International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 314-321

https://doi.org/10.70517/ijhsa46426

Research on real-time processing technology of agricultural sensor data under 5G network

Weidong Zhang[1,*] and Huadi Tan[2]
[1] Information Center, Changzhou College of Information Technology, Changzhou, Jiangsu, 213164, China
[2] School of Cyberspace Security, Changzhou College of Information Technology, Changzhou, Jiangsu, 213164, China
Corresponding authors: (e-mail: zhangweidong@czcit.edu.cn).

Abstract The increasing improvement and maturity of 5G network technology provides a new developable direction for real-time processing of agricultural production data. For the monitoring of sensor data in the agricultural production process, this paper proposes a data collection network (GCN) system consisting of a perception layer, a network layer, and an application layer as a method for collecting research data. After obtaining the research data, a GCN with a forest-like network structure is used to construct the GCN components for data exchange between the data collection system and the gateway as well as the optimization of large-scale data transmission between networks. The ARMA time series data model with high accuracy and short training time was selected to perform short-term time series data prediction in agricultural production scenarios by describing the dynamic characteristics of time series data. AMRA modeling is performed on the experimental data, and the calibration accuracy of the model is 88.81% on average for 10 runs, and it can be as high as 89.16% and as low as 88.27%.

Index Terms sensor data monitoring, GCN components, ARMA time-series data modeling, agricultural production

I. Introduction

The quality of traditional Chinese agricultural production is limited by the experience and quality of practitioners, which affects crop yield and quality to a certain extent, and causes different degrees of waste of agricultural resources [1], [2]. With the improvement of people's living standards and the growing concern for food safety, the traditional farming production model has been unable to effectively meet people's demand for high-quality agricultural products, and smart agriculture has gradually been pushed onto the stage of history [3]. Smart agriculture is a new production model formed by the rapid integration of a new generation of information technology such as Internet of Things, big data, cloud computing, artificial intelligence and other aspects of the agricultural production process [4]. This model through the sensor real-time collection of farmland crop growth environment, growth, water and fertilizer information, analyzed by the agricultural information processing system to give the decision-making reference for the supply of water, fertilizer, medicine, and other terminal equipment such as water, fertilizer, medicine, and other elements of the growth of crops to carry out scientific and reasonable regulation [5]-[8].

Sensor-based smart agriculture model can realize the purpose of water saving, fertilizer saving, pesticide reduction, effectively reduce agricultural production costs, and provide technical support for the green, low-carbon and high-quality development of agriculture [9], [10]. However, the current smart agriculture faces many bottlenecks and challenges. On the one hand, the data of smart agriculture are not well collected and utilized, which leads to difficulties in data services for some farmers, and also makes it difficult for farmers to obtain accurate information [11], [12]. On the other hand, digital technology programs for smart agriculture are relatively lagging behind, which has led to some unsatisfactory technology applications for smart agriculture [13], [14]. Therefore, how to improve the management and collection of agricultural sensor data is an obvious issue that must be considered.

For these aspects that need to be improved, the application of 5G technology can be regarded as an important technical support point. 5G technology, as a representative of a new generation of communication technology, has a number of significant advantages, such as ultra-high speed, low latency, large capacity, and so on, which are the technical characteristics needed for smart agriculture [15]. With the help of 5G new generation mobile communication technology to develop the data collection and monitoring system of agricultural sensor network, to realize the reliable and stable transmission of collected data, which provides a good reference for the application of new generation information technology in agricultural production [16]-[18].

This paper firstly briefly analyzes the architecture and overall workflow of agricultural data collection network. GCN is introduced as the data transmission medium, the purpose and working principle of GCN are elaborated in detail, and the construction process of GCN components is designed. Then the process of ARMA modeling and processing of agricultural time series data and the information criteria for model selection and evaluation are explained. Subsequently, based on the production data of K Modern Agriculture Company, we analyze the impact of data segment length on the anomaly detection performance of the GCN system, and measure the correlation between multiple indicator targets. Finally, the data are differentiated into smooth series, and ARMA prediction model is constructed to evaluate the accuracy performance of the prediction model.

II. Agricultural Sensor Data Acquisition and Real-Time Processing

II. A. Data acquisition network structure

The system adopts a three-layer network structure design scheme according to the function. The first layer is the sensing layer, which is mainly composed of sensors and LoRa wireless communication module, and completes the data collection of air temperature and humidity, light intensity, soil temperature and humidity, soil pH, soil nitrogen, phosphorus and potassium content. The collected data are sent to the LoRa gateway through the LoRa wireless communication module. The second layer is the network layer, and the LoRa gateway is the core device with the functions of multi-channel data reception and transmission, as well as format conversion and transmission rate matching. This layer completes the transmission of collected data to the cloud server in 5G communication packet format. The third layer is the application layer, and the smart agriculture data management platform accesses the cloud server to transfer the collected data to the database. It completes data analysis and visualization, provides historical data query, network node status query, and gives commands to the water, fertilizer and medicine supply terminal control system according to the analysis and query results to realize the automatic control of water supply, fertilizer and medicine application. The system network structure is shown in Figure 1.

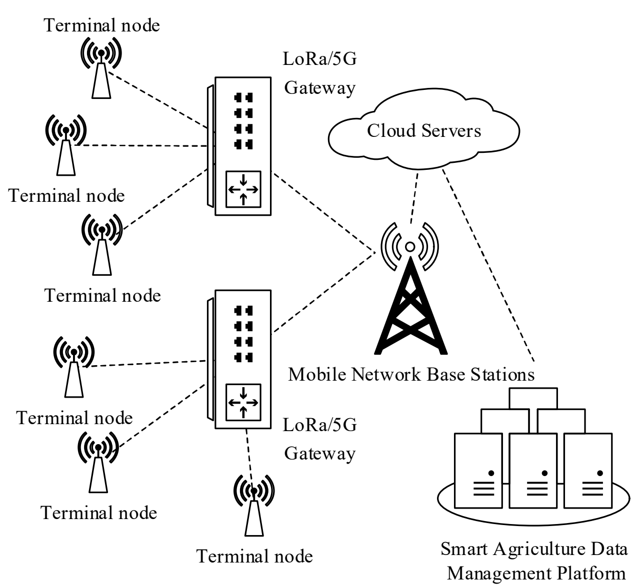


Figure 1: System network architecture

II. B. Construction of GCN

II. B. 1) Purpose and working principle of GCN

One of the purposes of constructing the GCN is to filter the TCP packets of the data at the network layer through the MSDAS gateway device in the GCN, and if it is found that the type of the data packet is the same as the type of the data packet generated by the MSDAS gateway device, the TCP packet of the data is intercepted and encapsulated together with its own data to form a new data packet to be sent to the MDSS, otherwise the TCP Otherwise, the TCP packet sequence of the data is forwarded normally, thus reducing the number of packets in the GCN, reducing the resource consumption of the GCN for packet forwarding, and reducing the pressure on the data access side of the MDSS. Another purpose of GCN is to eliminate the occurrence of HSG in the network, because GCN is a forest-like structure, it is inevitable that when the amount of data concurrency is large, it will cause some MSDAS gateway devices close to the root node to produce delays in high concurrency data forwarding, so GCN discovers the existence of HSG in time through the monitoring of the data traffic of MSDAS gateway devices, and changes the routing information of some of its sub-node devices, so that the path of data routing and forwarding changes, so as to eliminate HSG.

On each MSDAS gateway device runs a Linux system and MSDAS gateway program customized for the needs of the GCN, where the customized Linux is specifically programmed for the LinuxKernel network interface to achieve data filtering at the network layer according to the needs of the GCN. The MSDAS gateway program, in addition to being responsible for the acquisition, parsing and repackaging of sensor data, runs an Agent program component that monitors and manages the MSDAS gateway device and is responsible for the GCN, which is responsible for the initialization of the GCN, setting up the IPtable, sending the data to the MDSS, controlling it by the MDSS, and controlling the sensors in the reverse direction.

The MSDAS gateway program has root privileges on the Linux system, and the MDSS builds the network topology of the GCN as well as eliminates the HSG by sending control messages to the GCNAgent in the MSDAS gateway program, configuring the IPtable list and routing table of the MSDAS gateway device for routing and forwarding data.

II. B. 2) GCN construction process

GCN is mainly responsible for the data exchange between MSDAS gateway devices and MDSS, and is also responsible for optimizing the data transmission between networks, which is manifested in the existence of a proxy forwarding relationship between MSDAS gateways and MSDAS gateways, and part of the data of the child node MSDAS gateways can be screened and cached by the parent node MSDAS gateway, so that the parent node MSDAS gateway can repackage the child node's MSDAS gateway's useful data with its own data for re-encapsulation, and then pass it to the MDSS. In this process, if some MSDAS gateway nodes become HSGs, the MDSS reduces the pressure of HSG data forwarding by changing the routing information of some child node MSDAS gateway devices of the HSG.

II. C. Time series data model ARMA

II. C. 1) ARMA modeling process

The purpose of $ARMA(p,q)$ modeling of the time series data is to determine the parameters p and q in the model, and in the process of determining the parameters may be obtained more than one p and q combinations, which need to be optimally selected, so as to obtain a higher fit of the time series data model. The specific modeling process is shown in Figure 2, and each step in the modeling process is described below.

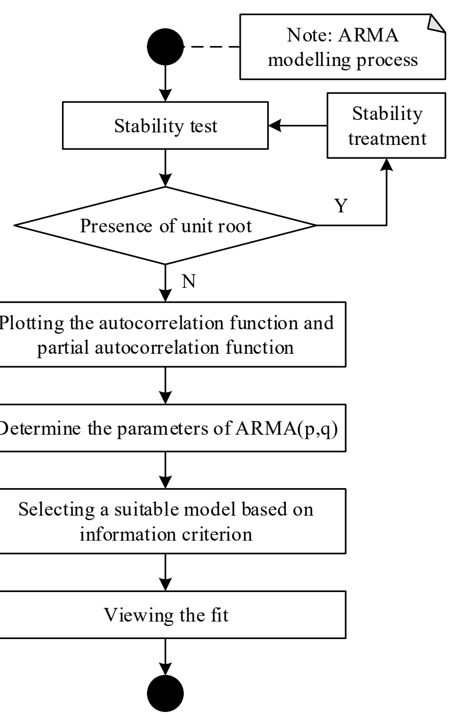


Figure 2: The ARMA modeling process

(1) Stability test of time series data

Before ARMA modeling needs to determine whether the time series data to meet the requirements of stability, the stability of the time series data refers to the statistical characteristics of the time series data, such as expectation, variance, etc. does not change with the change of time t, the stability of the time series data is a prerequisite for the modeling of the ARMA model, the stability of the process of testing is also known as the unit root test, the commonly used method is the ADF test.

In the ADF test, hypothesis testing by t-statistics, when the t-statistics of the concomitant probability of p value is less than the level of significance to reject the original hypothesis: the existence of a unit root, so as to determine the time-series data is stable, i.e., does not contain a unit root.

(2) Time series data stability processing

The general time series data can be composed of three basic parts: deterministic trends, seasonal variations and irregular fluctuations, expressed as equation (1):

$$y_t = T_t + S_t + I_t \tag{1}$$

where y_t is the time series data at moment t, T_t denotes the deterministic trend component at moment t, S_t denotes the seasonal variation component at moment t, and I_t denotes the irregular fluctuation at moment t. The smoothness treatment of time series data is to eliminate the deterministic trend and seasonal variation and retain the irregular fluctuations for modeling in ARMA model.

(3) Determination of ARMA model parameters

ARMA model only needs to determine the p, q two parameters, p represents the order of the autoregressive model, q represents the order of the moving average model, the determination of the two model parameters is determined by the autocorrelation coefficient of the time-series data and partial autocorrelation coefficient of the trailing and truncation of the judgment.

The autocorrelation coefficient is a function of the time interval s, which is used to reflect the correlation of the time series data with an interval of s, and its function is defined as equation (2):

$$\rho_s = \frac{cov(y_t, y_{t-s})}{var(y_t)} \tag{2}$$

Since the premise of ARMA modeling is that the time series data needs to remain stable, the autocorrelation coefficient is equal to the correlation coefficient $corr(y_t, y_{t-s})$ between y_t and y_{t-s}. The stability condition shows that $\mathrm{var}(y_t) = \mathrm{var}(y_{t-s})$, which leads to Eq. (3):

$$\rho_s = \frac{\mathrm{cov}(y_t, y_{t-s})}{\mathrm{var}(y_t)} = \frac{\mathrm{cov}(y_t, y_{t-s})}{\sqrt{\mathrm{var}(y_t)\cdot \mathrm{var}(y_{t-s})}} = corr(y_t, y_{t-s}) \tag{3}$$

So the meaning of autocorrelation coefficient can also be obtained from the meaning of correlation coefficient.

The partial autocorrelation coefficient is the conditional correlation between y_t and y_{t-s} given $y_{s-1}, y_{s-2}, \cdots, y_{t-s-1}$. The partial autocorrelation coefficient at s order lag is defined as equation (4):

$$\varphi_{ss} = \begin{cases} \rho_1 & s = 1 \\ \dfrac{\rho_s - \sum_{j=1}^{s-1} \varphi_{s-1,j}\rho_{s-j}}{1 - \sum_{j=1}^{s-1} \varphi_{s-1,j}\rho_j} & s > 2 \end{cases} \tag{4}$$

Selection and combination of p, q parameters by trailing and truncation of autocorrelation and partial autocorrelation coefficients.

II. C. 2) Selection of ARMA models

When there are more than one ARMA models identified by ACF and PACF, the obtained ARMA models are usually subjected to model evaluation in order to select the best model. Among the information criteria for model evaluation, AIC (Akaike Information Criterion) and SBIC (Schwarz Bayesian Information Criterion) are the most commonly used, which are defined as Eqs. (5)-(6):

$$AIC = T \ln SSR + 2n \tag{5}$$

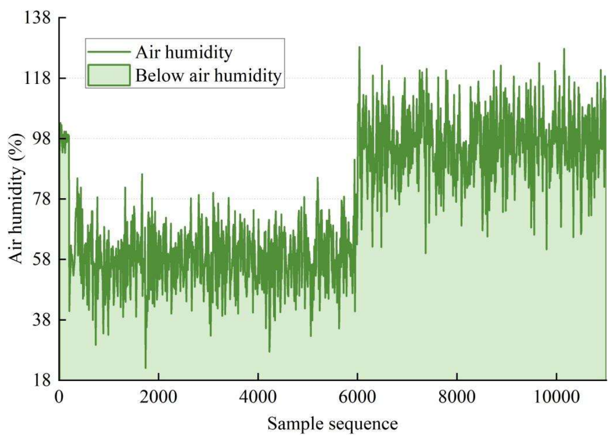
$$SBIC = T \ln SSR + n \ln T \tag{6}$$

where SSR is the sum of squared residuals, n is the number of explanatory variables, and T is the number of available observations. Based on the definition of the information criterion, the model with the smallest value of the information criterion is selected to complete the model selection.

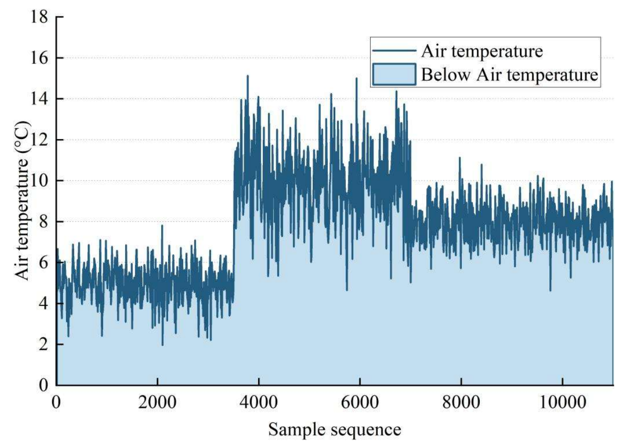
III. Performance analysis of agricultural data collection and processing methods

III. A. Experimental materials

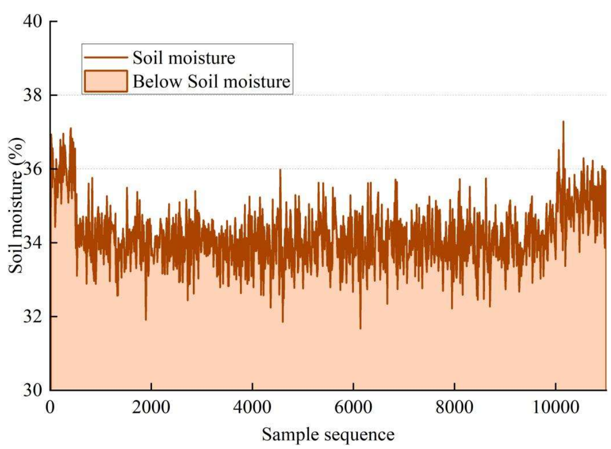
The data collection network system was used to obtain air humidity, air temperature, soil humidity and soil temperature data from the base of K Modern Agriculture Company, with the time coverage from October 1, 2020 to November 11, 2020, with a sampling interval of 5 min, and a total of 11,520 data were obtained. In order to facilitate the comparative analysis of detection performance, a set of anomalies with random failure rate (13%~38%) was introduced into the data, and the dataset is shown in Fig. 3. There are differences in the periodicity and trend characteristics of the four types of data. The air humidity data fluctuates greatly in the value domain, and the periodicity characteristics are not significant. The air temperature data is cyclical with less fluctuation. Soil humidity data has obvious periodicity characteristics, and the range of values is smaller. Soil temperature data has similar periodicity characteristics as air temperature data and has a smaller fluctuation in the value range.



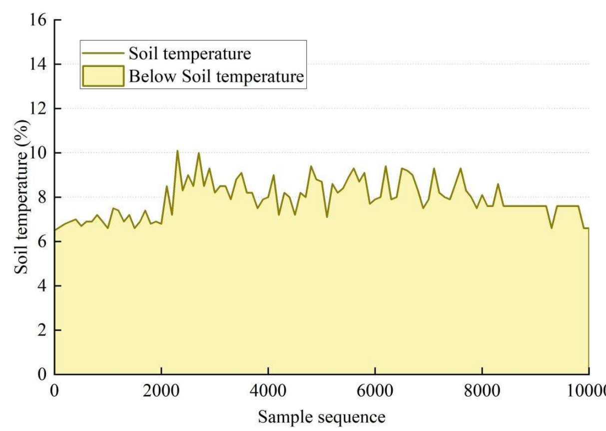
(a) Air humidity



(b) Air temperature



(c) Soil moisture



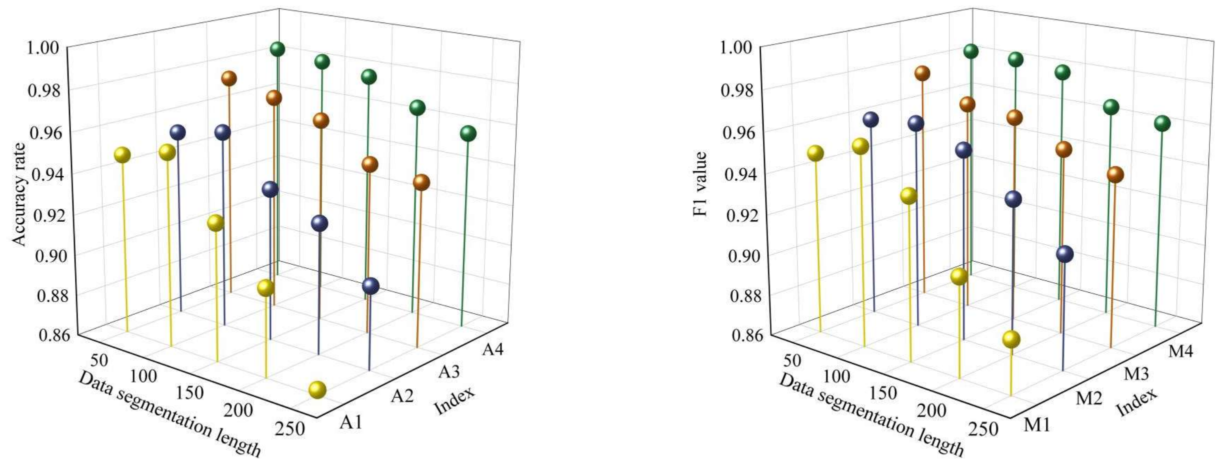
(d) Soil temperature

Figure 3: Experimental date sets

III. B. Performance Evaluation of the Data Acquisition Network (DAN) System

III. B. 1) Effect of data segment length on anomaly detection performance

In order to evaluate the influence of the length of the data segment on the abnormal data detection performance of the data acquisition network system, the air temperature and humidity and soil temperature and humidity datasets of 50-250 data segment sizes were used for experiments, and the abnormal data detection results of the data acquisition network system are shown in Figure 4, where "M1" is the air humidity, "M0" is the air temperature, "M3" is the soil moisture, and "M4" is the soil temperature.



(a) Accuracy rate (b) F1 value

Figure 4: Accuracy and F1score of anomaly detection of different data segment lengths

The data acquisition network system used in this paper achieves good anomaly detection accuracies on the four different segment length data, but there are differences in the trend of the anomaly detection accuracies with the change of the data segment lengths for the data with different characteristics. The anomaly detection accuracy and F1 value decrease more when the data segmentation length is greater than 100 in the air humidity and temperature datasets, and the best accuracy is obtained when the data segmentation length is around 100. The soil moisture dataset has the best detection performance, followed by the soil temperature dataset, and the average values of accuracy and F1 value for both are 0.9811, 0.9769 and 0.9542, 0.9583, respectively, and the detection performance is higher than that for the air temperature and humidity dataset due to less volatility than that of the air temperature and humidity dataset, and the optimal detection performance is obtained with a data segment length of 50, which is smaller than that for the air temperature and humidity dataset. This is due to the weaker volatility and only a small amount of data is needed to learn the distribution characteristics. The soil temperature and humidity data have good detection results in the interval of [50,150].

III. B. 2) Correlation analysis

Correlation analysis is to measure the correlation data of several variables with the target data, so as to delete some variables with small degree of correlation, to reduce the input features, to reduce the training time of the model, and to facilitate the prediction results. The selected attributes are (B1) wind speed, (B2) lighting, (B3) air temperature, (B4) air humidity, (B5) soil temperature (20cm), (B6) soil humidity (20cm), (B7) soil temperature (40cm), (B8) soil humidity (40cm), (B9) soil temperature (60cm), and (B10) soil humidity (60cm) Each of the The correlation coefficients between the attributes are shown in Fig. 5. The correlation coefficients are in [-1,1], with 1, -1, and 0 indicating positive, negative, and no correlation, respectively. Based on the correlation results in Figure 5, attributes with small correlation coefficients were excluded.

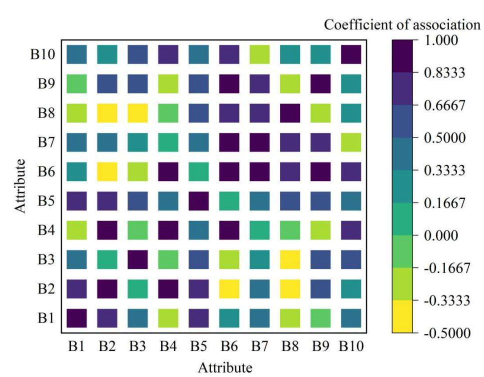
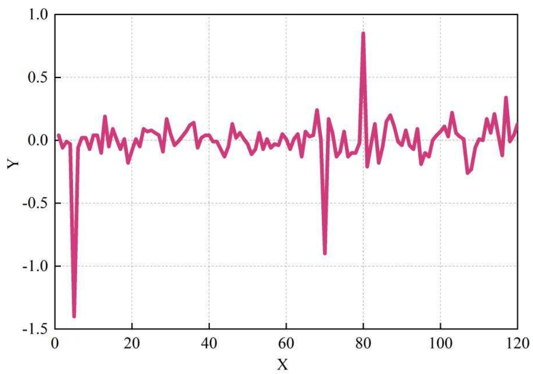


Figure 5: Correlation coefficients between attribute

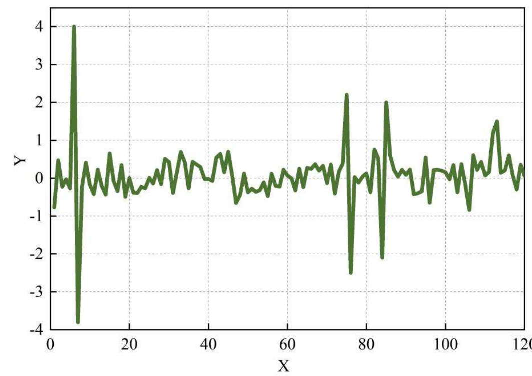
III. C. Simulation of ARMA model

III. C. 1) Model modeling

Before making model predictions, it is important to determine whether the data series is smooth or not. For the sample sequence of this data, it is judged to be non-smooth after visual observation. For accuracy, the unit root test (ADF) is used to verify the above conclusion. The test result shows that the unit root exists, then the sequence is non-smooth, and needs to be differentiated into a smooth sequence before the next step of prediction. Figure 6 shows the image of the sample sequence after differencing.



(a) First-order difference sequence



(b) Second-order difference sequence

Figure 6: The image after the differential of the sample sequence

Observe Figure 6, the data stream used in this experiment is stabilized by the second-order differencing of the sequence image, the ADF test value is 0, which can determine the differential order. After finding the differential order and smoothing the sequence, the autocorrelation function of the sequence is made to determine the value of the sum is shown in Fig. 7, which gives q=5, so the prediction model is obtained as ARIMA(5).

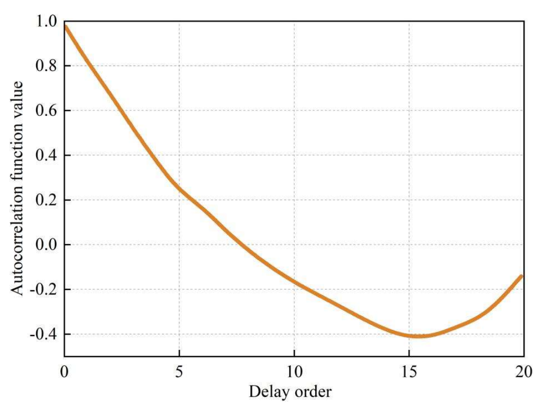


Figure 7: Autocorrelation function

III. C. 2) Predictive modeling accuracy

In order to prevent one experiment from having chance, the validation model is made to run 10 times, and the resulting correction accuracy is shown in Table 1.The evaluation indexes selected are: the correct rate of anomalous data detection (ACCabnormal), the accuracy of temporal data prediction (ACCpred), and the accuracy of anomalous data correction (ACCcor).As can be seen in Table 1, the first of the anomalous data correction methods of the constructed RRMA model 10 correction accuracy average is 88.81%, the highest is 89.16% and the lowest is 88.27%. For time series data prediction, the average value of the first 10 times accuracy is even higher than 96.65%, and the 10 times accuracy is between [0.9615,0.9749].

Table 1: accuracy of abnormal data correction method

Number of experiment	$ACC_{abnormal}$	ACC_{pred}	ACC_{cor}
1	0.8963	0.9715	0.8916

2	0.898	0.9632	0.8857
3	0.8879	0.9648	0.8771
4	0.8905	0.9749	0.8889
5	0.8702	0.9715	0.8656
6	0.9029	0.9548	0.8827
7	0.8852	0.9598	0.8699
8	0.9053	0.9615	0.8913
9	0.8963	0.9699	0.8901
Average value	0.8974	0.9665	0.8881

IV. Conclusion

In this paper, a data collection network system with a three-tier structure is proposed as a real-time monitoring method for agricultural production data. The GCN components are also designed to optimize data transmission and fast short-term model fitting using ARMA model for time series data.

The constructed data collection network system has a maximum accuracy of 0.9811 in detecting multiple indicator data of different lengths.The ARMA model constructed on the basis of the smooth series production data of K Modern Agricultural Company has a calibration accuracy of 88.81% on average and 89.16% on maximum, and the prediction accuracy of the time-series data is in the range of [0.9615,0.9749].

Funding

This work was supported by Key Laboratory of Changzhou College of Information Technology, Future Intelligent Network (TSTIRNet) (Grant No. KYPT202102Z).

References

[1] Azadi, H., Moghaddam, S. M., Burkart, S., Mahmoudi, H., Van Passel, S., Kurban, A., & Lopez-Carr, D. (2021). Rethinking resilient agriculture: From climate-smart agriculture to vulnerable-smart agriculture. Journal of Cleaner Production, 319, 128602.
[2] Quy, V. K., Hau, N. V., Anh, D. V., Quy, N. M., Ban, N. T., Lanza, S., ... & Muzirafuti, A. (2022). IoT-enabled smart agriculture: architecture, applications, and challenges. Applied Sciences, 12(7), 3396.
[3] Ray, P. P. (2017). Internet of things for smart agriculture: Technologies, practices and future direction. Journal of Ambient Intelligence and Smart Environments, 9(4), 395-420.
[4] Zhu, N., Liu, X., Liu, Z., Hu, K., Wang, Y., Tan, J., ... & Guo, Y. (2018). Deep learning for smart agriculture: Concepts, tools, applications, and opportunities. International Journal of Agricultural and Biological Engineering, 11(4), 32-44.
[5] Ullo, S. L., & Sinha, G. R. (2021). Advances in IoT and smart sensors for remote sensing and agriculture applications. Remote Sensing, 13(13), 2585.
[6] Garlando, U., Bar-On, L., Avni, A., Shacham-Diamand, Y., & Demarchi, D. (2020). Plants and environmental sensors for smart agriculture, an overview. 2020 IEEE SENSORS, 1-4.
[7] Soussi, A., Zero, E., Sacile, R., Trinchero, D., & Fossa, M. (2024). Smart sensors and smart data for precision agriculture: a review. Sensors, 24(8), 2647.
[8] Liu, W. (2023). Smart sensors, sensing mechanisms and platforms of sustainable smart agriculture realized through the big data analysis. Cluster Computing, 26(5), 2503-2517.
[9] Priya, P., & Kaur, G. (2021). Smart sensors for smart agriculture. In Artificial Intelligence and IoT-Based Technologies for Sustainable Farming and Smart Agriculture (pp. 175-191). IGI Global Scientific Publishing.
[10] Shaikh, F. K., Karim, S., Zeadally, S., & Nebhen, J. (2022). Recent trends in internet-of-things-enabled sensor technologies for smart agriculture. IEEE Internet of Things Journal, 9(23), 23583-23598.
[11] Li, C., & Niu, B. (2020). Design of smart agriculture based on big data and Internet of things. International Journal of Distributed Sensor Networks, 16(5), 1550147720917065.
[12] Amr, M. E., Al-Awamry, A. A., Elmenyawi, M. A., Adly, S., & Eldien, T. (2022). Design and Implementation of a Low-cost IoT Node for Data Processing, Case Study: Smart Agriculture. J. Commun., 17(2), 99-109.
[13] Saiz-Rubio, V., & Rovira-Más, F. (2020). From smart farming towards agriculture 5.0: A review on crop data management. Agronomy, 10(2), 207.
[14] Roukh, A., Fote, F. N., Mahmoudi, S. A., & Mahmoudi, S. (2020). Big data processing architecture for smart farming. Procedia Computer Science, 177, 78-85.
[15] Liu, J., Shu, L., Lu, X., & Liu, Y. (2023). Survey of intelligent agricultural IoT based on 5G. Electronics, 12(10), 2336.
[16] Damsgaard, S. B., Marcano, N. J. H., Nørremark, M., Jacobsen, R. H., Rodriguez, I., & Mogensen, P. (2022). Wireless communications for internet of farming: An early 5G measurement study. IEEE Access, 10, 105263-105277.
[17] Tang, Y., & He, Y. (2023). Application of 5G Communication Technology in Precision Agriculture. In Encyclopedia of Digital Agricultural Technologies (pp. 70-75). Cham: Springer International Publishing.
[18] Murugamani, C., Shitharth, S., Hemalatha, S., Kshirsagar, P. R., Riyazuddin, K., Naveed, Q. N., ... & Batu, A. (2022). Machine Learning Technique for Precision Agriculture Applications in 5G‐Based Internet of Things. Wireless Communications and Mobile Computing, 2022(1), 6534238.