PERFORMANCE OF K* ALGORITHM IN YIELD PREDICTION AS A DECISION SUPPORT TOOL TO DERIVE SITE-SPECIFIC NUTRIENT MANAGEMENT RECOMMENDATIONS FOR MAIZE PRODUCTION
#9431

O. Ennaji, S. Baha, L. Vergutz and A. El allali
Mohammed VI Polytechnic University, Benguerir, Morocco
e-mail: achraf.elallali@um6p.ma; tel: 00212 662 096 347

ABSTRACT

Agriculture is the main source of food and income for rural communities in developing countries, especially in Africa. Given the current population growth, pressures on agricultural systems will continue to increase. Many countries have agricultural economies that are highly dependent on agricultural productivity. For example, several variables can influence fertilization for optimal grain yields. Quantifying the effects and relative importance of soil properties such as soil type, pH, climate, regional soil variability, and policy could provide a basis for optimizing fertilization to increase economic benefits and reduce nutrient losses.

For decades, maize (Zea mays L.) management decisions have been a conundrum in smallholder agriculture in tropical climates. Alternative management practices aim to achieve an optimal balance between environmental and economic outcomes. In this context, farmers need advice on predicting their future crop productivity and analysis to help them maximize production in their fields. Nutrient management is an important issue in agriculture. Despite significant recent advances, actual, practical, and accurate application is challenging. Conventional methods have many limitations related to both the actual problem and the problem itself. The selection of meaningful parameters is critical. Machine learning (ML) as an emerging technology can help identify patterns in huge datasets. This method extrapolates predictions directly from the data provided. ML Algorithms can predict production based on genetic information, environmental factors, land management, and fertilizer rates. Our study aims to develop a decision support system based on machine learning and integrated datasets. To achieve our goal, field variables (yield, varieties, and fertilization) were combined with environmental variables from soil mapping (soil texture, pH, and organic matter concentration) and gridded meteorological datasets (precipitation, temperature). The performance of K* algorithm to predict yield using these variables as input is explored. To address this problem, precision agriculture soon will likely require the merging of different disciplines and expertise, and the development of hybrid systems that incorporate different ML and agronomic techniques.

Keywords: Machine learning, Fertilization, Site-specific, Yield prediction, Maize.

INTRODUCTION

Machine learning approaches are widely used to study the factors that influence yield and crop yield prediction (Barbosa, et al. 2020; Qin, et al. 2018; Coulibali, et al. 2020). Integrating these methods into food production could increase the precision of nutrient management, sustainable cultivation, and food security. These tools could help conserve biodiversity while increasing crop yields under different cropping systems, soil, and climate conditions (Barbosa, et al. 2020).

Indeed, crop yield prediction is one of the key elements for sustainable cultivation and optimal use of mineral resources, but this prediction is extremely complex and influenced by several factors. This complexity makes it difficult for researchers, let alone farmers, to predict the economic benefits in both the short and long term. Corn (Zea mays L.) is considered a staple crop that plays an important role in the economy not only as food, but also as feed and fuel. Growth and yield of maize cobs are affected by various variables and environmental dynamics (Jiang, et al. 2017; Ogutu, et al. 2018). Conventional statistical models, field surveys, drones and simulation models have been used in different associations to predict and explain crop yield dynamics (Liakos, et al., 2018; Chlingaryan, et al. 2018). It has been reported that each of these techniques seems to look at complexity from different perspectives.

To gain further insight and understand many of these limitations, maize has been used as a model species in many studies. Yield and fertilization aspects have been analyzed in different experimental setups and locations around the world. Recent studies have integrated these methods into different hybrid combinations.

In this study, we investigate the performance of K* algorithm in predicting yield to gain a better understanding of these constraints. This algorithm has demonstrated a great ability to analyze and predict yields using high-dimensional data.

MATERIALS AND METHODS

Study area

China is the second largest producer and importer of maize in the world (FAO 2020). The demand for maize is continuously increasing due to the increased consumption of animal products caused by the modernization of human food systems. Therefore, corn production in China is critical to global supply and demand. Better knowledge of the impact of climate on maize production in China is therefore crucial. The hybrid Zhengdan 958 has been selected as one of the leading corn hybrids currently grown in China. Zhengdan958 has higher yield, planting density, and stress tolerance than most hybrids, but few studies evaluating yield parameters were found in the literature (Lai, et al. 2010; Li, et al. 2017; Li, et al. 2015).

Plant material and data collection

Summer maize was selected for this case study because it is important for the country's food security and is widely grown in larger areas of China. According to the conventional geographical classification based on the characteristics and regional distribution of maize growing systems, the summer maize growing areas in the study area were defined as the summer maize growing areas of the North China Plain (Wu, et al. 2014), based on their climatic, growing, topographic and soil conditions. Fig. 1 shows the occurrence of the data points used in this study. The database used for this study includes maize field experiments conducted in China from 2005 to 2010 using soil testing and fertilizer recommendations (Yan, et al. 2021). that we complemented by seven Weather features from a global weather API.

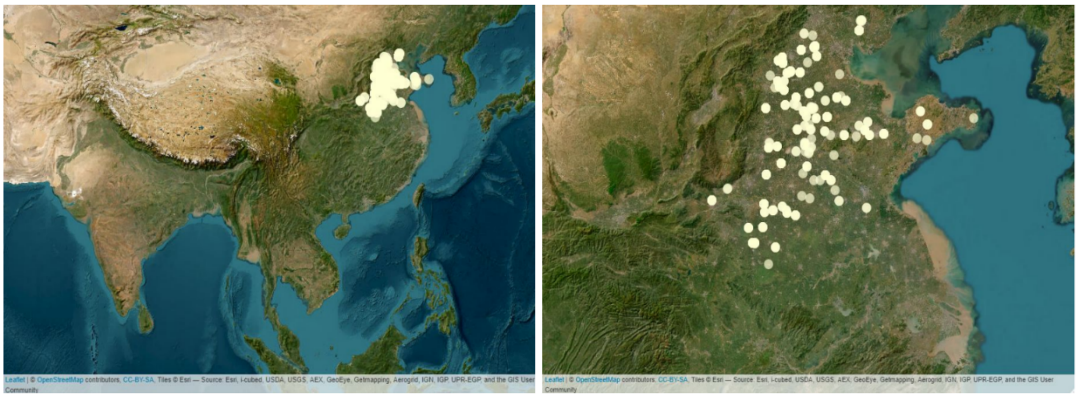

Fig. 1. The study zone in China, including the occurrence of the data points used in this study.

Data analysis

The K* algorithm was used to describe the collected and cleaned data. This machine learning algorithm was also used to evaluate the factors that influenced yield prediction (SOM (g kg^{-1}), Olsen P (mg kg^{-1}), Ava-K (mg kg^{-1}), N input (kg ha^{-1}), K_2O input (kg ha^{-1}), P_2O_5 input (kg ha^{-1}), temperature at 2 meter elevation (C), temperature at 2 meter elevation maximum, temperature at 2 meter elevation minimum, temperature at 2 meter elevation, relative humidity at 2 meter elevation, precipitation corrected (mm/day), and surface pressure (kPa) on corn yield (kg ha^{-1}). Farmers apply a mix of technologies to improve declining soil fertility and reduce yield loss. This implies that the decision to apply technologies is inherently multivariate and that attempting machine learning modeling would incorporate useful decision support functions to derive site-specific recommendations for nutrient management in corn production. Ignoring these interdependencies can lead to conflicting policy recommendations (Marenya and Barrett, 2007). Therefore, the use of multivariate models is essential. The hyperparameters of the algorithm were tuned to the best result of the 10-fold cross-validation by searching the hyperparameter space (grid search).

The performance of the trained models was determined using an independent testing dataset including 20% of the total number of data using MAE and RMSE.

RESULTS AND DISCUSSION

A comparison of actual maize yields and predicted values, including new weather features derived from the weather API, is shown in Fig. 2. This scatter plot shows that a good correlation was achieved. These results prove that integrated experimental datasets collected from various published studies can be used to perform more complex analyzes using machine learning techniques (such as the instance-based classifier K*) than in standard meta-analyzes, which are usually based on linear models. K* showed good classification performance, with values of relative absolute error and relative root mean square error of 7.05% and 14.41%, respectively. This algorithm achieved a correlation coefficient of 0.98 MAE and RMSE were 73.31 and 188.48 (kg ha^{-1}), respectively. The feature selection results emphasized the effect of independent weather features by including precipitation and temperature in the best prediction configurations. This algorithm and the proposed pipeline have shown that it can be adapted as a decision support tool for the analysis of numerous outcomes or land management systems, even with large and spatially diverse data points.

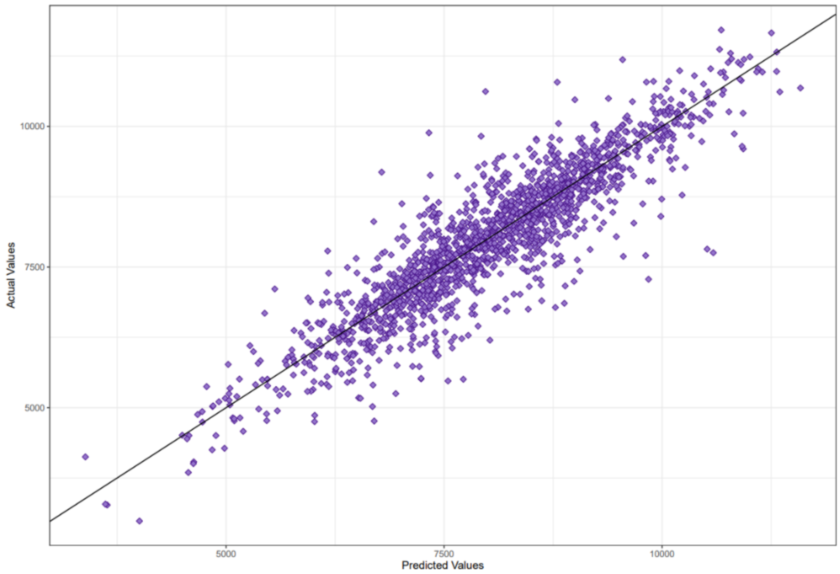

Fig. 2. Scatter-plot of measured against predicted yield (kg/ha) by K*.

CONCLUSIONS

Many maize data sets from China were used to summarize the changes in yield responses under different fertilizer rates, growing regions, and soil properties. The summer corn planting system type and the variety ZhengDan958 were used as case study plants. This study demonstrated that it is possible to generate spatially explicit yield responses for this hybrid using machine learning. This technique will aid in the development of decision support systems for smallholder farmers to make site-specific yield predictions, improve nutrient use efficiency, and increase yield, increase the economic return on fertilizer investments while reducing environmental problems. Fig. 1 provides detailed geographic information on the distribution of data points used in this study. This wide spatial diversity alongside the proposed pipeline shows that this algorithm can be adapted as a decision support tool for analyzing numerous outcomes or land management systems.

REFERENCES

Barbosa, A., Trevisan, R., Hovakimyan, N., Martin, N.F. 2020. Modeling yield response to crop management using convolutional neural networks. Computers and Electronics in Agriculture, 170(February):105197

Chlingaryan, A., Sukkarieh, S., Whelan, B. 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Computers and Electronics in Agriculture, 151(May):61–69

Coulibali, Z., Cambouris, A.N., Parent, S-E. 2020. Site-specific machine learning predictive fertilization models for potatocrops in Eastern Canada. PLOS ONE, 15(8):e0230888

Jiang, W., Liu, X., Qi, W., Xu, X., Zhu, Y. 2017. Using QUEFTS model for estimating nutrient requirements of maize in the Northeast China. Plant, Soil and Environment, 63(11):498–504

Lai, J., Li, R., Xu, X., Jin, W., Xu, M., Zhao, H., Xiang, Z., Song, W., Ying, K., Zhang, M. 2010. Genomewide patterns of genetic variation among elite maize inbred lines. Nature genetics, 42(11):1027–1030

Li, H., Yang, Q., Gao, L., Zhang, M., Ni, Z., Zhang, Y. 2017. Identification of heterosis-associated stable qtls for earweight-related traits in an elite maize hybrid zhengdan 958 by design iii. Frontiers in plant science, 8:561

Li, H., Liu, T., Cao, Y., Wang, L., Zhang, Y., Li, J., Wang, H., Tang, B. 2015. Transcriptomic analysis of maize mature embryos from an elite maize hybrid zhengdan958 and its parental lines. Plant Growth Regulation, 76(3):315–325

Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D. 2018. Machine learning in agriculture: A review. Sensors (Switzerland), 18(8):1–29

Marenya, P.P., Barrett, C.B. 2007. Household-level determinants of adoption of improved natural resources management practices among smallholder farmers in western Kenya. Food policy, 32(4):515–536

Ogutu, G.E.O., Wietse, H.P.F., Supit, I., Omondi, P., Hutjes, R.W.A. 2018. Probabilistic maize yield prediction over east africa using dynamic ensemble seasonal climate forecasts. Agricultural and forest meteorology, 250:243–261

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

Wu, L., Chen, X., Cui, Z., Zhang, W., Zhang, F. 2014. Establishing a regional nitrogen management approach to mitigate greenhouse gas emission intensity from intensive smallholder maize production. PLoS One, 9(5):e98481

Yan, X., Chen, X., Ma, C., Cai, Y., Cui, Z., Chen, X., Wu, L., Zhang, F. 2021. What are the key factors affecting maize yield response to and agronomic efficiency of phosphorus fertilizer in China? Field Crops Research, 270:108221