The Use of Fractionally Autoregressive Integrated Moving Average for the Rainfall Forecasting

H. P. T. N. Silva[1](✉), G. S. Dissanayake[2], and T. S. G. Peiris[3]

[1] Department of Social Statistics, Faculty of Humanities and Social Science, University of Sri Jayewardenepura, Nugegoda, Sri Lanka
thanuja@sjp.ac.lk
[2] University of Sydney, Sydney, Australia
[3] Department of Mathematics, Faculty of Engineering, University of Moratuwa, Moratuwa, Sri Lanka

Abstract. A study of rainfall pattern and its variability in South Asian countries is vital as those regions are frequently vulnerable to climate change. Models for rainfall have been developed with different degrees of accuracy, since this key climatic variable is of importance at local and global level. This study investigates the rainfall behaviour using the long memory approach. Since the observed series consists of an unbounded spectral density at zero frequency, a fractionally integrated auto regressive model (ARFIMA) is fitted to explore the pattern and characteristics of the weekly rainfall in the city of Colombo. The maximum likelihood estimation (MLE) method was utilized to obtain estimates for model parameters. To evaluate the suitability of the method for parameter estimation, a Monte Carlo simulation was done with various fractionally differenced parameter values. Model selection was done based on the minimum of the mean absolute error and validated by the forecasting performance that was evaluated using an independent sample. The experimental result yielded a good prediction accuracy with a best fitted long range dependency model and a coverage probability of 95% in terms of prediction intervals that resulted in closer nominal coverage.

Keywords: Rainfall · Fractional differencing · Long-memory
Maximum likelihood estimators · Forecasting

1 Introduction

Modelling rainfall is a challenging task for researchers due to the high degree of uncertainty in atmospheric behaviour. Observational evidence indicates that the climate change has significantly affected global community at a different level. Climate vulnerabilities are expected to be critical in Sri Lanka in the various sectors as agriculture, fisheries, water, health, urban development, human settlement, economic infrastructure, biodiversity and ecosystem in the country [22].

V. Kreinovich et al. (Eds.): ECONVN 2019, SCI 809, pp. 567–580, 2019.
https://doi.org/10.1007/978-3-030-04200-4_40

Information on key climatic variable predictions allow to various stakeholders to prompt themselves for action in order to reduce adverse impacts and enhance positive effects of climatic variation. Rainfall is the one of the most important climatic variable to tropical country like Sri Lanka and this is the variable which give erratic variation at any time in the country. Sri Lanka receives rainfall during the year, with a mean annual rainfall varying from 900 mm in the dry zone to over 5000 mm in the wet zone. Annual rainfall pattern in many parts of Sri Lanka are bimodal and predominantly governed by a seasonally varying monsoon system. Sri Lanka needs to address climate change adaptation to ensure the economic development by the careful investigating of the information on rainfall pattern and its variability which resulting from the predictions of the best fitted rainfall models in various regions. Rainfall analysis is not only important for agricultural areas but also for the urban areas since those areas engage with many activities such as construction, industrial planning, urban traffic, sewer systems, health, rainwater harvesting and climate monitoring. Rainfall is the main source of the hydrological cycle and provides practical benefits through its analysis. Thus, modelling rainfall is one of the key requirements in the country, some of the researchers made attempt to analyse weekly rainfall in Sri Lanka using percentile bootstrap approach to identify the extreme rainfall events [24]. Another study was carried out by the Silva and Peiris [25] to identify the most likelihood time period to form the extreme rainfall events during the South west moons time span by fitting best probability distribution for the weekly rainfall percentiles. Since the Sri Lanka is a developing country which hasn't high technology to sensitive to some important climatic information with related to rainfall is one of the reason cause to low prediction accuracy. However, researchers made effort to model rainfall of the country with increasing degree of accuracy using different techniques. Silva and Peiris [26] discussed problems faced in modelling rainfall which showed positive skewed distribution with longer tail to the right. Rainfall is one of the most difficult variables of the hydrological cycle to understand and model due to its high variability in both space and time [13]. However, several modelling strategies have been applied for the forecasting of rainfall in different areas all over the world. Box-Jenkins autoregressive integrated moving average (ARIMA) model has been widely used for rainfall modelling ([11,20,29,30]). Some of the researchers have made attempts to model rainfall using artificial neural networks ([10,18]). However, very few studies on rainfall in context of long memory can be identified in literature. Granger and Joyeux [15] and Hosking [17] initially proposed a long memory class of models, known as the fractionally integrated autoregressive moving average (ARFIMA) process for stochastic processes. The model defined as ARFIMA (p, d, q) allows the parameter "d" to take fractional values for differencing. There is a fundamental change in the correlation structure of the ARFIMA model, when compared with the correlation structure of the conventional ARIMA model ([6]). According to Granger and Joyeux [15], the slowly decaying autocorrelation exhibited in long range dependency or long memory models differ from stationary ARIMA models that decay exponentially. Many researchers proposed different methods to estimate the fractional differencing parameter. Porter-Hudak and Geweke [14] proposed a method for estimating the long memory differencing parameters based on a simple linear regression of the log periodogram. An approximate maximum likelihood method for parameter "d" was proposed by Fox and Taqqu [12]. An exact maximum likelihood estimation method for differencing parameter was introduced by Sowell [27]. Chen et al. [6] developed a regression type estimator of "d" using lag window spectral density estimators. Number of studies were carried out by comparing various properties of the ARFIMA model based on the estimation method used for the fractionally differencing parameter. (See [2,3,7,16,23]). Dissanayake [9] established a methodology to find an optimal lag order of a standard long memory ARFIMA series within a short process time duration and applied the theory to Nile river data.

Though short memory models have been developed for rainfall still there is a noticeable gap modeling persistent rainfall in view of long memory. The main goal of this study is to fit an ARFIMA model for a weekly rainfall data series in the city of Colombo by capturing the long range dependency features. The paper outline is shaped as follows. In Sect. 2, the long memory ARFIMA model is introduced and some properties of the model are discussed. The model parameter estimation procedure is also described within the section. The results of the Monte Carlo simulation which was used to evaluate the suitability and reliability of the parameter estimation procedure is presented in Sect. 3. Section 4 provides brief details on prediction intervals for forecasting values relevant to the utilized series. The results of weekly rainfall modelling are presented in Sect. 5. Final section, comprises of the conclusion and proposed suggestions.

2 ARFIMA Long Range Dependency Model

ARFIMA is a natural extension of the Box and Jenkins model with non-integer values assigned for d. The ARFIMA (p, d, q) model of a process $\{Y_t\}_{t\in Z}$ is given by the formula

$$\phi(B)\nabla^d(Y_t - \mu) = \psi(B)\varepsilon_t \tag{1}$$

Where μ is the mean of the process, $\{\epsilon_t\}_{t\in Z}$ is a white noise process with zero mean and variance σ_ϵ^2. B is the backward shift operator such that $y_{t-n} = B^n y_t$, $\phi(B)$ and $\theta(B)$ are autoregressive and moving average polynomials of order p and q respectively.

$$\phi(B) = \sum_{i=1}^{p} \phi_i B^i \quad 1 \leq i \leq p \tag{2}$$

$$\psi(B) = \sum_{j=1}^{P} \psi_j B^j \quad 1 \leq j \leq q \tag{3}$$

where d is called as the long memory parameter and differencing operator ∇^d is defined as,

$$\nabla^d = (1-B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \tag{4}$$

Where $\binom{d}{k} = \frac{\Gamma(1+d)}{\Gamma(1+k)\Gamma(1+d-k)}$.

If $d > -0.5$ then the process is invertible and if $d < 0.5$ then the process is stationary. Therefore $d \in (-\frac{1}{2}, \frac{1}{2})$ shows that the process is stationary and invertible. The spectral density function of the $\{Y_t\}_{t \in Z}$ is $f(\omega)$ that can be written as

$$f(\omega) = (2sin\frac{\omega}{2})^{-2d} \quad 0 < \omega \leq \pi \tag{5}$$

$f(\omega) \approx \omega^{-2d} \quad \omega \to 0$

The spectral density function $f(\omega)$ is unbounded when the frequency is near zero. Also, the autocovariance function and correlation function of the process can be expressed as follows

$$\gamma_k = \frac{(-1)^k(-2d)!}{(k-d)!(-k-d)!} \tag{6}$$

$$\rho_k = \frac{d(1+d)...(k-1+d)}{(1-d)(2-d)(3-d)...(k-d)}(k = 1, 2, 3, 4...) \tag{7}$$

Hosking (1981) showed that the auto correlation function of the process satisfies the expression $\rho_k \approx k^{2d-1}$ when $0 < d < 1/2$. Thus, the autocorrelation of the ARFIMA process decays hyperbolically to zero as $k \to \infty$ and in contrast, the auto correlation function of the ARIMA process has a exponential decay. The process with $d = 0$ reduces to a short memory ARMA model.

Let Z denote a series of "n" observations with mean μ and variance σ_Z^2. If the decay parameter is considered as α, then the natural fractional differencing parameter "d" can be written as $d = (1 - \alpha)/2$.

The log likelihood function of the Exact Gaussian can be written as

$$l(\alpha, \sigma_y^2) = -\frac{1}{2}(log \quad det(\Gamma_n) + Z'\Gamma_n^{-1}Z') \tag{8}$$

The arfima package (See [28]) in R optimized the log likelihood function and obtained the exact maximum likelihood estimators. Two algorithms namely Durbin-Levinson and Trench algorithms were utilized to maximize the likelihood and obtain optimal simulation and forecasting results.

3 Result of the Monte Carlo Simulation

A number of Monte Carlo experiments were carried out to evaluate the performance of the maximum likelihood method used for parameter estimation. The simulation was done based on various fractional differencing parameter values with 1000 replications. The four different series lengths ($n = 100$, $n = 200$, $n = 500$ and $n = 1000$) were considered for the simulation. The simulation results provided fractionally differenced parameter estimates and corresponding standard and mean square errors. Monte Carlo experiment was conducted on a simulated ARFIMA(0,d,0) series with parameter values: $d = 0.1$, $d = 0.15$, $d = 0.3$ and $d = 0.45$.

The simulation was carried out using the R programming Language (Version 3.4.2) utilizing a HP11(8 GB, 64 bit) computer. The standard errors of the estimates $SD(\hat{d})$ and mean square error of the estimates $MSE(\hat{d})$ can be expressed as;

$$SD(\hat{d}) = \sqrt{\sum_{r=1}^{R}(\hat{d}_r - \hat{d})/R} \qquad MSE(\hat{d}) = \sum_{r=1}^{R}(\hat{d}_r - d)^2/R$$

Where $\hat{d}_r$ is the MLE of d for the r^{th} replication. The value R denotes the number of replications ($R = 1000$ for all tabulated simulation results of this paper). Tables 1, 2, 3 and 4 present the average of the estimated d, corresponding standard error and MSE of the estimator.

According to the results in Tables 1, 2, 3 and 4, the performance of the maximum likelihood estimator is reasonably accurate. It can be clearly seen that the parameter bias has decreased with the increase in sample size. Furthermore,

Table 1. MLE of d for a generating process of ARFIMA(0,d,0) with d = 0.1. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	SD($\hat{d}$)	MSE($\hat{d}$)
100	0.0517	0.0912	0.0106
200	0.0748	0.0626	0.0045
500	0.0885	0.0367	0.0014
1000	0.0949	0.0254	0.0006

Table 2. MLE of d for a generating process of ARFIMA(0,d,0) with d = 0.15. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	SD($\hat{d}$)	MSE($\hat{d}$)
100	0.1048	0.0915	0.0104
200	0.1265	0.0593	0.0040
500	0.1408	0.0367	0.0014
1000	0.1456	0.0254	0.0006

Table 3. MLE of d for a generating process of ARFIMA(0,d,0) with d = 0.3. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	SD($\hat{d}$)	MSE($\hat{d}$)
100	0.2493	0.0877	0.0102
200	0.2726	0.0575	0.0040
500	0.2892	0.0362	0.0014
1000	0.2947	0.0251	0.0006

Table 4. MLE of d for a generating process of ARFIMA(0,d,0) with d = 0.45. The results are based on 1000 Monte Carlo replications

n	$\hat{d}$	SD($\hat{d}$)	MSE($\hat{d}$)
100	0.3774	0.0695	0.0101
200	0.4079	0.0477	0.0040
500	0.4310	0.0314	0.0013
1000	0.4435	0.0270	0.0007

the results provide evidence that the parameters become consistent with the increase in series length. As we expected the standard deviation and the MSE of the estimators have decreased with the increase in series length.

4 Forecast and Prediction Intervals

Forecasts are obtained based on the best fitted long memory model. However, predicting of future values along with their prediction intervals become more beneficial in long memory time series analysis. The lower (L) and upper (U) boundaries covering the forecast values with known probability are simply called prediction intervals of the form [L, U]. A detailed review of approaches in calculating interval forecast using time series was described in Chatfield [5]. Charles et al. [4] made an effort to make prediction intervals to forecast US core inflation values that provided a unique fractional model. Prediction intervals were utilized to forecast tourism demand by Chu [8]. Zhou et al. [31] suggested a prediction interval method to predict aggregates of future values derived from a long memory model. A new bootstrap method for autoregressive models was proposed by Hwang and Shin [19]. Ali et al. [1] suggested a Sieve bootstrap approach to construct intervals for a long memory model. Prediction interval approach was utilized to measure the uncertainty about long-run predictions by Muller and Watson [21].

5 Application

Sri Lanka is a tropical country in South Asian region located at the latitudes of 5° 55 N and 9° 51 N and the longitudes of 79° 41 E and 81° 53 E with an area of 65610 km^2 and the Colombo city is the commercial capital of Sri Lanka. Daily rainfall data of Colombo were collected from 1990 to 2015 from the Department of Meteorology, Sri Lanka for this analysis. The daily rainfall (mm) data has been converted into weekly rainfall by dividing a year into 52 weeks such that week 1 corresponds to 1–7 January, Week 2 corresponds to 8–14 January and so on. The data during the time span from 1990 to 2014 was used to build the model while the rest was used for model validation. To examine the temporal

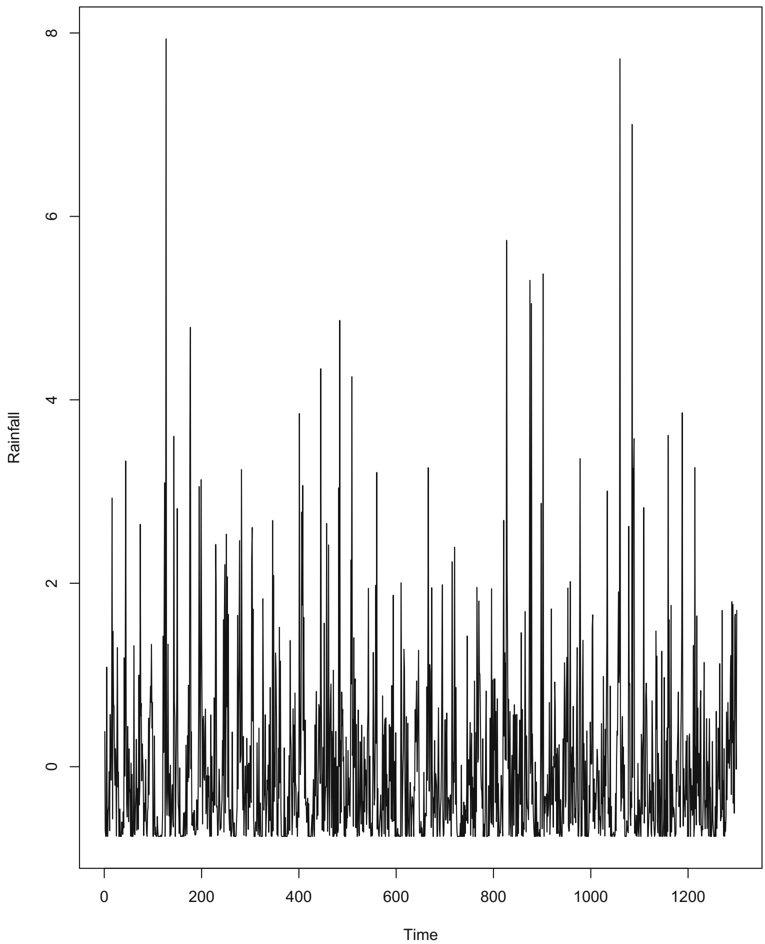

Fig. 1. Time series plot of weekly rainfall series from 1990 to 2014

variability of the rainfall series, time series plots were taken and presented in Fig. 1.

The time series plot explores the random behaviour of weekly rainfall during the considered time span from 1990 to 2014. In order to identify the correlation structure of the observed series, the autocorrelation and partial auto correlation plots were taken and those results are shown in Figs. 2 and 3 respectively.

In order to study the long memory features of the weekly rainfall series, the periodgram was obtained and presented in Fig. 4. The maximum spectrum density is 0.0385185 given at a frequency which is very close to zero. Based on those characteristics the series displays long memory. Thus, we conclude that the ARFIMA standard long memory model may be suitable for the observed weekly rainfall series. Long range correlation of observed data were considered

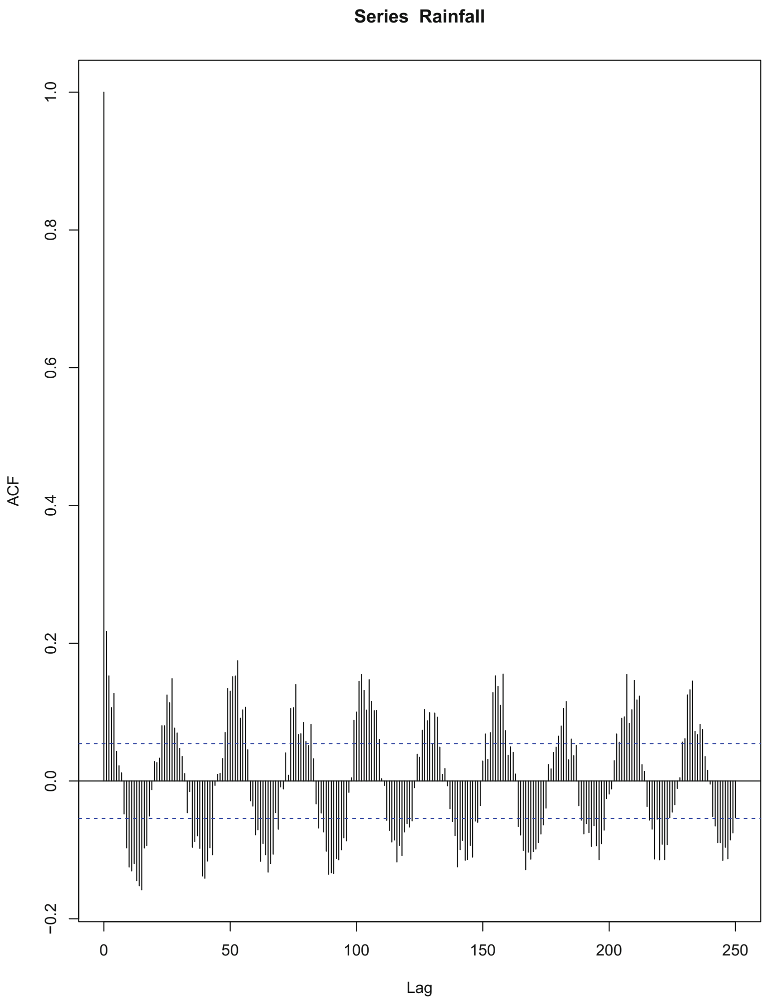

Fig. 2. Autocorrelation plot of the series from 1990 to 2014

in long memory modelling. Various ARFIMA models were fitted for the data set that vary from 1990 to 2014 (series length = 1300).

Those fitted models were employed to predict the weekly rainfall during the time span from 2014 to 2015 and best fitted model is selected with the minimum mean absolute error (MAE). The MAE can be written as,

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_i|$$

Where e_i is the forecasting error and n is the length of the forecasting series.

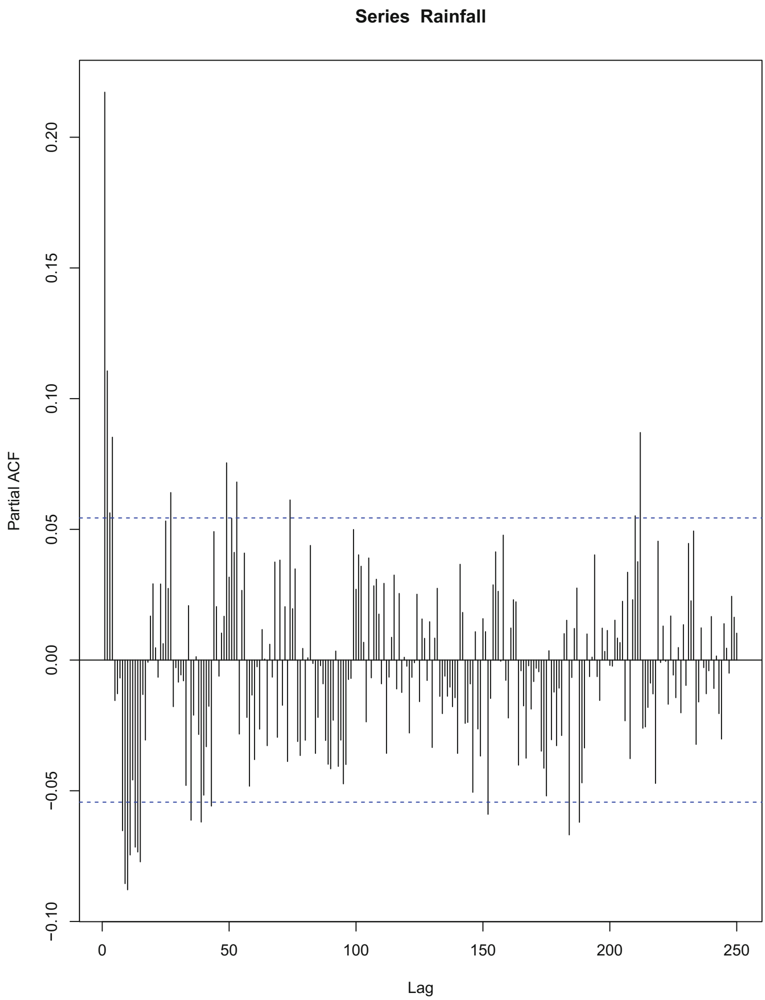

Fig. 3. Partial autocorrelation plot of the series from 1990 to 2014

The best fitted model and the corresponding parameter estimates are presented in Table 5. The ARFIMA (4,0,4) model was found to be the best fit for the weekly rainfall series returning the smallest MAE.

All model parameters except the constant are significant at the 0.05 level of significance. The residual analysis of the fitted model was performed and found the uncorrelated at a 5% level of significance. Furthermore, the model was tested for weekly rainfall data in 2015 and the result is presented in Table 6. Figure 5 illustrates the weekly rainfall over the year 2015 along with the predicted estimates.

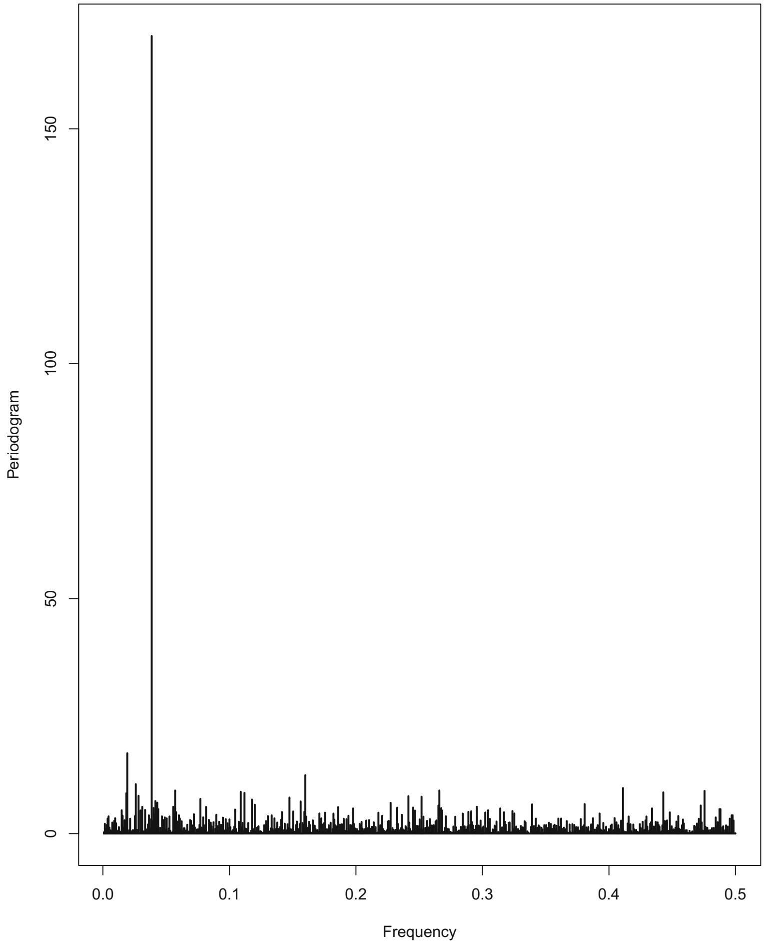

Fig. 4. The periodgram of the rainfall series from 1990 to 2014

Table 5. Fitted model for the weekly rainfall series ARFIMA (4,0,4) with p = 4, q = 4, d = 0.05792421

Coefficients	ϕ_1	ϕ_2	ϕ_3	ϕ_4	θ_1
Estimate	1.2059	−0.2493	0.5765	−0.6752	1.1243
Standard error	0.0242	0.0454	6.324e-07	6.324e-07	0.0231-Correct value (CV)
Z-value	4.9768e01	5.4903	9.1153e05	−1.0676e06	4.8638e01
Pr(>|Z|)	0.0000	0.0005	0.0000	0.0000	0.0000

Table 5. (*continued*)

Coefficients	θ_2	θ_3	θ_4	Constant	d
Estimate	−0.1131	0.5220	−0.6743	−0.0163	0.0579
Standard error	0.0365 (CV)	0.0354 (CV)	0.0215	0.0380	0.0276
Z-value	−3.0992	1.4735e01	−3.1363e01	−4.2907e-01	2.0950
Pr(>|Z|)	0.0019	0.0000	0.0000	0.6678	0.0361

Table 6. Absolute Forecast Error for independent sample (2015)

Absolute forecasting error in mm	ARFIMA number of weeks percentage
0–10	10(19.2)
11–15	6(11.5)
16–20	6(11.5)
21–25	4(7.7)
26–30	6(11.5)
31–35	1(1.9)
36–40	4(7.7)
41–45	1(1.9)
46–50	2(4.0)
More than 50	12(23.1)

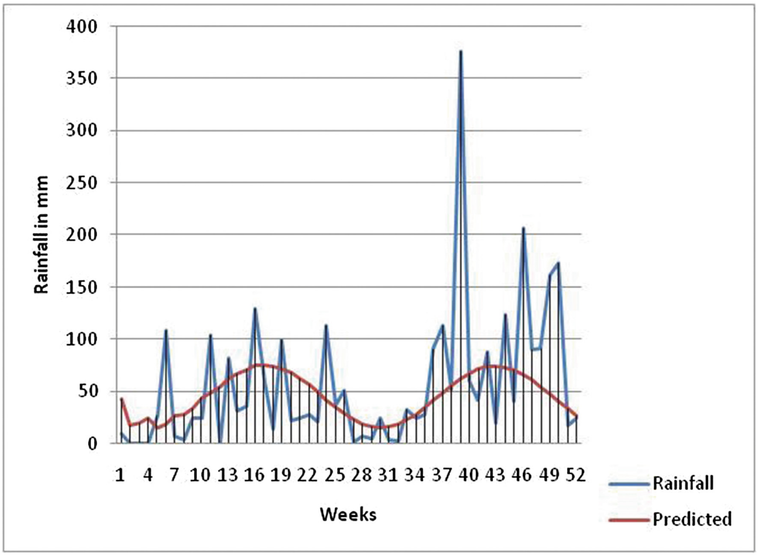

Fig. 5. Forecasted and actual weekly rainfall in 2015

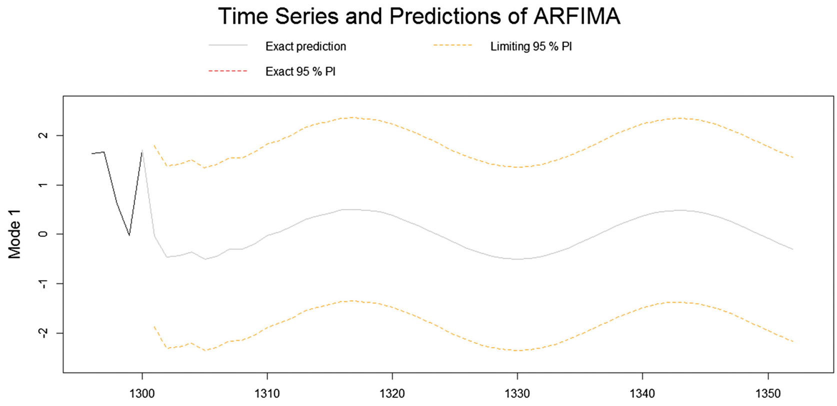

Fig. 6. Prediction intervals for forecasted rainfall values in 2015

According to Fig. 5, it can be seen that the predicted values are in considerable good agreement with the actual rainfall values. The result of the 95% prediction interval also provides encouraging prediction accuracy with a 93.23% coverage probability (Fig. 6).

6 Conclusion

Observed rainfall series illustrates long memory features with an unbounded spectral density. Therefore a standard long memory ARFIMA model was fitted to capture the rainfall pattern and its variability. The Monte Carlo simulation results prove the accuracy of the maximum likelihood method used to estimate the parameters of the model. Furthermore, it is noticed that the parameter bias has decreased and the parameters become consistent with the increase of the simulated series length. ARFIMA(4,0.0579,4) model was found to be the best fitted model that provided a minimum MAE. The out of sample prediction values give good conformity with the actual weekly rainfall in 2015. The 95% prediction intervals also give a promising result to capture the real dynamics of the persistent rainfall. For future work it is suggested that prediction intervals using the bootstrap re-sampling approach may forecast estimates with a higher degree of accuracy.

Acknowledgement. This study was partially funded by the University Research Grant, University of Sri Jayewardenepura, Sri Lanka under Grant (ASP/01/RE/HSS/2016/75).

References

1. Alamgir, A.A., Kalil, U., Khan, S.A., Khan, D.M.: A Sieve bootstrap approach to constructing prediction intervals for long memory time series. Res. J. Recent Sci. **4**(7), 93–99 (2015)

2. Beran, J., Feng, Y., Ghosh, S., Kulik, R.: Long memory Processes Probabilistic Properties and Statistic properties and Statistical Methods. Springer, Heidelberg (2013)
3. Chan, N.H., Palma, W.: Estimation of long-memory time series models. Adv. Econ. **20**, 89–121 (2006)
4. Charles, S.B., Franses, P.H., Ooms, M.: Inflation, forecast intervals and long memory regression models. Int. J. Forecast. **18**, 243–264 (2002)
5. Chatfield, C.: Calculating interval forecasts. J. Bus. Econ. Stat. **11**(2), 121–135 (1993)
6. Chen, G., Abraham, B., Peiris, S.: Lag window estimation of the degree of differencing in fractionally integrated time series models. J. Time Ser. Anal. **15**(5), 473–487 (1994)
7. Cheung, Y., Diebold, F.X.: On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. J. Econ. **62**, 301–316 (1994)
8. Chu, F.L.: A fractionally integrated autoregressive moving average approach to forecast tourism demand. Tour. Manag. **29**, 79–88 (2008)
9. Dissanayake, G.S.: Rapid optimal lag order detection and parameter estimation of standard long memory time series. In: Causal Inference in Econometrics, pp. 17–28. Springer, Cham (2016)
10. Dubey, A.D.: Artificial neural network models for rainfall prediction in Pondicherry. Int. J. Comput. Appl. **120**(3), 30–35 (2015)
11. Eni, D., Adeyeye, F.J.: Seasonal ARIMA modelling and forecasting of rainfall in Warri Town, Nigeria. J. Geosci. Environ. Protection **3**, 91–98 (2015)
12. Fox, R., Taqqu, M.S.: Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Ann. Stat. **14**(2), 517–532 (1986)
13. French, M.N., Krajewski, W.F., Cuykendall, R.R.: Rainfall forecasting in space and time using a neural network. J. Hydrol. **137**, 1–31 (1992)
14. Geweke, J., Hudak, S.P.: The estimation and application of long memory time series models. J. Time Ser. Anal. **4**, 221–238 (1983)
15. Granger, C.W.J., Joyeux, R.: An introduction to long-memory time series models and fractional differencing. J. Time Ser. Anal. **1**, 15–29 (1980)
16. Hauser, M.A.: Maximum likelihood estimators for ARMA and ARFIMA models: a Monte Carlo study. J. Stat. Plan. Inference **80**, 229–255 (1999)
17. Hosking, J.R.M.: Fractional differencing. Biometrika **68**, 165–176 (1981)
18. Hung, N.Q., Babel, M.S., Weesakul, S., Tripathi, N.K.: An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrol. Earth Syst. Sci. **13**, 1413–1425 (2009)
19. Hwang, E., Shin, D.W.: New bootstrap method for autoregressive models. Commun. Stat. Appl. Methods **20**(1), 85–96 (2013)
20. Momani, P.E.N.M.: Time series analysis model for rainfall data in Jordan: case study for using time series analysis. Am. J. Environ. Sci. **5**(5), 559–604 (2009)
21. Muller, U.K., Watson, M.W.: Measuring uncertainty about long-run predictions. Rev. Econ. Stud. **83**, 1711–1740 (2016)
22. National Climate Change Adaptation Strategy for Sri Lanka 2011 to 2016. Ministry of Environment (2010)
23. Palma, W.: Long Memory Time Series Theory and Methods. John Wiley and Sons, New Jersey (2007)
24. Silva, H.P.T.N., Peiris, T.S.G.: Analysis of weekly rainfall using percentile bootstrap approach. Int. J. Ecol. Dev. **32**(3), 97–106 (2017)

25. Silva, H.P.T.N., Peiris, T.S.G.: Statistical modeling of weekly rainfall: a case study in Colombo city in Sri Lanka. In: Proceedings of the 3rd Moratuwa Engineering Research Conference (MER Con), Sri Lanka, 29–31 May, pp. 241–246. IEEE (2017)
26. Silva, H.P.T.N., Peiris, T.S.G.: Accurate confidence intervals for Weibull percentiles using bootstrap calibration: a case study of weekly rainfall in Sri Lanka. Int. J. Ecol. Econ. Stat. **39**(3), 67–76 (2018)
27. Sowell, F.: Maximum likelihood estimation of stationary univariate fractionally integrated time series models. J. Econ. **53**, 165–188 (1992)
28. Veenstra, J.Q., McLeod, A.I.: Persistence and Anti-persistence: Theory and Software (Ph.D. thesis) (2013)
29. Wang, S., Feng, J., Liu, G.: Application of seasonal time series model in the precipitation forecast. Math. Comput. Model. **58**, 677–683 (2013)
30. Zakaria, S., Al-Ansari, N., Knutsson, S., Al-Badrany, T.: ARIMA models for weekly rainfall in the semi-arid Sinjar District at Iraq. J. Earth Sci. Geotech. Eng. **2**(3), 25–55 (2012)
31. Zhou, Z., Xu, Z., Wu, W.B.: Long-term prediction intervals of time series. IEEE Trans. Inf. Theory **56**(3), 1436–1446 (2010)