis N., 2021. Introducing digital twins to agriculture. Comput. Electron. Agric. 184, 105942 https://doi.org/10.1016/j.compag.2020.105942. ISSN 0168–1699. https://www.sciencedirect.com/science/article/pii/S0168169920331471.

Roberts, Michael J., Braun, Noah O., Sinclair, Thomas R., Lobell, David B., Schlenker, Wolfram, 2017. Comparing and combining process-based crop models and statistical models with some implications for climate change. Environ. Res. Lett. 12 (9) https://doi.org/10.1088/1748-9326/aa7f33. ISSN 17489326.

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

Shahhosseini, Mohsen, Rafael, A., Martinez-Feria, Hu, Guiping, Archontoulis, Sotirios V., 2019. Maize yield and nitrate loss prediction with machine learning algorithms. Environ. Res. Lett. 14 (12) https://doi.org/10.1088/1748-9326/ab5268. ISSN 17489326.

Shahhosseini, Mohsen, Hu, Guiping, Huber, Isaiah, Archontoulis, Sotirios V., 2021. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Sci. Rep. 11 (1) https://doi.org/10.1038/s41598-020-80820-1. ISSN 20452322.

Tait, Andrew, Henderson, Roddy, Turner, Richard, Zheng, Xiaogu, 2006. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. Int. J. Climatol. 26 (14), 2097–2115. https://doi.org/10.1002/joc.1350. ISSN 8998418.

Togliatti, Kaitlin, Archontoulis, Sotirios V., Dietzel, Ranae, Puntel, Laila, VanLoocke, Andy, 2017. How does inclusion of weather forecasting impact in-season crop model predictions? Field Crop. Res. 214, 261–272. https://doi.org/10.1016/j.fcr.2017.09.008. ISSN 3784290.

Verdouw, Cor, Tekinerdogan, Bedir, Beulens, Adrie, Wolfert, Sjaak, 2021. Digital twins in smart farming. Agric. Syst. 189, 103046 https://doi.org/10.1016/j.agsy.2020.103046. ISSN 0308521X.

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

Vogeler, Iris, Cichota, Rogerio, Snow, Val, 2013. Identification and testing of early indicators for N leaching from urine patches. J. Environ. Manag. 130, 55–63. https://doi.org/10.1016/j.jenvman.2013.08.047. ISSN 3014797.

Voosen, Paul, 2020. Europe builds 'digital twin' of Earth to hone climate forecasts. Science 370 (6512), 16–17. https://doi.org/10.1126/science.370.6512.16.

Waikato Regional Council, 2015. Nitrogen Fertiliser and Effluent. Technical report, 5. https://www.waikatoregion.govt.nz/assets/PageFiles/30016/factsheets/CNM%20factsheet%20get_most_fertiliser_effluent_nitrogen_8.pdf.

Whitehead, David Charles, 1995. Grassland Nitrogen. CAB international, ISBN 9780851989150.

Yang, Li, Jiachen, Yang, 2021. Meta-learning baselines and database for few-shot classification in agriculture. Comput. Electron. Agric. 182, 106055 https://doi.org/10.1016/j.compag.2021.106055. ISSN 0168–1699. https://www.sciencedirect.com/science/article/pii/S0168169921000739.

Zhang, Ruixi, Zen, Remmy, Xing, Jifang, Made Sri, Arsa Dewa, Saha, Abhishek, Stéphane, Bressan, 2020. Hydrological process surrogate modelling and simulation with neural networks. In: Lauw, Hady W., et al. (Eds.), Advances in Knowledge Discovery and Data Mining. Springer International Publishing, Cham, ISBN 978-3-030-47436-2, pp. 449–461. https://doi.org/10.1007/978-3-030-47436-2_34.