

APPLICATION PAPER 

Domain adaptation with transfer learning for pasture digital twins

Christos Pylianidis , Michiel G.J. Kallenberg and Ioannis N. Athanasiadis

Wageningen University & Research, Wageningen, The Netherlands

Corresponding author: Ioannis N. Athanasiadis; Email: ioannis.athanasiadis@wur.nl

Received: 20 February 2023; **Revised:** 08 December 2023; **Accepted:** 06 February 2024

Keywords: agriculture; digital twins; domain adaptation; transfer learning

Abstract


Domain adaptation is important in agriculture because agricultural 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 for domain adaptation in pasture digital twins. We use a synthetic dataset of grassland 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 and agricultural management practices data 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 fine-tuned model performance compared to adding more management practices. This is an interesting finding that is worth further investigation in future studies.

Impact Statement

This paper discusses domain adaptation with transfer learning to transfer field-level pasture growing knowledge between locations with diverse climates for nitrogen response rate prediction in the context of agricultural digital twins.

1. Introduction

Decision support systems are widely used in agriculture to convert data to practical knowledge (Rinaldi and He, 2014; Zhai et al., 2020). A paradigm of decision support systems that has recently found its way to agriculture is that of digital twins (Pylianidis et al., 2021). Digital twins are expected to merge the physical and virtual worlds by providing a holistic view of physical systems, through data integration, continuous monitoring, and adaptation to local conditions. They have started gaining traction with data architectures and applications for greenhouses (Howard et al., 2020; Ariesen-Verschuur et al., 2022), conceptual frameworks for designing and developing them (Verdouw et al., 2021), and case studies in aquaponics (Ghandar et al., 2021).

 This research article was awarded Open Materials badge for transparent practices. See the Data Availability Statement for details.

A factor differentiating digital twins from existing systems is their ability to adapt to local conditions (Blair, 2021). Following the digital twin paradigm, in contrast to generic models that apply global rules across all systems, we can create a blueprint that contains a high-level view of how a system works. This blueprint can then be instantiated as a digital twin in several systems, each with diverse local conditions, and further adjust to them, as more local data and feedback accumulate. In agriculture, adaptation to local conditions (or domain adaptation) is important because systems are affected by multiple local factors, and characterized by high uncertainty, also due to nature's variability. Decisions have to account for the variability in weather conditions, soil types, and agricultural management (i.e. fertilization, irrigation, crop protection actions). Examples of failure to adapt include wrong estimations of yield (Parkes et al., 2019), failure to detect plant drought stress (Schmitter et al., 2017), and expensive equipment that does not work the way it is supposed to be (Gogoll et al., 2020).

A challenge to apply domain adaptation techniques in agricultural digital twins lies with data-related issues. Process-based and machine learning (ML) models comprising the digital twins have difficulties operating with missing data or available data that do not conform with model 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. However, agricultural data are often sparse and noisy. Also, process-based models are typically numerical models that make predictions in small time intervals (from minutes to days). This can also be a problem as the prediction horizon is also short, or when those inputs are from future states of variables (e.g., weather and biophysical factors) and require 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 (Blanning, 1975). 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 (Karpatne et al., 2017) and working with data of different resolutions and absence of future weather values in nitrogen response rate (NRR) prediction (Pylianidis et al., 2022). 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 (Purcell et al., 2022), and digital twins for adaption to climate change (Bauer et al., 2021). 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 (Angin et al., 2020). However, in other sectors, we find that transfer learning has been considered in several cases for digital twins (Xu et al., 2019; Voogd et al., 2022; Zhou et al., 2022). 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 NRR¹ 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:

¹ Additional kg/ha of dry matter harvested per kg of nitrogen fertilizer applied.

- 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 subquestions:

- 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?

2. Methodology

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 (**Figure 1**), soil types, and fertilization treatments. Based on these data, we pretrained ML metamodels in an *origin*

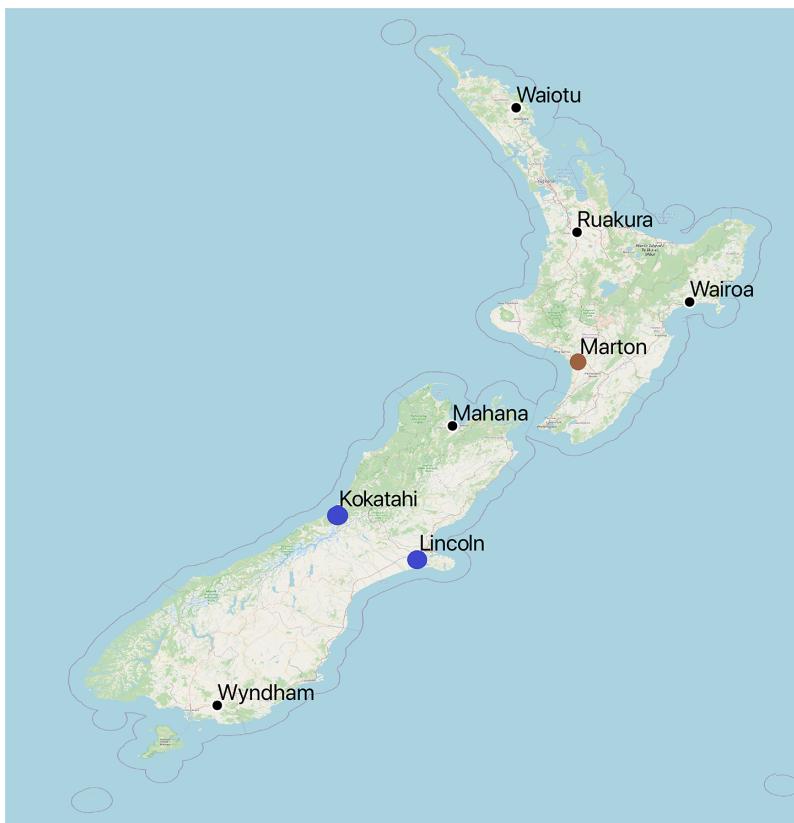


Figure 1. The sites contained in our dataset. With the brown color is the site in the origin climate (Marton, climate 1), and with the blue the sites in the target climates (Kokatahi and Lincoln, climates 2 and 3, respectively).

location and fine-tuned 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.

2.2. Data generation

The simulations comprising our dataset were generated with APSIM (Holzworth et al., 2014) using the AgPasture module (Li et al., 2011). This module has been proven to be an accurate estimator of pasture growth in New Zealand (Cichota et al., 2013, 2018). The simulation parameters covered conditions that are known to affect pasture growth. The full factorial (Antony, 2014) of those parameters was created and given as input to APSIM. The range of the parameters is shown in Table 1.

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, and May) and spring (September, October, and 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 2 months after fertilizer application per kg of nitrogen fertilizer applied.

Regarding the prediction scenario, we assumed to have weather and biophysical data only 4 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 (2 months later), we assumed that no data were available.

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:

- one type of soil and two types of fertilization treatments;
- three types of soil and five types of fertilization treatments.

Table 1. *The full factorial of the presented parameters was used to generate simulations with APSIM*

Parameter	Range
Weather	Daily weather from eight 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, nonirrigated
Fertilization year	1979–2018
Fertilization month	January–December
Fertilization day	5th, 15th, and 25th of the month
Fertilizer amount	0, 20, 40, 60, 80, and 100 kg N/ha

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 (Figure A3). 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 (Ramirez-Villegas et al., 2011) to be dissimilar with the origin location to varying degrees (see Figure A1). 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 is characterized by less frequent rainfall and a wider range of temperatures than location 1. The respective plots can be seen in Figure A2.

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 5 years. Also, to see how robust the models were, we trained each one of them five times with different seeds in each setup and sliding year.

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 2 months after fertilization. Then, data were filtered to contain only simulations for the nonirrigated case. After that, only daily weather data in a window of 4 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 that were not relevant to model for this study. From the remaining data, only the daily weather variables regarding precipitation, solar radiation, and minimum and maximum temperatures 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 50 cm 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.

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 (Pylianidis et al., 2022). 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

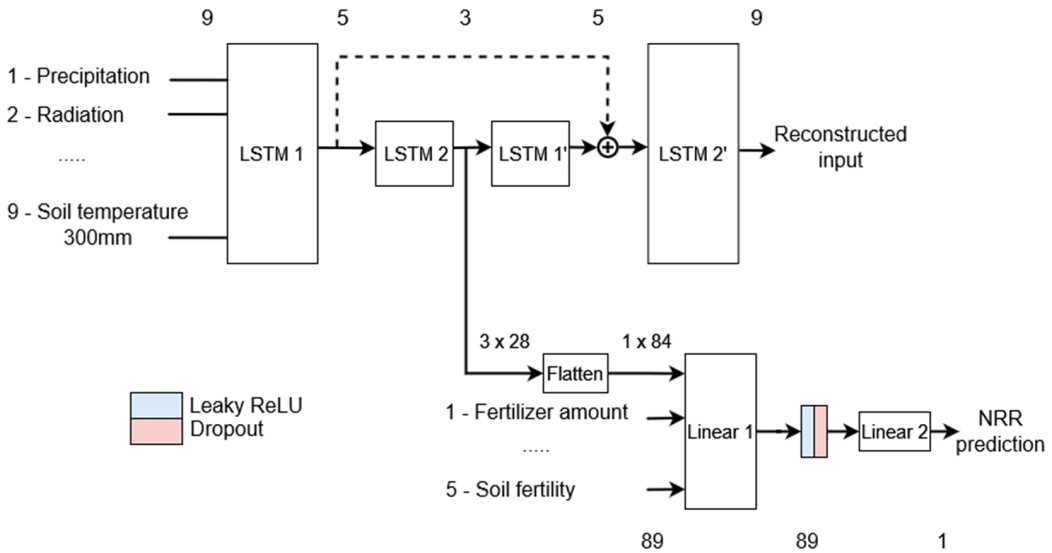


Figure 2. The autoencoder architecture used to pretrain and fine-tune the models. The numbers on the top and bottom of the architecture indicate the number of features in the input/output of each component. The inputs to the encoder were nine time-series variables. The compressed representation of those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

regression head with linear layers whose task is to predict the NRR (Figure 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 Tables A1 and A2.

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

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 how 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.

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 and s2, this behavior was consistent in both location 2 (Figure 3) and location 3 (Figure 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 (Figure 3d), and in location 3 minor improvements (Figure 4d). The standard deviations of the target models on the target location test sets were within the [0.01, 0.08] range (Figures A4 and A5).

Tuning also increased the average R^2 on the *origin location test set* for s1 (Figure 3a,b) and s2 (Figure 4a,b). 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

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 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 five seeds for each setup and year. On Figure 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 four sliding years, only the last year of each set is shown with gray color.

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	0.81	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 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 Figure 3a.

[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 (Figures 3a, 4a). The corresponding standard deviations were also high as shown in Figures A4a, A5a.

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–s2, s1–s3) but there were years like 2004–2005 in location 2, 2008–2009 on location 3 where R^2 dropped substantially. The standard deviations (Figures A4 and A5) also became lower across the years as more agromanagement variability was added.

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., Figure 3a,b), and from s3 to s4 throughout the years (e.g., Figure 3c,d). 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., Figure 4a–c), and from s2 to s4 (e.g., Figure 4b–d).

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 Figure A3), 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 60 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 60 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 [Figure A3](#), we see the number of samples in each setup. Adding more weather data (s1–s2, or s3–s4) doubles the samples included in the pretraining data. However, with the current experimental setup, adding soil types and fertilization treatments (s1–s3, or s2–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.

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 nonrepresentative 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. 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.

Acknowledgments. The authors are extremely grateful to Dr. Val Snow (AgResearch, NZ) for generating the synthetic dataset using APSIM and her contributions in designing the case study application (Pylidianidis et al., 2022). The authors would like to thank Dr. Dilli Paudel (Wageningen University & Research) for the constructive discussions during the validation phase of the experiments.

Author contribution. Conceptualization: C.P., I.A.; Data curation: C.P.; Data visualization: C.P., M.K.; Methodology: C.P., I.A.; Validation: C.P., I.A., M.K.; Writing—original draft: C.P.; Writing—review and editing: I.A., M.K. All authors approved the final submitted draft.

Competing interest. The authors declare none.

Data availability statement. Replication code can be found in <https://github.com/BigDataWUR/Domain-adaptation-for-pasture-digital-twins>.

Ethics statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Funding statement. C.P. has been partially supported by the European Union Horizon 2020 Research and Innovation program (Grant No. 810775, Dragon). M.K. has been partially supported by the European Union Horizon Europe Research and Innovation program (Grant No. 101070496, Smart Droplets).

References

- Angin P, Anisi MH, Göksel F, Gürsoy C and Büyükgülcü A (2020) Agrilora: A digital twin framework for smart agriculture. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 110(4), 77–96. <https://doi.org/10.22667/JOWUA.2020.12.31.077>.
- Antony J (2014) Full factorial designs. In Antony J (ed.), *Design of Experiments for Engineers and Scientists* (Second Edition). Oxford: Elsevier, pp. 63–85. <https://doi.org/10.1016/B978-0-08-099417-8.00006-7>.
- Ariesen-Verschuur N, Verdouw C and Tekinerdogan B (2022) Digital twins in greenhouse horticulture: A review. *Computers and Electronics in Agriculture* 199, 107183. <https://doi.org/10.1016/j.compag.2022.107183>.
- Bauer P, Stevens B and Hazeleger W (2021) A digital twin of earth for the green transition. *Nature Climate Change* 110(2), 80–83. <https://doi.org/10.1038/s41558-021-00986-y>.
- Blair GS (2021) Digital twins of the natural environment. *Patterns* 20(10), 100359. <https://doi.org/10.1016/j.patter.2021.100359>.
- Blanning RW(1975) The construction and implementation of metamodels. *SIMULATION* 240(6), 177–184. <https://doi.org/10.1177/003754977502400606>.
- Cichota R, Snow VO and Vogeler I (2013) Modelling nitrogen leaching from overlapping urine patches. *Environmental Modelling & Software* 41, 15–26. <https://doi.org/10.1016/j.envsoft.2012.10.011>.
- Cichota R, Vogeler I, Werner A, Wigley K and Paton B (2018) Performance of a fertiliser management algorithm to balance yield and nitrogen losses in dairy systems. *Agricultural Systems* 162, 56–65. <https://doi.org/10.1016/j.agsy.2018.01.017>.
- Ghandar A, Ahmed A, Zulfiqar S, Hua Z, Hanai M and Theodoropoulos G (2021) A decision support system for urban agriculture using digital twin: A case study with aquaponics. *IEEE Access* 9, 35691–35708. <https://doi.org/10.1109/ACCESS.2021.3061722>.

- Gogoll D, Lottes P, Weyler J, Petrinic N and Stachniss C** (2020) 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)*, pp. 2636–2642. <https://doi.org/10.1109/IROS45743.2020.9341277>.
- Holzworth DP, Huth NI, deVoil PG, Zurcher EJ, Herrmann NI, McLean G, Chenu K, van Oosterom EJ, Snow V, Murphy C, Moore AD, Brown H, Whish JPM, Verrall S, Fainges J, Bell LW, Peake AS, Poulton PL, Hochman Z, Thorburn PJ, Gaydon DS, Dalgliesh NP, Rodriguez D, Cox H, Chapman S, Doherty A, Teixeira E, Sharp J, Cichota R, Vogeler I, Li FY, Wang E, Hammer GL, Robertson MJ, Dimes JP, Whitbread AM, Hunt J, van Rees H, McClelland T, Carberry PS, Hargreaves JNG, MacLeod N, McDonald C, Harsdorf J, Wedgwood S and Keating BA** (2014) APSIM - evolution towards a new generation of agricultural systems simulation. *Environmental Modelling and Software* 62, 327–350. <https://doi.org/10.1016/j.envsoft.2014.07.009>.
- Howard DA, Ma Z, Aaslyng JM and Jørgensen BN** (2020) Data architecture for digital twin of commercial greenhouse production. In *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*. Ho Chi Minh City, Vietnam: IEEE. pp. 1–7. <https://doi.org/10.1109/RIVF48685.2020.9140726>.
- Karpatne A, Watkins W, Read J and Kumar V** (2017) Physics-guided neural networks (PGNN): An application in Lake temperature modeling. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 10. <http://arxiv.org/abs/1710.11431>.
- Li FY, Snow VO and Holzworth DP** (2011) Modelling the seasonal and geographical pattern of pasture production in New Zealand. *New Zealand Journal of Agricultural Research* 54(4), 331–352. <https://doi.org/10.1080/00288233.2011.613403>.
- Parkes B, Higginbottom TP, Hufkens K, Ceballos F, Kramer B and Foster T** (2019) Weather dataset choice introduces uncertainty to estimates of crop yield responses to climate variability and change. *Environmental Research Letters* 14(12), 124089. <https://doi.org/10.1088/1748-9326/ab5ebb>.
- Purcell W, Klipic A and Neubauer T** (2022) A digital twin for grassland management. In *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pp. 1–6. <https://doi.org/10.1109/ICECET55527.2022.9873446>.
- Pylianidis C, Osinga S and Athanasiadis IN** (2021) Introducing digital twins to agriculture. *Computers and Electronics in Agriculture* 184, 105942. <https://doi.org/10.1016/j.compag.2020.105942>.
- Pylianidis C, Snow V, Overweg H, Osinga S, Kean J and Athanasiadis IN** (2022) Simulation-assisted machine learning for operational digital twins. *Environmental Modelling & Software* 148, 105274. <https://doi.org/10.1016/j.envsoft.2021.105274>.
- Ramirez-Villegas J, Lau C, Kohler A-K, Jarvis A, Arnell NW, Osborne TM and Hooker J** (2011) Climate analogues: Finding tomorrow's agriculture today. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).
- Rinaldi M and He Z** (2014) Decision support systems to manage irrigation in agriculture. In Sparks DL (ed.), *Advances in Agronomy*, 123. Academic Press, pp. 229–279. <https://doi.org/10.1016/B978-0-12-420225-2.00006-6>.
- Schmitter P, Steinrücken J, Römer C, Ballvora A, Léon J, Rascher U and Plümer L** (2017) Unsupervised domain adaptation for early detection of drought stress in hyperspectral images. *ISPRS Journal of Photogrammetry and Remote Sensing* 131, 65–76. <https://doi.org/10.1016/j.isprsjprs.2017.07.003>.
- Verdouw C, Tekinerdogan B, Beulens A and Wolfert S** (2021) Digital twins in smart farming. *Agricultural Systems* 189, 103046. <https://doi.org/10.1016/j.agry.2020.103046>.
- Voogd K, Allamaa JP, Alonso-Mora J and Son TD** (2022) Reinforcement Learning from Simulation to Real World Autonomous Driving Using Digital Twin. <https://arxiv.org/abs/2211.14874> (accessed January 2023).
- Xu Y, Sun Y, Liu X and Zheng Y** (2019) A digital-twin-assisted fault diagnosis using deep transfer learning. *IEEE Access* 7, 19990–19999. <https://doi.org/10.1109/ACCESS.2018.2890566>.
- Zhai Z, Martínez JF, Beltran V and Martínez NL** (2020) Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture* 170, 105256. <https://doi.org/10.1016/j.compag.2020.105256>.
- Zhou X, Sbarufatti C, Giglio M and Dong L** (2022) A Fuzzy-Set-Based Joint Distribution Adaptation Method for Regression and its Application to Online Damage Quantification for Structural Digital Twin. <https://arxiv.org/abs/2211.02656>.

A. Appendix

A.1. Climate similarity

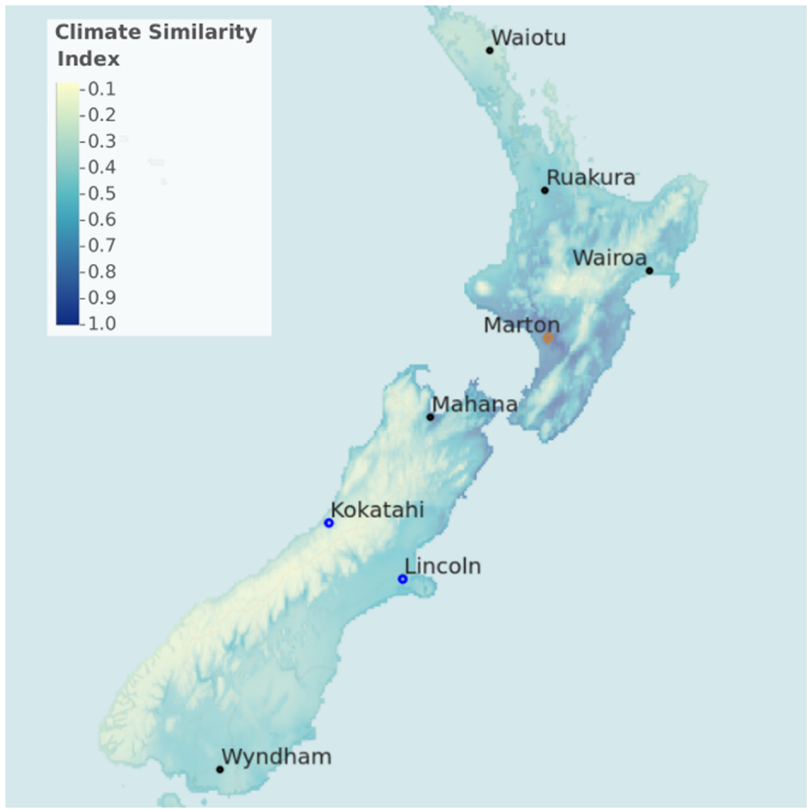


Figure A1. CCAFS similarity index across New Zealand. The weather parameters for the similarity were precipitation and average temperature. Location 1 (Marton) is colored in brown, and locations 2 (Kokatahi) and 3 (Lincoln) in blue. The darker the color on the map, the more similar the climate is to location 1. Location 2 had index value 0.354, and location 3 0.523.

A.2. Experimental setup simulation parameters and amount of samples

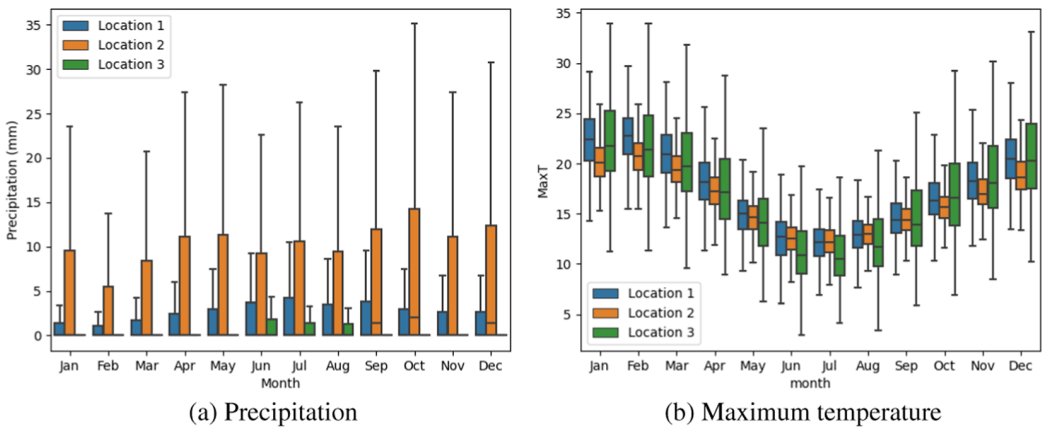


Figure A2. Weather parameters known to affect pasture growth for the climates included in this study. The parameters are presented across the months and are aggregated over the years.

A.3. Model hyperparameters

	Weather	Fertilization	Fertilization	Fertilizer	Soil	Soil water	Total
origin model (s1)	Years	days	months	amounts	Fertilities	capacities	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 A3. Number of parameters and total samples used in each training/validation/test set of each setup.

Table A1. The fixed hyperparameters of the origin models

Hyperparameter	Value
Learning rate	4×10^{-5}
Batch size	64
Epochs	60

Table A2. The search space for the hyperparameters of the target models

Hyperparameter	Values
Learning rate	$[4 \times 10^{-5}, 10^{-4}]$
Batch size	[2, 74]
Epochs	[7, 15, 30]

A.4. 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 A4. 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 A5. Standard deviations of the various setups for the origin models climate 1, and target models in climate 2.