

A Methodology for The Use of Long-Range Dependency and Machine Learning Algorithms for Rainfall Forecasting

K.D.H. Hasadara¹

INSBM Green University, Mahenwaththa, Pitipana, Homagama, 10206, Sri Lanka

huvini.h@nsbm.ac.lk

ABSTRACT

Forecasting rainfall is found to be an important indicator to the undisturbed continuation of social and economic activities. This derives the significance of predicting rainfall, providing better decision making in sectors such as agriculture, power, and tourism. This paper discusses the methodology behind applying long range dependency and machine learning in rainfall forecasting. It was determined that a time series should be established with data from past rainfall figures. This data will be analyzed through the Auto Correlation Factor (ACF) and the Partial Auto Correlation Factor (PACF) to check for seasonality and to identify the lag after which the other lags are insignificant. A time series process is categorized as long memory if the serial dependence or the ACF decays slower than an exponential decay. An exponentially decaying ACF can be categorized as having short memory. The spectral density of the time series will also be measured to gauge how far the peaks are from zero. The times series with

peaks closer to zero will be considered as standard long memory, while the further peaks will be categorized as generalized long memory. The Ljung-Box Test and the Brusch-Godfrey Test will be used to assess whether any group of autocorrelations are different from zero. Based on the outputs from this analysis, it is possible to apply the time series into models such as ETS and ARIMA for prediction. The residual values from the models can determine the best fit model. Which can be further verified through machine learning.

Keywords- Rainfall, Forecast, Time Series, Long Range Dependency

1. INTRODUCTION

Many countries have identified a linkage between the rainfall amount and the country's economic growth (Ojo & Ogunjo, 2022). Since excessive rainfall causes floods and insufficient rainfall causes drought, rainfall is a crucial consideration when planning and managing agriculture, and related industries globally. Surface runoff, which is a significant factor in the

transport of sediment, nitrate, phosphorus, and other chemical compounds in the watershed, as well as flood-related economic losses are indicators of how rainfall affects the economy of a country. Therefore, it is essential to develop suitable mechanisms to track and forecast rainfall for effective management, planning, and mitigation.

Forecasting future rainfall is crucial for decision makers working in irrigation, urban and rural water management, tourism, ecology, and agriculture as well as society. Despite the significance placed on the forecasting of rainfall, it has proven to be a complex process due to the inconsistent nature of rainfall. Adding to this, Scher & Messori (2019) have concluded that predictability of rainfall (specially in summertime) will decrease over time with global warming.

Despite the huge amount of research being conducted to forecast rainfall using various factors, Hudson et al. (2010) found that forecasting models created with climate drivers as potential predictors are unable to predict rainfall for more than a week. Therefore, more attention needs to be given to studies which take time series analysis into consideration. Due to the chaotic nature of rainfall, researchers have tried out various linear and nonlinear approaches to model the rainfall in different geographies and time scales.

Linear and non-linear modeling techniques are used widely for hydrological forecasting. Multiple Linear Regression (MLR) is widely used by researchers for variable forecasting; Rasel et al. (2015)

applied MLR to investigate the influence of ENSO and SAM on the South Australian rainfall, and Mekanik et al. (2013) investigated the influence of IOD and ENSO on the Victorian rainfall. On the other hand, non-linear modelling is also used for rainfall forecasting through applications such as Artificial Neural Networks (ANN). According to Maier and Dandy (2001), ANN can handle incomplete, noisy, and ambiguous data, which is consistent with the characteristics of the rainfall data.

Simple models such as linear regression models have been in use to model the monthly rainfall amounts (Westra & Sharma, 2009; Lloyd-Hughes & Saunders, 2002; Oettli & Camberlin, 2005). Bhakar et al. (2006) has used an additive time series decomposition model to predict the rainfall in the Kota region in India. A limitation of such simple models is the defectivity of the model where monthly rainfall does not follow a normal distribution. This is the case for Australian rain stations (Ghamariadyan & Imteaz, 2021).

In many cases, it is essential to have access to the monthly rainfall distribution rather than the mean seasonal value (Mehrotra et al., 2013). It is also very advantageous to predict the upcoming monthly rainfall some months in advance. A variety of computer models and observational data are used in rainfall prediction simulations by taking into account factors such as atmospheric conditions and historical weather patterns. These methods analyze historical data to identify trends, seasonality, cycles, and correlations, allowing researchers to develop models to predict future rainfall

based on observed patterns. Using regression techniques, researchers can derive mathematical equations that determine the relationship between meteorological variables and rainfall. This equation can be used to predict future rainfall events more quantifiably and accurately (Zanial et al., 2023). But the residual errors will not be independent. It is addressed through time series forecasting.

Going beyond using computational models to forecast rainfall, it is crucial to utilize much more innovative solutions such as artificial intelligence (AI). AI is a method for manipulating and using implicit, unused, and potentially valuable information from data. It is a broad and profound field that keeps growing and being applied in different fields. By utilizing the strength of sophisticated algorithms and machine learning techniques, AI offers promising solutions in the context of rainfall prediction and its impact on the economy. AI algorithms can find complex relationships and patterns that humans might miss by processing and learning from this data.

1.2 Problem Statement

It has been established that rainfall is one of the toughest hydrological variables to predict due to its volatility in time and space (French et al., 1992). Yet, due to the importance of rainfall forecasts to the economy and society, different researchers have tried various approaches to predict the rainfall in their respective study locations. Literature identifies several forecasting models such as the Box-Jenkins autoregressive integrated moving average

model and Artificial Neural Networks (ANN) being used for countries such as India, Jordan, Thailand, Iraq and Nigeria (Dubey, 2015; Eni & Adyeye, 2015; Hung et al., 2009; Momani, 2009; Wang et al., 2013; Zakaria et al., 2012).

Feng et al. (2020) developed probabilistic seasonal rainfall forecasting model for Australia using machine learning. This model managed to generate seasonal predictions for Queensland with higher accuracy than the governments Southern Oscillation Index (SOI) Phase model (Ghamariadyan & Imteaz, 2021). Ghamariadyan & Imteaz (2021) have used a hybrid wavelet ANN model using temperature and climate indices to predict the rainfall in Queensland with promising insights. Bagirov & Mahmood (2018) in their study mentioned that there is a research gap in Queensland for alternative models to forecast rainfall.

Despite the extensive research being done on rainfall forecasting using various models, little can be found in the context of time series memory. Gil-Alana et al., (2022) found out through their study that there is indeed a long memory dependency in rainfall data for the United States of America (USA). This is proven true for Malaysia as well through the research done by (Yusof & Kane, 2012). However, there is a significant research opportunity to set a guiding framework for future researchers to utilize when applying time series forecasting techniques to predict the future rainfalls. This study aims to provide a general methodology for researchers to follow were they to apply time series

forecasting methods to their selected research geography.

2. LITERATURE REVIEW

2.1 Time Series Analysis

The importance of time series data and the analysis of such data is growing by the day due to the higher leniency towards data-based decision making. It can be observed that time series models are being developed with increasing levels of accuracy over time. Most effective new models combine the two approaches of statistical and machine Learning techniques (Bagirov & Mahmood, 2018).

2.1.1 Long Memory Models

Since the introduction of the short memory autoregressive moving average (ARMA) model by Box and Jenkins (1970) the integer differenced time series model Autoregressive Integrated Moving Average (ARIMA) was introduced to the world of statistics by Granger and Joyeux (1980) along with Hosking (1984) (Mills, 2018). Many recent studies have shown that time series showcase Long Range Dependency (LRD) (Aks & Sprott, 2003; Delignieres, et al., 2006; Eke et al., 2002; Madison, 2004).

LRD is how fast the statistical dependence between two points degrades with the increase of time or spatial separation. A time series is LRD if the dependence has a slower decay than an exponential decay. On the contrary, if the time series shows short memory, the autocorrelation will decay faster at an exponential rate as the number of observations increases in time and space (Stroe-Kunold et al., 2009).

Unbounded spectral density closer to the origin in terms of the frequency along the x-axis, and hyperbolic decay of the Autocorrelation Function (ACF) can be established as key features of the ARIFMA model. It is identified as a standard long memory model.

Ideally any time series is required to be of a stationary nature for modelling. Yet, this is hardly the case in real life time series. The time series will have to be adjusted to make it fit the features of the model. Yet another unique feature of long memory models would be the hyperbolic decay of the Partial Autocorrelation Function (PACF) as opposed to an exponential decay of it in short memory models.

The time series will have to be adjusted to make it fit the features of the model. Apart from the hyperbolically decaying ACF, and a spectral density increasing limitlessly as the frequency gets closer to zero, the Hurst phenomenon is also a basic property of an LRD (Stroe-Kunold et al., 2009). The Hurst exponent (H) shows that the parameter showing the probability of an event occurring in a time series is continued with a similar event derived from 0.5. When $H=0.5$, the observations are independent (Hurst et al., 1965).

LRD can be captured in three major methods: (1) by means of H lying within the parameters of fractal analysis; (2) through power exponent or power spectrum function; (3) through the Autoregressive Fractionally Integrated Moving Average (ARFIMA) framework. Fractal processes are long-memory, self-similar processes. Mandelbrot & Wallis, (1969) implies that a

process will look the same or with comparable statistical features at different levels through the concept of self-similarity in statistics. These processes have self-similar autocorrelations and power spectra. The power-spectral density of a process with serial dependence is inversely proportional to the frequency for long-range serial dependence.

The power spectral density (PSD) analysis, also known as the periodogram method, is a technique for estimating in spectral analysis by determining the negative slope of the line connecting $\log S(f)$ and $\log f$ (Eke et al., 2002).

Granger, Joyeux, and Hosking (1980,1984) showed that by allowing the differencing parameter d of the Box-Jenkins ARIMA methodology to take on continuous values, self-similar hyperbolic decaying autocorrelations of fractional series could be parsimoniously modeled. In the context of the [ARFIMA (p, d, q)] framework, d is the parameter for simulating long-range dependence. The parameters p and q are used to capture the process's short-range dependencies. When $-0.5 < d < 0.5$, the process is both stationary and invertible. The range $[-0.5, 0]$ for the value of d suggests antipersistence. Long-term persistent processes have $d > 0$. As a result, $0 < d < 0.5$ can be used to model finite long memory. The process is nonstationary with noninvertible ARMA representations for $0.5 \leq d < 1$. The memoryless ARFIMA (0, 0, 0) process is Gaussian white noise.

2.1.2 ARFIMA Framework

In many time series, having a long memory or strong persistency is seen as stylized

facts. Modeling and subsequent estimation for these series are challenging due to the strong persistence that is on the one hand characterized by slowly decaying autocorrelations. The most common model for capturing strong persistence and long memory is Fractional Integration (FI) of order d . Estimated values of d range from 0 to 1, with $d = 1/2$ denoting the transition from stationarity to nonstationarity. Thus, modeling persistence can be done with a great deal of flexibility using FI.

In comparison to ARMA (1,1) models, (Brodsky & Hurvich, 1999) show that FI models are more effective at forecasting long memory time series. In comparison to several short memory rivals for medium sample sizes, (Bhardwaj & Swanson, 2006) report improvements in predictive performance when using FI models.

There is evidence that these gains increase noticeably for larger sample sizes. In terms of empirical evidence, Bhardwaj and Swanson (2006) again report that FI models are superior predictors of financial data, but that the evidence is less conclusive for macroeconomic data. They primarily attribute the much larger sample sizes in finance for this result, which resulted in a more accurate estimation of the long memory parameter. Dissanayake (2016a) provides an insight into a robust ARFIMA configuration for estimation and prediction. A stationary fractional ARIMA model was used successfully for rainfall prediction in Sri Lanka according to Silva, Peiris and Dissanayake (2018a) utilizing an ARFIMA series. An extension to it capturing the seasonality feature in addition to long memory is presented in Silva, Peiris and

Dissanayake (2018b) using a seasonal autoregressive fractionally integrated moving average (SARFIMA) model. The evidence is conflicting because some authors assert that short-memory substitutes, such as the Heterogeneous Autoregressive model (HAR) by (Corsi, 2009), have comparable predictive power. Yet it is found that FI models outperform a wide range of competing methods, including autoregressive (AR) models, exponential smoothing (ES) methods, and for the realized volatility series, also the HAR model, in systematic forecasting experiments using classical examples of strongly persistent time series from macroeconomics and finance, namely inflation and realized volatility, as well as simulated series. According to published research, adding a long memory component to the model used when forecasting a strongly persistent time series is advantageous in all tested scenarios, e. g. over all forecasting horizons and sample sizes. Extending the ARFIMA model to a much more generalized long memory time series version results in a Gegenbauer Autoregressive Moving Average (GARMA) model. It depicts long memory with unbounded spectral peaks away from the frequency origin and hyperbolically decaying ACF and PACF.

2.1.3 GARMA Framework

In certain texts GARMA is defined as Generalised Autoregressive Moving Average model. Extensive work with rich expositions related to fractional differencing and prediction using the GARMA model could be found in discussions presented in Dissanayake and

Peiris (2012), Dissanayake (2016b), Dissanayake, Peiris, and Proietti (2014), Dissanayake, Peiris, and Proietti (2016), Dissanayake, Peiris, and Proietti (2018), and Dissanayake (2023). The theoretical developments and empirical evidence found in such work would be utilized for the models to forecast rainfall if generalized persistence is found in the data under observation.

2.2 Machine Learning

Throughout the history researchers have been using data mining techniques (Kusaik et al., 2013; Chowdari et al., 2015; Tharun et al., 2018; Zainudin et al., 2016), big data analysis methods (Namitha et al., 2015; Manandhar et al., 2019), and machine learning algorithms to increase the accuracy of rainfall predictions (Arnav, 2019; Aswin et al., 2018; Zeelan et al., 2020; Vijayan et al., 2020). The current trends of rainfall forecasting are more lenient towards the machine learning (ML) techniques due to their higher accuracy and reliability (Namitha et al., 2015). The following section discusses a few machine learning algorithms this research could apply to the rainfall data of Queensland.

According to (Arnav, 2019), there are three algorithms which can be used to predict the rainfall patterns: Support Vector Machine (SVM), Support Vector Regression (SVR), and K-nearest Neighbor (KNN). Liyew & Melese (2021), chose algorithms Multivariate Linear Regression (MLR), Random Forest (RF) and XGBoost gradient descent to experiment for the most suitable algorithm to forecast rainfall.

2.2.1 Multivariate linear regression (MLR)

There are two types of linear regression: MLR, in which several independent variables are used as input features, and simple linear regression, in which there is just one independent or input feature. One dependent variable in each linear regression can be forecasted or predicted using the input features. In order to predict the dependent variable, daily rainfall amount, multiple environmental variables or features can be used, which will result in MLR. The MLR algorithm shows the intensity of the relationship between the used environmental variables and the rainfall.

2.2.2 Random Forest (RF)

As a supervised machine learning algorithm, RF regression performs regression using the ensemble learning approach and using input variable and features to predict outputs. The way RF operates is to construct a number of decision trees during the training phase, with each tree's prediction being the mean of the classes. This algorithm has been identified to be generating accurate and powerful outputs (Ojo & Ogunjo, 2022). The RF algorithm has proven to be useful for large datasets with large proportions of missing data (Kusaik et al., 2013).

2.2.3 XGBoost Gradient Descent

The XGBoost Gradient Descent method seeks out the best tree model by using precise approximations. The supervised machine learning problem, which uses data from multiple features is applied to forecast a target variable using XGBoost. Due to the algorithm's speed and precision in

predicting outcomes, most authors use XGBoost for various regression and classification issues (Ojo & Ogunjo, 2022). XGBoost is a potent algorithm that generates a solid solution quickly, efficiently using memory, parallel learning and distributed computing to produce robust solutions (Arnav, 2019).

Through their study, (Liyew & Melese, 2021), established that the XGBoost Gradient Descent method yielded the best performance.

2.2.4 Time series projections using ML paradigms.

In addition to using traditional time series methods the inclusion of the “Classification and REgression Training (CARET)” and “Time Series forecasting using general regression neural network (tsfGRNN)” library packages within the R programming environment enables much more robust time series forecasts in which unsupervised learning is utilized for time series model fit and selection. It will be an approach that will be adopted during the research endeavor. Additionally, the ML paradigm of “remote sensing” presented in Dissanayake and Ruymgaart (2000) will be utilized for rainfall analysis if required.

3. METHODOLOGY

The primary objective of the study is to establish and check whether there is a LRD between time series and the rainfall data. For this purpose, the data needs to be analysed through the ACF and the PACF. The ACF will check the data set for seasonality. PACF will create a regression of the time series onto itself. The PACF

will be used to identify the lag after which the other lags are insignificant (Crato & Ray, 1996).

A time series process is categorized as long memory if the serial dependence or the ACF decays slower than an exponential decay implying a higher level of persistence over time. An exponentially decaying ACF can be categorized as having short memory. This shows that in LRD time series, the ACF will decay hyperbolically and that there is an evident correlation between two points regardless of their distance. If the time series is indeed LRD, the autocorrelations will remain high for a greater number of lags, which is referred to as a “fat tail”. On the other hand, if the time series has short memory, the decay will be much quicker with most of the autocorrelations being closer to zero following a few lags (Kirchgassner, Wolters, & Hassler, 2013). A slowly decaying PACF will further solidify the claim of an LRD in the time series. This would indicate that there is an influence on the value even at longer lags. The PACF concentrates on the immediate impact of each lag, and for time series with short memory, this impact rapidly decreases the further you are from the current time point.

It is important to keep in mind that identifying long memory in a time series is not always simple, and that findings may vary depending on the lag order chosen, the amount of data available, and other time series-specific properties. To validate the presence of long memory the time series, the researcher has applied more tests to examine and validate the results from the

ACF and PACF. The Ljung-Box test is one such test used by the researcher to confirm the type of memory of the time series.

The Ljung-Box test will assess whether the time series has a long memory or a short memory. In summary it will check whether the autocorrelations of the time series are significantly differs from zero at multiple lags. The Ljung-Box test checks whether there is serial correlation amongst the data points (Granger & Joyeux, 1980).

The Ljung-Box test's null hypothesis is that there is no serial correlation since all of the time series' autocorrelations are equal to zero. The alternative hypothesis is that there is serial correlation because at least one of the autocorrelations does not equal zero. If the test rejects the null hypothesis, then it has long memory, and if does not reject the null hypothesis of no autocorrelation, the time series is considered as a short memory.

H₀: The autocorrelations at all lags up to a certain lag are equal to zero, implying no serial correlation.

H_a: At least one of the autocorrelations at those lags is not equal to zero, suggesting the presence of serial correlation.

The test statistic (Q) will be computed through the squared value of autocorrelations. The test statistics for each lag will be calculated. The degree of freedom associated with the test statistics (Q) chi-squared distribution is equal to the number of lags being considered. The researcher will have to compare the test statistic to the critical value from the chi-squared distribution at a significance level (alpha) of 0.05 to see whether it differs substantially from zero. If Q is higher than

the critical value, the null hypothesis will be rejected.

After determining whether the time series has long memory or short memory, the researcher can fit the appropriate models from the models discussed above in the literature review to the time series to check for the best fit model for the time series. The residuals of each model after applying the time series needs to be analysed to select the best fit model. This method is taken forth to assure that the selected model will capture all the underlying patterns of the time series. Models which align with the memory type of the time series will be considered for the fitting. Models such as ARFIMA and GARMA will be looked at for long memory models while ETS, ARIMA and SARIMA are common models applied for time series lacking long memory.

The historical data will be used to estimate the model parameters. The historical data for the selected time series will be fitted into the different models. The residuals are plotted against time to identify whether there are any trends or patterns. In an ideal conventional time series, residuals will appear as white noise. This can be interpreted as the residuals being uncorrelated, with a constant mean and variance. The Ljung-Box test was also applied to check for the autocorrelation between the residuals.

The residual results from the model fits will be analysed to then determine the best fit model. The objective of the test to select the best fit model is to have residuals with no identifiable patterns or autocorrelation,

which are as close to white noise as is practical. The researcher has selected the resulting model as the one that fits the time series predictions the best. Moreover, the researcher has applied cross-validation to further validate the selected model. This helped in providing more confidence in the model's applicability and robustness.

Machine learning algorithms has been utilized to validate the model selection. The input data is transformed into a higher-dimensional feature space through a set of radial basis functions (RBFs). The network determines the Euclidean distance between each input pattern in the training dataset and the center of each RBF. The activations of the RBFs are then calculated using these distances. The outputs of the hidden layer, which are typically Gaussian functions, are the activations of the RBFs in the pattern layer. The activation of the appropriate RBF increases with the proximity of the input pattern to a center. The input pattern's proximity to the RBF centers determines the weights of the weighted outputs of the RBFs from the pattern layer, which are then added together in the summation layer. It might simply be the sum of the outputs from the summation layer weighted.

The researcher simply feeds the input data into the network while utilizing a trained ML model to produce a prediction. The network computes the RBFs' activations in the pattern layer, adds them in the summation layer, and outputs the prediction. These predictions will be matched with the forecasts from the traditional time series models to further validate the selection of the best fit model.

4. CONCLUSION

This paper has explored and summarised the methodology to be followed when applying LRD to a time series to forecast the rainfall in any given geographic area. The need for rainfall prediction and the applicability of time series forecasting for rainfall predictions were established through the introduction of the paper. The literature review summarises the various models available to predict rainfall through time series. The methodology section brings together the findings of the literature review culminating in the framework a researcher can adopt to forecast rainfall using time series data. The methodology section provides a step-by-step guide which is backed by the literature findings of this study.

REFERENCES

- Abbot, J., & Marohasy, J. (2013). The potential benefits of using artificial intelligence for monthly rainfall forecasting for the Bowen Basin, Queensland, Australia. *WATER RESOURCES MANAGEMENT 2013*, 171, pp. 66-75.
doi:<http://dx.doi.org/10.2495/WRM130261>
- Aks, D. J., & Sprott, J. C. (2003). The role of depth and 1/f dynamics in perceiving reversible figures. *Nonlinear Dynamics, Psychology, & Life Sciences*, 7, 161-180.
- Arnav, G. (2019). Kanchipuram Tamil Nadu: Rainfall prediction using machine learning. *International Journal of Innovative Science and Research Technology*, 56-58.
- Aswin, S., Geetha, P., & Vinayakumar, R. (2018). Deep learning models for the prediction of rainfall. *2018 International Conference on Communication and Signal Processing (ICCSPP)* (pp. 657-661). New York: IEEE.
- Bagirov, A. M., & Mahmood, A. (2018). A comparative assessment of models to predict monthly rainfall in Australia. *Water Resources Management*, 32(5), 1777-1794.
- Bagirov, A. M., Mahmood, A., & Barton, A. (2017). Prediction of Monthly Rainfall in Victoria Australia: Clusterwise Linear Regression Approach. *Atmospheric Research*, 20-29.
- Bhakar, S. R., Singh, R. V., Chhajed, N., & Bansal, A. K. (2006). Stochastic Modelling of Monthly Rainfall. *ARNP Journal of Engineering and Applied Sciences*, 1(3), 36-44.
- Bhardwaj, G., & Swanson, N. R. (2006). An Empirical Investigation of the Usefulness of ARFIMA Models for Predicting Macroeconomic and Financial Time Series. *Journal of Econometrics*, 131, 539-578.
doi:<https://doi.org/10.1016/j.jeconom.2005.01.016>
- Brodsky, J., & Hurvich, C. M. (1999). Multi-step forecasting for long-memory processes. *Journal of Forecasting*.
doi:[https://doi.org/10.1002/\(SICI\)1099-131X\(199901\)18:1%3C59::AID-FOR711%3E3.0.CO;2-V](https://doi.org/10.1002/(SICI)1099-131X(199901)18:1%3C59::AID-FOR711%3E3.0.CO;2-V)
- Chowdari, K. K., Girisha, R., & Gouda, K. C. (2015). A study of rainfall over India using data mining. *2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)* (pp. 44-47). New York: IEEE.
- Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2), 174-196. doi:doi:10.1093/jffinec/nbp001
- Delignieres, D., Ramdani, S., Lemoine, L., Torre, K., Fortes, M., & Ninot, G. (2006). Fractal analyses for "short" time series: A reassessment of classical methods. *Journal of Mathematical Psychology*, 50, 525-544.
- Dissanayake, G.S. and Ruyngaert, F.H., 2000. Speed of convergence for estimators of finite translation mixture distributions. *Journal of nonparametric statistics*, 12(5), pp.727-736.
- Dissanayake, G.S., 2016a. Rapid optimal lag order detection and parameter estimation of standard long memory time series. *Causal Inference in Econometrics*, pp.17-28.
- Dissanayake, G.S., 2016b. Advancement of fractionally differenced Gegenbauer processes with long memory. *Bulletin of the Australian Mathematical Society*, 94(1), pp.173-174.
- Dissanayake, G.S., 2023. Time-Series Analysis for the State-Space Model with R/Stan, *Journal of the Royal Statistical Society Series A: Statistics in Society*, 186(1), pp.164-165, <https://doi.org/10.1093/jrsssa/qnad002>
- Dissanayake, G. and Peiris, S., 2012. Generalized fractional processes with conditional heteroscedasticity. *Sri Lankan Journal of Applied Statistics*, 12.
- Dissanayake, G.S., Peiris, M.S., Proietti, T., Ruiz, I.R. and Garcia, G.R., 2014. Estimation of generalized fractionally differenced processes with conditionally heteroskedastic errors. In *International Work Conference on Time Series*, Proceedings ITISE (Vol. 2014, pp. 871-890).
- Dissanayake, G.S., Peiris, M.S. and Proietti, T., 2016. State space modeling of Gegenbauer processes

- with long memory. *Computational Statistics & Data Analysis*, 100, pp.115-130.
- Dissanayake, G.S., Peiris, M.S. and Proietti, T., 2018. Fractionally differenced Gegenbauer processes with long memory: A review. *Statistical Science*, pp.413-426.
- Dubey, A. D. (2015). Artificial Neural Network Models for Rainfall Prediction in Pondicherry. *International Journal of Computer Applications*, 120(3), 30-35.
- Eke, A., Herman, P., Kocsis, L., & Kozak, L. R. (2002). Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement*, 23, R1-R38.
- Eni, D., & Adyeye, F. J. (2015). Seasonal ARIMA modelling and forecasting of rainfall in Warri Town, Nigeria. *Journal of Geoscience and Environment Protection*, 3, 91-98.
- Feng, P., Wang, B., Liu, D. L., Fei, J., Niu, X., Ruan, H., . . . Yu, Q. (2020). Machine learning-based integration of large-scale climate drivers can improve the forecast of seasonal rainfall probability in Australia. *Environmental Research Letters*, 15(8). doi:10.1088/1748-9326/ab9e98
- French, M. N., Krajewski, W. F., & Cuykendall, R. R. (1992). Rainfall Forecasting in space and time using a neural network. *Journal of Hydrology*, 137, 1-31.
- Ghamariadyan, M., & Imteaz, M. A. (2021). Monthly Rainfall Forecasting Using Temperature and Climate Indices through a Hybrid Method in Queensland, Australia. *Journal of Hydrometeorology*, 1259–1273.
- Gil-Alana, L. A., Gupta, R., Sauci, L., & Carmona-González, N. (2022). Temperature and precipitation in the US states: long memory, persistence, and time trend. *Theoretical and Applied Climatology*, 150, 1731–1744.
- Granger, C. J., & Joyeux, R. (1980). AN INTRODUCTION TO LONG-MEMORY TIME SERIES MODELS AND FRACTIONAL DIFFERENCING. *Journal of Time Series Analysis*, 1(1), 15-29. doi:https://doi.org/10.1111/j.1467-9892.1980.tb00297.x
- Hosking, J. M. (1984). Modeling Persistence in Hydrological Time Series Using Fractional Brownian Differencing. *Water Resources Research*, 20, 1898-1908. doi:https://doi.org/10.1029/WR020i012p01898
- Hung, N. Q., Babel, M. S., Weesakul, S., & Tripathi, N. K. (2009). An Artificial Neural Network model for rainfall forecasting in Bangkok, Thailand. *Hydrological Earth System Sciences*, 13, 1413-1425.
- Hurst, H. E., Black, R. P., & Simaika, Y. M. (1965). Long-term storage, an experimental study. London: Constable.
- Kusaik, A., Verma, A. P., & Roz, E. (2013). Modeling and prediction of rainfall using radar reflectivity data: a data-mining approach. *IEEE Transaction on Geoscience and Remote Sensing*, 2337-42.
- Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, 8.
- Lloyd-Hughes, B., & Saunders, M. A. (2002). A Drought Climatology for Europe. *International Journal of Climatology*, 22, 1571-1592. doi:https://doi.org/10.1002/joc.846
- Madison, G. (2004). Fractal Modeling of Human Isochronous Serial Interval Production. *Biological Cybernetics*, 90, 105-112.
- Manandhar, S., Dev, S., Lee, Y. H., Meng, Y. S., & Winkler, S. (2019). A data-driven approach for accurate rainfall prediction. *IEEE Transactions on Geoscience and Remote Sensing*, 9323-31.
- Mandelbrot, B. B., & Wallis, J. R. (1969). Computer experiments with fractional Gaussian noises. *Water Resources Research*, 228-267.
- Mehrotra, R., Sharma, A., Kumar, D. N., & Reshmidevi, T. V. (2013). Assessing future rainfall projections using multiple GCMs and a multi-site stochastic downscaling model. *Journal of Hydrology*, 488, 84-100. doi:https://doi.org/10.1016/j.jhydrol.2013.02.046
- Mills, T. C. (2018). Towards a relational understanding of the regression line. *Australian Senior Mathematics Journal*, 32(1), 13-17.
- Momani, P. N. (2009). Time series analysis model for rainfall data in Jordan: Case Study for using time series analysis. *American Journal of Environmental Studies*, 5(5), 559-604.
- Namitha, K., Jayapriya, A., & SanthoshKumar, G. (2015). Rainfall prediction using artificial neural network on map-reduce framework. *ACM*.
- Oettli, P., & Camberlin, P. (2005). Influence of topography on monthly rainfall distribution over East Africa. *Climate Research*, 28(3), 199-212. doi:http://dx.doi.org/10.3354/cr028199
- Ojo, O. S., & Ogunjo, S. T. (2022). Machine learning models for prediction of rainfall over. *Scientific African*, 16.
- Silva, H.P.T.N., Dissanayake, G.S. and Peiris, T.S.G., 2019a. The Use of Fractionally Autoregressive Integrated Moving Average for the Rainfall Forecasting. In *Beyond Traditional Probabilistic Methods in Economics 2* (pp. 567-580). Springer International Publishing.
- Silva, H.P.T.N., Dissanayake, G.S. and Peiris, T.S.G., 2019b. Modeling Persistent and Periodic Weekly Rainfall in an Environment of an Emerging Sri Lankan Economy. In *Structural Changes and their Econometric Modeling 12* (pp. 314-328). Springer International Publishing.
- Stroe-Kunold, E., Stadnytska, T., Werner, J., & Braun, S. (2009). Estimating long-range dependence in time series: An evaluation of estimators implemented in R. *Behavior Research Methods*, 41(3), 909-923. doi:doi:10.3758/BRM.41.3.909
- Tharun, V. P., Prakash, R., & Devi, S. R. (2018). Prediction of Rainfall Using Data Mining Techniques. 2018 Second International Conference on Inventive Communication and

- Computational Technologies (ICICCT) (pp. 1507-1512). IEEE Xplore.
- Vijayan, R., Mareeswari, V., Mohankumar, P., Gunasekaran, G., & Srikar, K. (2020). Estimating rainfall prediction using machine learning techniques on a dataset. *International Journal of Science and Technology Research*, 440-5.
- Wang, S., Feng, J., & Liu, G. (2013). Application of seasonal time series model in the precipitation forecast. *Mathematical and Computer Modelling*, 58, 677-683.
- Westra, S., & Sharma, A. (2009). Probabilistic estimation of multivariate streamflow using independent component analysis and climate information. *Journal of Hydrometeorology*, 10(6), 1479-1492.
- Yusof, F., & Kane, I. L. (2012). Short memory or long memory: an empirical survey of daily rainfall data. *Hydrology and Earth System Sciences Discussions*, 9(10), 12271-12291. doi:<http://dx.doi.org/10.5194/hessd-9-12271-2012>
- Zainudin, S., Jasim, D. S., & Bakar, A. A. (2016). Comparative analysis of data mining techniques for Malaysian rainfall prediction. *International Journal on Advanced Science, Engineering and Information Technology*, 1148-53.
- Zakaria, S., Al-Ansari, N., Knutsson, S., & Al-Badrany, T. (2012). ARIMA models for weekly rainfall in the semi-arid Sinjar District at Iraq. *Journal of Earth Science and Geotechnical Engineering*, 2(3), 25-55.
- Zanial, W. C., Malek, M. A., Reba, M. M., Zaini, N., Ahmed, A. N., Sherif, M., & Elshafie, A. (2023). Rainfall-runoff modelling based on global climate model and tropical rainfall measuring mission (GCM -TRMM): A case study in Hulu Terengganu catchment, Malaysia. *Heliyon*, 9(5). doi:<https://doi.org/10.1016/j.heliyon.2023.e15740>
- Zeelan, B. A., Bhavana, N., Bhavya, P., & Sowmya, V. (2020). Rainfall prediction using machine learning & deep learning techniques. *International Conference on Electronics and Sustainable Communication Systems (ICESC 2020)* (pp. 92-97). Middlesex: IEEE Xplore.