INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print), 2222-5234 (Online)
http://www.innspub.net
Vol. 22, No. 5, p. 25-34, 2023

RESEARCH PAPER OPEN ACCESS

Forecasting Volume of Corn Production through Neural Network Model: A Post-harvest Monitoring Tool

Ronald Lachica Aquino

Mathematics and Allied Disciplines Department, College of Arts and Sciences, Don Mariano Marcos Memorial State University (DMMMSU), South La Union Campus, La Union, Philippines

Key words: Maize, Neural network, Forecasting, Post-harvest, Monitoring, Time series

http://dx.doi.org/10.12692/ijb/22.5.25-34 Article published on May 04, 2023

Abstract

The paper deals with the forecasting of volume (in tonnes) of corn production relative to the harvested farmed area during the second semester of agricultural cropping. Time series data used were obtained from the open stat database published by the Philippine Statistics Authority from the second semester of 1987 to second semester of 2022. Artificial Neural Network (ANN) models were developed, trained and validated to forecast the volume of corn production. Statistical errors such as Root Mean Square Error (RMSE) were computed and compared to identify the most suitable model to forecast the corresponding volume. ANN (1:3:1) model was identified and used to forecast the volume of corn production. The two sets of data, namely actual and forecast volumes of production, were found to have statistically no significant difference, which implies that the model gives forecast values that are relatively close to the actual ones.

* **Corresponding Author:** Ronald L. Aquino ✉ raquino@dmmmsu.edu.ph

Introduction

The foundation of the global food supply system is agriculture, and the lack of available land in developing nations affects food security (Prosekov and Ivanova, 2018). The availability of food in a community is one of the concepts of food security. This is closely related and linked to efficient food import and production of food (Vedantu, 2022). Agricultural production and the price of key agricultural commodities, such as corn, rice, soybeans, and wheat, have always been volatile. As the world's population surges toward 8 billion, that volatility is more acute than ever (Gro-Intelligence, 2018).

The Philippines is still seen as a country that focuses heavily on agriculture. The majority of Filipinos still reside in rural areas, where agriculture is their primary source of income. Fisheries, farming, forestry, and livestock make up the four subsectors that make up the agricultural sector in the Philippines. The principal agricultural products of the Philippines are palay (rice), corn (maize), sugarcane, coconut, bananas, pineapple, mangoes, coffee, abaca (a plant that resembles a banana), and tobacco. Peanuts, cassava, camote (a sort of root crop), garlic, cabbage, onion, eggplant, rubber, calamansi (a variety of lemons), and cotton are examples of secondary crops (Agri Farming, 2021).

The Php 110 billion swine and poultry businesses depend primarily on locally farmed corn. The strong demand from the feed business has helped domestic prices of white maize and yellow flour rise steadily since early 2021 and reached record highs in September 2022 (FAO, 2022).

Due to favorable weather, the major maize crop for 2022 was harvested in September, and the yield is officially estimated at 3.9 million tonnes, which is slightly above average. The preliminary estimate for 2022 maize production is 8.3 million tonnes, which is 4% more than the five-year average. This figure includes the secondary crop for harvest in 2022 (FAO, 2022).

Aside from implementing new technologies and meeting increasing demand, producing high-quality food comes with its own set of difficulties. Farming is not without its challenges, though, even after the crops have been gathered. Providing consumers with high-quality food depends equally on what happens after harvest, which could be the difference between success and failure. The FAO estimates that 10% to 20% of the world's grain production is lost due to post-harvest losses, according to their research. Technologies for post-harvest management can improve food security primarily by lowering losses and waste. They can enhance the amount of food available to farmers and customers in rural and urban areas by reducing post-harvest handling loss. This method of reducing losses can cut food costs and increase access to food (GrainPro, 2023). Food security can be improved in the agricultural sector by reducing post-harvest losses, increasing the amount of food available for consumption by farmers and low-income rural and urban consumers, and reducing waste. Hence, increasing farm-level productivity and lowering post-harvest losses clearly complement each other in order to increase food security. In order to fight hunger, it is crucial to address the issue of food losses, which leads to increased income, and enhanced food security in the world's poorest nations (SourceTrace, 2019). Most of these post-harvest management systems or technologies focus on how to store the harvest to last longer and the safety of the agricultural product while in transit to the market. Note that the fight to ensure global food security also calls for more effective and efficient forecasting and monitoring. Yet the benchmark for the production and price of agricultural commodities has remained largely unchanged for nearly 40 years (Gro-Intelligence, 2018).

Forecasting agricultural systems has been done using a variety of methods. A model is applied to describe a direct mapping between inputs and outputs without detailed consideration of the internal structure of a biological process. This is called the" black box" approach. This approach can be applied to many practical situations as it lessens the effort and time required to develop, validate, and implement a model based on biological-fishery variables (Gutierrez-Estrada *et al.*, 2007). Based models that use the information on biometrical crop characteristics, weather factors, and agro meteorological data are one way to predict the future. Statistical forecast models like basic regression models, Multiple Linear Regression (MLR) models that use plant features, MLR based on weather indices, and Logistic Regression models for qualitative variables are some methods. For predicting area and crop production, time series models like Exponential smoothing models and Auto-Regressive Integrated Moving Average (ARIMA) models were also used. Others forecast agricultural production using probabilistic models like Markov chain models (Ramasubramanian, 2006).

"Black box models" like the Autoregressive Integrated Moving Average (ARIMA) model and the Artificial Neural Network (ANN) model are used in time series forecasting. ARIMA models are superior for data with seasonal trends because they efficiently capture linear patterns in a time series of data. However, it requires continuity from previous data, while ANN models do a good job of capturing a time series' non-linear trends. It also catches noise and extreme values and does not require previous data continuity. It is a class of statistical learning algorithms that are modeled after the biological neural networks of the human brain. An artificial neural network (ANN) is typically described as a system made up of a limited number of artificial neurons, also known as processing elements (PE), that are linked to coefficients, sometimes referred to as weights, that make up the neural architecture and are arranged in layers. The capability of brain calculations is provided by the interconnection of these neurons forming a network (Agatonovic-Kustrin and Beresford, 2000). An ANN is a non-linear mathematical model that can simulate intricate non-linear processes that link a system's inputs and outputs. It is said that a random input pattern can be translated into a random output pattern by neural networks (Kriegeskorte and Golan, 2019).

The purpose of this paper was to develop a model that could forecast volumes of corn production in metric tonnes using time series data. The mathematical model that was employed in this study is the artificial neural network model. The findings of this study are significant for the community, considering that mathematics plays an important role in social economics. The forecast values of the model could be used to monitor the volume of corn production relative to the harvested farmed area as a strategy for post-harvest management which will be utilized for decision and policy making (see Fig. 1).

Materials and methods

Several researchers have reported attempting to use the ANN method to predict agricultural crops such as soybean, which consistently gave more accurate yield predictions than regression models (Kujawa and Niedbala, 2021), sugar cane (Fernandes *et al.*, 2017), and rice damage from 1993 - 1996, leading to the conclusion that ANN can be an alternative or complement to current early warning methods to identify potentially degraded fields and propose agricultural management plans on these high-risk areas (Tourenq *et al.*, 1999).

Neural network models have been used in recent studies to detect water quality (Sarkar and Pandey, 2015), assessment of watersheds (Ahmed *et al.*, 2019; Samantaray *et al.*, 2019), biodiversity (Gagne *et al.*, 2020; Jahani and Saffariha, 2020; Morshed *et al.*, 2022), land-use changes (Islam *et al.*, 2018; Saputra and Lee, 2019; Yatoo *et al.*, 2020), and forecasting droughts (Ali *et al.*, 2017; Dehghani *et al.*, 2017; Dikshit *et al.*, 2022).

Data

This study used the data obtained from the Philippine Statistics Authority (PSA) official website under the OpenStat. PSA serves as the central statistical authority charged by the government with primary data collection and analysis and publishes statistical information relating to the country's social, demographic, economic, and general conditions of the people. The data cover the period of the second semester (wet season) from 1987 until the year 2022. The data consist of 36 data points concerning the corn production in tonnes and harvested farmed area in hectares without missing values.

Methodology

The researcher utilized the ANN model system in order to analyze the patterns possibly embedded in the time series during the wet season (second semester) and forecast the volume of corn production for the corresponding year. Statistical error measures such as Root Mean Square Errors (RMSE) for each model were observed in selecting the best model to forecast the volume of corn production. The model with the least value of the aforementioned forecast accuracy measure was selected as the final model. Furthermore, the actual and forecast values were compared to determine if there exists a significant difference between the two. A significant difference between the two sets of data may imply that the model might not give high accuracy level of forecasting. Exploratory data analysis (EDA) was conducted to determine the appropriate statistical test to be used in establishing a significant difference between the two sets of data and in establishing a direct relationship between harvested farmed area (in hectares) and the volume of corn production (in metric tonnes) (Jebb *et al.*, 2017).

Artificial Neural Network (ANN) is a system that is based on the simple mathematical model of the brain. It allows complex analysis of non-linear relationships between the response and predictive variables (Kriegeskorte and Golan, 2019). In this study, ANN model was trained using the previously observed values of the time series data x_t and a hidden layer with a number of processing units known as neurons. The hidden layer in artificial neural networks is a layer of neurons whose output is connected to the inputs of other neurons. Neurons simply calculate the weighted sum of inputs and weights, add the corresponding bias and execute an activation function. The hidden layer could contain several processing units or neurons. Hyperbolic tangent, $TanH$, was used as the activation function in this study for the following reasons: it is less susceptible to saturation in the later layers of a given network, and it results in a faster convergence compared to the standard sigmoid function (Cococcioni *et al.*, 2020). Neural Network models achieved better accuracy and Mean Square Error (MSE) by using the decimal scale normalization technique (Mustaffa and Yusof, 2011). In this study, time series data were then normalized using the decimal scaling method prior to the training and testing. The input data x_t is the farmed area at a given year and output y_t is the volume (in tonnes) of corn production for that particular year. Fig. 2 shows the corresponding ANN model architecture used in this study.

During the development process of the Multi-Layer Perceptron's (MLP) ANN model, data were trained and validated. The field of artificial neural networks is often called neural networks or multi-layer perceptron's, which is perhaps the most useful type of neural network (Di Persio and Honchar, 2016).

A perceptron is a single-neuron model that was a precursor to larger neural networks. Thirty-three percent (33%) of the data were randomly holdback for the validation process. The trained and validated ANN model was denoted by ANN ($n : p : q$), where $n$the number of input variables is, p is the number of processing units in the hidden layer, and q is the number of the output variable. In this study, there was only one (1) variable considered as input (n) which is the harvested farmed areas and output (q), which is the predicted volume of corn production. Hence, the model will be of the form ANN ($1{:}p{:}1$).

Results and discussion

The time series data were pre-processed by checking if there were missing values and if there were none. Results of exploratory data analysis for harvested farmed area and volume of corn production show no significant outliers (see Fig. 3a and 3b), show linear relationship (see Fig. 3c), and are approximately normally distributed as observed through Shapiro-Wilk statistic of 0.97 (p-value=0.488) and 0.94 (p-value=0.057), respectively. Thus, Pearson r was used to evaluate if the two variables, harvested farmed area and volume of corn production, which were assumed to be directly correlated (OECD.org, 2021) and were found to have a high positive correlation, $r = .80\ (p - \text{value} < .001)$, and was statistically significant. This initial finding leads the researcher to develop, test, and validate an artificial neural network model based. The relationship between the two variables suggests that the volume of corn production can be predicted using the corresponding harvested farmed area during the crop season. Table 1 shows the corresponding RMSE's of each candidate ANN model.

Table 1. Summary of Root Mean Square Error (RMSE) of the 5 ANN ($1\colon p\colon 1$) model.

Model	Training	Validation
ANN($1\colon 1\colon 1$)	0.0661051	0.0557006
ANN($1\colon 2\colon 1$)	0.0647064	0.0553406
ANN($1\colon 3\colon 1$)	0.0546463	0.0546004
ANN($1\colon 4\colon 1$)	0.0631528	0.0614655
ANN($1\colon 5\colon 1$)	0.0648189	0.0556405

Table 2. Actual and forecast volume of corn production using ANN ($1\colon 3\colon 1$) Model for the last 5 years.

Year	Actual volume of corn production (in m tonnes)	Forecast volume of corn production (in m tonnes)
2018	1701.00	1286.40
2019	1988.00	1909.71
2020	1920.00	1774.88
2021	1790.16	1567.20
2022	1847.82	1608.80

During the ANN model construction, five (5) candidate models were trained and validated. Table 1 shows the summary of RMSE for the 5 candidate models. It was observed that the ANN ($1\colon 3\colon 1$) model has the least RMSE. It indicates that three (3) neurons in the hidden layer were sufficient for the model to accurately forecast volumes of corn production relative to the corresponding harvested farmed area based on calculated RMSE. Hence, ANN ($1\colon 3\colon 1$) was selected and was the most suitable model to forecast the volume (in tonnes) of corn production.

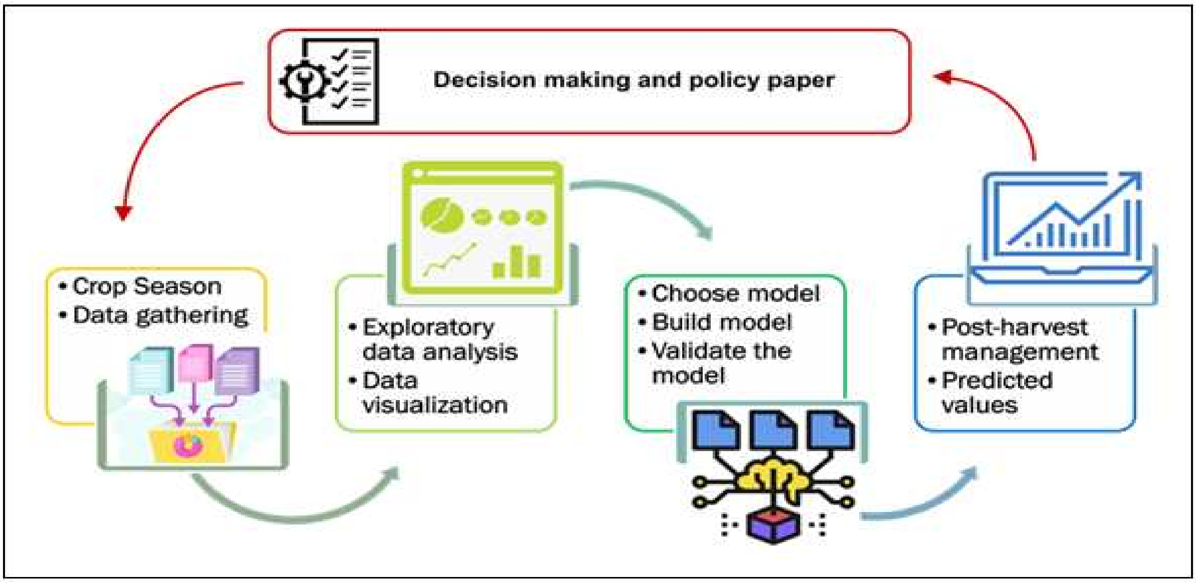

Fig. 1. Process of monitoring post-harvest crop production using mathematical model.

It can be observed in Table 2 that for the last 5 years, the forecast volume of production and the actual recorded data were relatively close.

Furthermore, after EDA on the sets of time series data, namely, the actual volume of production and forecast volume of production, which do not violate any of the assumptions of paired samples t-test (see Fig. 4, Shapiro-Wilk statistic of 0.98 (p-value$=0.661$)) and it was found out there is no significant difference between the two sets of data ($t(35)=-0.01, p-\text{value}=0.992$.

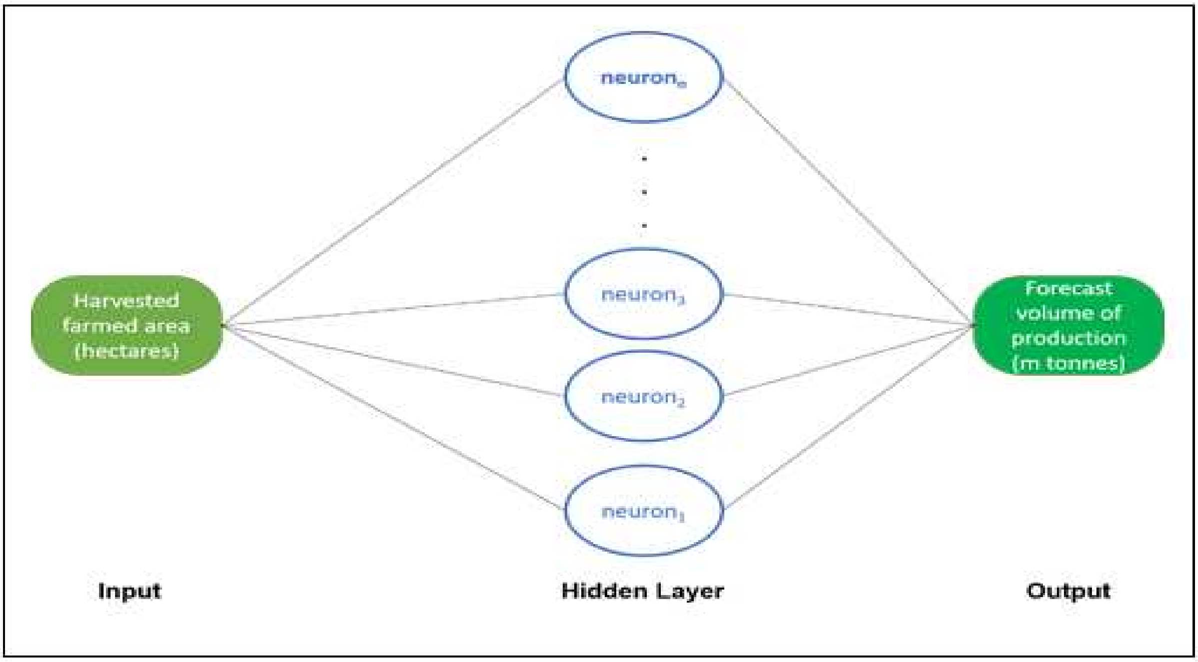

Fig. 2. Artificial Neural Network model architecture used in this study.

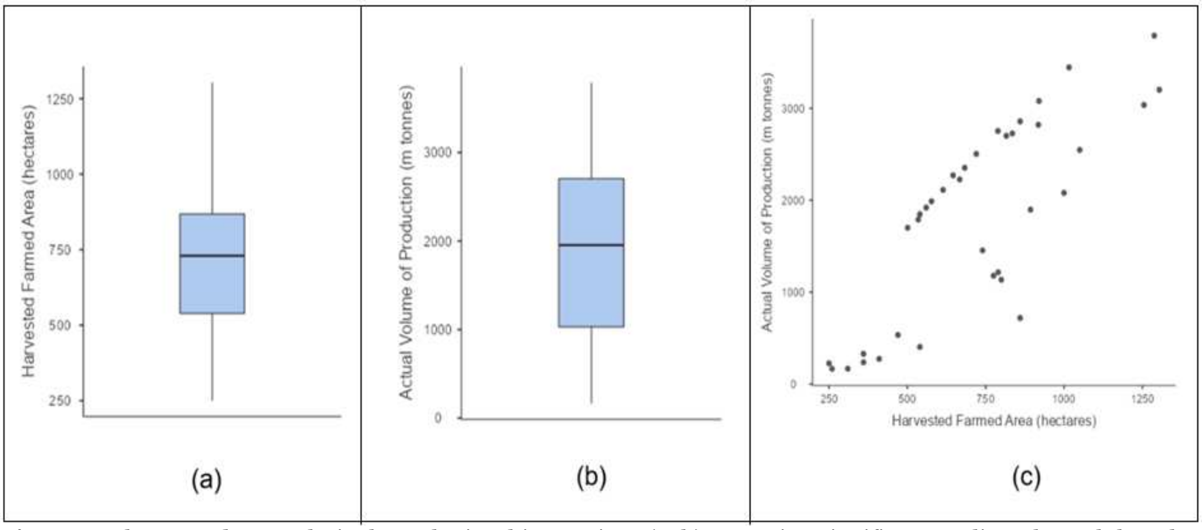

Fig. 3. Exploratory data analysis through visual inspection: (a, b) Detecting significant outliers through boxplot, (c) Detecting linear relationship between volume of production and harvested farmed area through scatterplot.

Fig. 5 shows the line graph of the two sets of time series data. The result conforms to the study conducted by Kaul *et al.* (2005), Mishra *et al.* (2016), and Palanivel and Surianarayanan (2019) in which ANN approach in the forecast of agricultural crops such as corn consistently produced more accurate yield predictions.

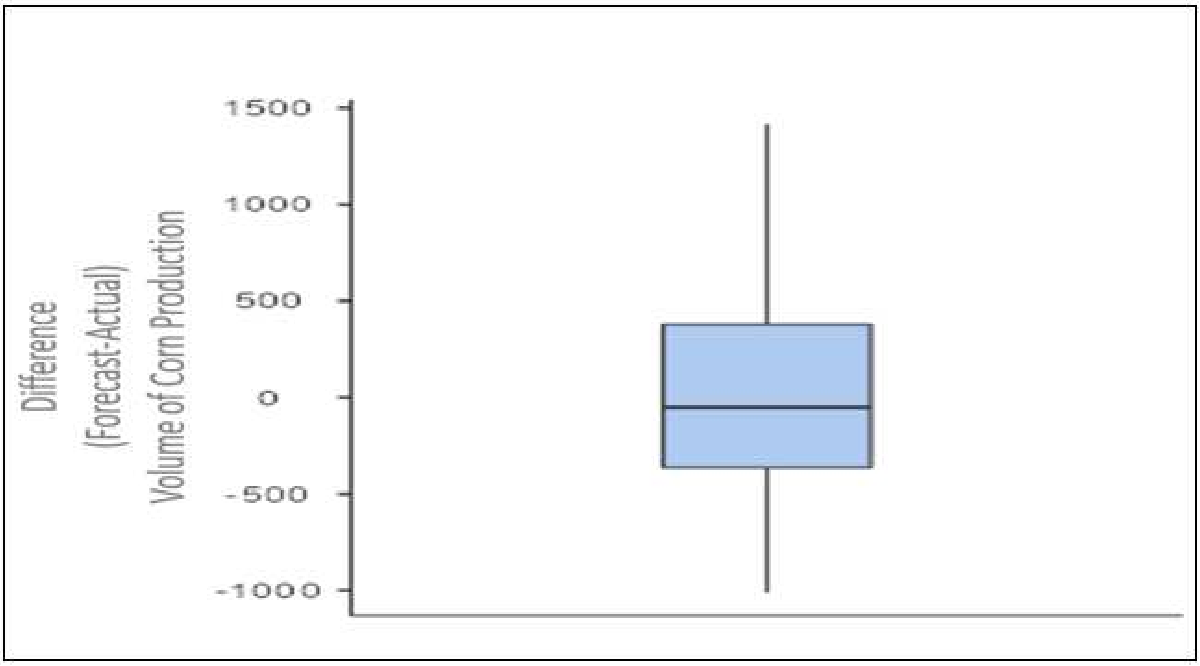

Fig. 4. Detecting significant outliers on the difference between the two data sets through boxplot.

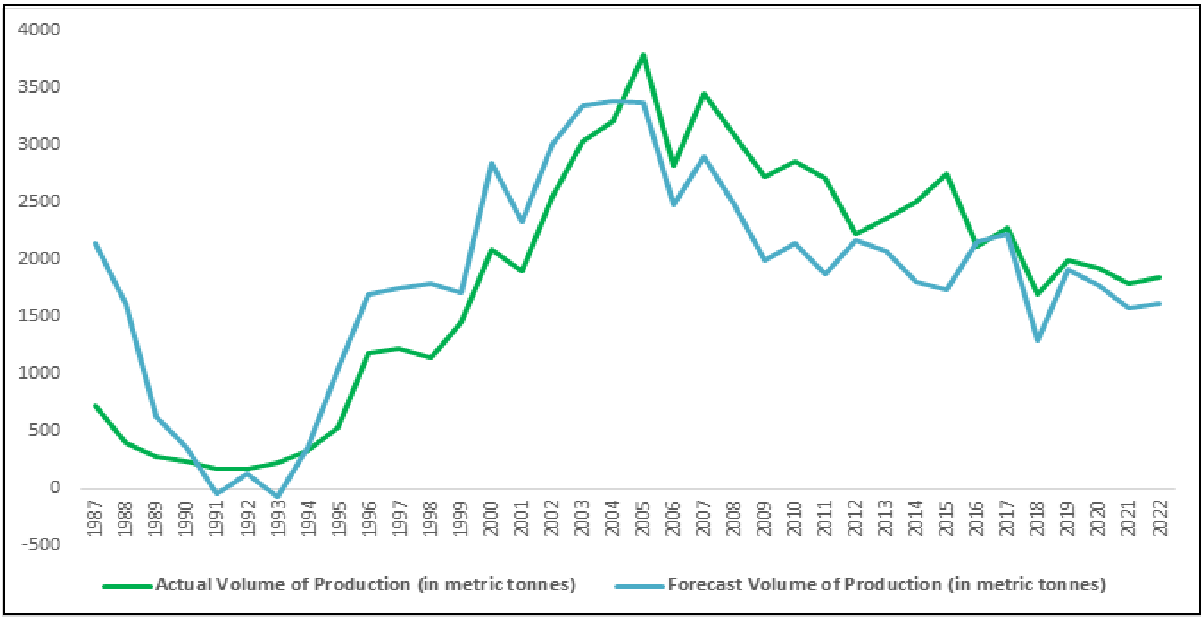

Fig. 5. Line chart visualization between the two time series data.

Recommendation (S)

Based on the findings, the volume (in tonnes) of corn production was highly correlated with the harvested farmed area. Based on the computed RMSE, a suitable neural network (ANN) model provides a good and useful forecast about the volume of corn production relative to the harvested farmed area. Furthermore, the two sets of data, namely actual and forecast volume of corn production, were found, statistically, not to have significant differences. This implies that the forecast values were relatively close to the actual volume of corn (maize) production. Researchers might consider forecasting volumes of production of other crops with the inclusion of some factors other than harvested farmed areas, such as precipitation. Also, it is recommended that the models be used by agencies concerned with monitoring post-harvest of agricultural crops. Agencies concerned should gather data and monitor farmers during the crop season and identify technologies and innovations used by the farmers for future reference. Such information could give vital data on the reasons for the increase or decrease in harvests. This information could benefit farmers. Furthermore, as data piles up, the models should be updated accordingly.

Acknowledgements

The researcher would like to express and extend his insightful gratitude to the following persons and institutions who manifested their untiring support in the completion and realization of this study: the members of the faculty of Mathematics and Allied Disciplines Department, the Philippine Statistics Authority for the data used in the study, the Don Mariano Marcos Memorial State University, the College of Arts and Sciences (CAS), and most especially to my wife, April, support and encouragement.

References

Agatonovic-Kustrin S, Beresford R. 2000. Basic concepts of artificial neural network (ANN) modeling. Journal of Pharmaceutical and Biomedical Analysis **22(5),** 717-727. https://doi.org/10.1016/S0731-7085(99)00272-1

Agri Farming. 2021. Agriculture in Philippines- Farming, Major Crops. Agri Farming: https://www.agrifarming.in/agriculture-in-philippines-farming-major-crops

Ahmed M, Sultan M, Elbayoumi T, Tissot P. 2019. Forecasting GRACE Data over the African Watersheds Using Artificial Neural Networks. Remote Sensing **11(15),** 1769. https://doi.org/10.3390/rs11151769

Ali Z, Hussain I, Faisal M, Nazir HM, Hussain T, Shad M, Gani S. 2017. Forecasting Drought Using Multilayer Perceptron Artificial Neural Network Model. Advance in Meteorology. https://doi.org/10.1155/2017/5681308

Cococcioni M, Rossi F, Ruffaldi E, Saponara S. 2020. Home Applications in Electronics Pervading Industry, Environment and Society Conference paper. Applications in Electronics Pervading Industry, Environment and Society **7,** 213-221.

Dehghani M, Saghafian B, Rivaz F, Khodadadi A. 2017. Evaluation of dynamic regression and artificial neural networks models for real-time hydrological drought forecasting. Arabian Journal of Geosciences.

Di Persio L, Honchar O. 2016. Artificial Neural Networks architectures for stock price prediction:. International Journal of Circuits, Systems and Signal Processing **10,** 403-413.

Dikshit A, Pradhan B, Santosh M. 2022. Artificial neural networks in drought prediction in the 21st century–A scientometric analysis. Applied Soft Computing **114,** 108080. https://doi.org/10.1016/j.asoc.2021.108080

FAO. 2022. Food and Agricultural Organization of the United Nation. GIEWS - Global Information and Early Warning System: https://www.fao.org/giews/countrybrief/country.jsp?code=PHL

Fernandes JL, Ebecken NF, Esquerdo JD. 2017. Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. International Journal of Remote Sensing **38(16),** 4631-4644. https://doi.org/10.1080/01431161.2017.1325531

Gagne TO, Reygondeau G, Jenkins CN, Sexton JO, Bograd SJ, Hazen EL, Van Houtan KS. 2020. Towards a global understanding of the drivers of marine and terrestrial biodiversity. *PloS One*. https://doi.org/10.1371/journal.pone.0228065

Grain Pro. 2023. *GrainPro*. Retrieved from The Benefits of Post-Harvest Management: https://news.grainpro.com/the-benefits-of-post-harvest-management

Gro-Intelligence. 2018. No Time to Waste: A New Model for Agricultural Forecasting. https://gro-intelligence.com/: https://gro-intelligence.com/insights/no-time-to-waste-a-new-model-for-agricultural-forecasting

Gutierrez-Estrada JC, Silva C, Yañez E, Rodriguez N, Pulido-Calvo I. 2007. Monthly catch forecasting of anchovy Engraulis ringens in. Fisheries Research **86(2),** 188-200.

Islam K, Rahman M, Jashimuddin M. 2018. Modeling land use change using Cellular Automata and Artificial Neural Network: The case of Chunati Wildlife Sanctuary, Bangladesh. Ecological Indicators **88,** 439-453. https://doi.org/10.1016/j.ecolind.2018.01.047

Jahani A, Saffariha M. 2020. Human Activities Impact Prediction in Vegetation Diversity of Lar National Park in Iran Using Artificial Neural Network Model. Integrated Environmental Assessment and Management. https://doi.org/10.1002/ieam.4349

Jebb AT, Pariggon S, Woo S. 2017. Exploratory data analysis as a foundation of inductive research. Human Resource Management Review **27,** 265-276. https://doi.org/10.1016/j.hrmr.2016.08.003

Kaul M, Hill RL, Walthall C. 2005. Artificial neural networks for corn andnsoybean yield prediction. Agricultural Systems **85,** 1-18. https://doi.org/10.1016/j.agsy.2004.07.009

Kriegeskorte N, Golan T. 2019. Neural network models and deep learning. Current Biology, **29(7).** https://doi.org/10.1016/j.cub.2019.02.034

Kujawa S, Niedbala G. 2021. Artificial Neural Networks in Agriculture. Agriculture **11(6),** 497. https://doi.org/10.3390/agriculture11060497

Mishra S, Mishra D, Santra G. 2016. Applications of Machine Learning Techniques in. Indian Journal of Science and Technology **9(38).** https://doi.org/10.17485/ijst/2016/v9i38/95032

Morshed SR, Fattah M, Haque M, Morshed S. 2022. Future ecosystem service value modeling with land cover dynamics by using machine learning based Artificial Neural Network model for Jashore city, Bangladesh. Physics and Chemistry of the Earth, Parts A/B/C **126,** 103021. https://doi.org/10.1016/j.pce.2021.103021

Mustaffa Z, Yusof Y. 2011. A comparison of normalization techniques in predicting dengue outbreak. International Conference on Information and Finance, 345-349.

OECD.org. 2021. Crop Production. Retrieved from OECD: https://data.oecd.org/agroutput/cropproduction.htm

Palanivel K, Surianarayanan C. 2019. An Approach for Prediction of Crop Yield Using Machine Learning and Big Data Techniques. International Journal of Computer Engineering and Technology, **10(3),** 110-118.

Prosekov AY, Ivanova SA. 2018. Food security: The challenge of the present. Geoforum **91,** 73-77. https://doi.org/10.1016/j.geoforum.2018.02.030

Ramasubramanian V. 2006. Forecasting Techniques in Agriculture. Journal of the Indian Society of Agricultural Statistics.

Samantaray S, Sahoo A., Ghose DK. 2019. Assessment of Runoff via Precipitation using Neural Networks:Watershed Modelling for Developing Environment in Arid Region. Pertanika Journal of Science & technology **27(4),** 2245-2263.

Saputra MH, Lee HS. 2019. Prediction of Land Use and Land Cover Changes for North Sumatra, Indonesia, Using an Artificial-Neural-Network-Based Cellular Automaton. Sustainability **11(11),** 3024. https://doi.org/10.3390/su11113024

Sarkar A, Pandey P. 2015. River Water Quality Modelling Using Artificial Neural Network Technique. Aquatic Procedia **4,** 1070-1077. https://doi.org/10.1016/j.aqpro.2015.02.135

Source Trace. 2019. *SourceTrace*. Cutting Post-Harvest Losses: The Bottom-up Approach to Increasing Food Security: https://www.sourcetrace.com/blog/cutting-post-harvest-losses-bottom-approach-increasing-food-security/

Tourenq C, Aulagnier S, Mesleard F, Durieux L, Johnson A, Gonzales G, Lek S. 1999. Use of artificial neural networks for predicting rice crop damage by greater flamingos in the Camargue, France. Ecological Modelling, **120(2-3),** 349-358. https://doi.org/10.1016/S0304-3800(99)00114-3

Vedantu. 2022. *Vedantu*. Retrieved from https://www.vedantu.com: https://www.vedantu.com/commerce/food-security

Yatoo SA, Sahu P, Kalubarme MH, Kansara B. B. 2020. Monitoring land use changes and its future prospects using cellular automata simulation and artificial neural network for Ahmedabad city, India. GeoJournal **87,** 765-786.