

Postprint of a Deep Learning-Based Forecasting Model for Supply and Demand of Multiple Agricultural Products

Authors: Zhuang Jiayu, Xu Shiwei, Li Yang, Xiong Lu, Liu Kebao, Zhong Zhiping

Date: 2023-02-17T00:00:00+00:00

Abstract

To further improve the simulation and estimation accuracy of agricultural product supply and demand processes, this study employs large-scale national and provincial agricultural data since 1980 as samples, fully considering the influences of factors such as agricultural product varieties, time, income, and economic development, to construct multiple agricultural product supply and demand forecasting models based on deep learning Long Short-Term Memory Neural Network (LSTM). Under the premise of fully considering mechanistic constraints, the model leverages the advantages of deep learning algorithms in nonlinear model analysis and forecasting to analyze and forecast the supply and demand of nine major agricultural products including rice, wheat, corn, soybeans, pork, poultry, beef, mutton, and aquatic products. The production forecasting results for 2019-2021 based on this model were compared with data published by the National Bureau of Statistics for validation, achieving a three-year average forecasting accuracy of 96.98%, which indicates that the forecasting model constructed in this study can efficiently reflect the impact of changes in implicit indicators on forecasting results. This model can provide intelligent technical support for multi-regional, cross-period agricultural outlook work through timely monitoring of agricultural operation data.

Full Text

Preamble

Supply and Demand Forecasting Model of Multi-Agricultural Products Based on Deep Learning

ZHUANG Jiayu^{1,2}, XU Shiwei^{1,2*}, LI Yang³, XIONG Lu^{1,2}, LIU Kebao³, ZHONG Zhiping^{1,2}

¹Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China

²Key Laboratory of Agricultural Information Service Technology, Ministry of Agriculture and Rural Affairs, Beijing 100081, China

³Institute of Agricultural Remote Sensing and Information, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China

Abstract: To further improve the simulation and estimation accuracy of agricultural product supply and demand processes, this study utilized extensive national and provincial agricultural data since 1980 as research samples. By fully considering factors such as product variety, time, income, and economic development, we constructed a multi-agricultural product supply and demand forecasting model based on Long Short-Term Memory (LSTM) neural networks. Under the premise of incorporating mechanistic constraints, the model leverages deep learning algorithms' advantages in nonlinear analysis and prediction to forecast supply and demand for nine major agricultural products: rice, wheat, corn, soybean, pork, poultry, beef, mutton, and aquatic products. Comparing the 2019-2021 production forecasts from this model with data published by the National Bureau of Statistics, the three-year average prediction accuracy reached 96.98%, demonstrating that the model effectively captures the impact of latent indicator changes on forecast results. By enabling timely monitoring of agricultural operational data, this model provides intelligent technical support for multi-regional, cross-temporal agricultural outlook work.

Keywords: deep learning; supply and demand forecasting model; LSTM; RNN; agricultural production; agricultural outlook

1 Introduction

“Food is the paramount necessity of the people,” and food security concerns national welfare and people' s livelihood. Monitoring and early warning represent advanced tools for modern agricultural management and crucial means for ensuring national food security. Agricultural outlook, as an important component of monitoring and warning systems, enhances government initiative in regulating agricultural product markets while improving market participants' ability to respond to changes [?]. However, the complexity of agricultural systems—with their numerous product varieties, diverse industrial influencing factors, and strong inter-variety correlations—poses significant challenges for real-time, precise quantitative analysis and early warning [?].

Agricultural product supply and demand analysis and forecasting help release market signals and guide production and consumption, attracting widespread attention from researchers worldwide. Studies generally fall into two categories: production forecasting and demand forecasting. Production forecasting research typically employs meteorological [?, ?], remote sensing [?, ?], and crop growth mechanism models [?, ?] combined with linear regression, time series [?], and ar-

tificial intelligence methods [?]. Demand forecasting research primarily utilizes per capita nutrition calculation methods [?, ?], trend and empirical estimation methods [?, ?], and structural model prediction methods [?, ?]. Agricultural systems exhibit strong nonlinearity, randomness, and dynamic characteristics [?], making effective model application dependent on rapid and accurate parameter estimation.

Many researchers have applied intelligent optimization algorithms to crop model parameter estimation, such as using individual-advantage genetic algorithms to optimize rice phenology model parameters, achieving automatic calibration and significantly improving tuning efficiency [?, ?]. Wang et al. [?] demonstrated that improved deep learning algorithms can effectively model nonlinear systems with excellent prediction results. Deléglis et al. [?] found deep learning suitable for processing multi-source heterogeneous data related to grain supply and demand with favorable prediction outcomes. Further studies by Emerson et al. [?], Schwalbert et al. [?], and Shahhosseini et al. [?] showed that intelligent learning methods like Artificial Neural Networks (ANN) outperform traditional time series and multiple linear regression models for crop yield prediction.

Intelligent model-based comprehensive analysis has become an effective tool for agricultural product supply and demand forecasting. Traditional statistical methods cannot fully quantitatively describe the entire supply and demand process. It is necessary to leverage deep models' advantages in temporal and spatial data continuity to further improve simulation and estimation accuracy. This study integrates multidisciplinary knowledge from agronomy, meteorology, economics, and computer science, building upon extensive data accumulated during the development of the *China Agricultural Outlook Report*. By fully considering factors such as product variety, time, income, and economic development, and utilizing LSTM's strengths in perceiving nonlinear features of multidimensional variables, we optimize traditional forecasting models and construct multi-temporal dimension models for different agricultural products. This approach enables intelligent allocation of model clusters across products and time scales, improving simulation accuracy while enabling real-time, precise monitoring, analysis, and early warning of supply and demand information, thereby providing effective technical support for multi-regional, cross-period agricultural outlook work.

2 Data and Methods

2.1 Data Sources and Preprocessing

2.1.1 Data Sources The agricultural product supply and demand forecasting model utilizes national and provincial data since 1980, covering indicators including: production, crop planting area, livestock product slaughter volume, food consumption, industrial consumption, feed consumption, seed consumption, loss and other consumption, imports/exports, prices, GDP, population,

urban population, rural population, and meteorological elements. Data sources include officially published statistical yearbooks and reports (*China Statistical Yearbook*, *National Agricultural Product Cost-Benefit Compilation*, Customs import-export database, etc.). Some data were estimated through surveys and backcasting methods, such as feed consumption, industrial consumption, seed consumption, and other consumption categories.

Specifically, crop production influencing factors include: rainfall, sunshine hours, temperature, cost-benefit conditions, expenses and labor inputs, fertilizer and seed inputs, support policies, protection policies, and technology policies. Live-stock production factors include: slaughter volume, inventory, age structure, cost-benefit conditions, and management factors. Consumption-related factors include: food consumption, industrial consumption, feed consumption, seed consumption, loss, per capita income, market prices, per capita GDP, industrial growth rate, livestock product output, feed-to-meat ratio, feed prices, input and output prices, seed usage per mu (for crops), and loss coefficients.

2.1.2 Data Preprocessing Bengio et al. [?] identified gradient vanishing (or explosion) as the fundamental factor making Recurrent Neural Network (RNN) training extremely difficult, necessitating data preprocessing before model calculation. Given the diverse sources and large volume of data involved in this study's forecasting model, we performed data standardization to eliminate dimensional effects across different indicators and ensure comparability.

Incomplete datasets were processed to ensure reliable analysis results using mean imputation, hot deck filling, k-nearest neighbors, and regression analysis methods. Data splitting and merging employed proportional division and trend analysis methods. Outlier detection and treatment included the Pauta criterion (3δ), Chauvenet's criterion, Dixon's criterion, Romanovski's criterion (t-test), and Grubbs' test. Preprocessing encompassed quality analysis, anomaly detection, filling, transformation, standardization, and normalization of agricultural supply and demand data. Transformation and normalization were autonomously selected within the model through parameters, while other processing methods were implemented during data import. Data transformation methods included:

1. **Difference calculation:** Subtracting previous time point data from current values for prediction using differenced values.
2. **Logarithmic transformation:** Calculating skewness for non-normally distributed data and applying logarithms to approximate normal distribution.

To accelerate LSTM model convergence, all feature data were normalized using min-max normalization, with predictions converted back to actual values using recorded maxima/minima.

2.2 Research Methods

2.2.1 LSTM Algorithm The deep neural network structure designed in this study features key components including InputGate, OutputGate, and Forget-Gate, implemented through Sigmoid functions with outputs between 0-1 that can be pointwise-multiplied with node $h(\cdot)$ to achieve “gating” functionality.

The update methods for input, output, and forget gates are shown in formulas (1)-(3):

$$in_t = \sigma(W_{in}[h_{t-1}, x_t] + b_{in}) \quad (1)$$

$$out_t = \sigma(W_{out}[h_{t-1}, x_t] + b_{out}) \quad (2)$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (3)$$

where in_t , out_t , and f_t represent input, output, and forget gates; σ is the activation function; C_t is the current LSTM layer output; C_{t-1} is the previous layer output; x_t is the current input; h_{t-1} is the input from the previous layer; and h_t is the output to the next layer.

The forward hidden state at time step t is calculated through formula (4):

$$h_t^f = \sigma(W_f[h_{t-1}^f, x_t] + b_f)$$

where h_t^f is the forward hidden state and b_f is the bias.

The backward hidden state is calculated using formula (5):

$$h_t^b = \sigma(W_b[h_{t-1}^b, x_t] + b_b)$$

Finally, the output at time step t is computed through formula (6):

$$out_t = [h_t^f, h_t^b]W_{h,out} + b_{out}$$

2.2.2 Model Establishment The agricultural product supply and demand forecasting model primarily considers China’s agricultural supply and demand characteristics, based on equilibrium principle assumptions: supply and demand jointly determine price, with total supply equaling domestic production plus imports plus beginning stocks, and total demand equaling domestic consumption plus exports plus ending stocks. For detailed consumption analysis, domestic consumption is subdivided by purpose. Grain consumption is categorized into food consumption, industrial consumption, feed consumption, seed consumption, loss, and other consumption; soybean consumption is divided into food consumption, crushing consumption, seed consumption, loss, and other consumption. Different consumption types are influenced by different factors:

food consumption is affected by population and urbanization levels; feed consumption primarily depends on livestock production; industrial consumption is influenced by industrial product output; seed consumption depends on seed usage per mu and planting area; and loss depends on production volume and loss coefficients.

The basic constraint equations for the agricultural product supply-demand balance model are shown in formulas (7)-(12):

$$\text{Total Supply} = \text{Total Demand} \quad (4)$$

$$\text{Total Supply} = \text{Production} + \text{Imports} + \text{Beginning Stocks} \quad (5)$$

$$\text{Total Demand} = \text{Consumption} + \text{Exports} + \text{Ending Stocks} \quad (6)$$

$$\text{Consumption} = \text{Food Consumption} + \text{Industrial Consumption} + \text{Seed Consumption} + \text{Loss} \quad (7)$$

$$\text{Beginning Stocks}_{\text{current period}} = \text{Ending Stocks}_{\text{previous period}} \quad (8)$$

$$\text{Production} = \text{Yield per unit area} \times \text{Planting area} \quad (9)$$

where bracketed terms “[]” indicate optional components.

The model covers 18 agricultural products: rice, wheat, corn, soybean, oil crops, cotton, sugar crops, potatoes, vegetables, fruits, pork, poultry, beef, mutton, eggs, dairy products, aquatic products, and industrial feed. In multi-variety operations, feed consumption for grain crops must be based on livestock production, so meat varieties (pork, beef, mutton, poultry) are prioritized in calculation groups. The model first computes joint balance sheet equations for meat products, then uses predicted meat production (or inventory, slaughter volume if predictions are unavailable) as input data for joint balance sheet equations of major grain crops (rice, wheat, corn, soybean) to calculate interrelated products with planting competition and consumption substitution.

The overall approach encapsulates the deep neural network training model as an I/O-open modular model with external data input control interfaces, enabling supply-demand indicator forecasting and balance sheet conversion. Key functions include automatic model training encapsulation, automatic parameter tuning, model interface parameter input, automatic balance sheet conversion, automatic balancing, and final output. This scheme enables automatic balancing of various crop, meat, and egg product balance sheets with inter-variety coupling, dynamically calculating conditions for the next 10 years. Users can select target products, historical training periods, and output future balance sheets (medium- to long-term forecasts) to configure prediction scenarios.

2.2.3 Model Data Interface The model input consists of n groups of historical data with m features, outputting k groups of forecast data with m features.

Generally, more output data requirements correspond to lower prediction accuracy. Therefore, the LSTM model's input-output data satisfies the format: $(X, Y) = [\text{samples, time steps, features}]$, where X represents vectors of various influencing factors requiring specific feature engineering based on their meanings, and Y represents vectors formed by balance sheet items. All features require standardization and normalization before network input, with boundary restrictions controlled through parameters for Y components. For multi-input, multi-step, multi-output LSTM data, this study employs Encoder-Decoder LSTM, which demonstrates good performance for sequence-to-sequence problems.

2.2.4 Model Evaluation Methods Model evaluation primarily uses Mean Absolute Deviation (MAE) and Mean Absolute Percentage Error (MAPE) to assess performance. During model cluster validation, historical data undergo sliding-window processing, using the most recent year's data as the validation set and earlier data for training. For example, to validate 2015 predictions, 1980-2014 data serve as the training set and 2015 actual data as the validation set; similarly for 2016, 1980-2015 data train the model validated against 2016 data, enabling sliding validation.

3 Results and Analysis

3.1 Single Variety Modeling and Prediction

Using the relatively decoupled egg supply-demand forecast (2021-2030) as an example, the prediction employs a dual-input model dividing sample data into historical egg balance sheet data and various influencing factors. Convolutional layers process influencing factor data for feature extraction. Fully connected layers receive balance sheet component data, combining them with extracted features to form new features. The model network employs a serial structure to enhance expressive power, using modular groups to repeatedly extract different feature quantities. Each module's basic structure consists of: a 3×3 convolution outputting $n/4$ channels, followed by a 2×2 convolution outputting $n/4$ channels for two feature transformations, and finally a 1×1 convolution to improve network expressive power while combining cross-channel information, outputting n channels (where n is the number of filters). Pooling layers synthesize features within initial regions and effectively reduce overfitting. Dropout layers randomly deactivate hidden neurons to prevent overfitting and improve generalization. Weighted cross-entropy serves as the loss function due to uneven sample distribution, with fully connected layers incorporating temporal information features.

The loss curve during training is shown in Figure 2 [Figure 2: see original paper], demonstrating decreasing loss during iterations. Forecasted egg supply-demand results for the next decade are presented in Table 1, with 2011-2020 showing historical balance sheets and 2021-2030 showing model predictions.

3.2 Multi-variety Cluster Modeling and Prediction

Similar to the egg forecasting method, the LSTM-optimized agricultural product supply-demand forecasting model enables multi-variety cluster modeling. For production forecasting, meteorological data, input factors, and management policies serve as primary influences. For consumption forecasting, differentiated prediction models incorporate population, product prices, and regional consumption differences to provide quantitative analysis for various consumption types. Price forecasting introduces multi-variety price correlation and competition mechanisms, establishing multi-dimensional agricultural product price prediction methods.

Applying this forecasting model to China's major agricultural products (rice, wheat, corn, soybean, pork, poultry, beef, mutton, and aquatic products) and comparing 2019–2021 production forecasts with National Bureau of Statistics data yields results shown in Figure 3 [Figure 3: see original paper]. The three-year overall MAPE is 3.02%, corresponding to an average prediction accuracy of 96.98% (96.10% in 2019, 98.26% in 2020, and 96.58% in 2021). The improved accuracy in 2020 and 2021 indicates that intelligent learning models perform better with increased sample sizes.

4 Summary and Outlook

Traditional statistical analysis methods cannot fully quantitatively describe the entire production, distribution, and consumption process of agricultural products. To leverage big data advantages of temporal and spatial continuity and improve multi-variety supply-demand simulation accuracy, this study proposes a deep learning-based LSTM agricultural product supply-demand forecasting model. This model enables multi-factor, high-frequency data collaborative analysis, effectively addressing “visual blind spots” caused by single information sources, representing an exploratory and pioneering approach to holographic and precise agricultural product supply-demand forecasting and early warning.

Based on parameter sensitivity analysis and calibration, and using China's agricultural product supply-demand balance sheet data since 1980 as foundational learning samples, this study constructed an LSTM and equilibrium-correlation-based forecasting model capable of quantitative simulation and estimation of supply and demand for major products including rice, wheat, corn, soybean, and eggs. By applying equilibrium theory to constrain supply-demand components while comprehensively considering historical characteristics across the agricultural product industry chain and influencing factors, the model avoids uncontrollable errors from manual intervention and optimizes parameters through recurrent neural networks to improve prediction performance.

As the core of China's agricultural outlook work, such models enhance technical support capabilities for “data-driven decision-making, management, and inno-

vation,” providing market warnings, policy evaluation, resource management, and public opinion analysis for agricultural production and macroeconomic regulation. However, the current model has limitations in responding to special events such as African swine fever and China-US trade friction, as major policies and emergencies represent occasional interference variables for which large sample features are difficult to obtain for deep learning-based intelligent modeling. Future agricultural product supply-demand forecasting models require 协同应用 with more models to better conduct quantitative analysis and prediction of supply-demand situations.

References

- [1] Ministry of Agriculture and Rural Affairs Market Early Warning Expert Committee. *China Agricultural Outlook Report (2021-2030)*[M]. Beijing: China Agricultural Science and Technology Press, 2021.
- [2] XU S, DI J, LI G, et al. The methodology and application of agricultural monitoring and early warning model cluster[J]. *Scientia Agricultura Sinica*, 2020, 53(14): 2859-2871.
- [3] ZHUANG J, XU S, LI G, et al. The influence of meteorological factors on wheat and rice yields in China[J]. *Crop Science*, 2018, 58(3): 837-852.
- [4] LAUDIEN R, SCHAUBERGER B, MAKOWSKI D, et al. Robustly forecasting maize yields in Tanzania based on climatic predictors[J]. *Scientific Reports*, 2020, 10: ID 19650.
- [5] JOHNSON M D, HSIEH W W, CANNON A J, et al. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods[J]. *Agricultural and Forest Meteorology*, 2016, 218-219: 74-84.
- [6] EUGENIO F C E, GROHS M, VENANCIO L P, et al. Estimation of soybean yield from machine learning techniques and multispectral RPAS imagery[J]. *Remote Sensing Applications*, 2020: ID 100397.
- [7] PAUDEL D, BOOGAARD H, DE WIT A, et al. Machine learning for large-scale crop yield forecasting[J]. *Agricultural Systems*, 2020, 187: ID 103016.
- [8] DE WIT A, BOOGAARD H, FUMAGALLI D, et al. 25 years of the WOFOST cropping systems model[J]. *Agricultural Systems*, 2019, 168: 154-167.
- [9] MASUDA T, GOLDSMITH P. World soybean production: Area harvested, yield, and long-term projections[J]. *International Food & Agribusiness Management Review*, 2009, 12(4): 143-161.
- [10] WANG G, HU H, CHEN J, et al. Application of Fourier model based on BP filter in crops yield prediction[J]. *Chinese Journal of Agrometeorology*, 2015, 36(4): 472-478.

- [11] XIAO Y, CHENG S, XIE G, et al. The balance between supply and consumption of the main types of grain in China[J]. *Journal of Natural Resources*, 2017, 32(6): 927-936.
- [12] XIE G, CHENG S, XIAO Y, et al. The balance between grain supply and demand and the reconstruction of China's food security strategy in the new period[J]. *Journal of Natural Resources*, 2017, 32(6): 895-903.
- [13] ZHAO X, SHAO Y. Analysis and forecast of China's grain supply and demand[J]. *Research of Agricultural Modernization*, 2014, 35(3): 277-280.
- [14] LIU Y, LUO Q, ZHOU Z, et al. Analysis and prediction of the supply and demand of China's major agricultural products[J]. *Strategic Study of CAE*, 2018, 20(5): 120-127.
- [15] LU W, NING L C, WEN X Q. Modeling the effects of urbanization on grain production and consumption in China[J]. *Journal of Integrative Agriculture*, 2017, 16(6): 1393-1405.
- [16] HUANG J. Recognition of recent and mid-long term food security in China[J]. *Issues in Agricultural Economy*, 2021(1): 19-26.
- [17] CHEN X, YANG C. Characteristic of agricultural complex giant system and national grain output prediction[J]. *Systems Engineering-Theory & Practice*, 2002(6): 108-112.
- [18] XU S. *Agricultural Information Analysis*[M]. Beijing: Higher Education Press, 2013.
- [19] GAO L. *Agricultural Modeling*[M]. Beijing: China Meteorological Press, 2019.
- [20] WANG Y, HAN H, GUO M. A nonlinear modeling method based on improved deep learning[J]. *Information and Control*, 2018, 47(6): 680-686.
- [21] DELÉGLISE H, INTERDONATO R, BÉGUÉ A, et al. Food security prediction from heterogeneous data combining machine and deep learning methods[J]. *Expert Systems with Applications*, 2022, 190: ID 116189.
- [22] EMERSON R A, DOS REIS J G M, VENDRAMETTO O, et al. Time series prediction with artificial neural networks: An analysis using Brazilian soybean production[J]. *Agriculture (Basel)*, 2020, 10(10): ID 475.
- [23] SCHWALBERT R A, AMADO T, CORASSA G, et al. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil[J]. *Agricultural and Forest Meteorology*, 2020, 284: ID 107886.
- [24] SHAHHOSSEINI M, HU G, HUBER I, et al. Coupling machine learning and crop modeling improves crop yield prediction in the US corn belt[J]. *Scientific Reports*, 2021, 11(1): 1-15.

[25] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. *IEEE Transactions on Neural Networks*, 1994, 5(2): 157-166.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.