Journal of Agriculture and Value Addition
December 2022, Vol. 5(2): 45-53
https://dx.doi.org/10.4038/ java.v5i2.47

RESEARCH ARTICLE

Appropriate conventional methods for estimating missing precipitation values in Sri Lanka

P.W. Jeewanthi[a]*, W. Wijesuriya[b], L.A.T.S. Liyanarachchi[c], L.H.P. Gunatathne[d]

[a]*Department of Export Agriculture, Uva Wellassa University, Badulla, 90 000, Sri Lanka*
[b]*Biometry Section, Rubber Research Institute of Sri Lanka, Dartonfield, Agalawatta,12 200, Sri Lanka*
[c]*Department of Soils and Plant Nutrition, Rubber Research Institute of Sri Lanka, Dartonfield, Agalawtta,12 200, Sri Lank*
[d]*Department of Agricultural Economics and Business Management, University of Peradeniya, 20 400, Sri Lanka*

Submitted: September 05, 2022; Revised: December 24, 2022; Accepted: December 26, 2022

*Correspondence: wasanap@uwu.ac.lk, ORCID: https://orcid.org/0000-0001-9023-3024

ABSTRACT

This study focused on finding the most appropriate conventional method for estimating missing precipitations values in different climate zones in Sri Lanka. The main location and surrounding locations were selected from each zone considering their data availability. Monthly precipitation data for each selected locations were collected from 1980 to 2020. Three temporal methods; Holt Winter's Exponential Smoothing, Seasonal Auto Regressive Integrated Moving Average and, Multiple Aggregation Prediction Algorithm and four spatial methods; namely, Arithmetic Mean Method (AMM), Normal Ratio Method (NR), Inverse Distance Weightage Method (IDW) and Aerial Precipitation Ratio (APR) method were used for estimating missing values. Root Mean Square Error, Mean Absolute Error, and Mean Absolute Percentage Error values were used for selecting the most appropriate method for estimating missing values for each location. It was difficult to name one method as the most appropriate method for all climate zones. Spatial methods have performed better than temporal methods in all seven climate zones. The missing values in the Low Country Wet Zone and Up Country Wet Zone could be estimated using AMM and NRM methods. APR method is the most suitable for estimating missing values in the Mid Country Wet Zone and Low Country Intermediate Zone. IDM method performed well in Mid Country Intermediate Zone, Up Country Intermediate Zones and Low Country Dry Zone. However, further improvement of both temporal and spatial methods for estimating missing values is necessary as none of these methods are capable of giving estimates with high accuracy.

Keywords*: Estimate, missing values, precipitation, spatial methods, temporal methods*

INTRODUCTION

Climate change and changes in the distribution of climate parameters are commonly discussed topics nowadays (Adedeji *et al.*, 2014). Many researchers and climatologists are working on forecasting, and recognizing the patterns of climate parameters as it is immensely helpful for planning agriculture and other activities with minimum risk (Basnayake, 2007; De Costa, 2008; Punyawardena *et al.*, 2013). Rainfall and temperature are key climate parameters that are commonly used for studying the changes in climate. In Sri Lanka, rainfall plays a major role in climate analysis as there is no much variation in temperature compared to the other countries in the world. Hence, rainfall analysis, including trends and changes in patterns, has become one of the most appealing research fields in recent years (Nisansala *et al.*, 2020).

The consistency and continuity of climate data are crucial in climate analyses such as trend analysis forecasting and studying changes in the behavior of climate parameters (Sattari *et al.*, 2014). Both consistency and continuity of data may be disturbed due to changes in observational procedure and incomplete records. Precipitation data are frequently incomplete due to many practical issues such as damaged measuring instruments, measurement error, changes of instruments over time, a change in the instrument site, and changes in data collectors. In the Sri Lankan context, the civil war was the main reason for missing records on climate parameters for a long period, especially in Northern and Eastern provinces. Moreover, these missing values and incomplete data series affect researchers' interest and create many difficulties while conducting climate studies with historical data. Hence, it is essential to fill in these missing observations before conducting research in order to maintain the continuity of the data series and also to increase the reliability of research findings.

There are different procedures and approaches to overcome the missing data problem. Estimating missing values using historical observations (Temporal methods) of the same location and estimating missing values using the information in surrounding stations (Spatial methods) are the two main categories that have been used commonly in climate research (Abbas *et al*, 2019; Sanusi *et al*, 2017).

Holt Winters Exponential Smoothing (ETS) method and stochastic time-series models are the commonly used temporal methods for estimating missing precipitation data. According to the literature, Seasonal Auto Regressive Moving Average (SARIMA) model is the most suitable time series model for capturing the variations of monthly precipitation values (Mishra *et al.*, 2007; Afrifa-Yamoah *et al,* 2016; Udumulla *et al,* 2018; Chandasekara *et al.*, 2021). Multiple Aggregation Prediction Algorithm (MAPA) (Kourentzes *et al*, 2014) is a novel method, which has been introduced recently for modeling time series data and there are many researches who have applied this method for modeling and forecasting time series data in different aspects (Wijesuriya *et al*, 2019; Petropoulos *et al*, 2021). However, there is no enough evidence on application of MAPA for estimating missing precipitation data. Hence, in this study MAPA technique was used as a novel temporal technique for estimating missing precipitation data in different climatic zones in Sri Lanka and the performance of MAPA technique was compared against the ETS and SARIMA methods.

Arithmetic Mean Method (AMM), Normal Ratio Method (NRM), Inverse Distance Weightage (IDW) method and Aerial Precipitation Ratio(APR) methods are the frequently used spatial techniques for estimating missing values with the information in surrounding rain gauges (De silva *et al*, 2007; Caldera *et al*, 2016). As per the literature, the performance of these methods are vary from region to region. Therefore, it is difficult to name one method as the best method for estimating missing precipitation data for all regions without conducting a proper study.

Even in Sri Lanka, few attempts have been taken for estimating missing precipitation data as it is an essential for having a complete data set in climatic research. De Silva *et al*. (2007) conducted a study to find the most appropriate method estimating missing precipitation data in major seven agro ecological regions in Sri Lanka. However, this study was focused only for spatial methods and conclude that it is difficult to name one method as the best for all agro ecological regions. Further, that study was suggest to validate their findings by further studies and the comparison between temporal and spatial methods were not address in this study. Another attempts on estimating missing precipitation data were record in 2016. Caldera *et al*. (2016) conducted the study only in upper catchment area of Badulu oya reservation and only spatial methods were used for the study. This research also emphasises the difficulty of naming one method as the best method for estimating missing precipitation data. At the same year, Wijemannage *et al*. (2016) studied the suitability of spatial methods on estimating missing precipitation data. Even though this study was conducted over Sri Lanka, the suitability of temporal methods was not addressed. As per the literature, it is evidenced that only few studies were carried out to study the suitability of different spatial methods but none of these studies were attempt on comparing the performances of temporal and spatial methods in a single study.

Under these circumstances, to fill this research gap this study was carried out for comparing the performance of both spatial and temporal methods on estimating missing precipitation data as well as selecting the most appropriate conventional method for estimating missing precipitation data for all climate zones in Sri Lanka.

MATERIALS AND METHODS

Selected locations for the study

The locations were selected to represent all seven climate zones in Sri Lanka, namely Low-Country Wet Zone (WL), Mid-Country Wet Zone(WM), Up-Country Wet Zone (WU), Low-Country Intermediate Zone (IL), Mid-Country Intermediate Zone(IM), Up-Country Intermediate Zone (IU) and Low-Country Dry Zone (DL). The main location and three or four surrounding locations for each climate zone were selected considering the availability and completeness of data (Table 1). Monthly precipitation data for all selected rain gauge stations were collected from the Department of Meteorology, Sri Lanka for the period, 1980 to 2020.

Table 1: Selected locations for the study.

Agro ecological zone	Primary station	Surrounding stations	Latitude (N)	Longitude (E)
Low Country Wet Zone (WL)	Rathnapura (6.68°, 80.40°)	Anhettiya	6.93°	80.37°
		Galatur	6.70°	80.28°
		Balangoda	6.65°	80.70°
		Aupolla	6.72°	80.58°
Mid Country Wet Zone (WM)	Kandy (7.33°, 80.63°)	Kobonella	7.35°	80.85°
		Kandy king's pavillion	7.30°	80.63°
Up Country Wet Zone (WU)	Kotagala (6.92°, 80.63°)	Katukithula	7.08°	80.68°
		Ginigathhena	7.00°	80.48°
		Nuwara Eliya	6.97°	80.77°
Low Country Intermediate Zone (IL)	Kurunegala (7.47°, 80.35°)	Delwita	7.53°	80.52°
		Egodagama	7.43°	80.42°
		Mellawa	7.32°	79.95°
		Nikaweratiya	7.75°	80.12°
Mid Country Intermediate Zone (IM)	Spring valley (6.91°, 81.10°)	Badulla	6.99°	81.05°
		Telbedda	6.94°	81.12°
		Ury	6.97°	81.08°
Up Country Intermediate Zone (IU)	Welimada (6.90°, 80.90°)	Haputhale	6.77°	80.93°
		Bandarawela	6.82°	80.99°
		Ella	6.86°	81.04°
Low Country Dry Zone (DL)	Ambalantota (6.12°, 81.02°)	Yala	6.37°	81.53°
		Hambantota	6.12°	81.13°
		Agunakolapelessa	6.17°	80.89°
		Kirama	6.22°	80.67°

Missing value estimation methods

Three (03) temporal methods and four (04) spatial methods were used for estimating missing precipitation data and R statistical package (4.1.0), Microsoft Excel and Arc GIS ver. 10.2 software (ESRI,2013) packages were used for the analyses. Holt winters method, Seasonal Auto Regressive Movig Average (SARIMA), Multiple Aggregation Prediction Algorithm (MAPA) methods were used as temporal methods and Arithmetic mean method (Chow *et al.*, 1988), Normal Ratio Method (Singh, 1994), Inverse distance weighting method (Lam, 1993) and Aerial Precipitation Ratio (APR) were used as spatial methods.

Evaluation procedure to compare efficiencies of different methods

The years from 2001 to 2020 were selected for random selection for the deletion of precipitation values for the selected locations. Once the location was selected, three random precipitation values per month were selected randomly for deletion. After deleting the first selected value, the estimation of the missing value was done employing the selected methods and the deleted (original) values and the estimated values were recorded. The measurement errors were calculated and repeated three times for each month with replacement. The same procedure was repeated for all twelve months as well as for all climate zones (Figure 1).

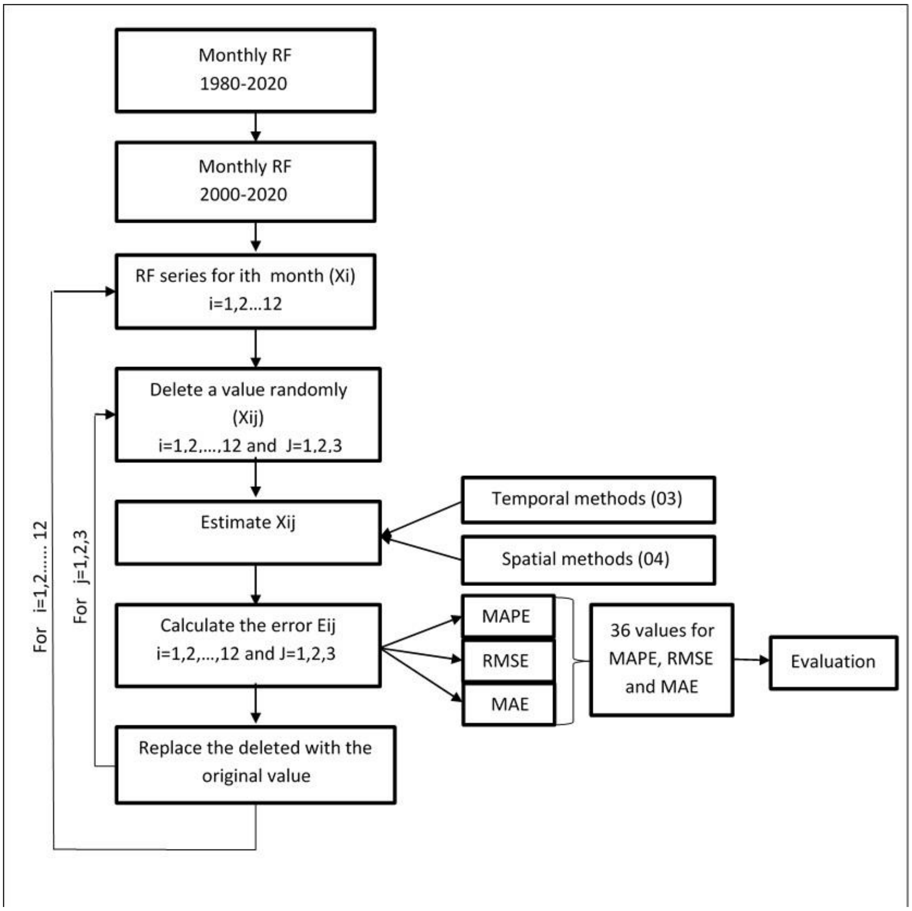

Figure 1: Flow chart of the estimation and evaluation process for a selected location

Selection of the most appropriate method for estimating missing values

Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) were used for selecting the most appropriate missing value estimating method among the methods mentioned above. The errors were calculated based on the estimated values and the actual values. The accuracy of estimates was discussed further using the classification of MAPE values (Lewis, 1982).

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100\% \qquad (01)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(A_t - F_t)^2}{n}} \qquad (02)$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} | A_t - F_t | \qquad (03)$$

where;
A_t = Actual precipitation value at time t
F_t = Estimated precipitation value at time t
n = number missing observations being estimated

RESULTS AND DISCUSSION

SARIMA and ETS methods are the commonly used temporal methods for estimating missing precipitation values. Performance of ETS method is poor compared to the SARIMA and MAPA model. Even though MAPA method is not used for estimating missing values, results revealed that the performance of MAPA is better than SARIMA in many locations except in WU and DL. SARIMA model is the most appropriate temporal method in WU and DL.

AMM, NRM, and IDW are commonly used for estimating missing precipitation under spatial methods. APR method was developed after making a few adjustments for IDW method. The results imply that the performance of these methods varies depending on the geographical location. All these methods required the precipitation data in surrounding locations while IDW and APR methods need some additional information such as distance and area covered by each location.

However, the AMM method is designed to estimate missing values when the annual precipitation of surrounding stations is within the range of 10% of the normal precipitation at the station which is going to be estimated and the NRM method is better when it exceeds the range of 10%. These two methods record the minimum error rate compared to other methods in WL and WU. Thus, either AMM or NRM method can be used for estimating missing data in WL and WU after considering whether the precipitations are within 10% or not.

Both IDM and APR methods perform alike in estimating missing precipitation values in all climate zones except WL and IU. However, based on the MAE, MAPE and RMSE, IDM method performed well in IM, and DL zones, while APR methods record the lowest error in WM, IU and IL zone. Nevertheless, neither MAPA nor SARIMA method is capable enough to compete with other spatial methods which are used for this study in estimating the missing values (Table 2).

Table 2: Performance of different methods in estimating missing precipitation data.

Method	Measurements of error	WL	WM	WU	IL	IM	IU	DL
MAPA	MAPE	34.05	94.47	67.31	175.18	98.52	218.80	108.82
	RMSE	71.82	67.90	149.47	92.09	65.24	62.31	55.63
	MAE	64.01	54.89	111.06	82.24	52.31	47.96	100.83
SARIMA	MAPE	36.07	124.66	56.32	230.29	132.52	334.22	107.19
	RMSE	75.63	78.11	144.76	112.99	76.32	88.60	51.68
	MAE	62.84	62.28	101.50	101.51	65.45	53.63	86.75
Holt Winter's method	MAPE	82.00	223.10	160.20	289.00	220.90	162.48	287.48
	RMSE	180.76	120.64	225.40	171.65	126.50	96.96	82.96
	MAE	134.00	90.60	146.70	115.50	90.90	66.00	60.52
AMM	MAPE	21.94	61.53	79.59	61.71	88.52	207.66	78.75
	RMSE	104.13	84.44	107.26	74.18	45.81	92.40	52.79
	MAE	67.27	51.69	86.30	46.90	34.26	69.55	36.54
NRM	MAPE	27.08	37.48	73.57	79.02	98.40	99.94	40.66
	RMSE	118.11	47.45	94.39	74.44	60.19	36.63	28.22
	MAE	81.38	33.43	77.35	52.18	44.53	27.65	19.69
IDM	MAPE	37.47	23.41	96.81	45.09	54.72	181.19	99.58
	RMSE	112.09	28.04	117.26	79.20	51.71	80.28	56.87
	MAE	70.32	20.98	137.06	48.04	37.08	60.10	39.59
APR	MAPE	45.13	23.62	78.45	52.45	44.38	108.57	96.88
	RMSE	122.01	27.94	116.19	80.25	53.27	49.91	59.73
	MAE	78.58	21.01	91.56	50.11	38.01	38.35	42.24

CONCLUSIONS

A single method cannot be designated as the most appropriate method for estimating missing precipitation values for all climate zones. The performance of spatial methods is better than the temporal methods in estimating missing precipitation values. AMM/NRM method is good for estimating missing values in WL and WU. APR method is most appropriate for estimating missing values in WM, IU and IL. Furthermore, IDM method performed well in IM and DL. However, both temporal and spatial methods should be improved further as none of these conventional methods are capable to estimate missing values with high accuracy.

ACKNOWLEGEMENT

The authors are thankful to the National Science Foundation for the financial assistance (NRC/15/071) given for purchasing data.

REFERENCES

Abbas, A.A., Ali, N.A., Abozeid, G. and Mohamed, H.I. (2019). Comparison of Different Methods for Estimating Missing Monthly Rainfall Data. In Twenty-Second International Water Technology Conference, IWTC22.

Adedeji, O., Reuben, O. and Olatoye, O. (2014) Global Climate Change. J. geosci. environ. 2, 114-122. DOI: 10.4236/gep.2014.22016.

Afrifa-Yamoah E. , Bashiru I. I. Saeed , Azumah Karim , Sarima Modelling and Forecasting of Monthly Rainfall in the Brong Ahafo Region of Ghana, World Environment, Vol. 6 No. 1, 2016, pp. 1-9. DOI: 10.5923/j.env.20160601.01.

Basnayake, B.R.S.B. (2007) Climate Change. In: The National Atlas of Sri Lanka. Survey Department of Sri Lanka, Colombo, Sri Lanka. pp 54-55.

Caldera, H.P.G.M., Piyathisse, V.R.P.C. and Nandalal, K.D.W., 2016. A Comparison of Methods of Estimating Missing Daily Rainfall Data. Engineer J. Inst. Eng. (Sri Lanka) 49(4), pp.1–8. DOI:http://doi.org/10.4038/engineer.v49i4.7232

Chow, V.T., Maidment, D.R. and Larry, W. (1988). Mays. Applied. Hydrol. International edition, MacGraw-Hill, Inc, 149.

De Costa, W.A.J.M. (2008) Climate change in Sri Lanka: myth or reality? Evidence from longterm meteorological data. J. Natl. Sci. Found. J Natl Sci Found., 36, 63–88. DOI: http://doi.org/10.4038/jnsfsr.v36i0.8048

De Silva, R.P., Dayawansa, N.D.K. and Ratnasiri, M.D. (2007). A comparison of methods used in estimating missing rainfall data. The J. Agri. Sci. 3(2), 101-108.

ESRI 2013.ArcGIS Desktop: Released 10.2 Redlands, CA: Environmental System Research Institute, USA.

Kourentzes, N., Petropoulos, F. and Trapero, J.R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. Int. J. Forecast. 30(2), 291-302.

Lam, N.S.N. (1983). Spatial interpolation methods: a review. J. Am. Cartogr., 10(2), 129-150.

Lewis, C. (1982). Industrial and business forecasting methods. London: Butterworth Scientific.

Nisansala, W.D.S., Abeysinghe, N.S., Islam, A. and Bandara, A.M.K.R. (2020). Recent rainfall trend over Sri Lanka (1987-2017). Int. J. Climatol. 40, 3417-3435. DOI: 10.1002/joc.6405

Petropoulos, F. and Spiliotis, E. (2021). The Wisdom of the Data: Getting the Most Out of Univariate Time Series Forecasting. Forecasting. 3(3), 478–497. DOI:10.3390/forecast3030029

Punyawardena, B.V.R., T. Dissanaike and A. Mallawatantri (2013) Spatial variation climate change induced vulnerability in Sri Lanka.; An analysis of the components of vulnerability at district level. Published by the Department of Agriculture in collaboration with UNDP/SGP, Sri Lanka

Sanusi, W., Wan Zin, W.Z., Mulbar, U., Danial, M. and Side, S. (2017). Comparison of the methods to estimate missing values in monthly precipitation data. Int. J. Adv. Sci. Eng. Inf. Technol. 7(6), 2168-2174

Sattari, M.T., Rezazadeh-Joudi, A. and Kusiak, A. (2017). Assessment of different methods for estimation of missing data in precipitation studies. J. Hydrol. Res. 48(4), 1032-1044. DOI:10.2166/nh.2016.364

Singh V P. (1994). Elementary Hydrology. Prentice Hall of India: New Delhi.

Udumulla, N., Madhubhashini, A. Hakmanage, N., Sandaruwani, I. and Rathnayaka, R.M.K. (2018). Time-series analysis of monthly rainfall data for the Nuwara Eliya District, Sri Lanka.

Wijemannage, A. L. K., Ranagalage, M., and Perera, E. N. C. (2016). Comparison of spatial interpolation methods for rainfall data over Sri Lanka. ACRS Proceedings.

Wijesuriya, W., Herath, H.M.L.K., lshani, P.G.N. and Sankalpa, J.K.S. (2019) 'Forecasting of Rubber Prices in the Sri Lankan market using Multiple Temporal Aggregation (MTA)', International Rubber Conference 2019, Myanmar.