DeepOryza: A knowledge-guided machine learning model for rice growth simulation

Jingye Han[1,2], Liangsheng Shi[1,*], Christos Pylianidis[2], Qi Yang[1], Ioannis N. Athanasiadis[2]
[1]State Key Laboratory of Water Resources and Hydropower Engineering Sciences, Wuhan University
[2]Wageningen University and Research, Wageningen, the Netherlands

*Corresponding author, E-mail: liangshs@whu.edu.cn

Abstract

Even though process-based crop models are widely used to simulate crop growth, the challenge of parameter calibration makes it difficult to use them in practice. Also, it is labor-consuming to improve the model by adjusting the model description. To address these issues, we propose a knowledge-guided machine learning model (DeepOryza) to directly learn the crop growth pattern from data. A synthetic dataset generated by a process-based model (ORYZA2000) was used to pre-train the DeepOryza. An observation dataset was used to finetune and evaluate the DeepOryza. The preliminary results showed that DeepOryza can perform equally or better than the well-calibrated ORYZA2000. To investigate the effect of the proposed knowledge-guided structure, we designed two DeepOryza models with different structures. Results showed that the knowledge-guided structure can improve the performance of DeepOryza when the synthetic dataset was generated by the uncalibrated ORYZA2000. This finding indicates that the knowledge-guided structure could potentially reduce the calibration requirement of the process-based model.

Introduction

In agriculture, crop models with the ability to describe interactions between crops and the environment, are widely used as physics-based methods to simulate crop growth and optimize crop production (Ewert et al., 2015; Keating et al., 2003). Studies over the past decades have developed crop models such as WOFOST (van Diepen et al., 1989), SWAP (van Dam, J. C., Huygen, J., Wesseling, J. G., Feddes, R. A., Kabat, P., van Walsum and P. E. V., Groenendijk, P., & van Diepen, 1997), and ORYZA (Bouman et al., 2001) based on different physical mechanisms. Despite their extensive use, these crop models have several limitations due to simplified representations of the physical processes (Feng et al., 2019; Karpatne et al., 2017) or challenges in parameter calibration (Seidel et al., 2018; Wallach et al., 2021). The limitations of physics-based models cut across discipline boundaries and are well-known in the scientific community (Lall, 2014).

To address these issues, various methods have focused on improving the accuracy of parameter estimation as well as reducing the model structure error. For parameter estimation, a model-independent software (e.g., PEST) (Doherty, 1994) or package (e.g., SOPTPY) (Houska et al., 2015) could be employed to calibrate models, but applications are limited by the difficulty of coupling the model and the software/package (Christian et al., 2020). As a result, most researchers used trial-and-error procedures to search for the best-fit parameters (Seidel et al., 2018) even though many other algorithms (least squares, generalized likelihood uncertainty estimation, Bayesian parameter estimation) could be used. However, the optimal parameters could change as the evaluation criteria change, which means that there will never be an optimal set of parameters that allow all simulated variables to be optimal at the same time. Regarding model structure error, it always arises from the imperfect description of crop growth mechanism. Thus, the only way to decrease the structure error of a process-based model is to adjust the model description, such as the model function, the model structure or involve additional inputs into the model. However, this procedure is labor-consuming and may increase the model complexity, making it more difficult to calibrate (Yin et al., 2021).

Data-driven methods, especially machine learning (ML) methods, have shown superior performance in various disciplines for their feature extraction capabilities and capturing nonlinear dynamics (Kashinath et al., 2021). For agricultural models, the application of purely data-driven methods is restricted by the lack of data, and also from the learned dynamics which may violate the laws of crop growth and physics. In recent years, research is being directed towards combining process-based crop models and data-driven methods (Everingham et al., 2016; Feng et al., 2020; Guzmán et al., 2018; Pagani et al., 2017; Shahhosseini et al., 2021). Prior studies consider crop models as tools for feature engineering to discover valuable information from raw data, but they do not truly integrate the two modeling approaches.

To fundamentally integrate scientific knowledge from process-based models and ML models, a new paradigm termed knowledge-guided machine learning (KGML) was proposed. There have been attempts to apply this new paradigm to several disciplines by transforming physical knowledge into different forms for integration into machine learning models, such as logic rules, constraints, simulated datasets, etc. A detailed review of these forms can be found in (Karpatne et al., 2017; Von Rueden et al., 2019; Willard et al., 2020). A study conducted by Read et al. (2019) used a neural network pre-trained on a synthetic dataset generated by a physical model to accurately predict lake water temperature for various conditions. Another research developed a model to estimate N_2O emissions using a knowledge-guided model structure that was designed according to the relationship of variables defined in equations of process-based models (Liu et al., 2021). However, to the best of our knowledge, there is no research exploring the potential of the KGML for crop growth simulation, which involves the interactions between crops and environment. As agricultural data acquisition becomes easier, it is necessary to build a crop growth model that can learn from both real observed data and knowledge embedded in crop model.

In this paper, we present a knowledge-guided machine learning model for crop growth. A baseline model (DeepOryza_baseline) and a knowledge-guided machine learning model (DeepOryza_kgml) are proposed and compared with a process-based crop growth model (ORYZA2000). DeepOryza was developed as an LSTM for multivariate time series forecasting of seven crop states. The baseline model does not consider any physical relationship between the outputs. In contrast, the knowledge-guided machine learning DeepOryza enforces known relationships between them. DeepOryza is demonstrated for a case study in China with field observations from two years. The DeepOryza model was trained with synthetic data generated using ORYZA2000.

The aim of this study is to investigate: 1) how a ML model performs against a well-calibrated process-based crop growth model; 2) the effect of knowledge-guided structures in ML models; 3) the necessity of calibration for the process-based model before generating a synthetic dataset.

Methodology and data

To develop DeepOryza and evaluate its performance, we design the following framework (Fig.1):

1 Synthetic dataset generation: The ORYZA2000 model was calibrated on observed data from YEAR-A. The calibrated and uncalibrated ORYZA2000 models were then used to generate synthetic datasets, respectively.

2 Model pretraining: The DeepOryza model was trained on the synthetic dataset to learn basic patterns of rice growth.

3 Model finetuning: The pre-trained DeepOryza model was finetuned on the observation dataset from Year-A and the Finetuned DeepOryza model was tested on the observation dataset from both Year-A and Year-B.

4 Model evaluation: The results of the uncalibrated ORYZA2000 (with default parameters), the calibrated ORYZA2000, and the finetuned DeepOryza were compared to evaluate model performance, the necessity for calibration, and the effect of knowledge-guided structures.

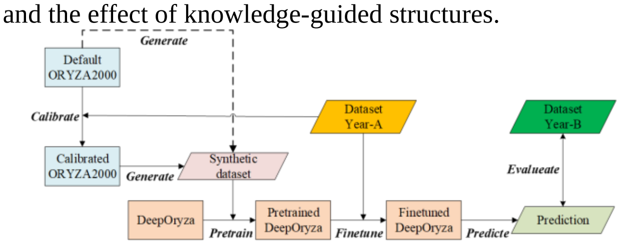

Fig. 1 The framework for building and evaluating the DeepOryza model.

The synthetic dataset

The ORYZA2000 model was selected for generating synthetic datasets due to its verification through a broad range of studies (Cao et al., 2017; Han et al., 2022, 2020; Li et al., 2013). After determining the model parameters, the date of seeding and transplanting, ORYZA2000 can simulate rice growth in daily steps for the rate of dry matter production and phenological development driven by the daily inputs (weather, irrigation, and fertilization) (Bouman et al., 2001). To reduce the complexity of the model, the irrigation and model parameters were not involved in the synthetic dataset. Also, only seven model outputs (development stage, plant area index, leaf/stem/grain biomass, above-ground biomass, final yield) which had the corresponding observations in the real datasets were selected for the synthetic dataset. Thus, the synthetic datasets consist of weather, management (fertilization, dates of seeding and transplanting), and the seven corresponding outputs of the ORYZA2000. Cligen (Nicks et al. 1995), a software that can generate random synthetic time-series weather data that match the pattern of historical weather of a specific site, was used to generate weather data. The management was randomly generated according to the schedule of local farmers. To calibrate the ORYZA2000 model, 14 parameters that are sensitive to the seven output variables were selected. The PEST (Doherty, 1994) software was then used to calibrate these parameters with the real dataset.

Finally, two synthetic datasets were created with the uncalibrated ORYZA2000 with default parameters (only adjusting the DVS-related parameters to make the growth timetable roughly consistent with the real data) and calibrated ORYZA2000. Each synthetic dataset included 250k samples, and was further divided into five sub-datasets to pre-train five models in order to reduce the effect of random factors. The results of DeepOryza were derived by averaging the results of the models in the five sub-datasets.

The observation dataset

The observation data were collected in a late-season rice experiment in a small region (23°5′52″~23°7′23″ N, 108°57′7″~108°58′34″ E) located in Binyang County of Guangxi in China. The meteorological data, including solar radiation, air temperature, relative humidity, and precipitation acquired from https://power.larc.nasa.gov/. There were 65 plots in 2018 and 40 plots in 2019 were used in this study. The management recording was obtained by performing surveys during the growing season. The observation values of DVS (development stage), PAI (plant area index), WLV (leaf biomass), WST (stem biomass), WSO (storage organ biomass), and AGB (above-ground biomass) were obtained by destructive sampling. The final yield was obtained by combine-harvesters. The details of the dataset can be found in Han et al., (2022).

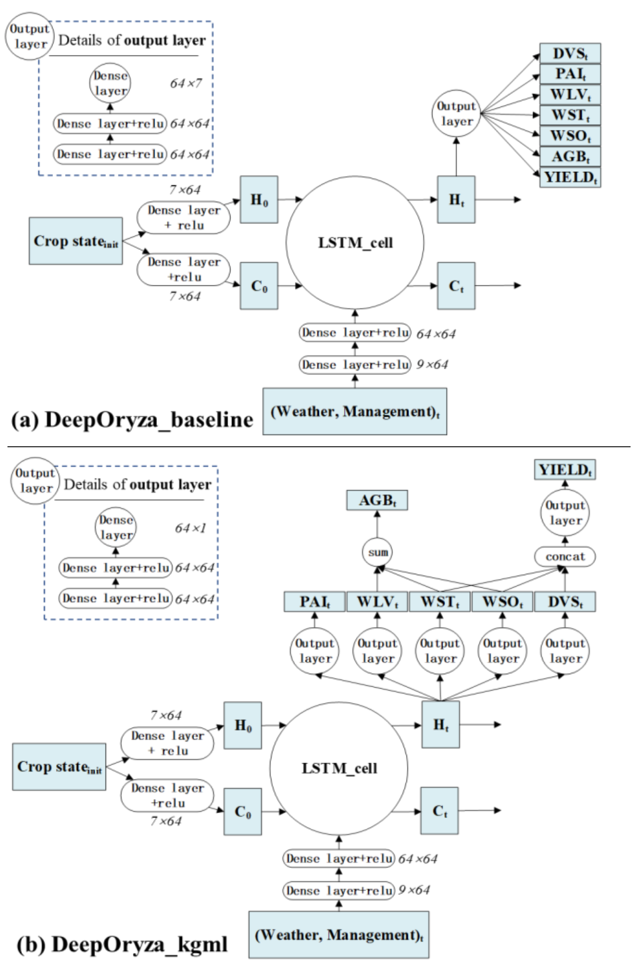

Fig. 2 The structure of the two DeepOryza models

The structure, training, and evaluation

The LSTM model, proposed by Hochreiter and Schmidhuber (1997), was employed to build the DeepOryza because it can learn longer dependencies between variables on long time series data. As shown in Fig.2, the hidden states of the LSTM cell were initialized from the values of the initial crop state, which could reduce the effect of random initialization. The time series of weather and management were used as input because they are known drivers of rice growth. The DeepOryza_baseline estimates seven crop states directly without considering their relationship. The structure of DeepOryza_kgml was designed by considering the relationship between the crop states: 1) AGB should be the sum of WLV, WST, and WSO; 2) the yield comes from the partition biomass of the storage organ and the translocation biomass of the stem over the stage after the flowering stage (Laza et al., 2003). These two relationships were also described by functions in ORYZA2000. Thus, the AGB was derived from WLV, WST, and WSO; the YIELD was derived from DVS, WST, and WSO in DeepOryza_kgml.

The input size was 8 (temperature_min, temperature _max, irradiance, vapor pressure, wind speed, precipitation, fertilization amount, if transplanting). The hidden state size of the LSTM cell was 64, and the other layer setting was shown in Fig.2. ADAM was used to optimize the model parameters. The loss function was defined as:

$$\mathrm{Loss} = \sum_{state} Loss_{state} \tag{1}$$

$$\mathrm{Loss}_{state} = \frac{\sum(\mathrm{prediction}_{state} - \mathrm{observation}_{state})^2}{N} \tag{2}$$

where state includes DVS, PAI, WLV, WST, WSO, AGB and YIELD; N is the number of observations for the corresponding state.

The model performance was evaluated by RMSE. The pretraining process was repeated 25 times (5 synthetic datasets × 5 random seeds) to make the results more robust. The RMSE was the average value of the 25 repeats.

Results

The results of Calibrated-Oryza2000, DeepOryza_baseline and DeepOryza_kgml

Since the ORYZA2000 was calibrated with the PEST software, its performance improved significantly (Case 5 and Case 6 in Table 1). As shown in Table 1, the DeepOryza models had better performance than calibrated-ORYZA2000 in Case 1 and Case 3. But the performances of the three models were similar in Case 2 and Case 4. A reason for that may be that ORYZA2000 model was well

Table 1
Performance of the models on the training and testing dataset

Case	Training set	Testing set	Synthetic dataset was generated by	Model	DVS -	PAI m^2/m^2	WLV kg/ha	WST kg/ha	WSO kg/ha	AGB kg/ha	YIELD kg/ha
					Mean of RMSE / Std of RMSE						
1	2018	2018	**Calibrated** ORYZA2000	**Calibrated**_ORYZA2000	0.099	1.05	464	962	926	1523	944
				DeepOryza_**baseline**	0.099/0.001	0.96/0.006	500/8	950/7	890/13	1515/7	922/26
				DeepOryza_**kgml**	0.097/0.002	0.94/0.006	488/8	952/11	886/12	1495/6	888/38
2	2019	2019		**Calibrated**_ORYZA2000	0.073	0.9	382	743	605	1186	660
				DeepOryza_**baseline**	0.077/0.001	0.87/0.007	407/4	682/9	594/9	1187/9	637/14
				DeepOryza_**kgml**	0.075/0.001	0.87/0.005	393/3	673/10	587/8	1176/8	639/14
3	2018	2019	**Calibrated** ORYZA2000	**Calibrated**_ORYZA2000	0.117	1.36	399	768	832	1173	894
				DeepOryza_**baseline**	0.110/0.006	1.00/0.022	387/4.4	665/8.3	765/24.4	1250/19.2	784/46
				DeepOryza_**kgml**	0.111/0.005	1.00/0.011	387/3.6	666/13.6	794/26.9	1220/12.3	804/57
4	2019	2018		**Calibrated**_ORYZA2000	0.115	1.01	521	1002	963	1569	1002
				DeepOryza_**baseline**	0.112/0.003	1.09/0.013	576/13	986/17	1058/32	1560/9	1056/22
				DeepOryza_**kgml**	0.111/0.003	1.11/0.011	549/8	973/13	994/25	1550/6	1036/30
5	2018	2019	**Uncalibrated** ORYZA2000	**Uncalibrated**_ORYZA2000	0.142	1.94	460	719	962	1395	749
				Calibrated_ORYZA2000	0.117	1.36	399	768	832	1173	894
				DeepOryza_**baseline**	0.101/0.005	0.97/0.019	433/12	726/21	867/40	1275/38	862/50
				DeepOryza_**kgml**	0.104/0.004	0.97/0.024	383/7	667/9	796/44	1245/28	836/74
6	2019	2018		**Uncalibrated**_ORYZA2000	0.108	1.46	470	1026	1674	1835	1267
				Calibrated_ORYZA2000	0.115	1.01	521	1002	963	1569	1002
				DeepOryza_**baseline**	0.119/0.005	1.01/0.023	533/20	977/26	1030/49	1587/23	1051/47
				DeepOryza_**kgml**	0.126/0.007	1.03/0.018	510/20	936/19	979/52	1559/26	1054/49

calibrated in the dataset of 2019, making it difficult to further improve the model performance with DeepOryza. Comparing the accuracy of WLV, WST, WSO and YIELD (Case 1-4 in Table 1), there was no obvious winner between the two DeepOryza models because the synthetic dataset was generated by the calibrated ORYZA2000 model, which had embedded the prior knowledge between the variables of the real dataset. Thus, the knowledge-guided structure did not improve the model performance.

The effect of knowledge-guided structure and the necessity of calibration for the process-based model

In Table 1, the DeepOryza models in Case 5 and Case 6 were pre-trained on the synthetic dataset generated by the uncalibrated ORYZA2000. The performance of DeepOryza_baseline of Case 5 was worse than that of Case 3. This is due to the lack of restriction of the knowledge-guided structure, which made the DeepOryza_baseline easier to overfit and violate crop growth pattern when fine-tuning. The DeepOryza_kgml of Case 3 had similar accuracy to that of Case 5, which meant that a reasonable structure, although not as detailed as the process model, could also have a positive effect on model training. Furthermore, the accuracy of biomass-related variables (except the YIELD in Case 6) of DeepOryza_kgml was better than that of DeepOryza_baseline, which indicated that the knowledge-guided structure could potentially reduce the requirement of calibration.

Conclusions and Limitations

In this study, we proposed a knowledge-guided machine model named DeepOryza for rice growth simulation. The LSTM model was used as the basic model. Its initialization and structure design were guided by the knowledge of the ORYZA2000 crop growth model. This is the first attempt of simulating multivariate time-series for crop growth. The preliminary results showed that the DeepOryza could perform equally or better than a well-calibrated process-based model. The knowledge-guided structure could improve the model performance and potentially reduce the calibration requirement of the process-based model.

However, the dataset is relatively small and requires more case studies and a variety of datasets to verify the described methods. Also, fundamental principles (e.g., mass conservation) are not embedded in the structure or loss function. The naïve LSTM cell makes the iterative process of crop growth a black box and weakens the interpretability. Therefore, a better structure and interpretation should also be further investigated in the future.

Acknowledgments

This study was supported by the National Natural Science Foundation of China Grant 51979200; the Open Research Fund of Guangxi Key Laboratory of Water Engineering Materials and Structures, Guangxi Institute of Water Resources Research, under grant NO: GXHRI-WEMS-2020-06 and the China Scholarship Council for 1.5 years' study at the Wageningen university & research.

References

Bouman, B.A.M., Kropff, M.J., Tuong To, P., Wopereis, M., Ten Berge, H.F.M., Van Laar, H.H., 2001. ORYZA2000 : modeling lowland rice, IRRI.

Cao, B., Hua, S., Ma, Y., Li, B., Sun, C., 2017. Evaluation of ORYZA2000 for simulating rice growth of different genotypes at two latitudes. Agron. J. 109, 2613–2629. https://doi.org/10.2134/agronj2017.03.0145

Christian, K., Zadeh, N., Maria, J.W., Kumar, A., Wit, D., Sabine, J., 2020. The chaos in calibrating crop models.

Doherty, J., 1994. PEST: Model independent parameter estimation. Fifth edition of user manual. Watermark Numer. Comput.

Everingham, Y., Sexton, J., Skocaj, D., Inman-Bamber, G., 2016. Accurate prediction of sugarcane yield using a random forest algorithm. Agron. Sustain. Dev. 36. https://doi.org/10.1007/s13593-016-0364-z

Ewert, F., Rötter, R.P., Bindi, M., Webber, H., Trnka, M., Kersebaum, K.C., Olesen, J.E., van Ittersum, M.K., Janssen, S., Rivington, M., Semenov, M.A., Wallach, D., Porter, J.R., Stewart, D., Verhagen, J., Gaiser, T., Palosuo, T., Tao, F., Nendel, C., Roggero, P.P., Bartošová, L., Asseng, S., 2015. Crop modelling for integrated assessment of risk to food production from climate change. Environ. Model. Softw. 72, 287–303. https://doi.org/10.1016/j.envsoft.2014.12.003

Feng, P., Wang, B., Liu, D.L., Waters, C., Xiao, D., Shi, L., Yu, Q., 2020. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. Agric. For. Meteorol. 285–286, 107922. https://doi.org/10.1016/j.agrformet.2020.107922

Feng, P., Wang, B., Liu, D.L., Waters, C., Yu, Q., 2019. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia. Agric. For. Meteorol. 275, 100–113. https://doi.org/10.1016/j.agrformet.2019.05.018

Guzmán, S.M., Paz, J.O., Tagert, M.L.M., Mercer, A.E., Pote, J.W., 2018. An integrated SVR and crop model to estimate the impacts of irrigation on daily groundwater levels. Agric. Syst. 159, 248–259. https://doi.org/10.1016/j.agsy.2017.01.017

Han, J., Shi, L., Yang, Q., Chen, Z., Yu, J., Zha, Y., 2022. Rice yield estimation using a CNN-based image-driven data assimilation framework. F. Crop. Res. 288, 108693. https://doi.org/10.1016/j.fcr.2022.108693

Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., Zhang, J., 2020. Prediction of winter wheat yield based on multi-source data and machine learning in China. Remote Sens. 12. https://doi.org/10.3390/rs12020236

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Houska, T., Kraft, P., Chamorro-Chavez, A., Breuer, L., 2015. SPOTting model parameters using a ready-made python package. PLoS One 10, 1–22. https://doi.org/10.1371/journal.pone.0145180

Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-guided data science: A new paradigm for scientific discovery from data. IEEE Trans. Knowl. Data Eng. 29, 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168

Kashinath, K., Mustafa, M., Albert, A., Wu, J.-L., Jiang, C., Esmaeilzadeh, S., Azizzadenesheli, K., Wang, R., Chattopadhyay, A., Singh, A., Manepalli, A., Chirila, D., Yu, R., Walters, R., White, B., Xiao, H., Tchelepi, H.A., Marcus, P., Anandkumar, A., Hassanzadeh, P., Prabhat, 2021. Physics-informed machine learning: case studies for weather and climate modelling. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 379, 20200093.

Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J.P., Silburn, M., Wang, E., Brown, S., Bristow, K.L., Asseng, S., Chapman, S., McCown, R.L., Freebairn, D.M., Smith, C.J., 2003. An overview of APSIM, a model designed for farming systems simulation. Eur. J. Agron. 18, 267–288. https://doi.org/10.1016/S1161-0301(02)00108-9

Lall, U., 2014. Debates-The future of hydrological sciences: A (common) path forward? One water. One world. Many climes. Many souls. Water Resour. Res. 50, 5335–5341. https://doi.org/10.1002/2014WR015402

Laza, M.R.C., Peng, S., Akita, S., Saka, H., 2003. Contribution of biomass partitioning and translocation to grain yield under sub-optimum growing conditions in irrigated rice. Plant Prod. Sci. 6, 28–35. https://doi.org/10.1626/pps.6.28

Li, T., Raman, A.K., Marcaida, M., Kumar, A., Angeles, O., Radanielson, A.M., 2013. Simulation of genotype performances across a larger number of environments for rice breeding using ORYZA2000. F. Crop. Res. 149, 312–321. https://doi.org/10.1016/j.fcr.2013.05.006

Liu, L., Xu, S., Jin, Z., Tang, J., Guan, K., Griffis, T., Erickson, M., Frie, A., Jia, X., Kim, T., Miller, L., Peng, B., Wu, S., Yang, Y., Zhou, W., Kumar, V., 2021. KGML-ag: A Modeling Framework of Knowledge-Guided Machine Learning to Simulate Agroecosystems: A Case Study of Estimating N₂O Emission using Data from Mesocosm Experiments. Geosci. Model Dev. Discuss. 1–29. https://doi.org/10.5194/gmd-2021-317

Nicks, A.D., Lane, L.J., Gander, G.A., n.d. Weather generator. Chapter 2 in USDA–Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation.

Pagani, V., Stella, T., Guarneri, T., Finotto, G., van den Berg, M., Marin, F.R., Acutis, M., Confalonieri, R., 2017. Forecasting sugarcane yields using agro-climatic indicators and Canegro model: A case study in the main production region in Brazil. Agric. Syst. 154, 45–52. https://doi.org/10.1016/j.agsy.2017.03.002

Read, J.S., Jia, X., Willard, J., Appling, A.P., Zwart, J.A., Oliver, S.K., Karpatne, A., Hansen, G.J.A., Hanson, P.C., Watkins, W., Steinbach, M., Kumar, V., 2019. Process-Guided Deep Learning Predictions of Lake Water Temperature. Water Resour. Res. 55, 9173–9190. https://doi.org/10.1029/2019WR024922

Seidel, S.J., Palosuo, T., Thorburn, P., Wallach, D., 2018. Towards improved calibration of crop models – Where are we now and where should we go? Eur. J. Agron. 94, 25–35. https://doi.org/10.1016/j.eja.2018.01.006

Shahhosseini, M., Hu, G., Huber, I., Archontoulis, S. V., 2021. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Sci. Rep. 11, 1–15. https://doi.org/10.1038/s41598-020-80820-1

van Dam, J. C., Huygen, J., Wesseling, J. G., Feddes, R. A., Kabat, P., van Walsum, P. E. V., Groenendijk, P., & van Diepen, C.A., 1997. Theory of SWAP version 2.0; Simulation of water flow, solute transport and plant growth in the soil-water-atmosphere-plant environment.

van Diepen, C.A., Wolf, J., van Keulen, H., Rappoldt, C., 1989. WOFOST: a simulation model of crop production. Soil Use Manag. 5, 16–24. https://doi.org/10.1111/j.1475-2743.1989.tb00755.x

Von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., Walczak, M., Garcke, J., Bauckhage, C., Schuecker, J., 2019. Informed machine learning - a taxonomy and survey of integrating knowledge into learning systems. arXiv.

Wallach, D., Palosuo, T., Thorburn, P., Hochman, Z., Gourdain, E., Andrianasolo, F., Asseng, S., Basso, B., Buis, S., Crout, N., Dibari, C., Dumont, B., Ferrise, R., Gaiser, T., Garcia, C., Gayler, S., Ghahramani, A., Hiremath, S., Hoek, S., Horan, H., Hoogenboom, G., Huang, M., Jabloun, M., Jansson, P.E., Jing, Q., Justes, E., Kersebaum, K.C., Klosterhalfen, A., Launay, M., Lewan, E., Luo, Q., Maestrini, B., Mielenz, H., Moriondo, M., Nariman Zadeh, H., Padovan, G., Olesen, J.E., Poyda, A., Priesack, E., Pullens, J.W.M., Qian, B., Schütze, N., Shelia, V., Souissi, A., Specka, X., Srivastava, A.K., Stella, T., Streck, T., Trombi, G., Wallor, E., Wang, J., Weber, T.K.D., Weihermüller, L., de Wit, A., Wöhling, T., Xiao, L., Zhao, C., Zhu, Y., Seidel, S.J., 2021. The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. Environ. Model. Softw. 145. https://doi.org/10.1016/j.envsoft.2021.105206

Willard, J.D., Jia, X., Xu, S., Steinbach, M., Kumar, V., 2020. Integrating physics-based modeling with machine learning: A survey. arXiv 1, 1–34.

Yin, X., Struik, P.C., Goudriaan, J., 2021. On the needs for combining physiological principles and mathematics to improve crop models. F. Crop. Res. 271, 108254. https://doi.org/10.1016/j.fcr.2021.108254