PERFORMANCE EVALUATION OF BOX-JENKINS' ARIMA, BOX-TIAO'S ARIMAX, AND MULTIPLE LINEAR REGRESSION MODELS FOR RAINFALL FORECASTING IN CABANATUAN, NUEVA ECIJA

MELENDREZ, JHOANA A.

An Undergraduate Thesis Submitted to the Faculty of the Department of Engineering Sciences, College of Engineering, Central Luzon State University, Science City of Muñoz, Nueva Ecija, Philippines
in Partial Fulfillment of the Requirements
for the Degree of

BACHELOR OF SCIENCE IN METEOROLOGY

MAY 2023

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF APPENDIX TABLES

LIST OF APPENDIX FIGURES

ABSTRACT

MELENDREZ, JHOANA A., Department of Engineering Sciences, College of Engineering, Central Luzon State University, Science City of Muñoz, Nueva Ecija, Philippines, **May 2023**, **PERFORMANCE EVALUATION OF BOX-JENKINS' ARIMA, BOX-TIAO'S ARIMAX, AND MULTIPLE LINEAR REGRESSION MODEL FOR RAINFALL FORECASTING IN CABANATUAN, NUEVA ECIJA**

Adviser: HENRISON C. SANCHEZ

Rainfall forecasting has gained increasing study interest in recent years because of its substantial impact on agriculture, public safety, and climate monitoring studies. Localized rainfall forecasting models are crucial as the Philippines is projected to have an increased number of heavy daily rainfall and extreme rainfall events by 2020 and 2050. Hence, this study aimed to develop Box-Jenkins' Autoregressive Integrated Moving Average (ARIMA), Box-Tiao's Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX), and Multiple Linear Regression (MLR) models to evaluate its performances and select the best rainfall forecasting model for Cabanatuan, Nueva Ecija. In the modeling process, the ARIMA model solely used accumulated rainfall data, the ARIMAX model was incorporated with external variables (temperature, relative humidity, wind direction, and wind speed), and the MLR model used predictor variables with a significant relationship to the response variable. The modeling procedures produced the ARIMA $(2,0,2)\ (1,0,0)_{12}$, ARIMAX $(0,0,0)\ (2,1,0)_{12}$, and $y = 1.9737x_1 - 73.7562x_2 - 89.9057x_3$ (MLR) models. The accuracy of these models was evaluated using Symmetric Mean Absolute Percentage Error (SMAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Normalized Root Mean

Square Error (NRMSE), and Coefficient of Determination (R^2). Results show that the ARIMAX (0,0,0) $(2,1,0)_{12}$ with exogenous variables of relative humidity, wind direction (cos and sin), wind speed, and temperature is the best-fitted model for monthly accumulated rainfall forecasting in Cabanatuan.

Keywords: ARIMA, ARIMAX, Multiple Linear Regression, Rainfall forecasting

LITERATURE CITED

Abbot, J., & Marohasy, J. (2018). Forecasting of Medium-term Rainfall Using Artificial Neural Networks: Case Studies from Eastern Australia. *Engineering and Mathematical Topics in Rainfall*, pp. 34–49.

Al-Saati, N., Omran, I., Salman, A.A., Al-Saati, Z., & Hashim, K. (2021). Statistical modeling of monthly streamflow using time-series and artificial neural network models: Hindiya Barrage as a case study. *Water Practice and Technology, 16*(2), 681–691. https://doi.org/10.2166/wpt.2021.012.

Amelia, R., Dalimunthe, D., Kustiawan, E., & Sulistiana, I. (2021). ARIMAX model for rainfall forecasting in Pangkalpinang, Indonesia. *IOP Conf. Ser.: Earth Environ. Sci. 926*(3). https://doi.org/10.1088/1755-1315/926/1/012034

Angco, R. J., Timtim, L., Ando, M., Leyson, C., & Villasin, C. R. (2021). Time-series approach on Philippines' three economic participation using ARIMA Model. *Technium Social Sciences Journal, 25*(1), 304–332. https://techniumscience.com/index.php/socialsciences/article/view/4937

Angelini, C. (2019). Regression Analysis. *Encyclopedia of Bioinformatics and Computational Biology, 2019*(1), 722–730. https://doi.org/10.1016/B978-0-12-809633-8.20360-9.

Anusha, N., Chaithanya, M., & Reddy, G. (2019). Weather Prediction Using Multi Linear Regression Algorithm. *IOP Conf. Series: Materials Science and Engineering, 590(2019)*, 1–9. https://doi.org/10.1088/1757-899X/590/1/012034.

Bari, S.H., Md. Sourov, M.H., & Rahman, M.T. (2015). Forecasting Monthly Precipitation in Sylhet City Using ARIMA Model. *Civil and Environmental Research*, *7*(1), 69–77.https://www.researchgate.net/publication/272744442_Forecasting_Monthly_Precipitation_in_Sylhet_City_Using_ARIMA_Model

Box, G., Jenkins, G., Reinsel, G., & Ljung, G. (2015). Time-series Analysis: Forecasting and Control. *John Wiley and Sons Inc., 5.* http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/504_05_Box_Time-Series-Analysis-Forecasting-and-Control-2015.pdf

Brouwer, C., Goffeau, A. and Heibloem, M. (1985) Irrigation Water Management: Training Manual No. 1-Introduction to Irrigation. https://www.fao.org/3/X5560E/x5560e02.htm.

Brownlee, J. (2020, December 9). *How to decompose time-series data into trend and seasonality.* MachineLearningMastery. Retrieved December 24, 2022, from https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/#:~:text=Time%20series%20decomposition%20involves%20thinking,time%20series%20analysis%20and%20forecasting

Chao, Z., Pu, F., Yin, Y., Han, B., & Chen, X. (2018). Research on Real-Time Local Rainfall Prediction Based on MEMS Sensors. *Hindawi Journal of Sensors, 2018*, 1–9. https://doi.org/10.1155/2018/6184713.

Chiu, T., Chiang, C., & Lee. T. (2016). Time-series modeling of biochemical oxygen demand at the upstream catchment of Feitsui Reservoir, Taiwan. *Hydrology Research, 47*(5), 1069–1085. https://doi.org/10.2166/nh.2016.136

Chou, C., Chiang, J., Lan, CW., Chung, C., Liao, Y., & Lee, C. (2013). Increase in the range between wet and dry season precipitation. *Nature Geosci, 6*(2013), 263–267. https://doi.org/10.1038/ngeo1744.

Cinco, T. A., De Guzman, R.G., Hilario, F.D., & Wilson, D. M. (2014). Long-term trends and extremes in observed daily precipitation and near-surface air temperature in the Philippines for the period 1951-2010. *Atmospheric Research, 145-146*(2014), 1226. https://www.sciencedirect.com/science/article/pii/S0169809514001495

Corporal-Lodangco, I. & Leslie, L. (2017). Defining Philippine Climate Zones Using Surface and High-Resolution Satellite Data. *Procedia Computer Science, 114*, 324-332. https://doi.org/10.1016/j.procs.2017.09.068

Crost, B. (2015, October 20). *Wetter wet seasons, drier dry seasons harm agriculture and Fuel Conflict in Asia.* PolicyMatters. Retrieved December 26, 2022, from https://policymatters.illinois.edu/wetter-wet-seasons-drier-dry-seasons-harm-agriculture-and-fuel-conflict-in-asia/

Cruz, R.V., et al. (2017). Philippine climate change assessment, impacts, vulnerabilities, and adaptation. NICCDIES. https://niccdies.climate.gov.ph/files/documents/Philippine%20Climate%20Change%20Assessment%20Working%20Group%202_%20Impacts%20Vulnerabilities%20and%20Adaptation.pdf Asian Development Bank. (2012). Country Partnership Strategy. Asian Development Bank. https://www.adb.org/sites/default/files/institutional-document/33024/files/cps-phi-2011-2016.pdf.

Dr. Chellam, A., Dr. Sangari, S., & Dr. Kumar, R. (2021). *Drought Prediction Using Ai Based Multiple Linear Regression Technique. JCDR, 12*(3), 892-898. https://doi.org/10.1016/B978-0-12-809633-8.20360-910.31838/jcdr.2021.12.03.118.

Einfalt, T., Denoeux, T., & Jacquet, G. (1990). A radar rainfall forecasting method designed for hydrological purposes. *Journal of Hydrology, 114*, 229-244.https://www.hds.utc.fr/~tdenoeux/dokuwiki/_media/en/publi/j_hydro90.pdf.

Fernandez, F.R., Montero, Po III, R., Addawe, R., & Diza, H.M. (2018). Forecasting manila south harbor mean sea level using seasonal arima models. *Journal of Technology Management and Business, 5*(1). https://doi.org/10.30880/jtmb.2018.05.01.001

Giles, J., Macandog, P.B., Nowak, A., Pablo, J., Sebiano Jr, B., Palao, L.K., Balanza, J.G., Hildebrand, J., and Grosjean, G. (2019). Climate-Resilient Agriculture in The Philippines: Climate Risk Profile, Luzon. *International Center for Tropical Agriculture (CIAT); Department of Agriculture - Adaptation and Mitigation Initiative in Agriculture, Government of the Philippines; The Food and Agriculture Organization of the United Nations (FAO), 32*. https://www.adb.org/sites/default/files/publication/722241/climate-risk-country-profile-philippines.pdf

Global Facility for Disaster Reduction and Recovery. (2011). Vulnerability, risk reduction, and adaptation to climate change Philippines. Climate Knowledge Portal. https://climateknowledgeportal.worldbank.org/sites/default/files/2018-10/wb_gfdrr_climate_change_country_profile_for_PHL.pdf.

Gornall, J., Betts, R., Burke, E., Clark, R., Camp, J., Willett, K., & Wiltshire A. (2010). Implications of climate change for agricultural productivity in the early twenty-first century. *Phil. Trans: R. Soc. B, 2010*(365), 2973–2989. http://doi.org/10.1098/rstb.2010.0158

Guijarro, JA (2011): Influence of network density on homogenization performance. *Seventh Seminar for Homogenization and Quality Control in Climatological Databases jointly organized with the Meeting of COST ES0601 (HOME) Action MC Meeting, 78*, 11-18. http://www.climatol.eu/MULTITEST/

Han, S., Xu, J., Yan, M., & Liu, Z. (2022). Using multiple linear regression and BP neural network to predict critical meteorological conditions of expressway bridge pavement icing. *Plos one, 17*(2), e0263539. https://doi.org/10.1371/journal.pone.0263539

Hodson, T. (2022). Root mean square error (RMSE) or mean absolute error (MAE): when to use them or not. https://doi.org/10.5194/gmd-2022-64.

Islam, F., & Imteaz, M. A. (2021, February). The effectiveness of ARIMAX model for prediction of summer rainfall in northwest Western Australia. In IOP Conference Series: Materials Science and Engineering (Vol. 1067, No. 1, p. 012037). IOP Publishing.https://iopscience.iop.org/article/10.1088/1757-899X/1067/1/012037/meta

Kritharas, P. (2013). Developing a SARIMAX model for monthly wind speed forecasting in the UK [Doctoral thesis, Loughborough University]. https://core.ac.uk/download/pdf/288377528.pdf

Lebedev, V. et al. (2019). *Precipitation Nowcasting with Satellite Imagery*. Arxiv. https://arxiv.org/abs/1905.09932#:~:text=Precipitation%20nowcasting%20is%20a%20short,map%20by%20the%20weather%20service.

Liu, Y., Zhao, Q., Yao, W., Ma, X., Yao, Y., & Liu, L. (2019). Short-term rainfall forecast model based on the improved BP–NN algorithm. *Scientific Reports*, *9*. https://doi.org/10.1038/s41598-019-56452-5

Makridakis, S. & Hibon, M. (1997). ARMA Models and the Box-Jenkins Methodology. *Journal of Forecasting, 16*(3), 147–163. https://doi.org/10.1002/(SICI)1099-131X(199705)16:3<147::AID-FOR652>3.0.CO;2-X

Md. Bagmar, S.H. & Md. Khudri, M. (2020). Application of box-Jenkins models for forecasting drought in northwestern part of Bangladesh. *Environmental Engineering Research, 26*(3). https://doi.org/10.4491/eer.2020.294

Montgomery, D., Peck, E., & Vining G. (2012). Introduction to linear regression analysis. https://scholar.google.com/scholar?q=.+Introduction+to+linear+regression+analysis.+John+Wiley+%26+Sons%2C+2012.

Ndamani, F. & Watanabe, T. (2015). Influences of rainfall on crop production and suggestions for adaptation. *International Journal of Agricultural Sciences 5*(2015), 367-374. https://www.researchgate.net/publication/270585608_Influences_of_rainfall_on_crop_production_and_suggestions_for_adaptation.

Olatayo, T. & Taiwo, A. (2019). *Statistical Modelling and Prediction of Rainfall Time-series Data*. Research Gate. https://www.researchgate.net/publication/338234358_Statistical_Modelling_and_

Prediction_of_Rainfall_Time_Series_Data_Statistical_Modelling_and_Prediction_of_Rainfall_Time_Series_Data

Osborne, Jason & Elaine Waters (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation, 8*(2), 1-5. Available online: http://PAREonline.net/getvn.asp?v=8&n=2.

Ozili, P. (2022). The Acceptable R-Square in Empirical Modelling for Social Science Research. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4128165.

PAGASA (2018). Observed and Projected Climate Change in the Philippines. *Philippine Atmospheric, Geophysical and Astronomical Services Administration*, pp.36.

Pajuelas, B.G. (2000). A study of rainfall variations in the Philippines: 1950-1996. *Science Diliman, 12*(1), 1–28. https://journals.upd.edu.ph/index.php/sciencediliman/article/view/227

Parveen, A. & Ahtisham, Y. (2021). Understanding and interpreting regression analysis. *Evidence-Based Nursing, 24*(4), 1-3. https://doi.org/ 10.1136/ebnurs-2021-103425.

Paz-Alberto, A., Dios, M., Alberto, R., & De Guzman, C. (2018) Climate Change Impacts and Vulnerability Assessment of Selected Municipalities and Agroecosystems to Support Development of Resilient Communities and Livelihoods in Nueva Ecija, Philippines. *American Journal of Climate Change, 7*, 295-335. https://doi.org/10.4236/ajcc.2018.72019.

Prabhakaran, S. (2021). ARIMA Model – Complete Guide to Time-series Forecasting in Python. https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/?fbclid=IwAR1rJG9DepKalN0BmwsV-oGDqb5NvlaPFszPJ2HVK-0MSsmvUeg87pIzhbo.

Ridwan, W.M., Sapitang, M., Aziz, A., Kushiar, K.F., Ahmed, A.N., & El-Shafie, A. (2021). Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. *Ain Shams Engineering Journal, 12*(2), 1651-1663. https://doi.org/10.1016/j.asej.2020.09.011.

Rodríguez, A. et al. (2016). Meteorological variables prediction through ARIMA models. *Agroscience, 50*(1), 1-13. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-31952016000100001&lng=en&tlng=en.

Schaffer, A.L., Dobbins, T.A. & Pearson, SA. Interrupted time-series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. *BMC Med Res Methodol 21*(2021), 58. https://doi.org/10.1186/s12874-021-01235-8

Schneider, A., Hommel, G., & Blettner, M. (2010). Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Dtsch Arztebl Int., 107*(44):776-82. https://doi.org/10.3238/arztebl.2010.0776.

Scott, G. (2021, June 29). *Box-Jenkins model.* Investopedia. https://www.investopedia.com/terms/b/box-jenkins-model.asp

Somlanare, R.K.& Badolo, F. (2019). Does rainfall variability matter for food security in developing countries? *Cogent Economics & Finance, 7*(1). https://doi.org/10.1080/23322039.2019.1640098

Swain, S., Patel, P.K., & Nandi, S. (2017). A multiple linear regression model for precipitation forecasting over Cuttack district, Odisha, India. *2017 2nd International Conference for Convergence in Technology, I2*(2017), 355-357. http://dx.doi.org/10.1109/I2CT.2017.8226150.

Talirongan, F., Talirongan, H., & Orong, M. (2021). Modeling National Trends on Health in the Philippines Using ARIMA. *J Health Med Informat, 11*(1), 1-5. https://doi.org/10.37421/jhmi.2020.11.342

Torres, M., Howitt, R., & Rodrigues, L. (2019). Analyzing rainfall effects on agricultural income: Why timing matters. *EconomiA, 20*(1), 1-14. https://doi.org/10.1016/j.econ.2019.03.006

Turner, P. (2019). Critical values for the Durbin-Watson test in large samples. *Applied Economics Letters. 27*(18), 1-5. https://doi.org/10.1080/13504851.2019.1691711.

Uba, E. & Bakari, H.R. (2015). An Application of Time-series Analysis in Modeling Monthly Rainfall Data for Maiduguri, North Eastern Nigeria. *Mathematical Theory and Modeling, 5*(11), 1-10. https://www.iiste.org/Journals/index.php/MTM/article/view/26365

United States Agency for International Development (2017). Climate Change Risk Profile Philippines.https://www.climatelinks.org/sites/default/files/asset/document/2017_Climate%20Change%20Risk%20Profile_Philippines.pdf

Vahdat, S.F., Sarraf, A., Shamsnia, A., & Shahidi, N. (2011). Prediction of monthly mean Inflow to Dez Dam reservoir using time-series models (Box-Jenkins). *International Conference on Environment and Industrial Innovation, 12*(2011). http://www.ipcbee.com/vol12/31-C10021.pdf

Valipour, M. (2012). Ability of Box-Jenkins Models to Estimate of Reference Potential Evapotranspiration (A Case Study: Mehrabad Synoptic Station, Tehran, Iran). *IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS). 1*(5), 1–11. https://www.iosrjournals.org/iosr-javs/papers/vol1-issue5/A0150111.pdf

Venema, V. et al. (2012). Benchmarking homogenization algorithms for monthly data. *Clim. Past*, *8*, 89–115. https://doi.org/10.5194/cp-8-89-2012

Villafuerte, M. Q. I., Matsumoto, J., Akasaka, I., Takahashi, H.G., Kubota, H. & Cinco, T. A. (2014). Long-term trends and variability of rainfall extremes in the Philippines. *Atmospheric Research, 137*(2014), 1–13. https://www.sciencedirect.com/science/article/pii/S0169809513002731

Villarin, J. T., Algo, J. L., Cinco, T. A., Cruz, F. T., de Guzman, R. G., Hilario, F. D., Narisma, G. T., Ortiz, A. M., Siringan, F. P., Tibig, L. V. (2016). 2016 Philippine Climate Change Assessment (PhilCCA): The Physical Science Basis. *The Oscar M. Lopez Center for Climate Change Adaptation and Disaster Risk Management Foundation Inc. and Climate Change Commission, 108*. https://climate.gov.ph/files/PhilCCAWG1.pdf?fbclid=IwAR1DqZiQj4LF5NM8JQXU0ZlLED6IPSXNfNEJMwi1r5OErpCBIeA_n9l30rg

WMO Commission for Climatology. (2022). *Web site of the Task Team on HOMOGENIZATION*. TT-Hom Web Site. Retrieved January 4, 2023, from http://www.climatol.eu/tt-hom/index.html

Yu, W., Nakakita, E., Kim, S., & Yamaguchi, K. (2015). Improvement of rainfall and flood forecasts by blending ensemble NWP rainfall with radar prediction considering orographic rainfall. *Journal of Hydrology, 531*, 494-507. https://doi.org/10.1016/j.jhydrol.2015.04.055

Zakaria, S. et al. (2012). ARIMA Models for weekly rainfall in the semi-arid Sinjar District at Iraq. *Journal of Earth Sciences and Geotechnical Engineering*, *2*(3), 22-55. https://www.academia.edu/34878392/ARIMA_Models_for_weekly_rainfall_in_the_semi_arid_Sinjar_District_at_Iraq

Zhou, K., Sun, J., Zheng, Y., & Zhou, Y. (2022). Quantitative Precipitation Forecast Experiment Based on Basic NWP Variables Using Deep Learning. *Adv. Atmos. Sci. 39* (2022), 1472–1486. https://doi.org/10.1007/s00376-021-1207-7

Data Source: Climate and Agrometeorological Data Section (CADS), Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA)