JGWR (ISSN: 2321- 4783), Vol.2, Issue1, June 2013 Published by AGGS, India

The Significance of Autocorrelation and Partial Autocorrelation on Univariate Groundwater Level Rise (Recharge) Time Series Modeling

S. Mohanasundaram[1], Balaji Narasimhan[2], G. Suresh Kumar[3]*

[1*]Research Scholar, EWRE division, Department of Civil Engineering, Indian Institute of Technology Madras, Chennai-600036
Email: mohanasundaram1986@gmail.com

[2]Assistant Professor, Department of Civil Engineering, Indian Institute of Technology Madras, Chennai-600036
Email: nbalaji@iitm.ac.in

[3]Associate Professor, Department of Ocean Engineering, Indian Institute of Technology Madras, Chennai-600036
Email: gskumar@iitm.ac.in

Abstract

A study has taken to model rainfall and water level rise time series for 36 observation wells in Adyar sub basin located in the north east coastal part of Tamil Nadu. As a part of simple linear regression analysis between rainfall and water level rise series, different ways of representing average rainfall methods namely simple arithmetic average (SAA), Thiessen polygon (TP) and Thiessen zone wise rainfall (TZR) have been discussed and corresponding coefficient of determination (R^2) values were compared across them. Three best correlated observation wells from regression analysis were selected for further modeling using univariate time series approach. Autocorrelation function (ACF) and partial autocorrelation function (PACF) were analyzed for raw water level rise time series data. Different combinations of seasonal autoregressive integrated moving average (SARIMA) models were fitted and best fitted model from them was identified based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values.

Keywords: Autocorrelation, Partial autocorrelation, Time series models, Linear regression models

1. Introduction

There are many methods in practice to assess the groundwater recharge such as physical, chemical and modeling based methods in which physical and chemical based methods are often laborious, expensive and time consuming processes. Modeling based methods are simple and effective ways of mimicking the real system in which simple linear regression modeling is very easy to understand the complex interacting system in terms of simple linear relationship based on few variables, parameters and constants. In rainfall-recharge studies, linear regression relationship can be constructed in terms of recharge (water level rise in this case) as dependent variable against rainfall as an independent variable. Relationships between rainfall and infiltration-recharge have been studied using groundwater regime data (Viswanthan, 1983, 1984; Rennolls et al., 1980; Venetis, 1971). Rainfall measurements over a given area is mostly point based measurements. Rain gauges are often used to measure the amount of rainfall for a particular location. Similarly, water level measurements are also based on particular location where it is measured from observation wells. In both the cases, spatially continuous measurement is not possible hence the point based measurements have to be converted to avoid spatial heterogeneities. To do this, there are many methods for averaging the precipitation over a given region such as simple arithmetic average (SAA), Thiessen polygon method (TP) and normal-ratio (NR) method (Chow et al. 1988). In each way of calculating average precipitation varies in magnitudes thus changing the strength of the regression relationship between rainfall and recharge.

Linear regression models and time series models assume that the residuals are independent and normally distributed (with mean=0 and variance=σ^2). But when the residuals are serially correlated, there may be chance for inappropriate estimation of the model parameters and thereby reducing the model performance significantly. When autocorrelation/serial correlation function (ACF) of lagged variables increases significantly either positively or negatively, it has to be accounted in the model properly by fitting with suitable number of parameters. Autocorrelation is inter-dependency of particular variable itself which gives the understanding of their correlation strength with lags within time series data. Partial Autocorrelation (PACF) is the amount of correlation between them which is not explained by their mutual correlations with a specified set of other variables. There are time series models to adjust for serially correlated errors. Autoregressive (AR), moving averages (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) are the common models available to correct the serially correlated errors and to fit the reasonable model for future forecasting. Box and Jenkins (1970) have described the direct relationship between the dynamic behavior of the physical system and their structural identification and parameter estimation by means of time series analysis. Based on the above discussion, this study has taken (1) to assess the rainfall-recharge regression model efficiency by inputting three different ways of representing rainfall in to the regression model and (2) to develop suitable SARIMA models for forecasting univariate water level rise (recharge) time series component by adjusting serially correlated residuals.

2. Study Location and Data Used

Adyar basin observation wells and rain gauge locations are shown in Fig.1. Both rain gauge and observation wells have sparsely distributed over the study region. Hence, different areal averaging rainfall methods were used in this study to deduce the representative rainfall values for modeling rainfall-recharge relationship.

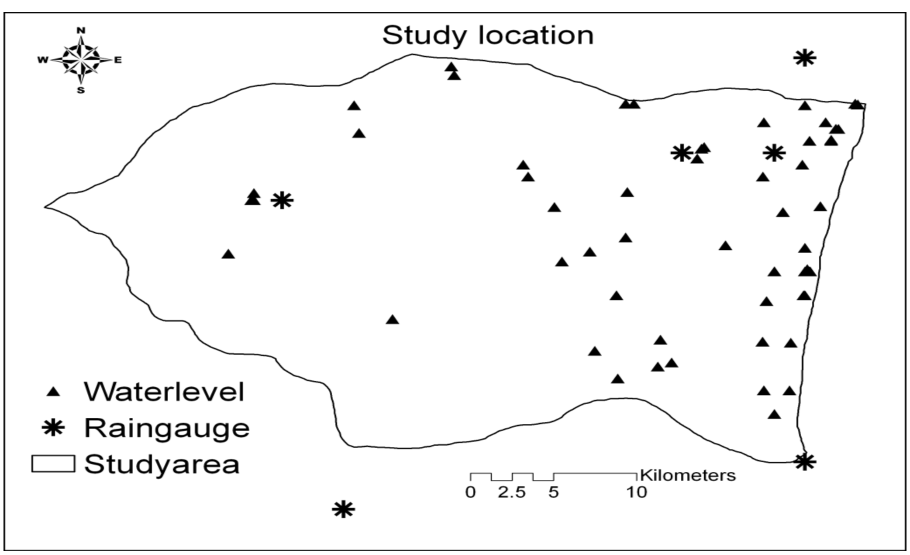

Fig.1. Observation wells and rain gauge locations in Adyar basin

2.1 Precipitation

Daily precipitation data of six rain gauge stations in and around Adyar basin was used in this study. Rainfall data was acquired from State Ground and Surface Water Resources Data Centre (SG&SWRDC), Taramani, over the period of 20 years (1988-2007). Rain gauge stations were located in Covelong, Guindy, Kattankulathur, Minambakkam, Nungambakkam and Sriperumbadur respectively.

2.2 Groundwater level

Monthly water level data was acquired from the same institute of SG&SWRDC for 36 observation wells over the same period of 20 years (1988-2007).

3. Methodology

3.1 Linear regression analysis

Linear regression modeling between rainfall and water level rise data points were attempted with different methods of representing rainfall as an independent variable in the model for the same area.

3.1.1. Simple arithmetic average method (SAA): Average precipitation is calculated by a simple arithmetic average of all stations precipitation data.

$$P_a = \frac{\sum_{i=1}^{N} P_i}{N} \tag{1}$$

where, P_a - Average precipitation; P_i -is the precipitation at ith station; N - Total number of stations

3.1.2. Thiessen polygon method (TP): Average rainfall of the entire area is estimated based on the weighted area precipitation of the nearby stations.

$$P_a = \frac{\sum_{i=1}^{N} P_i A_i}{\sum_{i=1}^{N} A_i} \tag{2}$$

where, A_i - is the area of the polygon belong to the ith station

3.1.3. Thiessen zone wise rainfall method (TZR): Instead of estimating one single areal average precipitation series, as similar to that of SAA and TP, for all 36 observation wells data to model independently, a zone wise approach was adopted where the zones were created by Thiessen polygon method in which corresponding zone rainfall series was taken with 100% weightage which was then correlated with that particular zone water level rise series.

Using the above three methods, average rainfall for each month was calculated. They were separately used in the regression analysis for model fitting and their efficiency assessment in terms of coefficient of determination (R^2). Based on the regression analysis, R^2 value was above 0.4 for three wells were selected for univariate time series modeling.

3.2 Univariate time series models

Time series modeling is a powerful tool for analyzing the dynamic behavior of time series variable in terms of their coefficients and constants. ARIMA and SARIMA models are very common in univariate time series modeling where the autocorrelation among the residuals are perfectly modeled with sufficient number of autoregressive (AR), moving average (MA), seasonal autoregressive (SAR) and seasonal moving average (SMA) coefficients.

3.2.1. AR models: The autoregressive model is one of a group of linear prediction formulas that attempt to predict an output of a system based on the previous outputs. AR(p) indicates an autoregressive model of order p.

$$\phi(B)y_t = c + \varepsilon_t \tag{3}$$

where,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

where, $B = \frac{y_{t-1}}{y_t}$ is the backshift operator, $\phi_1 \ldots \phi_p$ are the autoregressive parameters; c – constant; ε_t-white noise

3.2.2. MA models: MA(q) indicates the moving average model with order q

$$y_t = c + \theta(B)\varepsilon_t \tag{4}$$

where,

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$$

where, $\theta_1 \ldots \theta_q$ are the moving average parameters; ε_t -white noise terms

3.2.3. ARMA models: Autoregressive moving average model is denoted by ARMA (p, q) where p, q refers the autoregressive and moving average terms.

$$\phi(B)y_t = c + \theta(B)\varepsilon_t \tag{5}$$

where,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$$

3.2.4. ARIMA models: Autoregressive Integrated Moving average models can be specified by ARIMA (p,d,q) where, p-autoregressive term, d-non seasonal differencing term, q-moving average term. ARIMA models have one additional term with ARMA models which is d for differencing the time series. Higher order differencing can be applied when the series is extremely non stationary with clear trend in ACF and PACF values. It can be represented by,

$$\phi(B)(1-B)^D y_t = c + \theta(B)\varepsilon_t \tag{6}$$

Where,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$$

where, D is the order of non-seasonal differencing

3.2.5. SARIMA models: Multiplicative ARIMA models also called as SARIMA model. In addition to non-seasonal differencing, SARIMA model possesses seasonal differencing with seasonal autoregressive and moving average coefficients. It can be represented by,

$$\phi(B)(1-B)^D \Phi(B)(1-B^s)^{Ds} y_t = c + \theta(B)\Theta(B)\varepsilon_t \tag{7}$$

Where,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q$$

$$\Phi(B) = 1 - \Phi_{p1} B^{p1} - \Phi_{p2} B^{p2} - \ldots - \Phi_{ps} B^{ps}$$

$$\Theta(B) = 1 + \Theta_{q1} B^{q1} + \Theta_{q2} B^{q2} + \ldots + \Theta_{qs} B^{qs}$$

where, ps , qs are seasonal degrees; Ds is the order of seasonal differencing.

3.3. Box and Jenkins time series methodology

Basically, there are five major steps in univariate time series modeling as mentioned by Box and Jenkins (1994).

Step:1 Establishing stationary time series data: First the raw time series data is plotted for autocorrelation and partial autocorrelation functions. If the autocorrelation and partial autocorrelation values are not within the lower and upper bounds then the first differencing of time series is to be done.

Step:2 Identifying conditional mean models: Sample ACF and PACF plots can help in identifying the selection of models. For an autoregressive (AR) process, the sample ACF decays gradually, but the sample PACF cuts off after a few lags. Conversely, for a moving average (MA) process, the sample ACF cuts off after a few lags, but the sample PACF decays gradually. If both the ACF and PACF decay gradually, ARMA model is to be considered

Step:3 Specifying models and estimating model parameters: According to the ACF and PACF plots suitable AR or MA or ARMA models are to be selected. In this study since there is no particular pattern encountered in ACF and PACF plots as mentioned above in the second point, 3 models were adopted randomly and their corresponding parameters were estimated.

Step:4 Conducting goodness-of-fit: After, the model is selected and corresponding parameters are estimated, performance of the model is to be done. Again, ACF and PACF plots of the residuals are checked. a model is said to be perfect when the residuals are not correlated and normally distributed until the condition comes true, model specifications are to be revised. Comparison of models based on AIC, BIC values selects an appropriate model for the given data.

Step:5 Forecasting: Selected model is tested for forecasting of time series variable.

4. Results and Discussions

Linear regression R^2 value was ranging from 0-0.64. R^2 was found higher for most of the wells in TZR method compared to SAA and TP methods are shown in Fig.2.

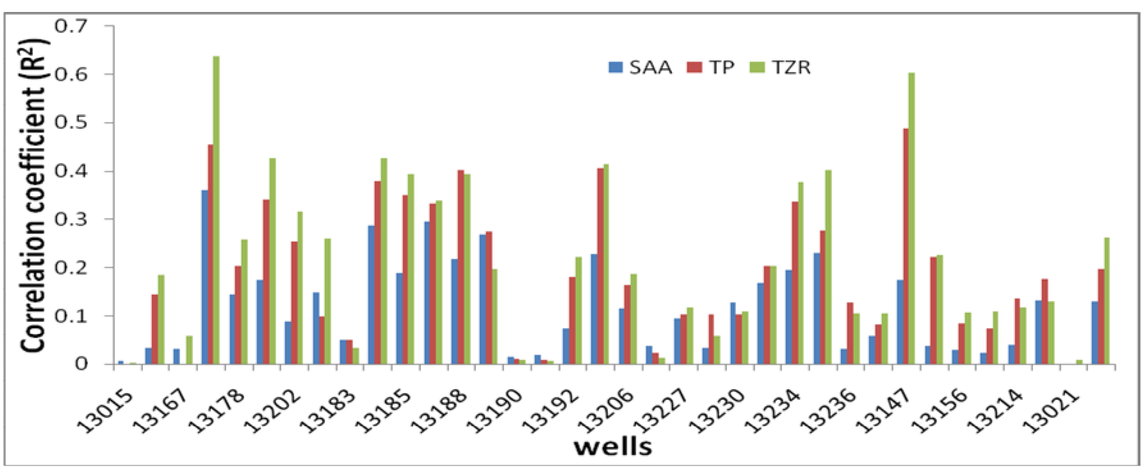

Fig.2. Effect of different methods of representing rainfall on linear regression analysis

Three maximally correlated wells among 36 wells were selected for univariate time series modeling. Time series plot of three selected wells are shown in Fig.3, Fig.4 and Fig.5. It was observed from the time series plots that the measurements possessed seasonality behavior and there is no trend in time series data since the groundwater hydraulic head values were subtracted from consecutive values by differencing process. Differenced series were having positive and negative values in which only positive values (recharge) were modeled after suppressing negative values (discharge) to 0. Seasonality behavior in the time series data is due to the measurements that were observed at monthly interval and recharge or water level rise is highly correlated with monsoon seasons. Maximum water level rise about 4.5 m was observed in the month of December 1998 for the well-13147 in the basin (Fig.3).

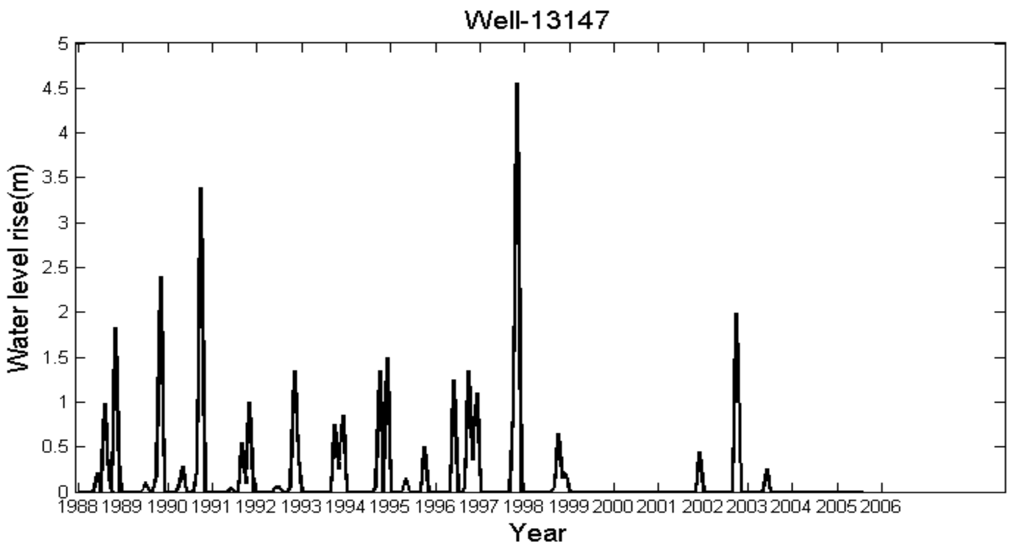

Fig.3. Water level rise time series data for the well 13147

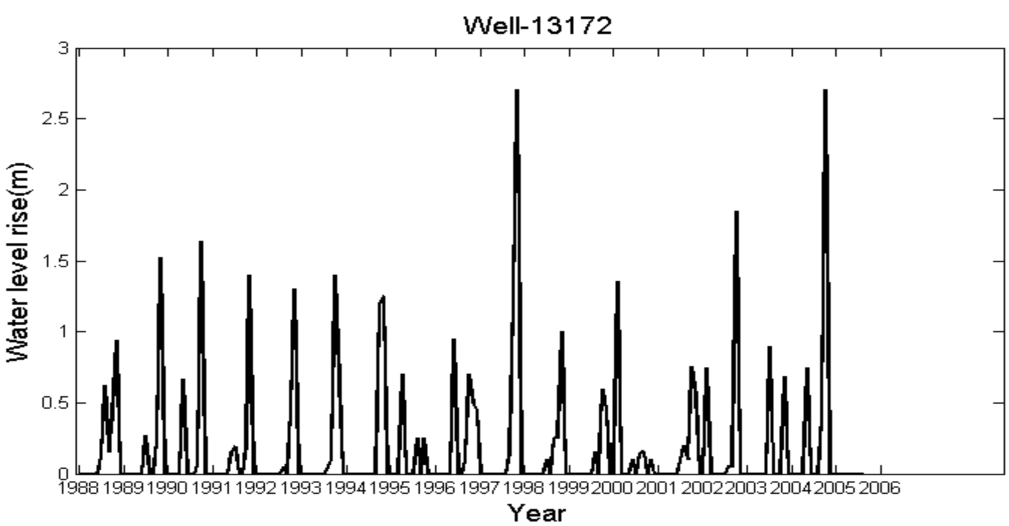

Fig.4. Water level rise time series data for the well 13172

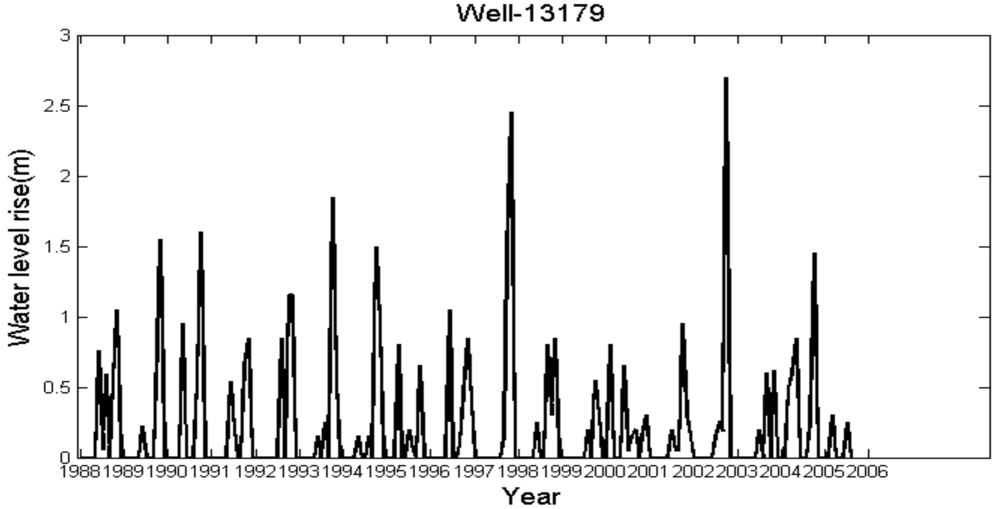

Fig.5. Water level rise time series data for the well 13179

4.1. ACF and PACF plots of water level rise data

Water level rise time series data was analysed with ACF and PACF plots (Fig.6, Fig.7 and Fig.8). It was observed from the ACF plots that there was significant serial correlation among the residuals at lags 11, 12 and 13 for all the wells. Significant positive serial correlation is the indication of repetition of either high-high values or low-low values in the overall time span. ACF plots also shows that the data are seasonal at lags 11, 12 and 13 due to strong correlation of water level rise phenomenon with seasonal north east monsoon rainfall in the basin.

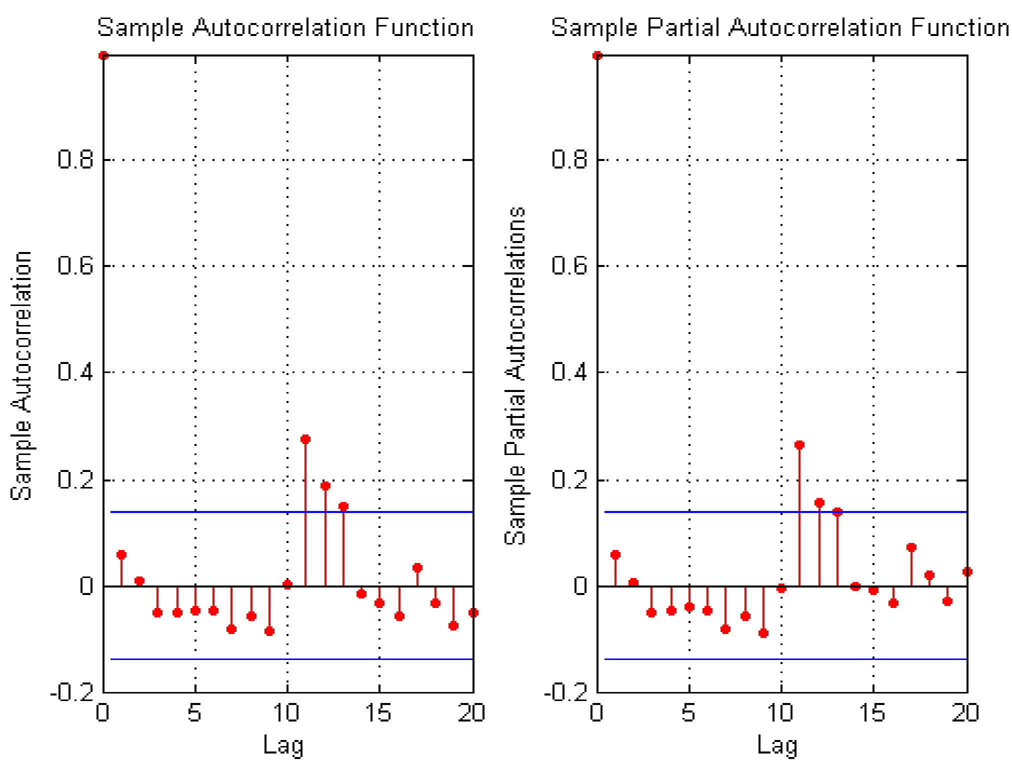

Fig.6. Sample ACF and PACF plots for the well 13147

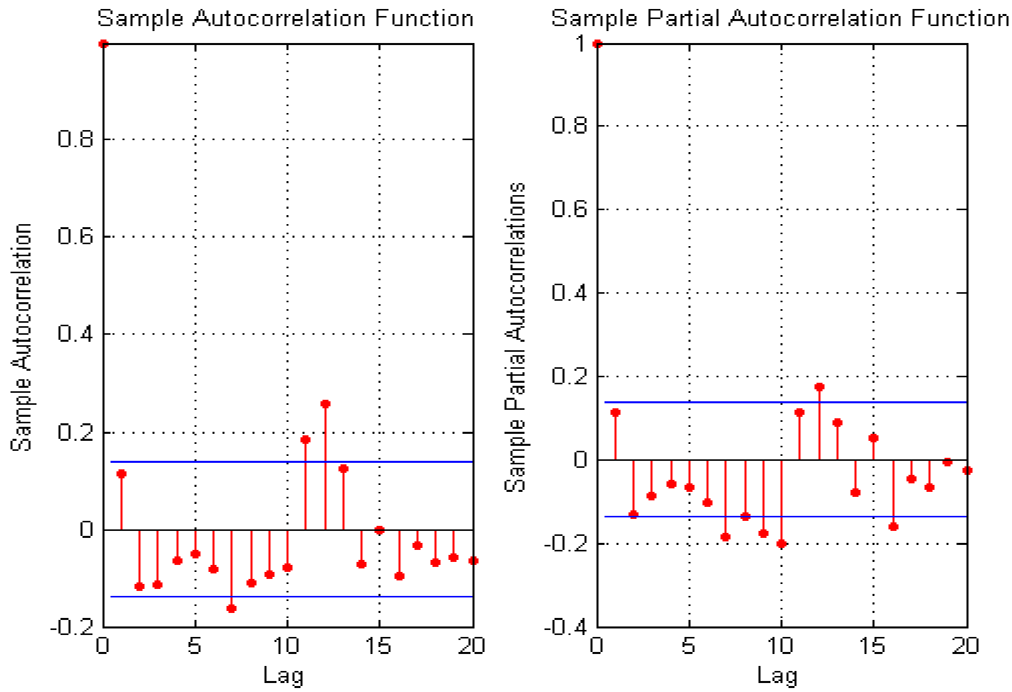

Fig.7. Sample ACF and PACF plots for the well 13172

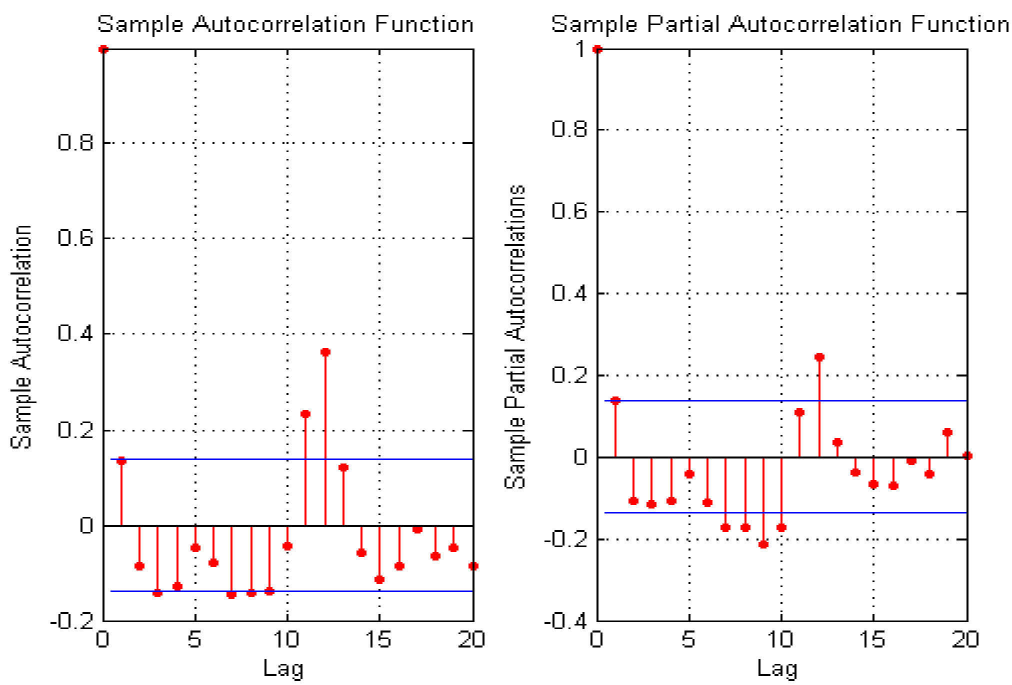

Fig.8. Sample ACF and PACF plots for the well 13179

4.2. Model specification and parameter estimation

SARIMA models were employed to make the series stationary and to make the residual with white noise or no serial correlation process. Models specifications were defined based on AR, MA, SAR and SMA processes as mentioned in methodology section. Since the data shows the seasonality behavior at lags 11 and 12 three different combinations of SARIMA models, SARIMA(0,0,1)(0,1,1)$_{12}$, SARIMA(0,0,1)(1,1,0)$_{12}$ and SARIMA(0,0,1)(1,1,1)$_{12}$, were fitted for each well separately and corresponding AIC and BIC values were compared across them. Parameter estimation was done using maximum likelihood technique and corresponding statistics have given in Table 1. It was found that the SMA parameter was very significant in the respective SARIMA models for all the wells.

Table 1. Parameter estimation and AIC, BIC values of SARIMA models for the wells 13147, 13172, 13179

Well	Models	Parameter	Value	SE	T-value	P	AIC	BIC
13147	SARIMA (0,0,1)(0,1,1)$_{12}$	Constant	-0.01	0.01	-1.34	0.09	317.16	330.59
		MA{1}	-0.17	0.05	-3.42	0.00		
		SMA{12}	-0.80	0.03	-28.69	0.00		
		Variance	0.25	0.01	26.26	0.00		
	SARIMA (0,0,1)(1,1,0)$_{12}$	Constant	-0.02	0.04	-0.42	0.34	371.41	384.84
		MA{1}	-0.12	0.05	-2.52	0.01		
		SAR{12}	-0.45	0.02	-22.61	0.00		
		Variance	0.33	0.01	33.73	0.00		
	SARIMA (0,0,1)(1,1,1)$_{12}$	Constant	-0.01	0.01	-1.34	0.09	319.15	335.93
		MA{1}	-0.16	0.05	-3.38	0.00		
		SAR{12}	-0.01	0.05	-0.20	0.42		
		SMA{12}	-0.79	0.03	-23.79	0.00		
		Variance	0.25	0.01	25.66	0.00		
13172	SARIMA (0,0,1)(0,1,1)$_{12}$	Constant	0.00	0.01	0.01	0.50	173.01	186.44
		MA{1}	-0.07	0.06	-1.22	0.11		
		SMA{12}	-0.87	0.04	-20.77	0.00		
		Variance	0.13	0.01	18.74	0.00		
	SARIMA (0,0,1)(1,1,0)$_{12}$	Constant	0.00	0.03	0.06	0.48	223.81	237.24
		MA{1}	0.00	0.06	0.00	0.50		
		SAR{12}	-0.62	0.03	-19.59	0.00		
		Variance	0.16	0.01	20.56	0.00		
	SARIMA (0,0,1)(1,1,1)$_{12}$	Constant	0.00	0.01	0.02	0.49	166.57	183.36
		MA{1}	-0.06	0.06	-0.98	0.16		
		SAR{12}	-0.24	0.06	-4.14	0.00		
		SMA{12}	-0.81	0.04	-20.85	0.00		
		Variance	0.12	0.01	18.94	0.00		
13179	SARIMA (0,0,1)(0,1,1)$_{12}$	Constant	0.00	0.01	-0.52	0.30	172.28	185.70
		MA{1}	-0.09	0.06	-1.55	0.06		
		SMA{12}	-0.83	0.03	-24.84	0.00		
		Variance	0.13	0.01	18.18	0.00		
	SARIMA (0,0,1)(1,1,0)$_{12}$	Constant	0.00	0.03	-0.04	0.48	229.39	242.82
		MA{1}	0.00	0.06	-0.04	0.48		
		SAR{12}	-0.52	0.04	-14.01	0.00		
		Variance	0.17	0.01	22.63	0.00		
	SARIMA (0,0,1)(1,1,1)$_{12}$	Constant	0.00	0.01	-0.49	0.31	172.05	188.83
		MA{1}	-0.09	0.06	-1.53	0.06		
		SAR{12}	-0.11	0.06	-1.84	0.03		
		SMA{12}	-0.81	0.04	-20.91	0.00		
		Variance	0.13	0.01	17.46	0.00		
SE-Standard Error								

4.3. Model goodness-of-fit

Model goodness-of-fit among three different SARIMA models were analyzed using two standard criteria's such as AIC and BIC. Least values of AIC and BIC yields the best model for the given data series. Least values of AIC and BIC have found from the models SARIMA $(0,0,1)(0,1,1)_{12}$, SARIMA$(0,0,1)(1,1,1)_{12}$ and SARIMA$(0,0,1)(0,1,1)_{12}$ for the wells 13147, 13172 and 13179 respectively (Table 1).

4.4. Residual analysis for the best selected SARIMA models

Residual serial correlation was analysed for the best fitted model with residual plot, residual ACF and residual PACF plots (Fig.9, Fig.10 and Fig.11). Zero crossing of residuals was ensured with the residual plot of the best selected models for all the wells. Similarly, insignificant serial correlations of residuals were obtained for the selected models. Residual ACF and PACF plots show that the residual serial correlations are insignificant and within the 95% confidence bounds.

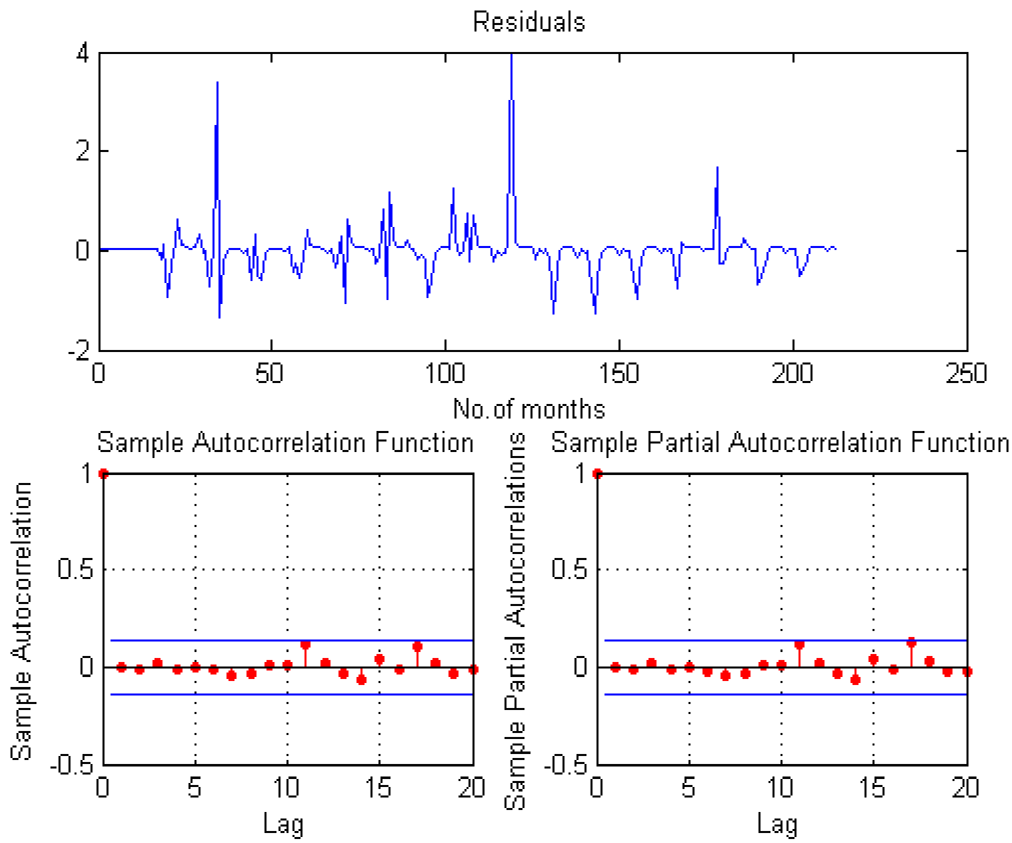

Fig.9. Residual plots of best chosen SARIMA model for the well 13147

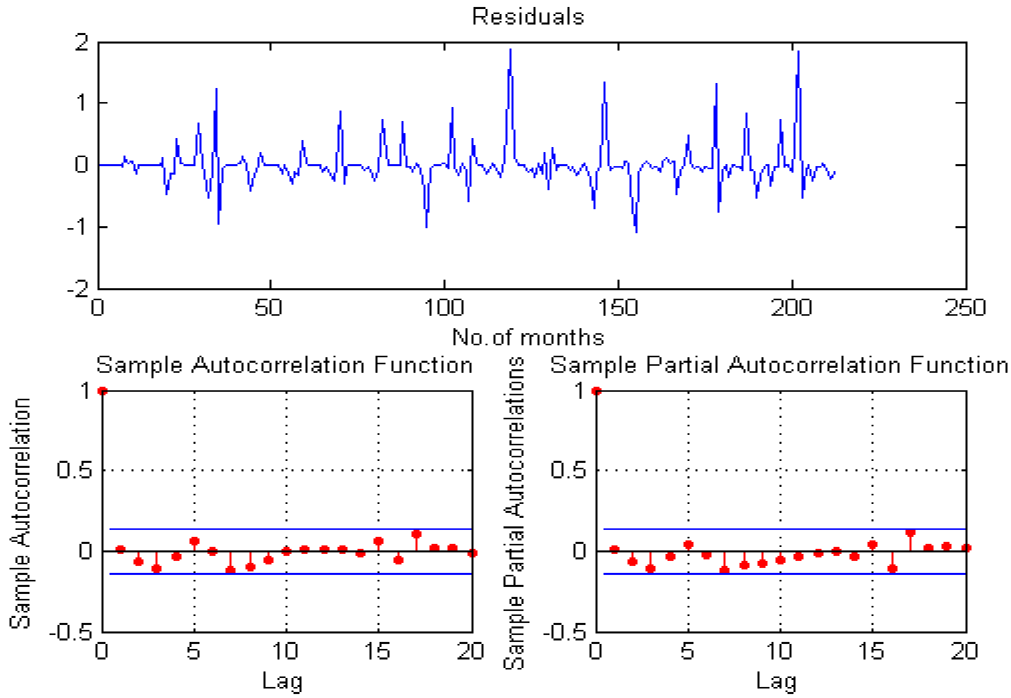

Fig.10. Residual plots of best chosen SARIMA model for the well 13172

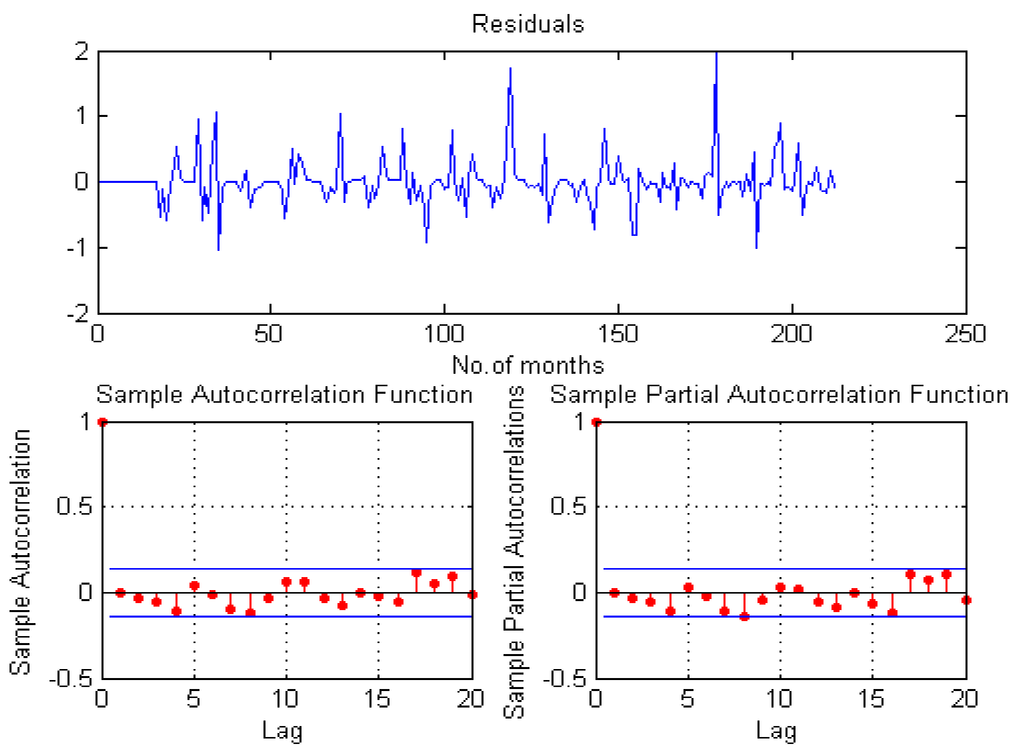

Fig.11. Residual plots of best chosen SARIMA model for the well 13179

5. Conclusion

Linear regression models have been improved analysing various ways of inputting rainfall data into the model. TZR method of representing rainfall in rainfall-recharge regression modeling was fitting better than the other methods. SARIMA models were proved as better models in reducing autocorrelation and partial autocorrelation among the residual series. Model goodness of fit was checked by AIC and BIC values based on maximum likelihood objective function calculated by the respective models. Residual plots of best selected models were analyzed for residual serial correlation and it was ensured with no serial correlation and normal distribution.

Acknowledgment

We would like to thank State Ground and Surface Water Resources Data Centre (SG&SWRDC), IWS, Taramani, Chennai for their help by sharing their rainfall and water level data covering Thiruvallur, Kancheepuram and Chennai districts of Tamil Nadu, India.

References

Box, G. E. P., and Jenkins, G. M. 1970. *Time series analysis: Forecasting and control*. Holden-Day.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. 1994. *Time series analysis: Forecasting and control*. 3rd ed. Upper Saddle River, NJ: Prentice-Hall

Chow, V.T., Maidment, D.R., and Mays, L.W. 1988. *Atmospheric water in Applied Hydrology*. 53-98. McGraw-Hill, Inc, New York.

Rennolls, K., Carnell, R., and Tee, V. 1980. A descriptive model of the relationship between rainfall and soil water table. *J. Hydrol.*, 47, 103-114.

Venetis, C. 1971. Estimating infiltration and/or the parameters of unconfined aquifers from groundwater level observations. *J. Hydrol.*, 12, 161-169.

Viswanthan, M. N. 1984. Recharge characteristics of an unconfined aquifer from the rainfall-water table relationship. *J. Hydrol.*, 70, 233-250.