Received: 18 June 2023 | Accepted: 5 September 2023

DOI: 10.1002/agj2.21470

NOTES AND INSIGHTS

Agronomy Journal

Integrating APSIM model with machine learning to predict wheat yield spatial distribution

Ahmed M. S. Kheir[1,2,3] | **Siyabusa Mkuhlani**[4] | **Jane W. Mugo**[4,5] | **Abdelrazek Elnashar**[6,7] | **Vinay Nangia**[8] | **Medha Devare**[4] | **Ajit Govind**[1]

[1]International Center for Agricultural Research in the Dry Areas (ICARDA), Maadi, Egypt

[2]Julius Kühn Institute (JKI)—Federal Research Centre for Cultivated Plants, Institute for Strategies and Technology Assessment, Kleinmachnow, Germany

[3]Soils, Water and Environment Research Institute, Agricultural Research Center, Giza, Egypt

[4]International Institute for Tropical Agriculture (IITA), c/o ICIPE, Nairobi, Kenya

[5]Department of Earth and Climate Science, University of Nairobi, Nairobi, Kenya

[6]Section of Soil Science, Faculty of Organic Agricultural Sciences, University of Kassel, Witzenhausen, Germany

[7]Department of Natural Resources, Faculty of African Postgraduate Studies, Cairo University, Giza, Egypt

[8]International Center for Agricultural Research in the Dry Areas (ICARDA), Rabat, Morocco

Correspondence

Ahmed M. S. Kheir, International Center for Agricultural Research in the Dry Areas (ICARDA), Maadi, 11728, Egypt. Email: drahmedkheir2015@gmail.com; a.kheir@cgiar.org

Assigned to Associate Editor Yao Zhang

Abstract

Traditional simulation models are often point based; thus, more research is needed to emphasize spatial simulation, providing decision-makers with fast recommendations. Combining machine learning algorithms with spatial process-based models could be considered an appropriate solution. We created a spatial model in R (APSIMx_R) to generate fine-resolution data from coarse-resolution data, which is typically available at the regional level. The APSIM crop model outputs were then deployed to train and test the artificial neural network, creating a hybrid modeling approach for robust spatial simulations. The APSIMx_R package facilitates preparing the required model inputs, executes the prediction, processes, and analyzes the APSIM crop model outputs. This note demonstrates the use of a new approach for creating reproducible crop modeling workflows with the spatial APSIM next-generation model and machine learning algorithms. The tool was deployed for spatial and temporal simulation of potential wheat yield under different nitrogen rates and various wheat cultivars. The spatial APSIMx_R was validated by comparing the simulated yield at 100 kg N ha^{-1} to the analogues' actual yield at the same grid points, which showed good agreement (d = 0.89) between the spatially predicted and actual yield. The hybrid approach increased such precision, resulting in higher agreement (d = 0.95) with actual yield. When the interaction between cultivars and nitrogen levels was considered, it was found that the novel cultivar Sakha95 is nitrogen voracious, exhibiting a larger drop in yield (65%) under minimal nitrogen treatment (0 kg N ha^{-1}) relative to the potential yield.

Abbreviations: ANN, artificial neural network; APSIM, agricultural production system simulation; GEM, genotype × environment × management; ML, machine learning; RB, relative bias.

1 | INTRODUCTION

New technology will transform farming and agribusiness, making it more profitable (Popescu et al., 2022). Crop models have been widely used in cropping system simulations for a variety of purposes, including agricultural water management (Kheir et al., 2021), food security and nutrition, and genotype × environment × management (G × E × M) interactions (Cooper et al., 2021), and narrowing yield gaps (Asseng et al., 2018; Getnet et al., 2022). Crop models have recently demonstrated the ability to predict genetic yield potential despite significant difficulties (Guarin et al., 2022). The majority of crop models now in use are created for field-scale modeling and simulate agronomic factors using homogenous (average) field conditions (You et al., 2022). However, simulation at the field scale is no longer sufficient to address precision agriculture concerns (Pasquel et al., 2023), confirming the necessity of spatial simulation, which shifts crop model simulation scale from field scale to finer scale (Pasquel et al., 2022). In addition, if models are parameterized with coarse-scale information, such as an entire country, and then used to address consequences at a finer geographical scale, there is a greater possibility of generating incorrect inferences due to substantial biases (Ogle et al., 2006). Spatial simulation can also help in determining the yield heterogeneity at regional scale and contribute to understanding the nature of recent yield progress (Lobell & Azzari, 2017; van Ittersum et al., 2013; van Wart et al., 2013). One of the agricultural system models that has been developed through many years of research and has been used to comprehend G × E × M effects on yield under present changing climatic scenarios is the Agricultural Production System Simulation (APSIM) model (Asseng et al., 2002; D. P. Holzworth et al., 2014; Keating et al., 2003; Pasquel et al., 2023). The APSIM initiative recently upgraded APSIM to APSIM Next Generation, which can be run on different operating systems. APSIM Next Generation contains several lines of code in different programming languages, integrates multiple disciplines, includes more complex farming systems, and runs faster for larger simulation analysis (D. Holzworth et al., 2018). Most crop modelers use the APSIM graphical user interface to prepare the input files and execute the model, but in large-scale studies, this method is time consuming, especially when big data and complex management practices are used. The most ideal solution for this issue is to develop ad hoc scripts in any appropriate programming language (i.e., Python, R, and SAS) to automate various stages of the simulations. This will allow users to quickly generate input files, perform simulations, and make the scripts easier and more available for other user groups. Although there are various packages developed in R (Alderman, 2020) and Python (X. He et al., 2015) for running the Decision Support System for Agrotechnology Transfer Cropping Systems Model (DSSAT; Hoogenboom, Porter, Boote et al., 2019; Hoogenboom, Porter, Shelia et al., 2019; Jones et al., 2003), no packages were developed with the APSIM next-generation model for the spatial simulation, indicating the state-of-the-art of the developed script in this application note. The generic framework of APSIMx_R can manipulate all inputs of the model, such as soil features, weather datasets, management practices, and genetic parameters, in a customized grid cell resolution. Machine learning (ML) algorithms could be used for simulations in approaches similar to biophysical models by creating links between input parameters (production factors) and target variables (Reichstein et al., 2019). Execution of ML in the cloud, as Google colaboratory (colab), is much better than in any other environment (Elnashar et al., 2020; Zeng et al., 2022). ML outperforms crop models because it first considers new features such as yield trend (technology advances) that most crop models do not have, works well in large datasets, extracts the important features (Falconnier et al., 2023; Kheir et al., 2022a, 2022b , Attia et al., 2022), and considers a biotic stress factors (Pradhan et al., 2023) and other features such as topography. Meanwhile, crop models have distinct advantages when it comes to mimicking plant physiological processes, demonstrating the importance of integrating both MLs and crop models to improve prediction accuracy (Attia et al., 2022). In addition, more research is needed to integrate crop model-based programming scripts with ML algorithms. This application note shows a stepwise process on how to develop APSIMx_R and integrate with an artificial neural network (ANN), an ML model for simulating crop management practices system spatial simulations at various spatial scales.

Core Ideas

- Crop models are frequently point based, while developing spatial models is required.
- We developed a spatial Agricultural Production System Simulation model in R to generate fine-resolution data.
- The spatial model-based R was integrated with an artificial neural network, creating a hybrid approach.
- The developed approach is used to determine the yield heterogeneity at scale.
- The hybrid approach's simulated yield correlated positively with farmer yield.

2 | APSIMX_R_ML APPROACH

A field experiment was conducted at the Menoufya location (latitude 30.7 and longitude 31) in the Nile Delta during two

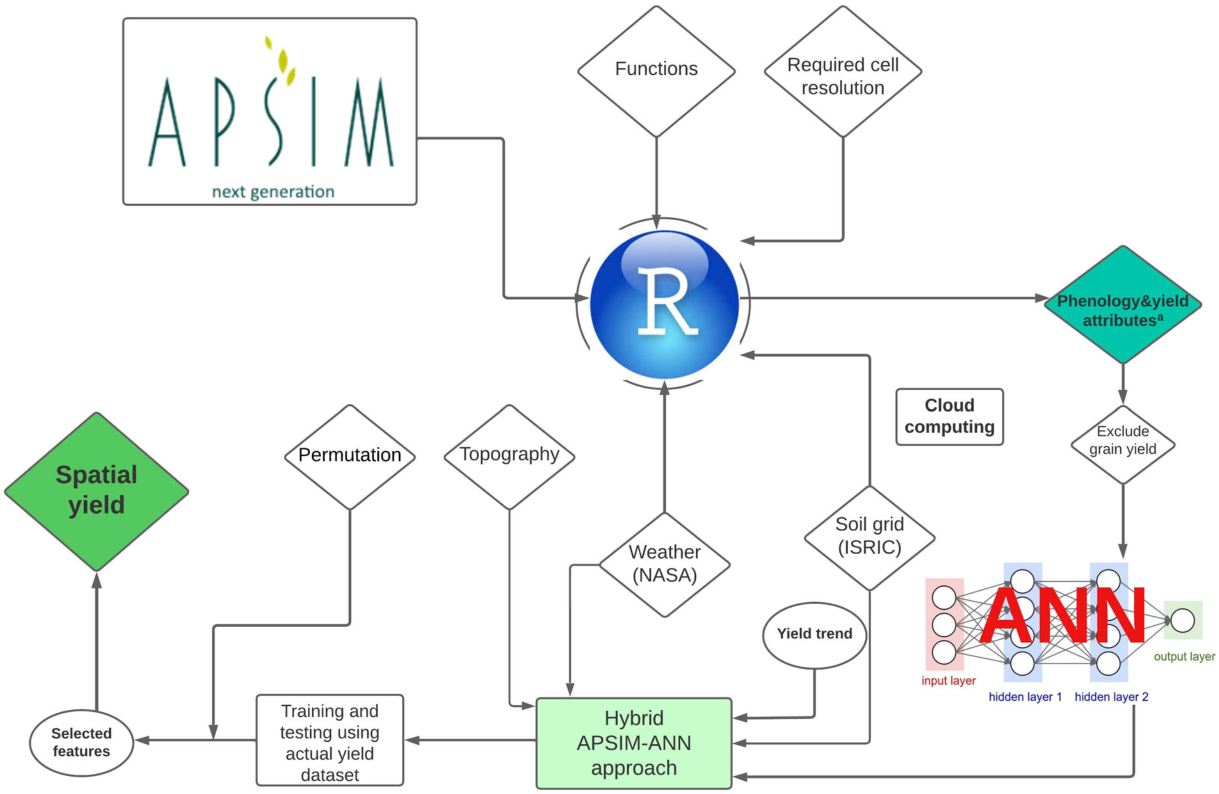

FIGURE 1 Flowchart of the process and steps required to form the hybrid APSIMx_R_ML approach and key outputs. First, APSIMx was executed in R using different functions, daily weather data from NASA power, and soil data from International Soil Reference and Information Centre (ISRIC) to simulate wheat phenology, yield, and yield attributes of three calibrated cultivars at the required cell resolution. The agricultural production system simulation (APSIM) inputs (soil and weather) and outputs excluding grain yield such as phenology, leaf area index, grain size, grain protein, and grain number in unit area were integrated with other datasets of yield trend and topography (elevation, slope, and aspect) to train and test artificial neural network (ANN). Actual yield was used as dependent variable in training and testing ANN. The default method by dividing the entire data set into 80% training and 20% testing using a random selection approach was used to train and test ANN. [a]Phenology and maturity dates, yield attributes include leaf area index, grain protein, grain size, and grain number per unit area, were used as inputs (independent variables) to ANN.

growing seasons (2019/2020 and 2020/2021) with different wheat (*Triticum aestivum*) cultivars (Sakha95, Giza171, and Misr3). Various datasets were obtained from this experiment to calibrate and evaluate the APSIM model. Soil physical and chemical properties, daily climatic dataset, soil initial irrigation, and fertilization conditions before each growing season, grain yield, biomass yield, leaf area index at anthesis, number of grains per m^2, grain size, anthesis date, maturity date, nitrogen content in grains, irrigation water applied and dates of application, and fertilization (time and doses) were all measured in this dataset. The manual calibration method was used, and Table S1 shows the calibrated parameters for each cultivar. The APSIM R package necessitates first a recall to the APSIMX file, which contains its main components in the interface of the target directory (Figure 1). The source code contains the APSIM spatial function, working directory, selected spatial resolution, target country shapefile, number of years required for the weather dataset, crop type, simulation time, start and end of simulation, target cultivar, and installation of various packages (Supporting Information). Egypt is the case study (Figure S1), and wheat is the target crop because of its importance to food security and the fact that Egypt is the world's largest importer of wheat. APSIMx_R was used to spatially predict potential yield and yield under three nitrogen fertilizer levels (0, 100, and 200 kg N ha^{-1}) in Egypt at 1° grid resolution for three novel calibrated cultivars of wheat (Sakha95, Giza171, and Misr3) over 30 years (1991–2020). The model outputs included wheat phenology (anthesis and maturity dates), aboveground biomass, nitrogen content in biomass, grain yield, grain protein, grain size, and grain number. The outputs except grain yield with other inputs to APSIMx_R, such as weather data and soil data, were combined with other datasets, including technology trend and topography dataset (elevation and slope), to train and validate the ANN algorithm using the actual dataset of yield over 30 years, creating a hybrid approach (Figure 1). The observed yield increased over time from 1991 to 2020, except in the warmest years (2010), demonstrating

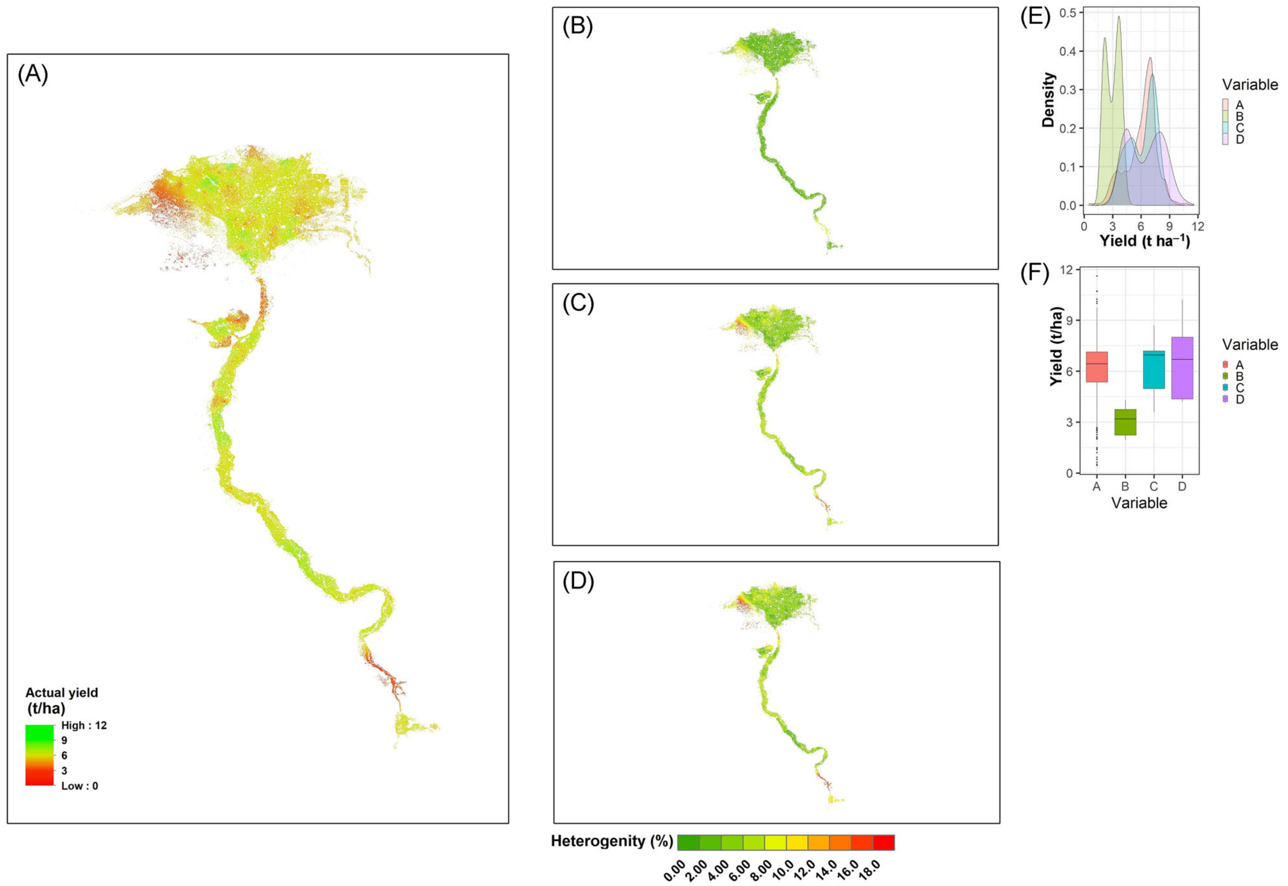

FIGURE 2 Actual spatial yield (A) and heterogeneity of simulated yield under zero application of nitrogen (B), 100 kg N ha^{-1} (C), and yield at 200 kg N ha^{-1} (D) over Egypt. Simulation has been done by the developed approach APSIMx_R_ANN at resolution of 1.0° averaged over 30 years (1990–2020) and three cultivars. (E) represents the distribution of yield for different variables as actual yield (A), yield at zero application of nitrogen (B), yield at 100 kg N ha^{-1} (C), and yield at 200 kg N ha^{-1} (D). Boxplots (F) demonstrate a visual indication of how a dataset's mean, median, mode, minimum, maximum, and outlier values are spread out and compared to each other for the different variables actual yield (A), yield at zero application of nitrogen (B), yield at 100 kg N ha^{-1} (C), and yield at 200 kg N ha^{-1} (D).

the importance of including the technology trend in the simulation. The increased production was mostly ascribed to technological advances in farm management, genetic enhancement programs, agricultural mechanization, and other technologies (Günay et al., 2021; Moeinizade et al., 2020). Despite the relevance of crop models in simulating crop phenology and physiology over ML, integrating other inputs that crop models (CMs) cannot examine, such as technology trend and geography, highlights the importance of building a hybrid approach from crop models and ML. The application of permutation importance in ANN was used to find the most significant features to enhance prediction accuracy while excluding non-important ones (Altmann et al., 2010).

The ANN algorithm was built in Python and executed in Google Colab as a cloud computing environment, and the source code is hosted on an open-source project on GitHub (https://github.com/DrAhmedKheir/tANN-.git). We selected ANN rather than other ML algorithms because deep learning has different hidden layers and can learn and model nonlinear and complex relationships, which is critical because many of the relationships between inputs and outputs in real life are nonlinear and complex (Kheir, Ammar, Attia et al., 2022; Kheir et al., 2023). ML approaches have some advantages over crop models, including the capacity to involve additional input variables that crop models cannot. In the current case, technology trend and topography (i.e., elevation, slope, and aspect) were included as external variables in ML, highlighting the outperformance of ML over CMs. The hybrid approach was used to predict potential yield and yield under different N levels due to the importance of nitrogen to crop growth and development (Bhattarai et al., 2021; Figure 2).

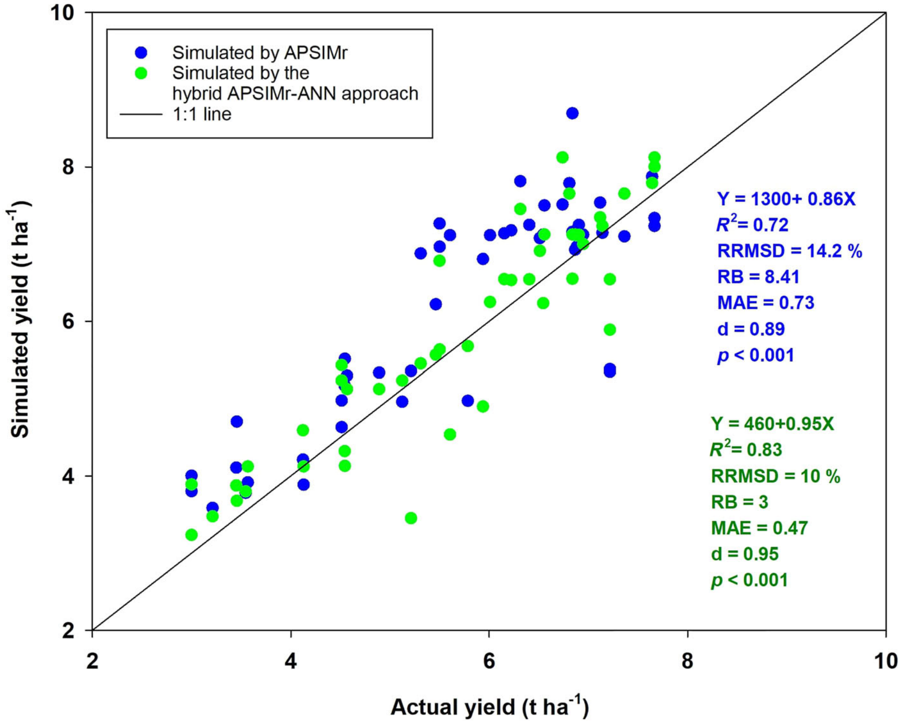

FIGURE 3 Actual and spatial simulated yield at 100 kg N ha^{-1} by APSIMx_R (blue), and hybrid APSIMx-R_ANN approach (green) averaged over 30 years at 1° resolution. Determination coefficient (R^2), relative root mean square deviation (RRMSD), relative bias (RB), mean absolute error (MAE), and Willmott degree of agreement (*d*) represent the correlation significance of both approaches with the actual yield.

3 | VALIDATION AND APPLICATION OF THE DEVELOPED APPROACH IN SPATIAL SIMULATION OF WHEAT YIELD

Egypt's Ministry of Agriculture and Land Reclamation (MALR, 2021) provided the actual on-farm wheat production. These datasets for wheat grain yield from 50 locations covering all agroclimatic zones were collected during a 30-year period (1991–2020) and used to train and validate the developed approach. The developed approach was used for the spatial simulation of wheat yield in the Excellence in Agronomy (EiA), a CGIAR-funded initiative specifically under the Government of Egypt use case at cell resolution 1.0° for different cultivars under various nitrogen fertilizer application practices. The simulation was conducted over 30 years and included potential yield, yield without application of N fertilizer, yield at 100 kg N ha^{-1} (common application with smallholders), and 200 kg N ha^{-1} (Figure 2). The total county yield heterogeneity (Figure 2B–D) was computed as the difference between the 95th percentile yield and the mean yield (Lobell & Azzari, 2017). The spatial pattern of predicted yield was in agreement with actual yield, particularly under moderate and high nitrogen doses (Figure 2E). The highest producing areas were primarily in the northern parts (Nile delta), with yields reaching up to 12 t ha^{-1}, while yields dropped in the western and southern parts due to salinity and temperature stressors, respectively. Yield heterogeneity ranged from 0% to 18% of mean yield, suggesting that smallholder yield may be raised by 15% on average if all farmers achieved the 95th percentile of present yield (Figure 2B–D). The greater heterogeneity occurred in the western and southern parts, which coincided with lower yielding counties. The density distribution of the spatially simulated yield under 100 kg N ha^{-1} and 200 kg N ha^{-1} was in agreement with the actual yield, while yield under 0 kg N ha^{-1} showed left skewness and a lower yield (Figure 2E). The application rate of 100 kg N ha^{-1} was selected in the simulation because it is the most common application rate among smallholder farmers in Egypt, allowing us to validate our approach by comparing simulated yield to actual farmer yield (Figure 3). In this regard, APSIMx_R's long-term simulation of wheat yield revealed a good agreement ($d = 0.89$) with actual yield at the same nitrogen fertilizer application (Figure 3). When the hybrid approach of APSIMx_R_ANN was used (Figure 3), such agreement increased ($d = 0.95$) and the relative bias (RB) decreased to 3%. In addition to correlation (R^2) and Willmott degree of agreement (*d*), different statistical indicators, such as relative root mean square deviation (RRMSD), RB, and mean absolute error (MAE), were considered and confirmed the synchronization with actual yield, demonstrating

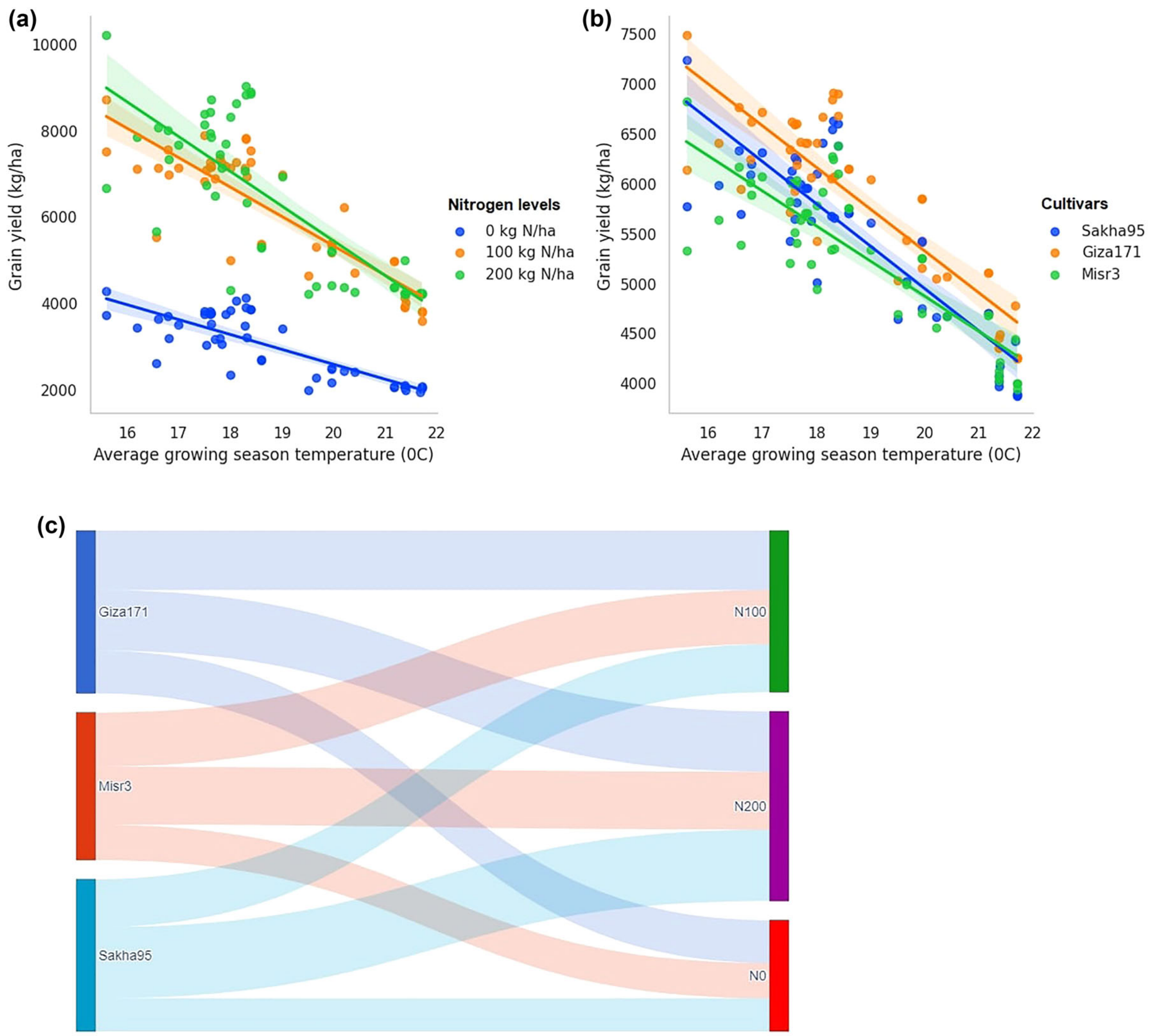

FIGURE 4 Predicted grain yield by the hybrid approach under different nitrogen (N) levels with average growing season temperature (A), different cultivars (B), Sankey plot of yield subjected to nitrogen levels, and cultivars (C). Panel (C) visualizes the flow from cultivars to nitrogen application rates to explore the best cultivar under higher N rates and vice versa.

the importance of the developed approach for mapping yield gap analysis in similar or different environments.

Wheat grain yield declined with increasing growing season temperatures when fertilization, cultivar interactions, and average growing season temperatures were included (Figure 4A,B). The yield decline with increasing growing season mean temperature was greater with zero nitrogen fertilization than with a higher N rate (Figure 4A). Giza171 surpassed other cultivars in terms of yield increase with rising growing season mean temperatures, followed by Sakha95 and Misr3 (Figure 4B). Considering the sensitivity of the cultivars investigated to N levels, Figure 4C revealed that Sakha95 is the most responsive cultivar to nitrogen fertilization, recording the lowest yield with zero nitrogen fertilization and the highest yield under 200 kg N ha^{-1} treatment. Under low nitrogen conditions, however, Giza171 outperformed all other cultivars. As a result, Giza171 is the best cultivar for hot temperatures and low nitrogen levels, whereas Sakha95 is the best for normal conditions. Accordingly, Sakha95 showed the lowest yield reduction (48% and 65%) under 0 kg N ha^{-1}, and 100 kg N ha^{-1}, respectively, relative to the potential yield (Figure 5). The hybrid APSIMx_R_ML approach predicted spatial potential yield and yield under varied nitrogen levels for several genotypes, proving its utility in assessing yield gaps under a wide variety of GEM interactions.

4 | DISCUSSION AND FUTURE DIRECTIONS

ML techniques were recently used in site-specific recommendations (Qin et al., 2018), but further application in an integrated approach with spatial modeling has received less

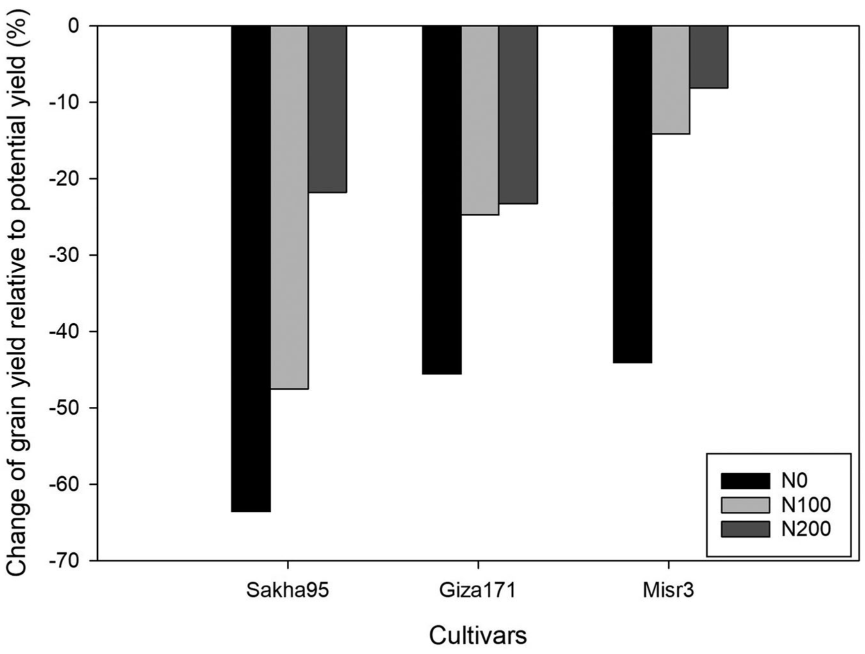

FIGURE 5 Change of grain yield for different cultivars under various nitrogen (N) levels relative to the potential yield.

attention so far. Field survey data of actual wheat yields from multi-farms covering the most heterogeneous soils, management practices, and temperature gradients would be considerably superior to using present statistics on yield in limited areas (50 locations) in verifying the developed approach. This is the main limitation of the current study, which may be overcome by conducting a large-scale survey to collect actual farmers' yields, but this would necessitate more financing. The target cultivars (Sakha95, Giza171, and Misr3) were calibrated and evaluated in APSIMx using a field dataset from two Nile delta growing seasons. Meanwhile, the best calibrations necessitate datasets from different locations and growing seasons to assure heterogeneity in climate, soils, and topography (Coudron et al., 2021; Kheir, Ammar, Attia et al., 2022), which adds another limitation to the current work. Estimating model parameters with little observational data raises the possibility of discovering several parameter value combinations for which the model output fits the observations equally well. This phenomenon, also known as equifinality, prevents the estimation of unique and recognizable parameter values (D. He et al., 2017). However, preparing crop growth and phenology for a large number of fields using standard ground-based sample methods can be inconveniently time-consuming, opening the way for the use of remote sensing datasets as a quick and low-cost method (Ko et al., 2006; Maas, 1993; Xia et al., 2021). The integration of the developed APSIMr package with ANN acts as a scientific modeling workflow that provides fast, accurate, and low-cost spatiotemporal simulation, which is very important for food security and nutritional analysis. Further advancements may include improvements in biophysical functions that involve a multi-model ensemble as well as the use of deep learning algorithms in conjunction with ML. The developed approach should be applied under broad range of GEM interactions using remote sensing observations. This will aid in identifying the yield gap of the targeted crops as well as the innovations that will fill these gaps and support food security, especially in light of climate change, rapid population growth, and limited natural resources. Furthermore, attention should be paid to use the developed approach in quantifying greenhouse gas emissions and the impact of cover crops on soil health (Joshi et al., 2022).

AUTHOR CONTRIBUTIONS

Ahmed Kheir: Conceptualization; data curation; formal analysis; writing—original draft; writing—review and editing. **Siyabusa Mkuhlani**: Resources; software. **Jane Mugo**: Investigation; methodology. **Abdelrazek Elnashar**: Data curation; formal analysis; writing—original draft. **Vinay Nangia**: Investigation; validation; visualization; writing—original draft. **Medha Devare**: Funding acquisition; validation. **Ajit Govind**: Supervision; writing—review and editing.

ACKNOWLEDGMENTS

The authors would like to thank the CGIAR Excellence in Agronomy-Egypt Use Case (https://www.cgiar.org/initiative/11-excellence-in-agronomy-eia-solutions-for-agricultural-transformation/) where this work was conceptualized and is currently being undertaken.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Ahmed M. S. Kheir https://orcid.org/0000-0001-9569-5420
Jane W. Mugo https://orcid.org/0000-0003-3836-9113
Vinay Nangia https://orcid.org/0000-0001-5148-8614
Medha Devare https://orcid.org/0000-0003-0041-4812
Ajit Govind https://orcid.org/0000-0002-0656-0004

References

Alderman, P. D. (2020). A comprehensive R interface for the DSSAT Cropping Systems Model. *Computers and Electronics in Agriculture*, *172*, 105325. https://doi.org/10.1016/j.compag.2020.105325

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Asseng, S., Kheir, A. M. S., Kassie, B. T., Hoogenboom, G., Abdelaal, A. I. N., Haman, D. Z., & Ruane, A. C. (2018). Can Egypt become self-sufficient in wheat? *Environmental Research Letters*, *13*(9), 094012. https://doi.org/10.1088/1748-9326/aada50

Asseng, S., Turner, N. C., Ray, J. D., & Keating, B. A. (2002). A simulation analysis that predicts the influence of physiological traits on the potential yield of wheat. *European Journal of Agronomy*, *17*(2), 123–141. https://doi.org/10.1016/S1161-0301(01)00149-6

Attia, A., Govind, A., Qureshi, A. S., Feike, T., Rizk, M. S., Shabana, M. M. A., & Kheir, A. M. S. (2022). Coupling process-based models and machine learning algorithms for predicting yield and evapotranspiration of maize in arid environments. *Water*, *14*(22), 3647. https://doi.org/10.3390/w14223647

Bhattarai, D., Kumar, S., & Nleya, T. (2021). Nitrogen and sulfur fertilizers effects on growth and yield of *Brassica carinata* in South Dakota. *Agronomy Journal*, *113*(2), 1945–1960. https://doi.org/10.1002/agj2.20501

Cooper, M., Voss-Fels, K. P., Messina, C. D., Tang, T., & Hammer, G. L. (2021). Tackling G × E × M interactions to close on-farm yield-gaps: Creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. *Theoretical and Applied Genetics*, *134*(6), 1625–1644. https://doi.org/10.1007/s00122-021-03812-3

Coudron, W., Gobin, A., Boeckaert, C., De Cuypere, T., Lootens, P., Pollet, S., Verheyen, K., De Frenne, P., & De Swaef, T. (2021). Data collection design for calibration of crop models using practical identifiability analysis. *Computers and Electronics in Agriculture*, *190*, 106457. https://doi.org/10.1016/j.compag.2021.106457

Elnashar, A., Zeng, H., Wu, B., Zhang, N., Tian, F., Zhang, M., Zhu, W., Yan, N., Chen, Z., Sun, Z., Wu, X., & Li, Y. (2020). Downscaling TRMM monthly precipitation using Google Earth Engine and Google Cloud computing. *Remote Sensing*, *12*(23), 3860. https://doi.org/10.3390/rs12233860

Falconnier, G. N., Leroux, L., Beillouin, D., Corbeels, M., Hijmans, R. J., Bonilla-Cedrez, C., van Wijk, M., Descheemaeker, K., Zingore, S., Affholder, F., Lopez-Ridaura, S., Malézieux, E., Makowski, D., Rurinda, J., van Ittersum, M. K., Vanlauwe, B., Giller, K. E., Lammoglia, S.-K., & Waha, K. (2023). Increased mineral fertilizer use on maize can improve both household food security and regional food production in East Africa. *Agricultural Systems*, *205*, 103588. https://doi.org/10.1016/j.agsy.2022.103588

Getnet, M., Descheemaeker, K., van Ittersum, M. K., & Hengsdijk, H. (2022). Narrowing crop yield gaps in Ethiopia under current and future climate: A model-based exploration of intensification options and their trade-offs with the water balance. *Field Crops Research*, *278*, 108442. https://doi.org/10.1016/j.fcr.2022.108442

Guarin, J. R., Martre, P., Ewert, F., Webber, H., Dueri, S., Calderini, D., Reynolds, M., Molero, G., Miralles, D., Garcia, G., Slafer, G., Giunta, F., Pequeno, D. N. L., Stella, T., Ahmed, M., Alderman, P. D., Basso, B., Berger, A. G., Bindi, M., … Asseng, S. (2022). Evidence for increasing global wheat yield potential. *Environmental Research Letters*, *17*(12), 124045. https://doi.org/10.1088/1748-9326/aca77c

Günay, E. E., Okudan Kremer, G. E., & Zarindast, A. (2021). A multi-objective robust possibilistic programming approach to sustainable public transportation network design. *Fuzzy Sets and Systems*, *422*, 106–129. https://doi.org/10.1016/j.fss.2020.09.007

He, D., Wang, E., Wang, J., & Robertson, M. J. (2017). Data requirement for effective calibration of process-based crop models. *Agricultural and Forest Meteorology*, *234–235*, 136–148. https://doi.org/10.1016/j.agrformet.2016.12.015

He, X., Peng, L., & Sun, H. (2015). *pyDSSAT documentation release 1.0*. http://xiaoganghe.github.io/pyDSSAT/doc/_downloads/pyDSSATdoc.pdf

Holzworth, D., Huth, N. I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Verrall, S., Herrmann, N. I., Zheng, B., & Snow, V. (2018). APSIM next generation: Overcoming challenges in modernising a farming systems model. *Environmental Modelling & Software*, *103*, 43–51. https://doi.org/10.1016/j.envsoft.2018.02.002

Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P. M., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., … Keating, B. A. (2014). APSIM—Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, *62*, 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009

Hoogenboom, G., Porter, C. H., Boote, K. J., Shelia, V., Wilkens, P. W., Singh, U., White, J. W., Asseng, S., Lizaso, J. I., Moreno, L. P., Pavan, W., Ogoshi, R., Hunt, L. A., Tsuji, G. Y., & Jones, J. W. (2019). The DSSAT crop modeling ecosystem. In K. J. Boote (Ed.), *Advances in crop modeling for a sustainable agriculture* (pp. 173–216). Burleigh Dodds Science Publishing. https://doi.org/10.19103/AS.2019.0061.10

Hoogenboom, G., Porter, C. H., Shelia, V., Boote, K. J., Singh, U., White, J. W., Hunt, L. A., Ogoshi, R., Lizaso, J. I., Koo, J., Asseng, S., Singels, A., Moreno, L. P., & Jones, J. W. (2019). *Decision support system for agrotechnology transfer (DSSAT) version 4.7.5*. DSSAT Foundation. https://DSSAT.net

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, *18*(3–4), 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7

Joshi, D. R., Clay, D. E., Clay, S. A., Moriles-Miller, J., Daigh, A. L. M., Reicks, G., & Westhoff, S. (2022). Quantification and machine learning based N2O–N and CO2–C emissions predictions from a decomposing rye cover crop. *Agronomy Journal*, In press. https://doi.org/10.1002/agj2.21185

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., … Smith, C. J. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, *18*(3), 267–288. https://doi.org/10.1016/S1161-0301(02)00108-9

Kheir, A. M. S., Alrajhi, A. A., Ghoneim, A. M., Ali, E. F., Magrashi, A., Zoghdan, M. G., Abdelkhalik, S. A. M., Fahmy, A. E., & Elnashar, A. (2021). Modeling deficit irrigation-based evapotranspiration optimizes wheat yield and water productivity in arid regions. *Agricultural Water Management*, *256*, 107122. https://doi.org/10.1016/j.agwat.2021.107122

Kheir, A. M. S., Ammar, K. A., Amer, A., Ali, M. G. M., Ding, Z., & Elnashar, A. (2022a). Machine learning-based cloud computing improved wheat yield simulation in arid regions. *Computers and Electronics in Agriculture*, *203*, 107457. https://doi.org/10.1016/j.compag.2022.107457

Kheir, A. M. S., Ammar, K. A., Attia, A., Elnashar, A., Ahmad, S., El-Gioushy, S. F., & Ahmed, M. (2022b). Cereal crop modeling for food and nutrition security. In M. Ahmed (Ed.), *Global agricultural production: Resilience to climate change* (pp. 183–195). Springer International Publishing. https://doi.org/10.1007/978-3-031-14973-3_6

Kheir, A. M. S., Elnashar, A., Mosad, A., & Govind, A. (2023). An improved deep learning procedure for statistical downscaling of climate data. *Heliyon*, *9*(7), E18200. https://doi.org/18210.11016/j.heliyon.12023.e18200

Ko, J., Maas, S. J., Mauget, S., Piccinni, G., & Wanjura, D. (2006). Modeling water-stressed cotton growth using within-season remote sensing data. *Agronomy Journal*, *98*(6), 1600–1609. https://doi.org/10.2134/agronj2005.0284

Lobell, D. B., & Azzari, G. (2017). Satellite detection of rising maize yield heterogeneity in the U.S. Midwest. *Environmental Research Letters*, *12*(1), 014014. https://doi.org/10.1088/1748-9326/aa5371

Maas, S. J. (1993). Within-season calibration of modeled wheat growth using remote sensing and field sampling. *Agronomy Journal*, *85*(3), 669–672. https://doi.org/10.2134/agronj1993.00021962008500030028x

Ministry of Agriculture and Land Reclamation, MALR. (2021). (*Winter crops 2021* (Agricultural Bulletin Part 1)). Ministry of Agriculture and Land Reclamation, The Egyptian Economic Affairs Sector, Dokki, Egypt.

Moeinizade, S., Wellner, M., Hu, G., & Wang, L. (2020). Complementarity-based selection strategy for genomic selection. *Crop Science*, *60*(1), 149–156. https://doi.org/10.1002/csc2.20070

Ogle, S. M., Breidt, F. J., & Paustian, K. (2006). Bias and variance in model results associated with spatial scaling of measurements for parameterization in regional assessments. *Global Change Biology*, *12*(3), 516–523. https://doi.org/10.1111/j.1365-2486.2006.01106.x

Pasquel, D., Cammarano, D., Roux, S., Castrignano, A., Tisseyre, B., Rinaldi, M., Troccoli, A., & Taylor, J. A. (2023). *Downscaling the APSIM crop model for simulation at the within-field scale*. https://doi.org/10.2139/ssrn.4394507

Pasquel, D., Roux, S., Richetti, J., Cammarano, D., Tisseyre, B., & Taylor, J. A. (2022). A review of methods to evaluate crop model performance at multiple and changing spatial scales. *Precision Agriculture*, *23*(4), 1489–1513. https://doi.org/10.1007/s11119-022-09885-4

Popescu, G. C., Popescu, M., Khondker, M., Clay, D. E., Pampana, S., & Umehara, M. (2022). Agricultural sciences and the environment: Reviewing recent technologies and innovations to combat the challenges of climate change, environmental protection, and food security. *Agronomy Journal*, *114*(4), 1895–1901. https://doi.org/10.1002/agj2.21164

Pradhan, U. K., Meher, P. K., Naha, S., Rao, A. R., Kumar, U., Pal, S., & Gupta, A. (2023). ASmiR: A machine learning framework for prediction of abiotic stress-specific miRNAs in plants. *Functional & Integrative Genomics*, *23*(2), 92. https://doi.org/10.1007/s10142-023-01014-2

Qin, Z., Myers, D. B., Ransom, C. J., Kitchen, N. R., Liang, S.-Z., Camberato, J. J., Carter, P. R., Ferguson, R. B., Fernandez, F. G., Franzen, D. W., Laboski, C. A. M., Malone, B. D., Nafziger, E. D., Sawyer, J. E., & Shanahan, J. F. (2018). Application of machine learning methodologies for predicting corn economic optimal nitrogen rate. *Agronomy Journal*, *110*(6), 2596–2607. https://doi.org/10.2134/agronj2018.03.0222

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

van Ittersum, M. K., Cassman, K. G., Grassini, P., Wolf, J., Tittonell, P., & Hochman, Z. (2013). Yield gap analysis with local to global relevance—A review. *Field Crops Research*, *143*, 4–17. https://doi.org/10.1016/j.fcr.2012.09.009

van Wart, J., Kersebaum, K. C., Peng, S., Milner, M., & Cassman, K. G. (2013). Estimating crop yield potential at regional to national scales. *Field Crops Research*, *143*, 34–43. https://doi.org/10.1016/j.fcr.2012.11.018

Xia, Y., Guan, K., Copenhaver, K., & Wander, M. (2021). Estimating cover crop biomass nitrogen credits with Sentinel-2 imagery and sites covariates. *Agronomy Journal*, *113*(2), 1084–1101. https://doi.org/10.1002/agj2.20525

You, Y., Tian, H., Pan, S., Shi, H., Bian, Z., Gurgel, A., Huang, Y., Kicklighter, D., Liang, X.-Z., Lu, C., Melillo, J., Miao, R., Pan, N., Reilly, J., Ren, W., Xu, R., Yang, J., Yu, Q., & Zhang, J. (2022). Incorporating dynamic crop growth processes and management practices into a terrestrial biosphere model for simulating crop production in the United States: Toward a unified modeling framework. *Agricultural and Forest Meteorology*, *325*, 109144. https://doi.org/10.1016/j.agrformet.2022.109144

Zeng, H., Elnashar, A., Wu, B., Zhang, M., Zhu, W., Tian, F., & Ma, Z. (2022). A framework for separating natural and anthropogenic contributions to evapotranspiration of human-managed land covers in watersheds based on machine learning. *Science of the Total Environment*, *823*, 153726. https://doi.org/10.1016/j.scitotenv.2022.153726

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kheir, A. M. S., Mkuhlani, S., Mugo, J. W., Elnashar, A., Nangia, V., Devare, M., & Govind, A. (2023). Integrating APSIM model with machine learning to predict wheat yield spatial distribution. *Agronomy Journal*, 1–9. https://doi.org/10.1002/agj2.21470