IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 December 2024, accepted 3 February 2025, date of publication 7 February 2025, date of current version 12 February 2025.
Digital Object Identifier 10.1109/ACCESS.2025.3539775

RESEARCH ARTICLE

# Precise Assimilation Prediction of Short-Term and Long-Term Maize Irrigation Water Based on EnKF-DSSAT and Fuzzy Optimization-DSSAT Models

**YANSHU YU[1], YOUXI LUO[2], XINHANG WANG[2], XINRAN WANG[3], AND CHAOZHU HU[2]**
[1]School of Computer Science, Hubei University of Technology, Wuhan 430000, China
[2]School of Science, Hubei University of Technology, Wuhan 430000, China
[3]School of Economics and Management, Hubei University of Technology, Wuhan 430000, China

Corresponding author: Chaozhu Hu (huchaozhu0035@hbut.edu.cn)

This work was supported in part by the National Social Science Fund under Grant 24BTJ068, in part by the National Science Fund under Grant 11701161, and in part by the Key Humanities and Social Science Fund of Hubei Provincial Department of Education under Grant 20D043.

**ABSTRACT** With the progress of information technology, precision irrigation technology has become the core of modern agriculture. In particular, technologies such as Internet of Things (IoT), Big Data and Artificial Intelligence (AI) have provided strong support for the intelligence of agricultural production. This paper focuses on the precise prediction of irrigation water use in the maize industry through Ensemble Kalman Filter (EnKF) and fuzzy optimization methods combined with the DSSAT (Decision Support System for Agrotechnology Transfer) model. We use the remote sensing data of land moisture and leaf area in the Yellow Huaihai Plain provided by Google Earth Engine (GEE), as well as the maize market data released by the Ministry of Agriculture (MOA), to make predictions through the EnKF-DSSAT and fuzzy optimization-DSSAT models. The results showed that these models achieved high accuracy of 98.11% and 97.78% in short-term and long-term forecasts, respectively, which were significantly better than the traditional models. We also introduce a Boltzmann machine-based fusion algorithm to improve the model convergence speed and prediction accuracy. Ultimately, this paper verifies the important influence of policy factors on long-term irrigation prediction and proposes an adaptive prediction model and policy recommendations, which provide innovative methods and technical support for the implementation of precision irrigation technology.

**INDEX TERMS** Data assimilation, DSSAT model, ensemble Kalman filter, fuzzy optimization, precision irrigation.

## I. INTRODUCTION

China is a major agricultural country, with a strong focus on building national agricultural big data resources. Research shows that the country faces significant issues with high irrigation water consumption, waste, and low utilization of water resources. In recent years, technological innovations have driven agricultural modernization and intelligence. To address water and fertilizer loss as well as soil degradation from traditional methods, scholars have conducted extensive research on precision irrigation, exploring control strategies, irrigation methods, monitoring systems, and predictive models to improve crop yields.

Most scholars agree that precision irrigation relies on modern monitoring tools to digitally, networked, and intelligently monitor crop growth stages and environmental factors [1], [2] By leveraging big data, AI, and other advanced technologies, precision irrigation monitors and predicts soil moisture and

The associate editor coordinating the review of this manuscript and approving it for publication was Liandong Zhu.

climate conditions [3], [4] Based on these results, the most accurate irrigation equipment is used for efficient fertilization and irrigation, ultimately achieving intelligent, efficient, and precise agricultural irrigation methods [2].

Most scholars adopt sensor-based and wireless transmission methods for control strategies. Zhang et al. [5] used wireless data transmission to control solenoid valves based on plant water demand and soil moisture, automating precision irrigation with a self-developed controller. Wei et al. [6] employed intelligent circuits to collect soil moisture, star network structure, and hexagonal deployment for data processing, integrating sensors and wireless transmission in the design of a precision irrigation control system.

Existing studies on monitoring systems mainly utilize IoT, wireless sensor networks (WSN), cloud platforms, and wireless communication technologies. Ferrández-Pastor et al. [7] used WSN and IoT to monitor soil, plants, and weather in real time. Hamouda and Elhabil [8]employed distributed WSN nodes for real-time monitoring of precision agriculture metrics. Jayaraman et al [9]. highlighted the role of IoT cloud servers in improving decision-making for high-quality food production. Elijah et al. [10], Jawad et al. [11], Tzounis et al. [12], and Bitella et al. [13] emphasized IoT and WSN as key tools for wide-area agricultural monitoring.

Remote sensing data have been applied to crop simulation models for forcing functions or steering [14]. Ines and Hansen [15] used remotely sensed evapotranspiration to reparameterize a pseudo-regional SWAP model with soil, crop, and water management parameters. Kamilaris and Prenafeta-Boldú [16]noted that agricultural systems are complex and unpredictable. Prediction models mainly include classical control models, AI-based intelligent optimization models, and biological models.

First, classical control models primarily include on/off, P, PI, and PID control, though they are often time-consuming and cumbersome [17].Huang et al. [18] enhanced PID control with fuzzy reasoning, using fuzzy rules for self-tuning parameters, and added multi-factor gray prediction to address time lag. This system effectively predicts crop water demand for accurate irrigation [18]. From this study, we got the insight that the use of fuzzy rules can be very helpful for parameter adjustment. Therefore, fuzzy optimization was used for parameter updating in constructing the long-term prediction model of this paper.

Second, when crop models are used for early yield prediction, two main uncertainties arise: climate and modeling [19]. Climate-related uncertainty is higher early in the season, but can be reduced with skillful climate prediction, while model instability can be mitigated by assimilating remotely sensed data [20], [21], [22]. This paper finds that existing research on forecasting models only focuses on collecting natural data to input into the model to forecast crop yields without considering human factors, so this paper not only considers natural factors but also considers human factors to forecast the model. This paper also found that the existing forecasting research mainly focuses on short-term forecasting, and there is little long-term forecasting. Therefore, this paper proposes two models for short-term and long-term forecasting. Current intelligent optimization models based on AI and machine learning include the EnKF model, fuzzy set theory, DSSAT, neural networks, and machine learning models. The existing Enkf model found that the prediction accuracy of the model can be effectively improved if the leaf area index is added to the data. In this paper, we changed the calculation of the leaf area index, not only obtaining the value of the leaf area index from the remotely sensed data, but also combining the deep soil moisture data into the calculation of the leaf area index through biological modeling. Grubert [23] developed an intelligent irrigation strategy using fuzzy logic and neural network control based on deficit regulation theory. Abioye et al. [24] evaluated the DSSAT model with four crops, including spring wheat in Jilin Province, and found it accurately fits indicators like crop fertility period and leaf area index. The DSSAT model can well simulate the growth and development and final yield of different crops under specific environments, so it is also chosen for prediction in this paper. Gavasso-Rita et al. [25] showed that EnKF can enhance leaf area index simulation accuracy Montaldo et al. [26] used UAV multispectral data and support vector machine classification to develop a quantitative soil water content prediction model through ridge regression and extreme learning machines. Wei et al. [27] and Kumar et al. [28]designed intelligent irrigation systems based on improved Elman neural networks and fuzzy control for irrigation prediction.

Third, Existing biological models primarily use two-way loose coupling approaches. Dan et al. [29] coupled a crop growth model with a hydrological model, establishing a system to quantify water demand and evapotranspiration during crop growth, and predict yields under varying environmental and climatic conditions. Li et al. [30] demonstrated that the Hydrus 2D model could effectively predict salinity distribution in drip-irrigated soils based on brackish water irrigation experiments. When using remote sensing data to replace state variables or infer soil-plant-atmosphere-continuum characteristics, it is assumed that data errors are either absent or manageable [31]. Thorp et al. [32] used mechanisms to assimilate leaf area index (LAI) in the DSSAT CSM wheat model, enhancing its predictive performance [33], [34], [35]. Existing studies have shown that combining two models together is likely to be able to improve the accuracy of the predictions, so this paper draws inspiration from this and utilizes the strengths of multiple models for the construction of a joint model. In the short-term prediction model, the advantages of the Enkf assimilation model are used to assimilate soil surface moisture and leaf area index, and the simulation advantages of DSSAT are used to make predictions. In the long-term prediction model, data from two natural directions and one human direction were assimilated using the advantage of fuzzy optimization that can well integrate data from multiple sources and thus update the parameters, and the simulation advantage of DSSAT was used to make the prediction jointly.

Summarizing the research on precision irrigation by domestic and international scholars, it primarily focuses on four areas: control strategies, irrigation methods, monitoring systems, and prediction models. Agricultural systems are inherently complex and challenging to predict [32], [33], [34], [35]. Current studies mostly focus on short-term predictions and overlook human factors and deep soil moisture prediction. To address these issues, this paper leverages big data and AI technologies to aid precision irrigation. By introducing meteorological and human parameters, it designs an assimilation prediction scheme, integrating ensemble Kalman filter, fuzzy optimization, the DSSAT system, and Boltzmann machine technology. The system predicts maize irrigation water requirements in both short- and long-term dimensions, considering human and natural factors. The results provide targeted recommendations for precision irrigation strategies and integrated water resource planning in agriculture. This scheme can offer a scientific basis for precision maize irrigation, enabling more accurate irrigation calculations, improving water resource utilization, and promoting agricultural modernization and crop yield improvement.

## II. ASSIMILATION MODEL AND ALGORITHM DESIGN

As shown in FIGURE 1, for the short-term prediction of accurate irrigation water, this paper uses data from two natural directions, fuses the remote sensing data of soil surface moisture and leaf area index using the EnKF assimilation model, assimilates the real-time data in short-term time steps as the time dimension continues to extend, and uses the assimilated error matrix as a priori knowledge to guide the DSSAT assimilation model for irrigation water short-term prediction. For the long-term prediction of accurate irrigation water, this paper uses two natural data and one humanistic data, fuses the remote sensing data of soil surface moisture, leaf area index and government subsidy data using a fuzzy optimization assimilation model, assimilates the real-time data in a long term time step as the time dimension continues to be extended, and uses the predicted values of the assimilated values of the three data as a priori knowledge to guide the DSSAT assimilation model for the short-term prediction of irrigation water. assimilation model for long-term prediction of irrigation water.

### A. PRECISE IRRIGATION PREDICTION SYSTEM BASED ON ENKF-DSSAT ASSIMULATION MODELING

Data assimilation algorithm is the core part of the data assimilation system, and the two main data assimilation algorithms are: parametric algorithm and filtering algorithm. According to the research, the analysis accuracy of filtering algorithm in short assimilation window is better than parametric algorithm [36]. Common filtering algorithms include Ensemble Kalman Filter (EnKF) and Particle Filter (PF) [37]. Common filtering algorithms include EnKF and PF. Based on the good adaptability of EnKF in dealing with a large number of uncertainties and linear systems, we consider implementing EnKF in the crop model to control the crop model operation, assimilate the remotely sensed data and update the model state variables. This paper improves on the original EnKF model, first of all, the output is not directly the prediction result, but an error matrix which will be rooted in different prediction models, respectively short-term and long-term time step using Boltzmann probability to update the parameters.

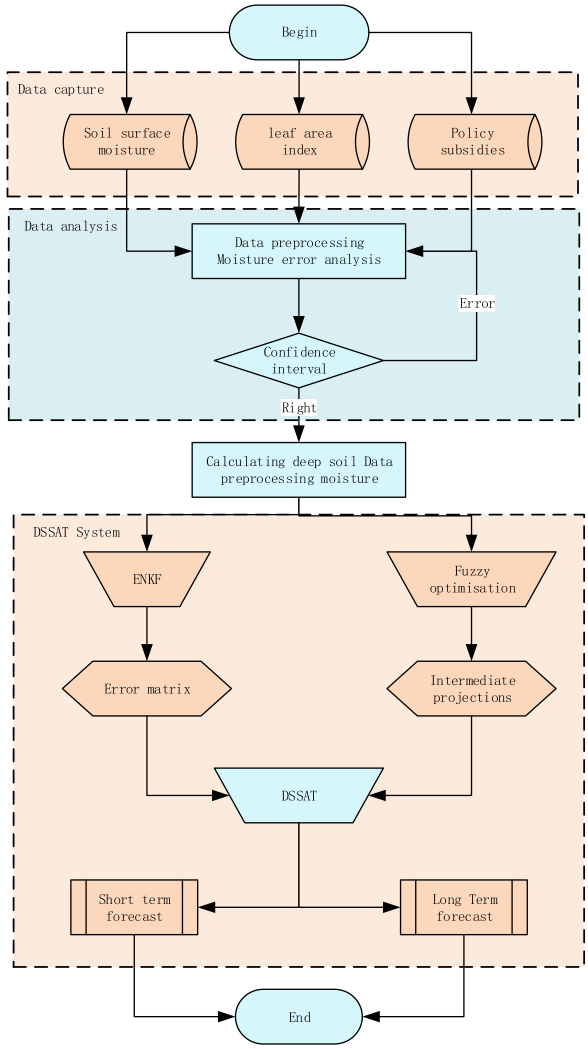


FIGURE 1. Research framework.

#### 1) ENKF ASSIMILATION MODEL

The traditional EnKF predicts and updates variables based on state transfer equations and observation equations through a nonlinear system to solve the problem of error covariance matrix forecasting. By defining as the observation vector, it is assumed that the observed values are related to the true state as follows:

$$y = H\chi_t + \varepsilon \tag{1}$$

where $\varepsilon$ is a Gaussian random error vector with mean 0 and observed error covariance of $R$ and $H$ is an operator that maps the model variables to the observation space. Based on this formula the analysis of estimated states and error covariances are considered to find the desired results needed.

#### 2) ENKF-DSSAT ASSIMILATION MODEL

The DSSAT model, as one of the traditional crop growth models in agronomy, is widely used in fine agriculture, agroecology, and many other fields related to agricultural production and research. Using the DSSAT model, the growth, development and final yield of different crops under specific environmental conditions can be simulated. In this paper, the DSSAT model is applied to realize the interaction between soil and climate by taking soil moisture and leaf area index into account, so as to simulate the crop growth conditions for accurate prediction.

As shown in FIGURE 2, We trained the two data, soil moisture and leaf area index, by assimilating them through the EnKF assimilation model to obtain the error matrix, which was updated in open-loop with two times steps, short-term and long-term, while performing the assimilation training. The short-term is in two days time step and the long-term is in one month time step. The error matrix was updated with Boltzmann probability to the parameters in the DSSAT with long and short term time steps respectively, and finally the prediction of corn irrigation water was obtained from the output port of the DSSAT system.

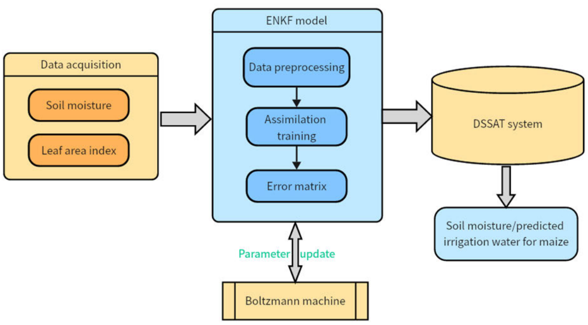


**FIGURE 2.** **Flowchart of the EnKF-DSSAT model.**

When DSSAT is combined with EnKF, it is able to utilize real-time data from remote sensing to dynamically update the model state through data assimilation techniques, thus enhancing the prediction accuracy and utility of the model. In this case, DSSAT serves as the model framework for generating model states and predictions, while EnKF updates the predicted states based on data changes. Our contribution is mainly in two parts, the first part is for the source of leaf area index, and the second part is the joint prediction of EnKF model and DSSAT. For the data part, when calculating the leaf area index, the original model only obtains the data of leaf area index through remote sensing data and then uses them, but when calculating the leaf area index, we not only use the remote sensing data but also use the data of deep soil moisture, which is extrapolated by the biological model and then get the data of the leaf area index, and we integrate the data of the leaf area index obtained by the deep soil moisture extrapolated by the biological model. We integrated the data of leaf area index obtained from deep soil moisture through biological modeling and remote sensing directly. For the joint modeling part, we mainly change the output of EnKF model, the original EnKF output is directly the prediction result, while the current EnKF output is the error matrix. And we update this error matrix in two time steps, long term and short term, in real time using Boltzmann probability. This approach improves the flexibility and responsiveness of crop models in changing environments, and the specific algorithms are implemented as follows.

**Algorithm 1** EnKF-DSSAT assimilation model

**Inputs:** SM-Soil Moisture, LAI-Leaf Area Index
**Output:** LID-Irrigation Demand
1: Preprocess data (SM, LAI)
2: Ensemble KF←Initialize EnKF (SM, LAI)
3: while not convergence do
4: Ensemble KF.Predict()
5: Ensemble KF.Update()
6: DSSAT Parameters = Adjust Parameters (Ensemble KF.Error Covariance)
7: Irrigation Demand = DSSAT Model (DSSAT Parameters)
8: if Time Step is short-term then
9: Update Parameters with Boltzmann (Irrigation Demand, short-term)
10: else if Time Step is long-term then
11: Update Parameters with Boltzmann (Irrigation Demand, long-term)
12: end if
13: Print Result (Irrigation Demand)
14: end while
15: end

### B. OTHER RECOMMENDATIONS PRECISE IRRIGATION PREDICTION SYSTEM BASED ON FUZZY OPTIMIZATION-DSSAT ASSIMILAION MODEL

In the process of constructing the EnKF model, we mainly considered the prediction of irrigation water for crops when natural factors dominate. However, the influence of human factors was not discussed, and it became a new challenge to better integrate human factors for macro prediction of irrigation water in the long-term time dimension. We explored this issue and found that replacing the EnKF assimilation model with a fuzzy optimization assimilation model and adding a model prediction scheme can make the model prediction better regulated by the human indicators and more accurate macro prediction of water resources in the long-term time dimension.

### 1) FUZZY OPTIMIZATION ASSIMILATION MODEL

In this model, we added the factor of government subsidies as a constraint parameter and assimilated the multi-source data using real-time data, combined with the previous factors as input variables added to the objective function to participate in the prediction. Our contribution is mainly in two parts, the first part is for the source of obtaining the leaf area index, and the second part is that the fuzzy optimization model and the DSSAT have made a joint prediction. For the first part, most of the existing studies mainly used natural data, while this paper used humanistic data and natural data, a form of multi-source data for prediction. For the, joint model part, we mainly changed the output of the fuzzy optimization model, the output of the original fuzzy optimization model is directly the prediction result, while the output of the current fuzzy optimization is the error matrix. And we update this error matrix in two time steps, long term and short term, in real time using Boltzmann probability.

Assuming that the soil moisture is represented by the remote sensing data S, the leaf area index is represented by L, and the government subsidy is B, their corresponding affiliation functions are defined as $f_S$, $f_L$ and $f_B$ respectively. The closer the values of these functions are to 1, the closer the current soil moisture, leaf area index and government subsidy are to the ideal state. When these functions are applied to the predictive control model, the objective optimization function can be expressed as:

$$\min_{u(k),\ldots,u(k+N_u-1)} \left[ \begin{array}{l} \sum_{j=1}^{N_y} (f_y(\hat{y}(k+j) - r(k+j)))^p + \\ \sum_{j=1}^{N_u} ((f_u(\Delta u(k+j-1)))^p \\ +f_S(S(k)))^p + f_L(L(k)))^p + f_B(B(k)))^p) \end{array} \right] \tag{2}$$

The constraints are:

$$\hat{y}(k+j) = f\left(\hat{y}(k+j-1), \ldots, u(k+j-1), \ldots\right),\ j = 1, \ldots, N_y \tag{3}$$

where $\hat{y}(k+j)$ denotes the predicted output at the time of $k+j$, $r(k+j)$ is the reference trajectory, $\Delta u(k+j-1)$ is the change in the control action, and $p$ is a specific positive number, usually 2, representing the use of a quadratic form of the affiliation function. $f_y$ and $f_u$ are the tracking error of the output and the affiliation function of the control action, respectively. These affiliation functions are usually determined from specialized knowledge data and express the ideal state or desired range of the corresponding variable.

The use of this assimilation model integrates the actual crop growth, environmental factors, and policy support to better handle the dynamic uncertainty in the decision-making process through a fuzzy optimization assimilation model.

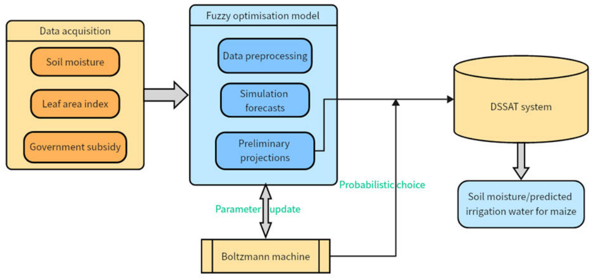


**FIGURE 3.** Flowchart of fuzzy optimization-DSSAT model.

**Algorithm 2** Prediction algorithm for fuzzy optimization-DSSAT assimilation model

**Inputs:** SM-Soil Moisture, LAI-Leaf Area Index, GS-Growth Stage
**Output:** LID-Irrigation Demand
1: Begin
2: Preprocess data (SM, LAI, GS)
3: Fuzzy Optimizer ← Initialize Fuzzy Optimizer(SM, LAI, GS)
4: while not convergence do
5: Predicted Values ← Fuzzy Optimizer. Optimize ()
6: DSSAT Input ← Prepare Input(Predicted Values)
7: if Time Step is short-term then
8: DSSAT Output ← DSSAT Model (DSSAT Input, short-term)
9: Update Predictions with Boltzmann (DSSAT Output, short-term)
10: else if Time Step is long-term then
11: DSSAT Output ← DSSAT Model (DSSAT Input, long-term)
12: Update Predictions with Boltzmann (DSSAT Output, long-term)
13: end if
14: Print Result (DSSAT Output)
15: end while
16: end

### 2) FUZZY OPTIMIZATION-DSSAT ASSIMILATION MODEL

As shown in FIGURE 3, In order to be able to make long term predictions of the results, we take the natural factors - soil moisture and leaf area index and the human factor - government subsidies, three multi-source data and train them through fuzzy optimization model to get the predicted values, and train them in assimilation with short-term and long-term Two cases of open-loop updating are performed to update **the predicted** values with Boltzmann probability to update the state in DSSAT respectively, and finally the predicted results of soil moisture are obtained from the output port of DSSAT system.

When DSSAT is combined with a fuzzy optimization model, it is able to combine information from both human and natural factors. The specific implementation program is as follows Algorithm 2 shown.

## III. ASSIMILATION PREDICTION OF WATER USE FOR PRECISION IRRIGATION FOR MAIZE IN THE TELLOW HUAIHAI PLAIN

### A. DATA COLLECTION

We selected three different areas in the Yellow Huaihai Plain production areas as sampling sites with two times steps, long-term and short-term, and collected soil surface moisture image data and leaf area index remote sensing data in the Yellow Huaihai Plain from 2014 to 2024 (data source: GEE Remote Sensing Database). Soil surface moisture and Leaf Area Index (LAI) are crucial variables in irrigation water demand forecasting, and their selection has a clear scientific basis. First of all, soil surface moisture is the core index that characterizes the soil moisture status, which directly affects the availability of water in the crop root zone and is an important basis for measuring the timing and intensity of irrigation. Studies have shown that dynamic changes in soil moisture are key drivers of fluctuations in crop water requirements. Secondly, leaf area index, as a comprehensive characterization of plant canopy cover, reflects the growth status of vegetation and the intensity of transpiration, and can indirectly indicate the level of crop water consumption [23]. Combining these two variables not only accurately captures the spatial and temporal heterogeneity of crop water demand, but also improves the accuracy of irrigation demand prediction through multi-source data fusion. In addition, existing studies have confirmed that the synergistic effect of soil moisture and leaf area index can significantly optimize irrigation regulation strategies, thus achieving agricultural water use efficiency [24]. Therefore, the selection of these two variables has important theoretical value and practical significance for accurate irrigation prediction. Taking into account the characteristics of irrigation demand in China's national context, this study not only selected soil surface moisture and leaf area index as the core variables, but also incorporated humanistic data in order to comprehensively consider the influence of regional actual conditions and socioeconomic factors on irrigation demand. This design has important scientific value and practical significance. As the main grain-producing region in China, agricultural irrigation accounts for more than 60% of the total water use in the Yellow and Huaihai regions, which have long been facing the problems of water scarcity and inefficient utilization. By introducing humanistic data such as population and economy, farmland distribution and cash crop cultivation, the dynamics of irrigation demand in the region can be reflected more comprehensively. For example, the irrigation habits and economic return orientation of farmers in different regions can also significantly affect the allocation and utilization efficiency of water resources. Therefore, based on the combination of humanistic data and ecological variables, this study realizes an accurate prediction of irrigation water demand, which provides a scientific basis and policy reference for solving the water shortage problem in the Huanghuaihai region.

The selection of input variables directly affects the performance and reliability of irrigation demand prediction. In this study, soil surface moisture and Leaf Area Index (LAI) were selected as the core input variables, and their selection was based on the scientific basis and the key contribution to the prediction process.

Soil surface moisture is a direct indicator of available water in the soil, which is closely related to crop evapotranspiration and water uptake, and can reflect the current balance of soil water supply and demand, guiding the timing and intensity of irrigation. Leaf area index (LAI), on the other hand, as an important variable reflecting the growth and transpiration potential of vegetation, indirectly characterizes the water demand intensity of crops. The combination of the two provides dynamic ecological information for the model, making the prediction more relevant and practical.

In addition, although the core variables of this study have been able to fulfill the main objective of irrigation demand forecasting, the possible effects of other potential variables on the forecasting performance can still be discussed theoretically. For example, weather patterns (including precipitation, temperature, evapotranspiration, etc.) can provide more comprehensive information about the environmental context for irrigation forecasting, while soil characteristics such as soil texture and organic matter content can further refine the understanding of water retention and release capacity. These variables, although not included in the core analytical framework of this study, merit further exploration in future research for their potential value to irrigation practices.

We acquired soil moisture and leaf area index data for a total of 3,742 days from January 2014 to April 2024, resulting in approximately 560,000 individual data points extracted across three temporal resolutions: 1 hour, 2 days, and 1 month. The dataset was updated in real-time, with April 31, 2024, as the final assimilation point. To ensure the robustness of the model and reduce biases in evaluation, we divided the data into training, validation, and test sets in an 8:1:1 ratio. Specifically, 448,000 data points were allocated to the training set for model fitting, 56,000 data points were used in the validation set to tune hyperparameters and monitor performance during training, and the remaining 56,000 data points constituted the test set for final performance evaluation. To prevent overfitting, several strategies were employed. First, L2 regularization was applied to the model's loss function, penalizing overly complex parameter weights. Additionally, dropout layers with a rate of 0.3–0.5 were used in the neural network components to randomly deactivate a fraction of neurons during each training iteration, thus enhancing generalization. Early stopping was implemented by monitoring validation loss, halting training once the loss ceased to improve for a predefined number of epochs, ensuring the model did not overfit to the training data. During model training, data normalization was performed to standardize input features, ensuring consistent scaling and

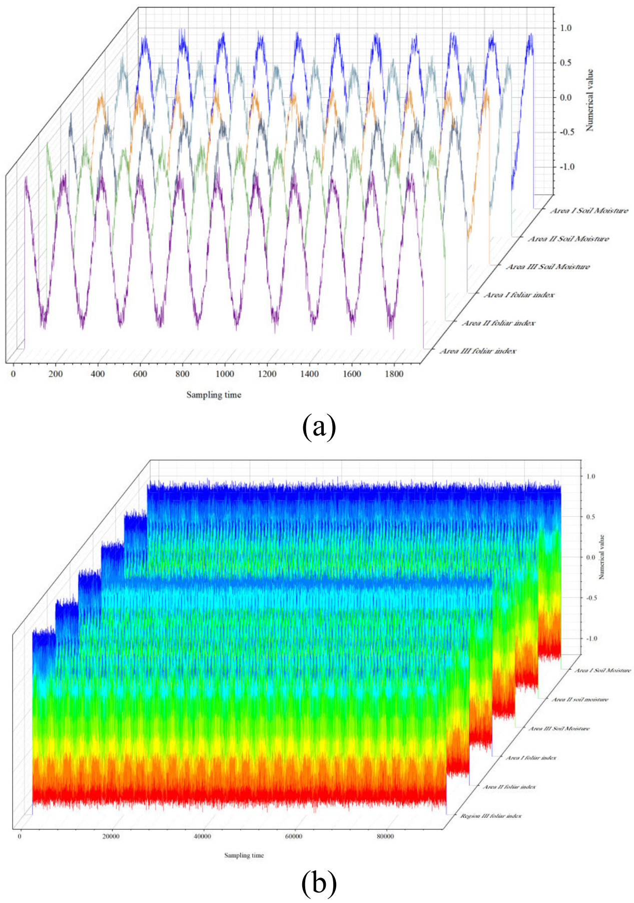


FIGURE 4. Distribution of corresponding data values for all sampling points. Sampling time: January 2014-April 2024 (a) data collection points with a step size of 2 days (12,000 data points in total); (b) data sampling points with a step size of 1 hour (543,000 data points in total).

facilitating faster convergence. Cross-validation was conducted to evaluate the stability of the model across different subsets of the data, with metrics such as root mean square error (RMSE) and mean absolute error (MAE) calculated to assess prediction accuracy. Model performance on the test set confirmed its ability to generalize to unseen data, with detailed results and data flow presented in FIGURE 4.

Soil surface moisture data are obtained from SMAP (Soil Moisture Active Passive) L-band radiometer measurements. Daily data are collected continuously from 6 a.m. to 6 p.m., which is the effective sunshine time for the crop. SMAP measures soil moisture every 2 days, making it possible to observe changes in soil moisture on different time scales. Soil moisture is mainly obtained by microwave sensors, which are highly penetrating to clouds, so there is no need for cloud culling operation, as shown in FIGURE 5.

Leaf Area Index (LAI) is provided by MODIS (Moderate Resolution Imaging Spectroradiometer) sensor measurements, which collect LAI data with a spatial resolution of 500 meters to 1 kilometer and are updated every 1 to 2 weeks to ensure accuracy and real-time availability. Since MODIS is based on data from visible or near-infrared sensors, the atmosphere has a strong absorption of infrared light and therefore needs to be cloud-cleared for use.

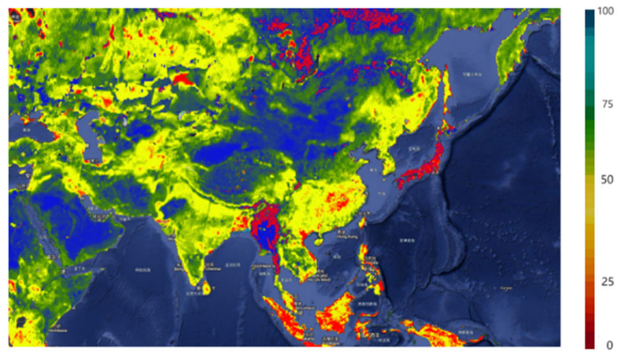


FIGURE 5. Satellite image of soil surface moisture.

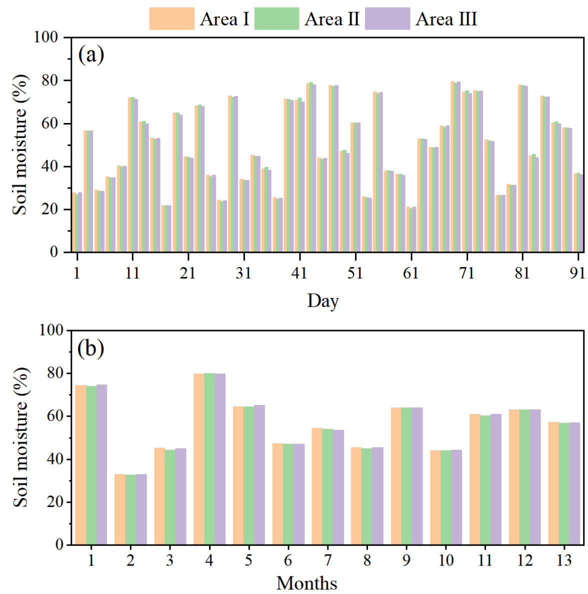


FIGURE 6. Soil moisture at a sampling site in the Yellow and Huaihai regions. (a) Soil moisture at the sampling site from January 1, 2024 to March 31, 2024; (b) Soil moisture from April 2023 to April 2023.

With the above data, we simulated the prediction of soil moisture in the relevant area from 2023 to 2024. Firstly, we collected the soil moisture at different time periods with different steps as the reference value of the predicted data as shown in FIGURE 6.

### B. PREDICTIONS BASED ON THE ENKF-DSSAT ASSIMULATION MODEL

#### 1) DATA PRE-PROCESSING

First, we used the method of removing extreme values to perform cloud culling on the remotely sensed images of leaf area index (LAI). Extreme LAI values are usually caused by cloud cover, which will lead to larger errors between remote sensing data and real data if they are not culled. For grasslands or crops, LAI generally ranges from 1 to 6, and we culled out

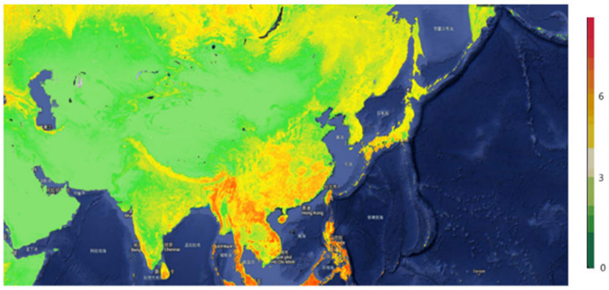


**FIGURE 7.** Image data without cloud rejection.

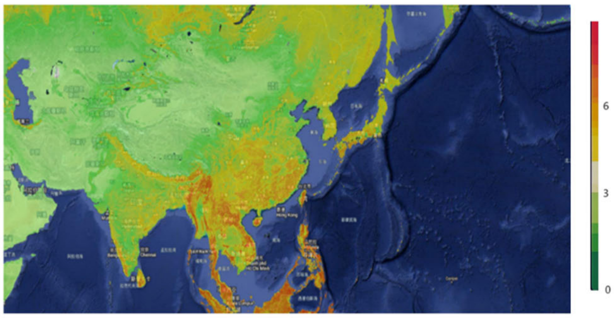


**FIGURE 8.** Satellite image of soil surface moisture.

**TABLE 1.** Error analysis of observed data.

| | Soil moisture | leaf area index (LAI) |
|---|---|---|
| systematic error | -0.2 | +0.1 |
| *random error* | 0.05 | 0.08 |

0 and extreme LAI values above 6. Take two leaf area index impact data on April 20, 2024 as an example.

By FIGURE 7 it can be seen that the images without cloud rejection are brightly colored, which indicates that the data have extreme values at many locations, resulting in large variance in the data, which is not conducive to model analysis; from FIGURE 8 it can be seen that the color of the culled data is significantly less vivid, which indicates that the extreme values have been culled, and the data has less variance, has a reasonable LAI value, and is more conducive to data assimilation by the EnKF model.

Second, the EnKF assimilation model has high requirements for data systematic and random errors, and only data that pass the error test can have better prediction performance after assimilation by the EnKF model. We performed an error analysis on soil moisture and leaf area index after cloud rejection, which included both systematic error and random error. By calculation, we derived the errors as follows TABLE 1 shown:

We analyze the sources of error, which may include instrumental measurement errors, changes in observation conditions and approximations during data processing. It can be seen that the standard value of the systematic error of

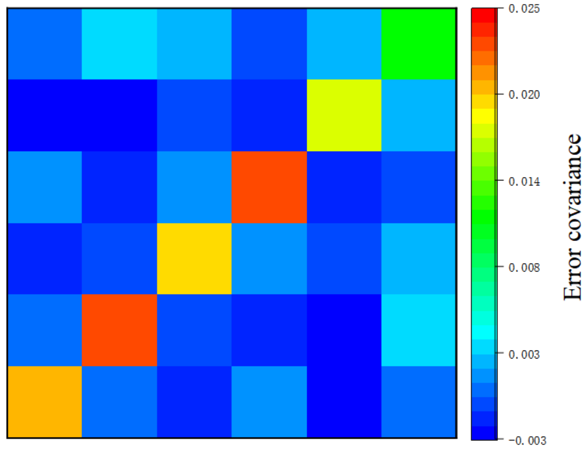


**FIGURE 9.** EnKF model error covariance matrix image.

soil moisture is 0.2 lower than the true value, while the leaf area index is 0.1 higher than the true value, and the random errors of soil moisture and leaf area index are within the 95% confidence interval, which indicates that the error range of our model is consistent with the acceptable accuracy for the input data of the EnKF assimilation model.

Meanwhile, for our three different sampling points, the soil moisture at the sampling points was brought into the EnKF assimilation model, and the mean value of the updated state vectors was measured by calculating the results after passing through the vector matrix as follows:

$$\begin{bmatrix} Soil \quad moisture \\ LAI \end{bmatrix} = \begin{bmatrix} 0.5109 & 1.1217 & 0.9627 \\ 0.7201 & 0.2481 & 0.5486 \end{bmatrix} \tag{4}$$

We initially consider delivering the EnKF prediction results to the DSSAT system for data simulation prediction. Before that, we first calculate the error covariance matrix of the set of data through the EnKF assimilation model as a way to determine whether the data is eligible for substitution into the DSSAT model. Through the calculation, we get the error covariance matrix image as follows FIGURE 9.

From FIGURE 9, it can be seen that the absolute value of the maximum value of the error covariance of the whole matrix is lower than 0.025, and the level of error is small, so it can be considered that the output of EnKF meets the input conditions of the DSSAT system, and is able to be substituted into the DSSAT system and be able to obtain a relatively ideal value.

### 2) ASSIMILATION MODEL PROGECTIONS

We simulated the prediction of the data using the EnKF-DSSAT assimilation model with time steps of 2 days and 1 month corresponding to the short and long term, respectively. As shown in FIGURE 10 and FIGURE 11, the predicted values of soil moisture for the two times steps are obtained, and it can be seen that the predicted values of soil moisture are very good compared to the real values when using the model. At the same time, we found that the error of

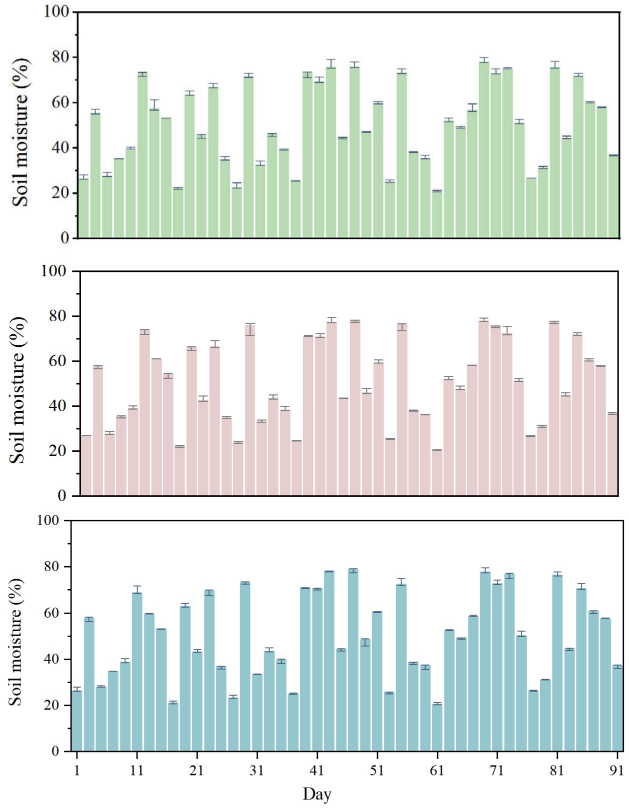


**FIGURE 10.** Short-term soil moisture predictions from the EnKF-DSSAT assimilation model. Prediction period: from January 1, 2024 to March 31, 2024. (a) Predicted soil moisture in area I, (b) predicted soil moisture in area II, and (c) predicted soil moisture in area III.

the short-term prediction is lower than the long-term prediction error, which indicates that the EnKF-DSSAT assimilation model is more effective in predicting the soil moisture in the short term than in the long term. This may be due to the fact that the influence on soil conditions in the short term mainly originates from natural factors, and when the prediction period is extended, the soil conditions will be disturbed by other factors. The results provide an important method for accurate prediction of soil moisture in the short term.

To effectively measure the accuracy of our predictions, we calculated the weighted average of the relative errors of the predicted values (weighted by the reference value of soil moisture) as a technical indicator with the following formula:

$$\begin{aligned} &\textit{average weighted error} \\ &= \sum_{i=1}^{n} \left( \frac{|c_i^*|}{\sum_{i=1}^{n} |c_i^*|} \cdot \frac{c_i - c_i^*|}{|c_i^*|} \right) \\ &= \frac{\sum_{i=1}^{n} |c_i - c_i^*|}{\sum_{i=1}^{n} |c_i^*|} \times 100\% \end{aligned} \tag{5}$$

where $c_i^*$ is the reference value of soil moisture and is the calculated value. Therefore, we can use average weight error as the prediction accuracy.

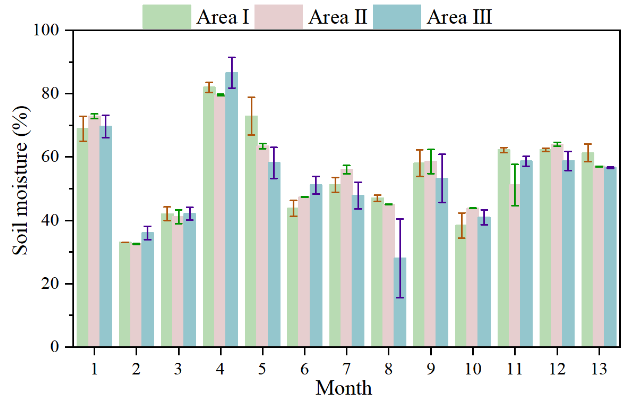


**FIGURE 11.** Long-term soil moisture predictions from the EnKF-DSSATassimilation model. Prediction period: from April 2023 to April 2023.

**TABLE 2.** Prediction accuracy of EnKF-DSSAT integrated model.

| | Region I | Region II | Region III | average value |
|---|---|---|---|---|
| short-term | 97.90% | 98.20% | 98.23% | 98.11% |
| long term | 93.84% | 96.72% | 90.06% | 93.45% |

We calculated the accuracy of the model as shown in Table 2, it can be seen that the EnKF-DSSAT integrated model has a strong short-term prediction ability, with a prediction accuracy of more than 98%, and its long-term prediction ability is weaker compared to the short-term prediction.

### C. PREDICTION BASED ON FUZZY OPTIMIZATION-DSSAT ASSIMILATION MODELS

#### 1) FUZZY OPTIMIZATION PRELIMINARY PREDICTION

We first obtained the remotely sensed soil moisture values and leaf area index before the sampling time point, while adding the policy factors in the use of only fuzzy optimization model to process the data, through the initial prediction, a series of soil moisture prediction values such as FIGURE 12 and 13 shown. Through the error analysis of the predicted values and the actual values, we found that the data processed only by fuzzy optimization had a large error and could not get a good positive result. In order to improve the prediction accuracy, we consider the results of the fuzzy optimized data as the input port to access to the DSSAT system for secondary prediction.

We further calculated the prediction accuracy of a single fuzzy optimization prediction model as shown in Table 3. It can be seen that the prediction of single fuzzy optimization is not satisfactory, so we improved it using fuzzy optimization-DSSAT assimilation model.

#### 2) PREDICTION BASED ON FUZZY OPTIMIZATION-DSSAT ASSIMILATION MODEL

Based on the preliminary prediction results of fuzzy optimization, we found that the accuracy of the single prediction

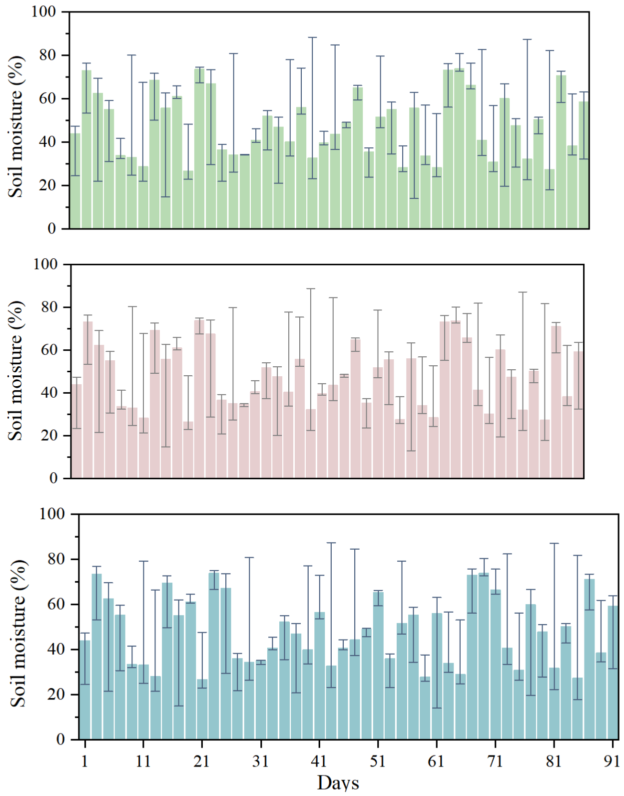


**FIGURE 12.** Short-term soil moisture predictions under the fuzzy optimization model. Time: January 1, 2024 to March 31, 2024.
(a) Predicted values of soil moisture in area I, (b) predicted values of soil moisture in area II, and (c) predicted values of soil moisture in area III.

**TABLE 3.** Prediction accuracy of EnKF-DSSAT integrated model.

| | Region I | Region II | Region II | average value |
|---|---|---|---|---|
| short-term | 61.14% | 30.91% | 60.92% | 60.96% |
| long term | 59.38% | 59.20% | 59.48% | 59.35% |

method is low, which means that the reference value of the prediction results is not high. In order to improve the prediction accuracy, we used a fuzzy optimization-DSSAT integrated model to predict the soil moisture in the same time period, as shown in FIGURE 14 and FIGURE 15. We found that the prediction accuracy of the fuzzy optimization-DSSAT assimilation model was higher compared to the single fuzzy optimization model. Through analysis, we believe that it is due to the characteristic of the fuzzy optimization model's strong dependence on constraints, by combining the fuzzy optimization model with the DSSAT system, simultaneously considering the effects of natural and human factors on crop production conditions, and combining time-sensitive and reliable multi-source data, which leads to an effective improvement of the model's time scale. The fuzzy optimization-DSSAT assimilation model effectively reduces the time complexity of the algorithm by fusing multi-source

**TABLE 4.** Fuzzy optimization-DSSAT assimilation model prediction accuracy.

| | Region I | Region II | Region III | average value |
|---|---|---|---|---|
| short-term | 93.42% | 93.53% | 93.49% | 93.48% |
| Long-term | 94.31% | 94.77% | 92.27% | 93.78% |

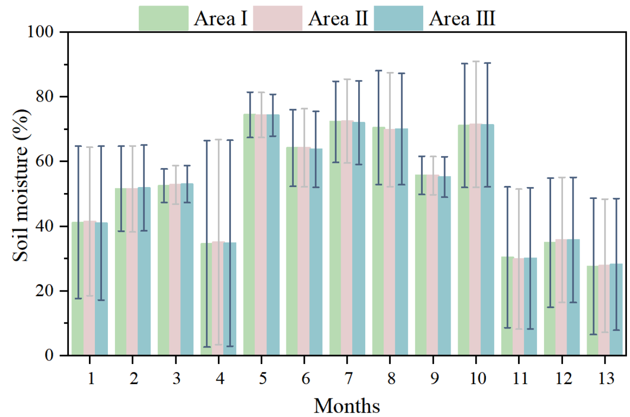


**FIGURE 13.** Long-term soil moisture predictions under the fuzzy optimization model. Time: April 2023 to April 2023.

real-time data, increasing the update step size, and narrowing the amount of data computation.

By calculation we conclude that the prediction accuracy of the fuzzy optimization-DSSAT assimilation model is shown in Table 4. It can be seen that compared with the single prediction model, the prediction accuracy of the assimilation model is significantly improved, and for the long-term prediction, the accuracy is improved to more than 96.27%.

### D. ABLATION EXPERIMENTS-SELECTION AND DETERMINATION OF OPTIMAL DATA SOURCES

In order to verify that multi-source data can make more accurate predictions compared to single-source data, rather than variability in predicted values due to the model, we used ablation experiments. Our single-source-DSSAT group input only single-source data-soil surface moisture into DSSAT for prediction, while the multi-source-DSSAT group input multi-source data-soil surface area moisture and leaf area index together into the DSSAT model into DSSAT for prediction.

In order to objectively evaluate and test the accuracy of our prediction model, we used three indicators, root mean square error, average absolute error, and Nash efficiency coefficient, as references for the purpose of effectively measuring the prediction accuracy.

#### 1) ROOT MEAN SQUARE ERROR

Root Mean Square Error (RMSE) indicates the average degree of deviation between the predicted value and the true value. The formula is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{mod\ el,i})^2}{n}} \tag{6}$$

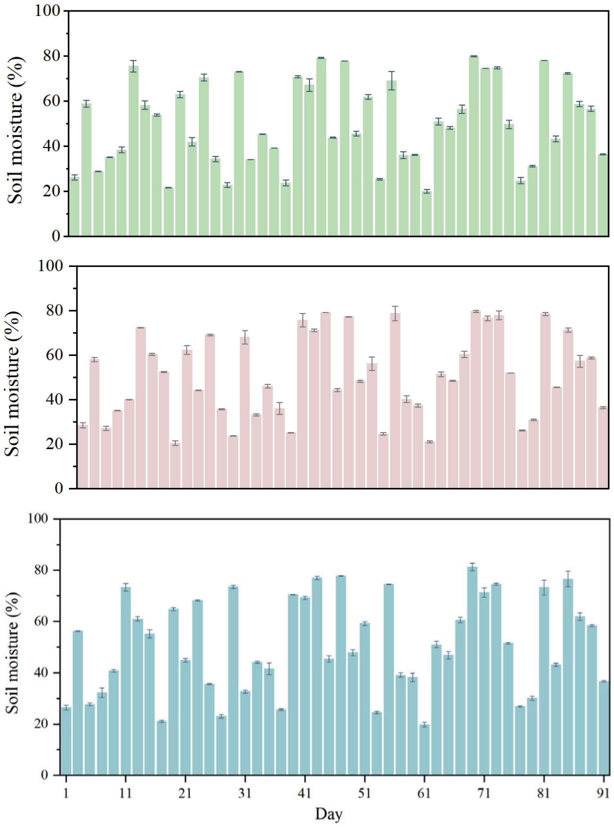


**FIGURE 14.** L Short-term soil moisture predictions under the fuzzy optimization-DSSAT assimilation model. Prediction period: from January 1, 2024 to March 31, 2024. (a) Predicted soil moisture in area I, (b) predicted soil moisture in area II, and (c) predicted soil moisture in area III.

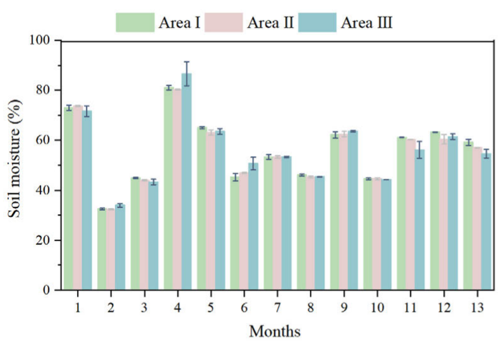


**FIGURE 15.** Long-term soil moisture predictions under the fuzzy optimization-DSSAT assimilation model. Prediction period: from April 2023 to April 2023.

where $X_{obs,i}$ is the model predicted value, $X_{\mathrm{mod}\ el,i}$ is the model observed value, and is the total number of samples.

### 2) MEAN ABSOLUTE ERROR

The Mean Absolute Error (MAE) is the average of the absolute values of the deviations of all individual observations from the arithmetic mean. The formula is as follows:

$$MAE = \frac{1}{n}\sum_{1}^{n}\left|X_{obs,i} - X_{mod\ el,i}\right| \tag{7}$$

**TABLE 5.** Short-term prediction errors for multi-source-DSSAT and single-source-.

| | Single source (soil moisture) - DSSAT | Multi-source (soil moisture + LAI) - DSSAT |
|---|---|---|
| Area 1 RMSE | 5.48 | 1.78 |
| Area 1 MAE | 4.87 | 1.07 |
| Area 1 NSE | 0.80 | 0.95 |
| Area 2 RMSE | 5.01 | 1.80 |
| Area 2 MAE | 4.63 | 0.92 |
| Area 2 NSE | 0.82 | 0.94 |
| Area 3 RMSE | 5.38 | 1.61 |
| Area 3 MAR | 4.82 | 0.90 |
| Area 3 NSE | 0.83 | 0.92 |

**TABLE 6.** Long-term prediction errors for multi-source-DSSAT and single-source-dssat.

| | Single source (soil moisture) - DSSAT | Multi-source (soil moisture + LAI) - DSSAT |
|---|---|---|
| Area 1 RMSE | 5.94 | 1.17 |
| Area 1 MAE | 5.97 | 2.48 |
| Area 1 NSE | 0.79 | 0.89 |
| Area 2 RMSE | 4.99 | 1.00 |
| Area 2 MAE | 5.72 | 1.84 |
| Area 2 NSE | 0.76 | 0.93 |
| Area 3 RMSE | 5.97 | 2.86 |
| Area 3 MAR | 5.33 | 1.61 |
| Area 3 NSE | 0.78 | 0.69 |

### 3) NASH EFFICIENCY FACTOR

The Nash efficiency coefficient (NSE) is used to quantify the predictive accuracy of the simulation model with the following formula:

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(X_{mod\ el,i} - X_{obs,i}\right)^2}{\sum_{i=1}^{n}\left(X_{mod\ el,i} - \bar{X}_{mod\ el}\right)} \tag{8}$$

where $\bar{X}_{\mathrm{mod}\ el}$ is the average of the model observations. The closer the value of NSE is to 1, the better the quality of the simulation and the higher the confidence of the model.

We conducted long-term and short-term experiments for the models of single-source-DSSAT and multi-source-DSSAT at three sampling points, and measured the RMSE, MAE, and NSE coefficients for each sampling point, respectively, and the results are as follows Tables 5 and 6.

By Table 5, the Table 6 shows that the three indicators of multisource-DSSAT are significantly better than single-source-DSSAT in areas 1, 2, and 3, both in the long term and the short term. the RMSE of multisource-DSSAT is the lowest at 1.00, while the RMSE of single-source-DSSAT is the lowest at 4.99, and the MAE of multisource-DSSAT is the lowest at 1.61, while the MAE of single-source-DSSAT is the lowest at 5.33. This indicates that the deviation between the predicted and actual values of multisource-DSSAT is smaller, which can improve the prediction accuracy of the model using multisource data. indicates that the deviation between the predicted and actual values of Multi-source-DSSAT is smaller, and the use of multi-source data can improve the prediction accuracy of the model. For prediction accuracy, Multi-source-DSSAT accuracy can reach up

**TABLE 7.** Error comparison of multi-model models for short-term forecasting.

| | SVM | Transformer | EnKF-DSSAT | Fuzzy Optimization-DSSAT1 | Fuzzy Optimization-DSSAT2 |
|---|---|---|---|---|---|
| Area 1 RMSE | 2.95 | 4.20 | 1.29 | 1.99 | 1.78 |
| Area 1 MAE | 2.15 | 3.42 | 1.07 | 1.54 | 1.32 |
| Area 1 NSE | 0.93 | 0.92 | 0.95 | 0.92 | 0.95 |
| Area 2 RMSE | 2.90 | 4.03 | 1.14 | 1.82 | 1.80 |
| Area 2 MAE | 2.40 | 3.40 | 0.92 | 1.33 | 1.26 |
| Area 2 NSE | 0.94 | 0.94 | 0.95 | 0.90 | 0.95 |
| Area 3 RMSE | 2.91 | 3.86 | 1.09 | 1.73 | 1.61 |
| Area 3 MAR | 2.32 | 3.05 | 0.90 | 1.30 | 1.27 |
| Area 3 NSE | 0.94 | 0.90 | 0.94 | 0.92 | 0.95 |

**TABLE 8.** Comparison of errors in multiple models for long-term forecasting.

| | SVM | Transformer | EnKF-DSSAT | Fuzzy Optimization-DSSAT1 | Fuzzy Optimization-DSSAT2 |
|---|---|---|---|---|---|
| Area 1 RMSE | 2.45 | 5.03 | 4.18 | 1.66 | 1.17 |
| Area 1 MAE | 1.99 | 4.00 | 3.48 | 1.89 | 0.96 |
| Area 1 NSE | 0.94 | 0.83 | 0.85 | 0.91 | 0.94 |
| Area 2 RMSE | 2.35 | 4.98 | 3.19 | 1.43 | 1.00 |
| Area 2 MAE | 1.98 | 4.23 | 1.84 | 1.88 | 0.69 |
| Area 2 NSE | 0.92 | 0.82 | 0.90 | 0.92 | 0.94 |
| Area 3 RMSE | 2.25 | 5.01 | 7.05 | 1.76 | 2.86 |
| Area 3 MAR | 1.78 | 4.23 | 5.61 | 1.77 | 2.10 |
| Area 3 NSE | 0.94 | 0.81 | 0.65 | 0.85 | 0.93 |

to 0.94, while Single-source-DSSAT accuracy reaches up to 0.79. Accuracy NSE = 1 indicates perfect prediction, while NSE = 0 indicates that the model prediction is not as good as the average value. This results in a 0.15 improvement in model prediction accuracy for multi-source-DSSAT compared to single-source-DSSAT. It is concluded that multi-source-DSSAT can improve the prediction accuracy of the model.

### E. COMPARATIVE EXPERIMENTS

Based on the ablation experiments, it is known that multi-source data is beneficial to improve the accurate prediction of irrigation water. To use multi-source data to participate in the DSSAT model for assimilation prediction, two multi-source data assimilation models are used in this paper, which are EnKF and fuzzy optimization assimilation model. In order to explore the effectiveness of the assimilation models and to compare the multi-source data assimilation effect between the models, we conducted a comparison test. The fuzzy optimization-DSSAT1 assimilation model and the EnKF-DSSAT assimilation model use the same natural direction of the two multi-source data for assimilation prediction. The fuzzy optimization-DSSAT2 assimilation model uses not only two natural direction data, but also one human direction data, government allowance, for assimilation prediction. The EnKF model, the fuzzy optimization model, and the DSSAT model are all well recognized as effective prediction models in biology and agronomy. Transformer and Support Vector Machine (SVM) are the more effective prediction models in the field of computers, and we conducted a cross sectional comparison experiment of these five models in the field of irrigation water prediction. The results are shown in Tables 7 and 8.

By comparing the EnKF-DSSAT assimilation model and the fuzzy optimization-DSSAT1 assimilation model with the same data source, we can find that for the EnKF-DSSAT assimilation model, the short-term data performs better, and the RMSE and MAE at each location are lower, with a minimum of 1.09 and 0.90, and the accuracy of the predicted values of the model reaches a maximum of 0.94, which suggests that the that the model prediction effect is more accurate and has a good fit with the actual data. The fuzzy optimization-DSSAT1 assimilation model performs better in long-term prediction, with RMSE and MAE reaching a minimum of 1.00 and 0.69, which indicates that the prediction results have a small error with the actual values. The accuracy of the predicted values, NSE, can reach a maximum of 0.94, which indicates that the model predicted values almost perfectly predicted the actual data.

By comparing the Fuzzy Optimization-DSSAT1 model with the Fuzzy Optimization-DSSAT2 model, we can find that the Fuzzy Optimization-DSSAT2 model with the humanities data added to the government subsidies has an improvement close to 0.3 in all three indicators. The errors of RMSE and MAE of the Fuzzy Optimization-DSSAT2 model in the long-term data decreased by more than 0.4, and the NSE was closer to 1. Therefore, the inclusion of humanities metrics can better predict irrigation water in the long-term. It is hypothesized that humanistic factors may have a large impact on irrigation water in the long-term time dimension.

By comparing artificial neural networks in machine learning - Transformer, SVM model and our EnKF-DSSAT, Fuzzy Optimization-DSSAT2 assimilation model. We find that the EnKF-DSSAT assimilation model has an NSE metric of 0.95 in most cases, which is closer to perfect prediction than both Transformer and SVM. This is because the EnKF-DSSAT assimilation model assimilates real-time data in two times steps, long-term and short-term, and substantially improves the reliability and accuracy of the model predictions by continuously adding accurate real-time data to the model. The EnKF-DSSAT assimilation model, on the other hand, mainly utilizes natural data, so its prediction effect in the short term is better. The prediction accuracy NSE index of the fuzzy optimization-DSSAT assimilation model is mainly 0.94, which is also better than Transformer and SVM models. This is also due to the fact that the fuzzy optimization-DSSAT assimilation model utilizes real-time data for real-time adaptive adjustment and updating of model predictions. It is due to this adaptive parameter tuning and updating that the Fuzzy Optimization-DSSAT assimilation model is more robust and yields excellent prediction accuracy even when long-term irrigation water predictions are made. The prediction performance of both the EnKF-DSSAT model and the fuzzy optimization-DSSAT model due to the conventional model, both in the long and short term, is attributed to the fact that they have continuously fused multiple sources of real-time data for data assimilation, which greatly improves the accuracy and reliability of the predictions.

This study incorporates policy factors into the irrigation demand forecasting model by introducing government subsidy data, and through comparative experiments, it is found that the inclusion of humanistic factors can indeed effectively improve the forecasting accuracy, which leads to the following conclusions. Specifically, government subsidies for efficient water-saving technologies (e.g., drip irrigation, sprinkler irrigation) can significantly reduce farmers' equipment acquisition costs and improve the problem of water wastage in traditional irrigation methods. For example, in the Huanghuaihai Plain, subsidizing a certain amount of water-saving technology per mu of land can directly increase the proportion of farmers adopting drip irrigation and reduce the reliance on groundwater resources in traditional flood irrigation. In addition, differentiated subsidy policies can more precisely target the planting needs of different cash crops and food crops. For example, for areas where fruit crops with high economic value are grown, appropriately increasing the proportion of subsidized high-efficiency irrigation equipment can not only increase farmers' earnings, but also enhance the attractiveness and implementation effect of the policy.

However, subsidy policies may also face a series of specific challenges in implementation, such as insufficient policy coverage or inefficient allocation in some regions, which makes it difficult for small-scale or low-income farmers to enjoy subsidy support; at the same time, some farmers may fail to make full use of the subsidies due to insufficient technological suitability. In addition, it is important to strengthen the dynamic feedback mechanism of the subsidy policy and optimize the policy layout of water resource management and water-saving technology promotion in the region through regular monitoring and adjustment of the subsidy effect, so as to achieve more scientific irrigation regulation and resource allocation.

## IV. DISCUSSION

In this study, we used an accurate assimilation prediction of irrigation water for maize based on the Ensemble Kalman Filter and Fuzzy Optimization-DSSAT model, which showed excellent prediction accuracy in experiments in the Yellow Huaihai Plain. To assess the effectiveness of the model, we compared it with several other existing soil moisture prediction models, including the linear regression model proposed by Karan et al. [38] The artificial neural network (ANN) model developed by Agatonovic-Kustrin and Beresford [39]. The support vector machine (SVM) model proposed by Wang and Hu [40] and the deep learning regression network (DNNR) model cited in Cai et al. [41]. These models represent various techniques ranging from traditional statistics to deep learning-based techniques, respectively.

In our model comparison, we focus on key performance metrics such as mean square error (MSE), mean absolute error (MAE) and coefficient of determination ($R^2$) to quantitatively analyze the performance of different models. Our model performs well on all evaluation metrics, especially on the coefficient of determination ($R^2$) which is close to 1, which is much higher than the other models, which highlights the advantage of our model in terms of data fitting ability. In addition, the values of MAE and MSE are significantly lower than those of the compared models, indicating that our model has a higher predictive accuracy and lower error rate.

Our model has a higher data fitting ability compared to traditional models, thanks to the effective combination of ensemble Kalman filtering and fuzzy optimization techniques, which enables the model to capture the dynamic changes of soil moisture more accurately. In addition, the generalization ability of the model is also stronger, which is not restricted by geographical location and can be effectively applied in different regions. More importantly, the model can effectively integrate real-time data to provide immediate and accurate predictions for agricultural irrigation management, which is important for improving irrigation efficiency and crop yield.

Although our model has demonstrated many advantages, there is still room for further improvement. Future studies will explore further optimization of model parameters to improve prediction accuracy and reduce overfitting, as well as integration of more types of data sources, such as remotely sensed data and soil properties data, to further enrich the input characteristics of the model and enhance its predictive capability. In addition, testing the model's adaptability to different climates and soil types will help to validate the model's generalization ability and provide a solid foundation for its wider application. Although this study focuses on the Huanghuaihai Plain and has achieved some success in constructing an irrigation water demand prediction model by combining soil surface moisture, leaf area index and human data, the scalability and applicability of the model to other climate types or agricultural conditions have not been thoroughly explored. This issue is of great significance because evapotranspiration characteristics, precipitation patterns, and agricultural cropping structures in different climate zones may significantly affect the drivers and weights of irrigation demand. For example, in the humid south, abundant precipitation may reduce the importance of soil moisture as a predictor variable, whereas the arid north may need to pay more attention to the dynamics of deep soil moisture. In addition, differences in crop types, cropping cycles, and regional cash crop proportions under different agricultural conditions may also require adjustments to model parameters or the introduction of new variables. Therefore, future research should assess the applicability of the model under different climatic and agricultural conditions through cross-regional validation, and optimize the model structure according to specific regional characteristics to enhance its practical value and potential for dissemination. These explorations and improvements will provide more accurate and practical technical support for future agricultural water management.

While the technical aspects of this study have been thoroughly addressed, it is equally important to evaluate the economic and environmental implications of implementing the proposed model, as these factors are critical for its adoption in real-world scenarios. Economically, the use of a precise irrigation prediction model can significantly reduce water usage, thereby lowering costs for farmers and optimizing regional water resource allocation. Studies have shown that efficient irrigation management systems can reduce water expenditures by up to 30%, directly benefiting agricultural profitability (Zhang et al., 2022). In the context of the Yellow-Huai-Hai Plain, where water scarcity and agricultural productivity are closely intertwined, such cost savings can have a transformative impact on both smallholder and large-scale farming operations.

Environmentally, the model's implementation is expected to mitigate issues such as groundwater overexploitation and soil salinization, which are prevalent in over-irrigated areas. By optimizing irrigation schedules and volumes, the model can help maintain soil health and prevent long-term degradation of agricultural land. Moreover, a reduction in unnecessary water pumping will lower energy consumption, contributing to decreased carbon emissions and alignment with broader climate goals.

However, these benefits are not without challenges. Initial investments in data collection infrastructure, such as soil moisture sensors and remote sensing technology, may pose financial barriers, particularly for resource-constrained regions. Furthermore, the model's reliance on real-time data integration requires robust technological and institutional support, which could limit its scalability in underdeveloped areas.

Future research should quantitatively assess these economic and environmental trade-offs through cost-benefit analyses and scenario modeling, providing a more comprehensive understanding of the broader impacts of the proposed system. Such evaluations will be crucial for policymakers and stakeholders to make informed decisions regarding the adoption and implementation of this model in diverse agricultural settings.

Implementation of the precision irrigation system proposed in this study has significant potential in terms of cost savings, resource efficiency and environmental benefits. By introducing advanced Enkf-DSSAT and Fuzzy Optimization-DSSAT models, the system can significantly optimize irrigation decisions, improve water use efficiency, and reduce resource wastage during agricultural production.

The first can save potential costs. Through real-time data assimilation and long-term optimization planning, the precision irrigation system is able to effectively reduce unnecessary irrigation practices, thus reducing agricultural water costs. It is estimated that water-saving irrigation technology can reduce the amount of irrigation water by about 20-40% compared with the traditional high water diffusion irrigation method. In addition, due to the optimization of irrigation timing and frequency, the running time and energy consumption of pumps and other equipment are also significantly reduced, thus reducing electricity expenses. For regions such as the Yellow and Huaihai, where grain production is the mainstay, the system is expected to save irrigation costs ranging from 20-50 RMB per mu. These cost savings are especially important for small and medium-sized farmers, and help improve the economic efficiency of agricultural production.

Second, resource efficiency is improved. The system integrates multiple sources of data, such as soil moisture and leaf area index, to accurately calculate crop water requirements and avoid water wastage due to over-irrigation. Through a long-term optimization model (Fuzzy Optimization-DSSAT), the system is also able to dynamically allocate water resources in different seasons and crop growth stages to maximize the productivity of water resources. For example, in dry years or water-scarce areas, the system can prioritize the water needs of cash crops or high-yield crops, thus improving the resource utilization efficiency of the overall agricultural system.

The third has environmental benefits. The popularization of precision irrigation technology has multiple positive effects on environmental protection. First of all, reducing excessive irrigation can help alleviate the problem of over-exploitation of groundwater, thus maintaining the ecological balance of the regional water cycle. Secondly, reducing excess water infiltration can effectively prevent soil nutrient loss and salinization, and maintain the long-term productivity of the soil. In addition, by reducing energy consumption in the extraction and delivery of irrigation water, the system can indirectly reduce carbon emissions from agricultural production and contribute to the development of low-carbon agriculture.

## V. CONCLUSION

The multi-source data assimilation model proposed in this study realizes the real-time fusion and utilization of multi-source data such as soil moisture, leaf area index, and government subsidies by fusing the EnKF model and fuzzy optimization model. This new approach can effectively improve the accuracy of maize irrigation water demand prediction, especially in solving the problems of poor accuracy and low resource utilization of traditional irrigation systems, which shows significant advantages. Meanwhile, the innovation of the model is that it combines natural and human factors, and has been used to improve the reliability and accuracy of the predicted values by continuously fusing real-time multi-source data in both long-term and short-term time steps. It provides accurate short-term irrigation water forecasts for operators and reliable long-term water resource forecasts for the government.

Through experimental verification, we come to the following conclusions:

(i) Multi-source data assimilation can effectively improve the accuracy of model prediction.

Through ablation experiments, we found that the use of multi-source data can effectively improve the model's prediction accuracy of water. We integrated the multi-source data by EnKF model or fuzzy optimization model. The remote sensing data of soil moisture and leaf area index were integrated using the EnKF model. The fuzzy optimization model was used to integrate human and natural factors, and government subsidies and natural factors were integrated together as influencing factors for the preliminary prediction of irrigation water. The results of the ablation test proved that the data integrating multiple sources have certain advantages in the long-term prediction of accurate irrigation water resources. Thus, multi-source data helps us to help farmers to realize more accurate prediction with finer granularity.

(ii) Natural factors have a greater short-term impact on precision irrigation water.

By comparing the joint EnKF-DSSAT model with the fuzzy optimization- DSSAT model, we can see that the EnKF-DSSAT model performs better in short-term prediction, which may be due to the fact that natural factors have a greater impact on the prediction of irrigation water in short-term time (within one or two days), and the possibility of policy change is very small in the short term, so the policy impact is weak and negligible at this time. Therefore the EnKF-DSSAT model using only natural factors performs better and has an advantage in short-term prediction.

(iii) Human factors have a macro-regulatory function and have a greater impact on long-term forecasts.

Fuzzy Optimization- DSSAT model performs better on long-term forecasting. In the fuzzy optimization- DSSAT model, we added the variable of government subsidies of human factors, combined with two natural indicators to predict the model, and after the experiment, we found that the model has a better effect on the long-term prediction. The reason is summarized through the analysis that the policy factors become more influential at this time. Farmers' planting status and pre-investment are greatly affected by the national policy bonus allowance, so the human factor will have an impact on the prediction results. Therefore, the fuzzy optimization-DSSAT model with the consideration of human factors and government subsidies performs better.

(iv) The assimilation model can substantially improve the model prediction accuracy.

Both the EnKF-DSSAT assimilation model and the Fuzzy Optimization-DSSAT assimilation model demonstrate excellent prediction accuracy in comparative experiments. Compared to traditional machine learning models, the assimilation models are different in that they utilize real-time multi-source data for assimilation, which guides the model's parameter tuning and prediction in real-time, and thus can predict water resources more reliably and accurately.

The application of the Enkf-DSSAT and Fuzzy Optimization-DSSAT models to other geographic regions with different agricultural practices and environmental conditions requires adaptation to regional characteristics and agro-ecological differences. The Enkf-DSSAT model, as a short-term forecasting method based on assimilation of real-time data, is suitable for environmental conditions with frequent dynamic changes. When applied in different regions, the availability and quality of the data sources need to be ensured first, e.g. the spatial distribution of soil moisture sensors or weather station data may be adapted to local conditions. In addition, the regional calibration of model parameters is a key aspect, especially the variety characteristics, sowing period and fertilization method in the crop growth model, which need to be recalibrated with local experimental data to ensure the accuracy of the model prediction. Based on the real-time responsiveness of this model, it is especially valuable for application in regions with erratic precipitation or large fluctuations in crop water requirements (e.g., South Asian monsoon region or sub-Saharan Africa) to optimize irrigation timing and effectively respond to the impact of short-term weather extremes on agriculture.

On the other hand, the Fuzzy Optimization-DSSAT model focuses on long-term forecasting and decision optimization, which is mainly used for medium- and long-term planning of complex agroecosystems. When migrating to other geographic regions, the regional specificity of variables such as long-term climate patterns, land use changes and water supply patterns need to be fully considered. For example, in arid and semi-arid regions, the model needs to integrate local data on the sustainable use of water resources and reflect regional uncertainty characteristics through fuzzy optimization mechanisms. In addition, different agricultural management practices (e.g., crop rotation, intercropping, or water-saving irrigation techniques) may have different weights on long-term production objectives, which requires the model to optimize the construction of the objective function in conjunction with regional policy and economic constraints. With these adjustments, the Fuzzy Optimization-DSSAT model can provide scientific long-term planning support for agricultural production and resource management in different regions to achieve sustainable regional agricultural development.

## REFERENCES

[1] E. A. Abioye, O. Hensel, T. J. Esau, O. Elijah, M. S. Z. Abidin, A. S. Ayobami, O. Yerima, and A. Nasirahmadi, "Precision irrigation management using machine learning and digital farming solutions," *AgriEngineering*, vol. 4, no. 1, pp. 70–103, Feb. 2022.

[2] I. A. Lakhiar, H. Yan, C. Zhang, G. Wang, B. He, B. Hao, Y. Han, B. Wang, R. Bao, T. N. Syed, J. N. Chauhdary, and M. Rakibuzzaman, ''A review of precision irrigation water-saving technology under changing climate for enhancing water use efficiency, crop yield, and environmental footprints,'' *Agriculture*, vol. 14, no. 7, p. 1141, Jul. 2024.

[3] F. Fuentes-Peñailillo, K. Gutter, R. Vega, and G. C. Silva, ''Transformative technologies in digital agriculture: Leveraging Internet of Things, remote sensing, and artificial intelligence for smart crop management,'' *J. Sensor Actuator Netw.*, vol. 13, no. 4, p. 39, Jul. 2024.

[4] C. E. Hachimi, S. Belaqziz, S. Khabba, B. Sebbar, D. Dhiba, and A. Chehbouni, ''Smart weather data management based on artificial intelligence and big data analytics for precision agriculture,'' *Agriculture*, vol. 13, no. 1, p. 95, Dec. 2022.

[5] M. Zhang, T. Cao, and X. Zhao, ''Applying sensor-based technology to improve construction safety management,'' *Sensors*, vol. 17, no. 8, p. 1841, Aug. 2017.

[6] Y. Wei, Z. Wang, and T. Wang, ''Design of real-time soil moisture monitoring and precision irrigation systems,'' *Trans. Chin. Soc. Agricult. Eng.*, vol. 29, no. 17, pp. 80–86, 2013.

[7] F. Ferrández-Pastor, J. García-Chamizo, M. Nieto-Hidalgo, J. Mora-Pascual, and J. Mora-Martínez, ''Developing ubiquitous sensor network platform using Internet of Things: Application in precision agriculture,'' *Sensors*, vol. 16, no. 7, p. 1141, Jul. 2016.

[8] Y. E. M. Hamouda and B. H. Y. Elhabil, ''Precision agriculture for greenhouses using a wireless sensor network,'' in *Proc. Palestinian Int. Conf. Inf. Commun. Technol. (PICICT)*, Gaza, Palestine, May 2017, pp. 78–83.

[9] P. Jayaraman, A. Yavari, D. Georgakopoulos, A. Morshed, and A. Zaslavsky, ''Internet of Things platform for smart farming: Experiences and lessons learnt,'' *Sensors*, vol. 16, no. 11, p. 1884, Nov. 2016, doi: 10.3390/s16111884.

[10] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. N. Hindia, ''An overview of Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges,'' *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3758–3773, Oct. 2018, doi: 10.1109/JIOT.2018.2844296.

[11] H. Jawad, R. Nordin, S. Gharghan, A. Jawad, and M. Ismail, ''Energy-efficient wireless sensor networks for precision agriculture: A review,'' *Sensors*, vol. 17, no. 8, p. 1781, Aug. 2017, doi: 10.3390/s17081781.

[12] A. Tzounis, N. Katsoulas, T. Bartzanas, and C. Kittas, ''Internet of Things in agriculture, recent advances and future challenges,'' *Biosystems Eng.*, vol. 164, pp. 31–48, Dec. 2017, doi: 10.1016/j.biosystemseng.2017.09.007.

[13] G. Bitella, R. Rossi, R. Bochicchio, M. Perniola, and M. Amato, ''A novel low-cost open-hardware platform for monitoring soil water content and multiple soil-air-vegetation parameters,'' *Sensors*, vol. 14, no. 10, pp. 19639–19659, Oct. 2014, doi: 10.3390/s141019639.

[14] B. A. M. Bouman, C. A. van Diepen, P. Vossen, and T. van der Wal, ''Simulation and systems analysis tools for crop yield forecasting,'' in *Applications of Systems Approaches at the Farm and Regional Levels*, P. S. Teng, Ed., Dordrecht, The Netherland: Kluwer, 1997, pp. 325–340.

[15] A. V. M. Ines and J. W. Hansen, ''Bias correction of daily GCM rainfall for crop simulation studies,'' *Agricult. Forest Meteorol.*, vol. 138, nos. 1–4, pp. 44–53, Aug. 2006, doi: 10.1016/j.agrformet.2006.03.009.

[16] A. Kamilaris and F. X. Prenafeta-Boldú, ''Deep learning in agriculture: A survey,'' *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018, doi: 10.1016/j.compag.2018.02.016.

[17] L. N. Christofides, A. Polycarpou, and K. Vrachimis, ''Gender wage gaps, 'sticky floors' and 'glass ceilings' in Europe,'' *Labour Econ.*, vol. 21, pp. 86–102, Apr. 2013, doi: 10.1016/j.labeco.2013.01.003.

[18] M. Huang, M. Tian, Y. Liu, Y. Zhang, and J. Zhou, ''Parameter optimization of PID controller for water and fertilizer control system based on partial attraction adaptive firefly algorithm,'' *Sci. Rep.*, vol. 12, no. 1, p. 12182, Jul. 2022.

[19] D. Wallach and P. J. Thorburn, ''Estimating uncertainty in crop model predictions: Current situation and future prospects,'' *Eur. J. Agronomy*, vol. 88, pp. A1–A7, Aug. 2017.

[20] P. Hoefsloot et al., ''Combining crop models and remote sensing for yield prediction: Concepts, applications and challenges for heterogeneous smallholder environments,'' Tech. Rep., 2012.

[21] L. Chen, B. Han, X. Wang, J. Zhao, W. Yang, and Z. Yang, ''Machine learning methods in weather and climate applications: A survey,'' *Appl. Sci.*, vol. 13, no. 21, p. 12019, Nov. 2023.

[22] Mondal, ''Building a low-cost, low-power wireless sensor network for agriculture purpose,'' 2016.

[23] D. A. V. Grubert, ''A synergistic approach to sugarcane yield forecasting using machine learning, remote sensing, and process-based modeling,'' Universidade de São Paulo, São Paulo, Brazil, Tech. Rep., 2023.

[24] E. A. Abioye, M. S. Z. Abidin, M. S. A. Mahmud, S. Buyamin, M. H. I. Ishak, M. K. I. A. Rahman, A. O. Otuoze, P. Onotu, and M. S. A. Ramli, ''A review on monitoring and advanced control strategies for precision irrigation,'' *Comput. Electron. Agricult.*, vol. 173, Jun. 2020, Art. no. 105441.

[25] Y. L. Gavasso-Rita, S. M. Papalexiou, Y. Li, A. Elshorbagy, Z. Li, and C. Schuster-Wallace, ''Crop models and their use in assessing crop production and food security: A review,'' *Food Energy Secur.*, vol. 13, no. 1, p. 503, Jan. 2024.

[26] N. Montaldo, A. Gaspa, and R. Corona, ''Multiscale assimilation of sentinel and Landsat data for soil moisture and leaf area index predictions using an ensemble-Kalman-filter-based assimilation approach in a heterogeneous ecosystem,'' *Remote Sens.*, vol. 14, no. 14, p. 3458, Jul. 2022.

[27] G. Wei, Y. Li, Z. Zhang, Y. Chen, J. Chen, Z. Yao, C. Lao, and H. Chen, ''Estimation of soil salt content by combining UAV-borne multispectral sensor and machine learning algorithms,'' *PeerJ*, vol. 8, p. e9087, Apr. 2020.

[28] P. Kumar, S. H. Lai, N. S. Mohd, M. R. Kamal, A. N. Ahmed, M. Sherif, A. Sefelnasr, and A. El-Shafie, ''Enhancement of nitrogen prediction accuracy through a new hybrid model using ant colony optimization and an Elman neural network,'' *Eng. Appl. Comput. Fluid Mech.*, vol. 15, no. 1, pp. 1843–1867, Jan. 2021.

[29] L. Dan, J. Ji, and Y. Li, ''Climatic and biological simulations in a two-way coupled atmosphere–biosphere model (CABM),'' *Global Planet. Change*, vol. 47, nos. 2–4, pp. 153–169, Jul. 2005.

[30] Y. Li, Q. Zhou, J. Zhou, G. Zhang, C. Chen, and J. Wang, ''Assimilating remote sensing information into a coupled hydrology-crop growth model to estimate regional maize yield in arid regions,'' *Ecological Model.*, vol. 291, pp. 15–27, Nov. 2014.

[31] V. Phogat, M. A. Skewes, M. Mahadevan, and J. W. Cox, ''Evaluation of soil plant system response to pulsed drip irrigation of an almond tree under sustained stress conditions,'' *Agricult. Water Manage.*, vol. 118, pp. 1–11, Feb. 2013.

[32] K. R. Thorp, D. J. Hunsaker, and A. N. French, ''Assimilating leaf area index estimates from remote sensing into the simulations of a cropping systems model,'' *Trans. ASABE*, vol. 53, no. 1, pp. 251–262, Mar. 2010, doi: 10.13031/2013.29490.

[33] A. V. M. Ines and B. P. Mohanty, ''Near-surface soil moisture assimilation for quantifying effective soil hydraulic properties under different hydroclimatic conditions,'' *Vadose Zone J.*, vol. 7, no. 1, pp. 39–52, Feb. 2008, doi: 10.2136/vzj2007.0048.

[34] A. V. M. Ines and B. P. Mohanty, ''Parameter conditioning with a noisy Monte Carlo genetic algorithm for estimating effective soil hydraulic properties from space,'' *Water Resour. Res.*, vol. 44, no. 8, p. 06422, Aug. 2008, doi: 10.1029/2007wr006125.

[35] A. V. M. Ines and B. P. Mohanty, ''Near-surface soil moisture assimilation for quantifying effective soil hydraulic properties using genetic algorithms: 2. Using airborne remote sensing during SGP97 and SMEX02,'' *Water Resour. Res.*, vol. 45, no. 1, p. 01408, Jan. 2009, doi: 10.1029/2008wr007022.

[36] J. J. Ruiz, M. Pulido, and T. Miyoshi, ''Estimating model parameters with ensemble-based data assimilation: A review,'' *J. Meteorolog. Soc. Japan. Ser. II*, vol. 91, no. 2, pp. 79–99, 2013.

[37] B. Ait-El-Fquih and I. Hoteit, ''An efficient state–parameter filtering scheme combining ensemble Kalman and particle filters,'' *Monthly Weather Rev.*, vol. 146, no. 3, pp. 871–887, Mar. 2018.

[38] E. Karan, S. Asadi, R. Mohtar, and M. Baawain, ''Towards the optimization of sustainable food-energy-water systems: A stochastic approach,'' *J. Cleaner Prod.*, vol. 171, pp. 662–674, Jan. 2018, doi: 10.1016/j.jclepro.2017.10.051.

[39] S. Agatonovic-Kustrin and R. Beresford, ''Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research,'' *J. Pharmaceutical Biomed. Anal.*, vol. 22, no. 5, pp. 717–727, Jun. 2000, doi: 10.1016/s0731-7085(99)00272-1.

[40] H. Wang and D. Hu, ''Comparison of SVM and LS-SVM for regression,'' in *Proc. Int. Conf. Neural Netw. Brain*, vol. 1, Beijing, China, 2005, pp. 279–283.

[41] Y. Cai, W. Zheng, X. Zhang, L. Zhangzhong, and X. Xue, ''Research on soil moisture prediction model based on deep learning,'' *PLoS ONE*, vol. 14, no. 4, Apr. 2019, Art. no. e0214508.

**YANSHU YU** was born in Yichang, Hubei, China, in 2004. She is currently pursuing the degree with Hubei University of Technology, Hubei. In 2023, she finished an SCI paper on the topic, demonstrating her excellence in academic research. She is involved in several significant projects and holds the role of the Project Leader for a national-level research initiative on multimodal fusion, which focuses on integrating image and text data for various applications, including artificial intelligence and machine learning. She continues to actively contribute to the research community and plans to expand her work into new energy-related AI applications in the future. Her research focuses on large language models and text-related applications, with her work contributing to advancements in natural language processing and AI technologies.

**YOUXI LUO** is currently pursuing the Ph.D. degree in statistics from the Renmin University of China. He has been a Visiting Scholar with Emory University, USA, and Nanyang Technological University, Singapore. He completed his postdoctoral research with the Institute of Computing Technology, Chinese Academy of Sciences. His primary research interests include econometric modeling, high-dimensional data mining, MCMC algorithms, Bayesian analysis, and related fields.

**XINHANG WANG** was born in Songzi, Hubei, China, in 2004. He is currently pursuing the bachelor's degree in microelectronics science and engineering with Hubei University of Technology, China. He will pursue the Ph.D. degree in physics at the University of Chinese Academy of Sciences, in 2025. His research interests are related to process design in semiconductor industry.

**XINRAN WANG** was born in Zhijiang, Hubei, China, in 2002. She is currently pursuing the bachelor's degree in energy economics with Hubei University of Technology, Hubei. She will be recommended for a Master's degree in New Energy Science and Engineering at Huazhong University of Science and Technology in Hubei Province in 2024. Her research interests include carbon reduction effects, the current status and future trends of energy storage costs, and LCA related content.

**CHAOZHU HU** is currently pursuing the Ph.D. degree in mathematical statistics with the Huazhong University of Science and Technology, specializing in data mining, optimization algorithms, and related fields. She has led and participated in three projects funded by the National Science Foundation of China, the National Social Science Foundation, and the Humanities and Social Sciences Fund of the Ministry of Education.