Prediction of Sugarcane Yield in Guangxi Sugarcane Region Based on Multiple Machine Learning Algorithms: Postprint

Authors: Shi Jiefeng, Huang Wei, Fan Xieyang, Li Xiuhua, Lu Yangxu, Jiang Zhuhui, Wang Zeping, Luo Wei, Zhang Muqing

Date: 2023-08-14T00:00:00+00:00

Abstract

[Purpose/Significance] Analyze the relationship between sugarcane yield and meteorological factors in the main sugarcane-producing regions of Guangxi, predict sugarcane yield using meteorological data, and provide scientific data support for sugar mills and relevant management departments. [Methods] Utilizing yield data from sugarcane regions in five different prefecture-level cities in Guangxi from 2002 to 2019 and 14 types of daily meteorological data, correlation analysis was conducted between the mean values of each meteorological factor over 78 continuously increasing monthly periods each year and the yield. Key meteorological factors were identified based on the sensitive period analysis method, and the impact of each meteorological factor on yield during sensitive periods was analyzed. Single-region yield prediction models were established using BP Neural Network (BPNN), Support Vector Machine (SVM), Random Forest (RF), and Long Short-Term Memory (LSTM) networks respectively, with a control experiment using meteorological mean values from the entire growth period as model input. The Hodrick-Prescott Filter (HP Filter) was used to separate meteorological yield of sugarcane, data from the five sugarcane regions were combined, and universal multi-region meteorological yield prediction models were established using RF, SVM, BPNN, and LSTM respectively. [Results and Discussion] For single regions, the model prediction effect of the sensitive period analysis method was significantly superior to the method using meteorological mean values from the entire growth period. The LSTM model's prediction effect for both data processing methods was significantly better than the currently widely used BPNN, SVM, and RF models. The LSTM model based on the sensitive period analysis method achieved overall Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) of 10.34 t/ha and 6.85%, respectively, with a coefficient of determination (Rv2) of 0.8489. For multi-regions, the LSTM prediction results were poor, while the RF, SVM, and BPNN prediction models all achieved good performance. The best-performing BPNN model achieved RMSE and MAPE of 0.98 t/ha and 9.59%, respectively, with Rv2 of 0.965. [Conclusion] The key meteorological factors selected through the sensitive period analysis method showed significant correlation with yield, and the impact of each meteorological factor on yield could be accurately analyzed according to sensitive periods. Using the LSTM model to predict single-region yield and using the BPNN model to predict multi-region sugarcane meteorological yield are feasible approaches, with prediction errors within acceptable ranges.

Full Text

Yield Prediction Models in Guangxi Sugarcane Planting Regions Based on Machine Learning Methods

SHI Jiefeng[1], **HUANG Wei**[1], **FAN Xieyang**[1], **LI Xiuhua**[1,2*], **LU Yangxu**[1], **JIANG Zhuhui**[3], **WANG Zeping**[4], **LUO Wei**[1], **ZHANG Muqing**[2]

[1]School of Electrical Engineering, Guangxi University, Nanning 530004, China
[2]Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Nanning 530004, China
[3]Guangxi Sugar Industry Group Co., Ltd., Nanning 530022, China
[4]Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences, Nanning 530007, China

---

Abstract

[**Objective**] Accurate prediction of sugarcane yield in Guangxi can provide an important reference for government policy formulation and decision-making support for farmers, thereby improving both yield and quality while promoting the sustainable development of the sugarcane industry. This study aims to explore the relationship between sugarcane yield and meteorological factors in major production regions of Guangxi Zhuang Autonomous Region, using meteorological data to predict sugarcane yield and provide scientific data support for sugar mills and relevant management departments.

[**Methods**] The study area comprised five sugarcane planting regions located in five different prefecture-level cities across Guangxi. Average yield data per hectare for each region from 2002 to 2019 were provided by Guangxi Sugar Industry Group, which controls the sugar refineries in each region. Daily meteorological data encompassing 14 meteorological factors were acquired from the National Data Center for Meteorological Sciences. Recognizing that meteorological factors exert differential influences on sugarcane growth across various time spans, we defined a new type of factor combining meteorological variables with temporal dimensions—such as average precipitation in August or mean temperature from February to April. The inter-correlations among all meteorological factors across different time spans and their correlations with yield were analyzed to screen for key meteorological factors during sensitive periods. Four algorithms—BP neural network (BPNN), support vector machine (SVM), random forest (RF), and long short-term memory (LSTM)—were employed to establish sugarcane apparent yield prediction models for each individual region. Corresponding reference models based on annual average meteorological factors were also constructed for comparison. Additionally, meteorological yields for each region were extracted using HP filtering, and general meteorological yield prediction models were developed using mixed data from all five regions with RF, SVM, BPNN, and LSTM approaches.

[**Results and Discussion**] Correlation analysis revealed that different planting regions exhibited distinct sensitive meteorological factors and key time spans. The most representative factors included sunshine hours, precipitation, and atmospheric pressure. In Region 1, sunshine hours during October–November showed the strongest negative correlation with yield, while minimum relative humidity in November demonstrated the strongest positive correlation. In Region 2, average vapor pressure during February–March exhibited the maximum positive correlation, whereas precipitation in August–September showed the maximum negative correlation. In Region 3, precipitation from 20:00 to 20:00 (next day) during August–September had the strongest positive correlation, while sunshine hours in the same period showed the strongest negative correlation. In Region 4, precipitation from 20:00 to 20:00 during March–December demonstrated the strongest positive correlation, while maximum atmospheric pressure from August–December showed the strongest negative correlation. In Region 5, average vapor pressure during June–August exhibited the strongest positive correlation, while minimum atmospheric pressure in February–March showed the strongest negative correlation.

For each specific planting region, the apparent yield prediction models based on sensitive meteorological factors during key time spans significantly outperformed those using annual average meteorological values. The LSTM model demonstrated substantially better performance than the widely used classical BPNN, SVM, and RF models for both types of meteorological data (sensitive time spans vs. annual averages). The overall root mean square error (RMSE) and mean absolute percentage error (MAPE) for the LSTM model using key time spans were 10.34 t/ha and 6.85%, respectively, with a coefficient of determination (Rv^2) of 0.8489 between predicted and actual values. For general meteorological yield prediction across multiple sugarcane planting regions, the RF, SVM, and BPNN models all achieved excellent results, with BPNN performing best (RMSE = 0.98 t/ha, MAPE = 9.59%, $Rv^2 = 0.965$). In contrast, the LSTM model yielded RMSE = 0.25 t/ha, MAPE = 39.99%, and $Rv^2 = 0.77$, indicating it is less suitable for multi-region joint prediction.

[**Conclusion**] Sensitive meteorological factors identified during key time spans show significantly stronger correlations with yield than annual average meteorological factors. While the LSTM model excels at apparent yield prediction for individual regions compared to BPNN, SVM, and RF, the BPNN model demonstrates superior performance in predicting meteorological yield across multiple sugarcane planting regions. The prediction errors for both single-region and multi-region models fall within acceptable ranges, suggesting that the proposed methodology provides a valuable reference for regional sugarcane yield forecasting.

Keywords: meteorological factor; HP filter; sugarcane yield; BPNN model; LSTM model; machine learning

1 Introduction

Regional crop yield prediction is crucial for national food security assessment. As a strategic national commodity, sugar derives 87% of its production from sugarcane, with Guangxi ranking first in national output, accounting for approximately 70% of total production in recent years. Large-scale yield estimation for Guangxi sugarcane provides scientific data support for sugar mills and management departments. Beyond land productivity, field crop yields are primarily constrained by anthropogenic management factors (fertilization, irrigation, pest control) and climatic factors (rainfall, sunlight, wind speed). Geographic and management practices remain relatively stable, making climate the least controllable yield determinant.

Increasingly, researchers have focused on meteorological data-based crop yield prediction. Gao et al. established an early rice yield prediction model for Gaoyao District, Zhaoqing City, Guangdong Province, using multi-year meteorological data, achieving 80.23% average accuracy. Yu et al. employed a genetic algorithm-optimized neural network for sugarcane yield prediction, attaining $R^2 = 0.842$. Chen developed maize yield prediction models for Yangling, Heyang, and Changwu regions in Shaanxi using historical meteorological data and the CERES-maize model, with predictions closely approaching actual yields (mean absolute relative error $< 15\%$). Burdett and Wellen used multiple machine learning methods to predict maize and soybean yields, with random forest (RF) achieving the best performance ($R^2 = 0.85$ and 0.94, respectively).

Despite extensive research on meteorological yield prediction, few studies have examined monthly analysis of meteorological impacts on sugarcane yield or developed models adaptable to multiple sugarcane regions. This study utilizes daily meteorological observations and sugarcane yield data from five major production regions in Guangxi (2002–2019), analyzing correlations between yield and multi-month continuous meteorological means to identify optimal sensitive periods and assess meteorological impacts during these key time spans. Processed data were used with long short-term memory (LSTM), BP neural network (BPNN), support vector machine (SVM), and RF to establish comparative yield prediction models for individual regions. Additionally, the Hodrick-

Prescott (HP) filter separated meteorological yield from apparent yield, mixed data from all five regions, and developed general multi-region meteorological yield prediction models using BPNN, SVM, RF, and LSTM to enable large-area forecasting.

2 Materials and Methods

2.1 Data Sources The study encompassed five sugarcane planting regions across different prefecture-level cities in Guangxi, with areas ranging from 870 to 18,500 hectares. Most Guangxi sugarcane regions feature a subtropical monsoon climate, with annual mean temperatures of 16.5–23.1°C, >10°C accumulated temperatures of 5000–8300°C, annual precipitation of 1300–2000 mm, and sunshine duration of 1500–1800 h—providing favorable meteorological conditions for sugarcane growth.

Yield data (2002–2019, 18 crushing seasons) were provided by Guangxi Sugar Industry Group (formerly Guangxi State Farms Sugar Industry Group), including total annual production for each region (specific prefecture names withheld for confidentiality). Statistical summaries are presented in Table 1 .

Meteorological data were obtained from the National Meteorological Science Data Center (China Meteorological Data Network, http://data.cma.cn). Daily datasets (2002–2019) from the five nearest meteorological stations (within 40 km of each region) included 14 factors: precipitation (20:00–08:00, 08:00–20:00, and 20:00–next day 20:00), maximum wind speed, mean atmospheric pressure, mean 2-minute wind speed, mean temperature, mean vapor pressure, mean relative humidity, sunshine hours, minimum pressure, minimum temperature, maximum pressure, maximum temperature, and minimum relative humidity. To improve precipitation spatial resolution, we also incorporated the 1 km resolution monthly precipitation dataset for China (1960–2020) from Qu et al. published in Science Data Bank [16].

2.2 Data Processing

2.2.1 Yield Data Preprocessing To eliminate regional disparities in multi-region modeling, we separated meteorological yield (the yield component influenced by meteorological factors) from apparent yield. Crop yield is typically decomposed into three components: trend yield, meteorological yield, and random fluctuation, as shown in Equation (1) [17]:

$$Y = Y_t + Y_W + e$$

where Y is apparent yield (t/ha), Y_t is trend yield (t/ha) determined by production levels and land productivity, Y_W is meteorological yield (t/ha), and e is random noise.

We employed the HP filter [18] for yield separation. For a long time series $\{h_i\}$ $(i = 1, 2, \dots, n)$, it comprises a long-term trend component g_i (trend yield) and short-term fluctuation component c_i (meteorological yield), as expressed in Equation (2):

$$h_i = g_i + c_i$$

The HP filter minimizes the sum of squared deviations H between the long-term trend component g_i and the apparent yield series h_i, as shown in Equation (3):

$$H = \sum_{i=1}^{n}(h_i - g_i)^2 + \lambda \sum_{i=1}^{n}[(g_{i+1} - g_i) - (g_i - g_{i-1})]^2$$

No universal standard exists for λ; values vary by temporal scale. Based on sugarcane annual yield characteristics, we determined $\lambda = 100$ [19].

2.2.2 Meteorological Data Processing To validate our meteorological data processing methodology, we established yield prediction models for individual regions using processed meteorological data and compared results with previous methods.

Sugarcane has a 12–14 month growth period, with yield vulnerable to persistent rainfall, strong winds, and low temperatures, with impacts varying by growth stage. Previous studies typically used whole-growth-period or fixed-period meteorological means, ignoring objective differences in temporal durations and dominant factors across growth stages (e.g., 1–2 month seedling stage requiring appropriate moisture, 2–3 month elongation stage needing substantial water, 3–4 month maturity stage requiring controlled water to avoid affecting sugar accumulation, and persistent strong winds/low temperatures during late growth/maturity causing adverse effects).

We developed a Sensitive Period Analysis Method, constructing 78 continuous monthly-incremental time spans from annual (January–December) meteorological data (e.g., 1 month starting January, 2 months starting January, etc.; 1 month starting February, 2 months starting February, etc.). We calculated meteorological means S_{ij_t} for each span $(i, j = 1, 2, \dots, 12;\ t = 1, 2, \dots, 18)$ and analyzed correlations between 14 meteorological means across different spans and sugarcane yields for each region using Equation (4):

$$r_{ij} = \frac{\sum_{t=1}^{18}(S_{ij\_t} - \overline{S_{ij}})(Y_t - \overline{Y})}{\sqrt{\sum_{t=1}^{18}(S_{ij\_t} - \overline{S_{ij}})^2 \sum_{t=1}^{18}(Y_t - \overline{Y})^2}}$$

where t represents year, i and j represent months, S_{ij_t} is the mean of a meteorological factor from month i to j in year t, and r_{ij} is the correlation coefficient between the meteorological factor mean and yield (78 total). Higher correlation coefficients indicate greater influence on yield and suitability as model variables.

2.2.3 Data Standardization Meteorological data involve multiple indicators with inconsistent dimensions, affecting modeling. Additionally, significant yield differences among regions impact model development when using mixed data. To eliminate dimensional differences and regional yield disparities, we normalized both meteorological and yield data.

Common normalization methods include linear function normalization and zero-mean normalization. Considering negative values in sugarcane meteorological yield, we applied linear function normalization to scale data to $[-1,1]$ using Equation (5):

$$X_t = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

where X_t is the normalized result, X is the original data, and $X_{\max}$ and $X_{\min}$ are the maximum and minimum values in the original dataset.

2.3 Model Construction We employed four algorithms widely used in yield prediction: RF, SVM, BPNN, and the deep learning algorithm LSTM to establish both single-region apparent yield prediction models and multi-region meteorological yield prediction models.

2.3.1 BPNN The BPNN structure comprises an input layer, hidden layer, and output layer [20]. Prediction accuracy primarily depends on hidden layer structure, which we determined to be a single layer after multiple trials. The input layer uses selected sensitive meteorological factors from key time spans; hidden layer neuron counts were determined via trial-and-error; the output layer has one node representing yield. To prevent overfitting, we set training epochs to 200, dropout to 0.3, and batch size to 4. Using Region 1 as an example, training loss curves (Figure 1 [Figure 1: see original paper]) show decreasing training and validation losses with increasing epochs, indicating model convergence.

2.3.2 SVM SVM model construction requires careful kernel function selection [21]. We compared linear, polynomial, and radial basis function (RBF) kernels, ultimately selecting RBF for its superior nonlinear mapping across different data dimensions. Other parameters were optimized via grid search.

2.3.3 RF RF model development employed grid cross-validation to traverse parameter dictionaries for optimal parameters [22]. Key parameters include “n_{estimators}” (number of decision trees), “max_{depth}” (maximum tree depth), and “max_{features}” (maximum features considered for node splitting). Optimal parameters for each region are listed in Table 2 .

2.3.4 LSTM The LSTM model [23] contains one input layer, one hidden layer (25 neurons), and one output layer. Time step was set to 3, training epochs to 40, and batch size to 2.

3 Results and Discussion

3.1 Key Meteorological Factor Analysis Across Regions

3.1.1 Inter-factor Correlation Analysis Annual meteorological means for 14 factors across 78 time spans (14 × 78 data points) were correlated with yield. For Region 1, a representative correlation heatmap (Figure 2 [Figure 2: see original paper]) displays relationships among meteorological factors and yield (time span details omitted). High inter-factor correlations indicate multicollinearity, which is detrimental to modeling. We screened highly correlated factors, retaining only one representative factor, ultimately identifying key variables including sunshine hours, mean 2-minute wind speed, maximum wind speed, minimum relative humidity, and mean vapor pressure.

3.1.2 Key Meteorological Factors and Sensitive Time Spans Based on correlation analysis, key sensitive meteorological factors and their critical time spans for each region are summarized in Table 3 . Substantial differences exist across regions in both sensitive factors and key periods, with some factors showing opposite correlations in different regions—primarily because sugarcane abundance meteorological indicators do not fully align with physiological meteorological indicators [24].

Correlation results reveal that during the germination stage, minimum temperature shows significant positive correlation with yield, as appropriate minimum temperatures enhance enzyme activity in seed stems and accelerate germination. During the seedling stage, mean temperature exhibits significant positive correlation, with moderate warming benefiting seedling growth. During elongation and maturity stages, maximum and mean temperatures show significant negative correlations; abundant sunlight combined with persistent high temperatures exacerbates sugarcane desiccation, reducing yield. During elongation, minimum temperature shows significant negative correlation due to small diurnal temperature ranges that hinder photosynthesis and sugar accumulation; reducing nighttime temperatures during this stage benefits yield increase.

During tillering, sunshine hours correlate significantly positively with yield, as good light promotes tillering and increases production. From late elongation to maturity, sunshine hours correlate significantly negatively with yield; this period has less precipitation, and excessive light easily causes desiccation, hindering yield increase. From seedling to tillering stages, maximum wind speed correlates significantly negatively with yield, as strong winds easily break seedlings, severely affecting later growth and reducing yield.

During germination and elongation, mean vapor pressure correlates significantly positively with yield. Vapor pressure correlates significantly positively with precipitation, temperature, and humidity, influencing yield primarily through other meteorological factors and serving as a comprehensive indicator—making it valuable for assessing meteorological conditions and roughly estimating yield trends. During elongation to maturity, precipitation correlates significantly positively with yield, indicating that water demand during these stages is not met by current precipitation levels, and increasing precipitation benefits yield increase. During late maturity, minimum relative humidity correlates significantly positively with yield; drought during this period also reduces yield. Mean relative humidity affects yield throughout nearly the entire growth period, correlating significantly positively; it is primarily influenced by rainfall, and increased precipitation benefits yield increase. Thus, sugarcane has substantial water requirements, but excessive or insufficient precipitation during certain periods is detrimental, making appropriate precipitation particularly crucial.

3.2 Single-Region Apparent Yield Modeling Data for key sensitive meteorological factors and yields were standardized, and 18 years of yield data (2002–2019) were randomly split into training and testing sets at a 7:3 ratio. BPNN, SVM, RF, and LSTM models were constructed for each region and compared. To validate our sensitive period analysis method, we built reference models using annual mean meteorological factors following Li [10] and compared results.

3.2.1 BPNN Prediction Model BPNN apparent yield prediction results for each region are shown in Table 4 . Models based on our key time span sensitive meteorological factors significantly outperformed the reference method. Region 4 achieved the best results (RMSE = 5.04 t/ha, MAPE = 5.62%), while Region 1 showed the poorest performance (RMSE = 24.34 t/ha).

3.2.2 SVM Prediction Model SVM apparent yield prediction results are presented in Table 5 . Our data processing method significantly outperformed the reference approach. Regions 2 and 4 achieved MAPE < 10%, with slightly better overall precision than BPNN, though improvement was modest, and large prediction errors persisted for Regions 1 and 5.

3.2.3 RF Prediction Model RF apparent yield prediction results are shown in Table 6 . Our meteorological factor analysis produced high-precision RF models for all regions except 1 and 5 (RMSE = 3.5–7.33 t/ha, MAPE = 3.98%–7.48%). RF outperformed SVM and BPNN overall, though Region 1 and 5 errors remained substantial (RMSE = 31.83 t/ha and 11.22 t/ha, respectively). Our method demonstrated significant advantages over the reference approach.

3.2.4 LSTM Prediction Model LSTM apparent yield prediction results are summarized in Table 7 . Our meteorological factor analysis yielded excellent LSTM models for all five regions (MAPE < 10%), with substantially improved overall precision compared to the other three algorithms. The reference method also showed improved LSTM precision relative to other models, confirming LSTM's feasibility for regional yield prediction.

3.2.5 Comparative Model Performance Analysis Figure 3 [Figure 3: see original paper] presents scatter plots of predicted versus actual values for each algorithm across all regions. Table 8 compares yield prediction results across models. Our LSTM model achieved the lowest overall RMSE (10.34 t/ha) and MAPE (6.85%), with $\mathrm{Rv}^2 = 0.8489$, demonstrating superior trend fitting. Given the yield statistics in Table 1, these errors are acceptable. Reference models using annual meteorological factors [10] showed significantly lower precision. Our data processing methodology substantially improved prediction accuracy, with LSTM outperforming BPNN, SVM, and RF.

3.3 Multi-Region Meteorological Yield Modeling For large-area prediction, we separated meteorological yield from apparent yield using HP filtering [25], mixed data from all five regions, and established general meteorological yield prediction models using BPNN, SVM, RF, and LSTM. Results are shown in Table 9 . LSTM performed poorly for multi-region prediction, while RF, SVM, and BPNN achieved excellent results ($\mathrm{Rv}^2 > 0.94$). BPNN performed best (RMSE = 0.98 t/ha, MAPE = 9.59%, $\mathrm{Rv}^2 = 0.965$).

Our multi-region meteorological yield prediction model, built on mixed data from five regions, effectively expanded the dataset and significantly improved prediction performance. HP filtering eliminated regional differences caused by production conditions and socioeconomic factors, making the general model more scientifically robust. Our meteorological data processing methodology proved scientifically sound and unique, with results demonstrating that models based on processed data significantly outperformed alternatives.

4 Discussion

Single-region prediction results show that BPNN, SVM, RF, and LSTM models based on our key time span sensitive meteorological factors significantly outperformed the reference method [10]. LSTM achieved overall RMSE = 10.34 t/ha and MAPE = 6.85%, with $\mathrm{Rv}^2 = 0.8489$, clearly surpassing the other three widely-used models. While BPNN, SVM, and RF showed high precision overall, some regions exhibited lower accuracy. However, LSTM not only achieved the best overall performance but also maintained MAPE < 10% for all individual regions, confirming its suitability for regional yield prediction.

For multi-region prediction, SVM, RF, BPNN, and LSTM were used to predict meteorological yield from mixed samples. BPNN achieved the best performance (RMSE = 0.98 t/ha, MAPE = 9.59%, $\mathrm{Rv}^2 = 0.965$), outperforming SVM, RF, and LSTM, though all except LSTM produced excellent results. This indicates that LSTM is unsuitable for multi-region joint sugarcane yield prediction.

Our sensitive period analysis method, correlating 78 monthly continuous time span meteorological means with yield, identified key meteorological factors for Guangxi's five sugarcane regions: sunshine hours, vapor pressure, atmospheric pressure, temperature, and precipitation. Different factors have different key time spans, and the same factor may show opposite correlations across periods, underscoring the practical value of sensitive period analysis. Due to varying principles and characteristics, different models perform differently on different datasets, leading to prediction result variations.

In conclusion, establishing general multi-region meteorological yield prediction models using our methodology is feasible, with prediction errors within acceptable ranges for both single-region and multi-region scenarios. This approach provides valuable reference for regional sugarcane yield forecasting.

Conflict of Interest Statement: The authors declare no conflicts of interest related to this research.

References

[1] LI W, GU F X. Prediction of regional crop yield based on model[J]. Agricultural outlook, 2020, 16(3): 104-111.
[2] Expert Committee on Market Warning of Ministry of Agriculture and Rural Affairs. China agricultural outlook 2019−2028[M]. Beijing: China Agricultural Science and Technology Press, 2019.
[3] GAO J J, YUAN Y R, LIANG Y. Establishment of early rice yield prediction model in Gaoyao area[J]. Guangdong meteorology, 2022, 44(2): 50-52.
[4] YU Z Z, ZOU H F, YU D S, et al. Sugarcane yield GA-BP prediction model incorporating field water and heat factors[J]. Transactions of the Chinese society for agricultural machinery, 2022, 53(10): 277-283.
[5] CHEN S. Yield forecast and irrigation decision for maize based on historical weather data and the Ceres-maize model[D]. Yangling: Northwest A & F University, 2017.
[6] WANG E H, SONG X. Prediction model of peanut yield in Kaifeng city based on meteorological factors[J]. Shaanxi journal of agricultural sciences, 2012, 58(4): 31-33.
[7] HE H, WANG Q J, LI L, et al. Separating the effect of meteorology on maize yield from the impact of other factors in the Yellow River-water irrigated regions in Ningxia of China[J]. Journal of irrigation and drainage, 2022, 41(4): 1-7.
[8] GU Y W, YAO Y L, FU W D. Prediction model of apple yield in Aksu region based on key meteorological factors[J]. Xinjiang agricultural science and technology, 2021(2): 22-24.
[9] HE X J. Research on maize yield prediction model based on machine learning[D]. Changchun: Jilin Agricultural University, 2021.
[10] LI Y M. Wheat yield forecasting: A machine learning approach based on meteorological factors[D]. Zhengzhou: Henan Agricultural University, 2019.
[11] ZHAO Y X, XIAO D P, BAI H Z, et al. The prediction of wheat yield

in the North China plain by coupling crop model with machine learning algorithms[J]. Agriculture, 2022, 13(1): ID 99.
[12] CROCI M, IMPOLLONIA G, MERONI M, et al. Dynamic maize yield predictions using machine learning on multi-source data[J]. Remote sensing, 2022, 15(1): ID 100.
[13] OIKONOMIDIS A, CATAL C, KASSAHUN A. Hybrid deep learning-based models for crop yield prediction[J]. Applied artificial intelligence, 2022, 36(1): 1-18.
[14] DI Y, GAO M F, FENG F K, et al. A new framework for winter wheat yield prediction integrating deep learning and Bayesian optimization[J]. Agronomy, 2022, 12(12): ID 3194.
[15] BURDETT H, WELLEN C. Statistical and machine learning methods for crop yield prediction in the context of precision agriculture[J]. Precision agriculture, 2022, 23(5): 1410-1433.
[16] QU L S, ZHU Q A, ZHU C F, et al. 2022. Monthly precipitation data set with 1 km resolution in China from 1960 to 2020[DB/OL]. Science Data Bank. [2022-04-15]. http://www.scidb.cn/cstr/31253.11.sciencedb.01607.
[17] HUANG H X, ZHOU Y J, ZENG Y, et al. Meteorological forecast of sugarcane production in Guigang, Guangxi[J]. Journal of Chengdu university of information technology, 2020, 35(5): 554-559.
[18] XU X, MA Z W, XIONG S P, et al. Wheat yield forecast in Henan Province based on climate year type[J]. Journal of agricultural science and technology, 2022, 24(2): 69-77.
[19] WANG G Z, LU J S, CHEN K Y, et al. Exploration of method in separating climatic output based on HP filter[J]. Chinese journal of agrometeorology, 2014, 35(2): 195-199.
[20] ZHOU C H, WU Z Y, LIU C. A study on quality prediction for smart manufacturing based on the optimized BP-AdaBoost model[C]//2019 IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE). Piscataway, NJ, USA: IEEE, 2020: 1-5.
[21] KAZEMI A, BOOSTANI R, ODEH M, et al. Two-layer SVM, towards deep statistical learning[C]//2022 International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI). Piscataway, NJ, USA: IEEE, 2022.
[22] MIAH M O, KHAN S S, SHATABDA S, et al. Improving detection accuracy for imbalanced network intrusion classification using cluster-based under-sampling with random forests[C]//2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). Piscataway, NJ, USA: IEEE, 2019: 1-5.
[23] AKANDEH A, SALEM F M. Slim LSTM networks: Lstm_6 and LSTM_{C6}[C]//2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS). Piscataway, NJ, USA: IEEE, 2020: 630-633.
[24] OU Z R, TAN Z K, HE Y, et al. The key meteorological factors affecting the sugarcane yield in major production areas in China and their high-low yield indices[J]. Journal of Anhui agricultural sciences, 2008, 36(24): 10407-10410, 10415.

[25] LI Z Q, ZHANG X Y, TIAN H D. Prediction method of satellite telemetry data using HP filter[J]. Spacecraft engineering, 2021, 30(4): 23-30.
