

Received 04.04.2017  
Reviewed 22.05.2017  
Accepted 01.06.2017A – study design  
B – data collection  
C – statistical analysis  
D – data interpretation  
E – manuscript preparation  
F – literature search

## Application of SARIMA model to forecasting monthly flows in Waterval River, South Africa

Kassahun Birhanu TADESSE<sup>1)</sup> ABCDEF ✉, Megersa Olumana DINKA<sup>2)</sup> ABCDEF

<sup>1)</sup> University of Johannesburg, Faculty of Engineering and the Built Environment, Department of Civil Engineering Science, Auckland Park Campus Kingsway, 524 Johannesburg, South Africa; e-mail: ktadesse@uj.ac.za

<sup>1)</sup> Debre Markos University, College of Agriculture and Natural Resources, Department of Natural Resources Management, 269 Debre Markos, Ethiopia; e-mail: kbirhan@gmail.com

<sup>2)</sup> University of Johannesburg, Faculty of Engineering and the Built Environment, Department of Civil Engineering Science, Johannesburg, South Africa; e-mail: magarsol@yahoo.com

**For citation:** Tadesse K.B., Dinka M.O. 2017. Application of SARIMA model to forecasting monthly flows in Waterval River, South Africa. *Journal of Water and Land Development*. No. 35 p. 229–236. DOI: 10.1515/jwld-2017-0088.

### Abstract

Knowledge of future river flow information is fundamental for development and management of a river system. In this study, Waterval River flow was forecasted by SARIMA model using GRETL statistical software. Mean monthly flows from 1960 to 2016 were used for modelling and forecasting. Different unit root and Mann–Kendall trend analysis proved the stationarity of the observed flow time series. Based on seasonally differenced correlogram characteristics, different SARIMA models were evaluated; their parameters were optimized, and diagnostic check up of forecasts was made using white noise and heteroscedasticity tests. Finally, based on minimum Akaike Information (*AI*) and Hannan–Quinn (*HQ*) criteria, SARIMA (3, 0, 2) × (3, 1, 3)<sub>12</sub> model was selected for Waterval River flow forecasting. Comparison of forecast performance of SARIMA models with that of computational intelligent forecasting techniques was recommended for future study.

**Key words:** *heteroscedasticity, stationarity test, trend analysis, validation, white noise*

### INTRODUCTION

Accurate river flow forecasting provides information for any water resources planning systems like irrigation schemes, city and environmental plans, reservoir operation strategies, flood and drought analysis, and for designing and maintaining of hydraulic structures. Forecasting helps to anticipate river flow with use of historical river flow characteristics [BOX *et al.* 2008]. One of the many techniques used for forecasting is a time series model.

In time series models, historical traits are assumed to manifest themselves in the future [VAHDAT *et al.* 2011]. NAIL and MOMANI [2009] and ADHIKARY *et al.* [2012] noticed that a researcher having only a single time series of historical records may use the univariate Box–Jenkins method for predicting the coming observations without looking for the time se-

ries of other variables. Many univariate stochastic models like autoregressive integrated moving average (ARIMA) ones that use probability and statistics are in use for forecasting purposes [BROCKWELL, DAVIS 2002; WANG *et al.* 2008].

For ARIMA time series that have a tendency of showing a periodic behaviour after certain time intervals, its extension called seasonal ARIMA (SARIMA) modelling approach, is used [BOX *et al.* 2008]. Seasonal differencing in this model reduces the periodic intensity and makes a series static [SALAS *et al.* 1980].

MOEENI *et al.* [2017] showed that for each monthly stream flow and water temperature series, seasonal differencing in ARIMA models is the best stationarization method in terms of periodic effect elimination and model forecasting accuracy as compared to seasonal standardization and spectral analysis. The SARIMA model also performed best among

Thomas-Fiering and Spectral Analysis types of stochastic mathematical models in forecasting flow of five rivers in the Atrak basin, north-eastern Iran [TEYMOURI, FATHZADEH 2015].

Several studies on modelling of time series data using SARIMA models have been undertaken in hydrology, meteorology and other fields. PAPALASKARIS *et al.* [2016] employed a SARIMA model to perform short-term forecasts of monthly rainfall in Kavala city, Greece, which was aimed at identifying the potential patterns of flood and drought cycles occurring in this area. TARIQ and ABBASABD [2016] used a SARIMA model for Nyala station (Sudan), which was considered appropriate for forecasting monthly rainfalls.

VALIPOUR [2015] proved the appropriateness of SARIMA model for long-term runoff forecasting in the United States as compared to the ARIMA models. SARIMA model was also found to be appropriate to simulate water quality in the Danube River (Europe) under extreme discharges and water temperature conditions [PEKÁROVÁ *et al.* 2009]. KOMORNÍKOVÁ *et al.* [2008] used a hybrid forecasting model which combines separate nonlinear regime-switching time series for forecasting monthly inflows into Liptovská Mara reservoir (Slovakia).

GAUTAM and SINHA [2016] used SARIMA model for forecasting of mean monthly reference crop evapotranspiration in Bokaro (India). BELAYNEH and ADAMOWSKI [2013] indicated the best performance of hybrid wavelet neural network models in drought forecasting in the Awash River Basin (Ethiopia). TIWARI *et al.* [2016] forecasted water demand of the Calgary City (Canada) using extreme learning machines (ELM). This study showed a greater overall performance of hybrid wavelet transformed ELM (WA-ELM) model as compared to the ELM model. Using a boosting ensemble multi-wavelet extreme learning machine (Multi-WA-ELM) model improved water quality forecasting as compared to individual WA-ELM and ELM models [BARZEGAR *et al.* 2017].

The extreme learning machine model has shown better forecasting ability as compared to the support vector machine model for forecasting monthly groundwater levels at two observation wells located in Canada [YADAV *et al.* 2017]. A hybrid least square support vector regression-gravitational search algorithm (HLGSA) was successfully used for predicting monthly river flows in Astor and Shyok catchments (Pakistan) [ADNAN *et al.* 2017]. YANG *et al.* [2017] have demonstrated the importance of the use of different climate phenomenon indices together with different artificial intelligence techniques for reservoir monthly inflow forecasting.

ALI [2013] used a SARIMA model for rainfall forecasting in Baghdad, Iraq. In this study, rainfall forecast for four years was achieved showing similar trends compared to the original data. JANHABI and JHA [2013] used a SARIMA model for rainfall forecasting in the Mahanadi River Basin, India. CHANG *et al.* [2012] concluded that the SARIMA model is an appropriate method for modelling and predicting monthly

precipitation in Yantai, China. GERRETSADIKAN and SHARMA [2011] employed the SARIMA model for forecasting future values of monthly rainfall in the Ti-gray region, Ethiopia. SARIMA modelling was also used to forecast rainfall in Port-Harcourt, Nigeria [OSARUMWENSE 2013].

The Waterval River in South Africa supplies a significant amount of flow to the Vaal River, and hence to the Vaal Dam which is a principal provider of water to Gauteng province. Therefore, for planning and management of Waterval River system and optimal real-time adaptive operation of Vaal Dam reservoir, forecasting of the current and future Waterval River flows is vital. Based on extensive literature review there is no river flow forecasting study conducted on Waterval River.

In this study, monthly Waterval River flow was forecasted using SARIMA model with seasonal differencing. The developed model helps water resource managers and decision makers for development and management of Waterval River and optimal adaptive operation of Vaal Dam.

## MATERIAL AND METHODS

### RESEARCH AREA NARRATIVE

Waterval River is found in the Olifant River Basin with a catchment area of 188 km<sup>2</sup> at the gauging station that is located at 25.03594 S and 30.21941 E. The Waterval River catchment area (located between 26.32833 and 27.15611S and between 28.41361 and 29.47611 E) consists of the main river (the Waterval River) and its tributaries. The Waterval River flows in a south-westerly direction towards its destination, the Vaal River, which in turn, feeds into the Vaal Dam.

### DATA PREPARATION

The Waterval River flow data was accessed from the Department of Water and Sanitation of South Africa. Waterval River has 672 mean monthly flow observations from 3<sup>rd</sup> October 1960 to 17<sup>th</sup> September 2016. The first 612 mean monthly flows from October 1960 to September 2011 were used for model calibration. The rest 60 observations from October 2011 to September 2016 were used for model validation “in-sample” forecast. Finally, all 672 flow data were used for five years “out of sample” forecasting from October 2016 to September 2021. Gnu Regression, Econometrics and Time-series Library (GRETLL) [COTTRELL, LUCCHETTI 2017] computer program was used for river flow modelling and forecasting.

### SARIMA MODELLING

SARIMA model is the product of seasonal and non-seasonal polynomials and is designated by SARIMA  $(p, d, q) \times (P, D, Q)_s$ , where  $(p, d, q)$  and  $(P, D, Q)$  are non-seasonal and seasonal components, respectively with a seasonality ‘s’. SARIMA model

was defined at Equation 1 [BOX *et al.* 2008, CRYER, CHAN 2008; WANG *et al.* 2008]:

$$\Phi(B^S) \varphi(B)(1 - B^S)^D (1 - B)^d y_t = \Theta(B^S) \theta(B) \varepsilon_t \quad (1)$$

where:  $\Phi$  and  $\varphi$  = autoregressive (AR) parameters of seasonal and non-seasonal components, respectively;  $\Theta$  and  $\theta$  = moving average (MA) parameters of seasonal and non-seasonal components, respectively;  $B$  = backward operator,  $B(y_t) = y_{t-1}$ ;  $(1 - B^S)^D = D^{\text{th}}$  seasonal difference of season  $s$ ;  $(1 - B)^d = d^{\text{th}}$  non-seasonal difference;  $\varepsilon_t$  = an independently distributed random variable;  $P$  and  $p$  = the orders of the AR components;  $Q$  and  $q$  = the orders of MA components;  $D$  and  $d$  are difference terms.

Four sequential steps [BOX *et al.* 2008; WANG *et al.* 2008] as described below were followed for SARIMA modelling and forecasting.

**Model identification.** The first step for model identification is the examination of the river flows stationarity. For this purpose, graphical methods: sample autocorrelation function (ACF) and sample partial autocorrelation functions (PACF) [SINGH *et al.* 2012] were used. Moreover, unit root test has been made using Augmented Dickey–Fuller (ADF) [DICK- EY, FULLER 1979], Phillips–Perron (PP) [PHILLIPS, PERRON 1988], and Kwiatkowski–Philips–Schmidt–Shin (KPSS) [KWIATKOWSKI *et al.* 1992] at 0.5 significance level ( $\alpha = 0.05$ ). In addition, stationarity was checked by Mann–Kendall trend test (MK) [KENDALL 1975] using XLSTAT2017.

During model identification, a type (seasonal or non-seasonal) and order ( $p, q, P, Q$ ) of model parameters were determined on the basis of an ACF diagram [CRYER, CHAN 2008; SALAS *et al.* 1980]. According to BROCKWELL and DAVIS [2002], for seasonal components of SARIMA model,  $P$  and  $Q$  are usually less than three and  $D$  is seldom greater than one. Based on the stated principles, many SARIMA models were identified.

**Parameter estimation.** Conditional maximum likelihood called “conditional sum of squares” [HAMILTON 1994] was used to optimize SARIMA parameters.

**Diagnostic checking.** The residual correlograms (ACF and PACF), Ljung–Box  $Q$  Tests [LJUNG, BOX 1978] and Durbin–Watson test [DURBIN, WATSON 1951] were applied to test white noise (autocorrelation) of model forecasts. Whether the ACF and PACF of the residual values at various lags were settled within tolerance interval at 95% confidence limits was evaluated.

**Forecasting and performance evaluation of models.** As a final step for modelling, forecast precision was evaluated against performance measures (Equations 2 to 4): Akaike information ( $AI$ ) [AKAIKE 1974] and Schwarz Bayes ( $SB$ ) [SCHWARZ 1978] and Hannan–Quinn ( $HQ$ ) [HANNAN, QUINN 1979] criteria. Model performances were also checked by root mean square error ( $RMSE$ ) and mean absolute error ( $MAE$ ).

A model having the lowest  $AI$ ,  $SB$  and  $HQ$  was selected for river flow forecasting:

$$AI = 2k - \log L \quad (2)$$

$$SB = 2k \log n - \log L \quad (3)$$

$$HQ = -2 \log L + 2k \log \log n \quad (4)$$

where:  $k$  = the amount of estimable parameters;  $L$  = maximum likelihood;  $n$  = number of sample.

## RESULTS AND DISCUSSION

### STATIONARITY TESTS

The monthly time series plot of historical Waterval River flow from October 1960 to September 2016 is shown in Figure 1. No noticeable change about the mean is seen from the figure. But, periodic peaks (during the wet season) and lows (during the dry season) from the figure show the seasonality of the time series. Therefore, based on Figure 1, the river flow series looks stationary except for its seasonality. However, stationarity of river flow series was further proved by Mann–Kendall trend (MK) test; DF, PP and KPSS unit root tests at  $\alpha = 0.05$  (Tab. 1).

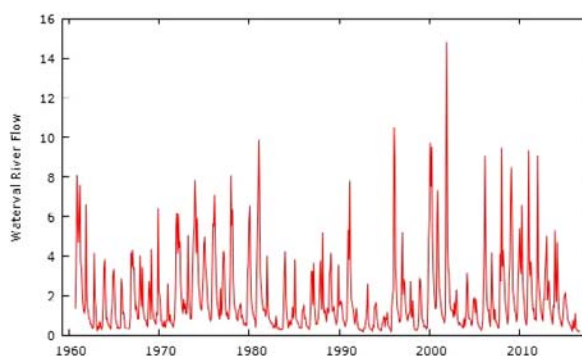


Fig. 1. Time series plot of observed mean monthly flows from 1960 to 2016; source: own study

**Table 1.** Mann–Kendall trend test and unit root tests for observed mean monthly flows

Parameter	Mann–Kendall trend test		Unit root and stationarity tests			
	Mann–Kendall test	seasonal Mann–Kendall test / period = 12	parameter	ADF	PP	KPSS
Kendall's	-0.013	-0.011	$\tau$ (observed value)	-6.968	-7.409	0.110
$S$	-2895.000	-209.000	$\tau$ (critical value)	-0.923	-1.941	0.149
$p$ -value**	0.619	0.671	$p$ -value*	<0.0001	<0.0001	0.124
$\alpha$	0.05	0.05	$\alpha$	0.05	0.05	0.05

\* one tailed; \*\* two tailed.

Explanations: ADF = Augmented Dickey–Fuller test, PP = Phillips–Perron test, KPSS = Kwiatkowski–Philips–Schmidt–Shin test. Source: own study.

From MK evaluation, the value of  $p$  is 0.619 (Tab. 1). This shows the absence of trend in general for Waterval River flow time series. Moreover, the  $p$ -value of 0.671 for seasonal Mann–Kendall test shows there is no trend not due to seasonality when 12 months seasonality was taken into account. Therefore, Waterval River has no trend based on Mann–Kendall trend tests.

From Table 1, the computed  $p$ -value of  $<0.0001$  for DF and PP test is lower than the significance level ( $\alpha = 0.05$ ). This rejects the nonstationary null hypothesis (i.e., there is unit root) at their level and proves the stationarity of the river flow. Moreover, the stationarity of the time series was further proved by KPSS test since the computed  $p$ -value of 0.124 was greater than the significance level ( $\alpha = 0.05$ ). Thus, based on all of the above stationarity analysis, the observed mean monthly flow time series was found to be stationary.

### MODEL IDENTIFICATION

The stationarity of the observed mean monthly flow series was evidently confirmed in section 3.1. Hence, no need of non-seasonal differencing ( $d = 0$ ) but seasonal differencing is required for seasonal stationarity. Moreover, as it is seen in Figure 2, significant spikes were observed at interval of every 12 months (12<sup>th</sup>, 24<sup>th</sup>, 36<sup>th</sup> lags...) at ACF plot and at lag 12<sup>th</sup> of PACF plot. These significant spikes revealed that observed mean monthly flow series is seasonally nonstationary.

Thus, seasonal differencing ( $D = 1$ ) of the observed flow data was done to make time series seasonally stationary. For further investigation, SARIMA ( $p, 0, q$ ) ( $P, 1, Q$ )<sub>12</sub> models were suggested. Initial, parameter  $p, q, P$  and  $Q$  of these models were determined based on the characteristics of seasonally differenced ACF and PACF plots (Fig. 3).

From ACF graph (Fig. 3), the seasonally differenced series has positive value significant spikes at 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> lags and ACF cuts off after 5<sup>th</sup> lag for non-seasonal component. ACF has also a negative significant spike at 12<sup>th</sup> lags and decayed after lag 23<sup>th</sup> for the seasonal component. Therefore, 1–4 moving average (MA) values for non-seasonal and one seasonal moving average (SMA) values were suggested for model identification.

Similarly, PACF of the seasonally differenced series has significant spikes at 1<sup>st</sup> lag for non-seasonal and at 12<sup>th</sup> and 24<sup>th</sup> lags for seasonal components. Then, the values of PACF cuts off up to lag 36. Hence one autoregressive (AR) parameters for non-seasonal and two seasonal autoregressive (SAR) parameters were suggested to be included in the SARIMA model.

Therefore, SARIMA (3, 0, 1) (3, 1, 1)<sub>12</sub> was identified as an initial model. However, as model identification involves much trial and error, maximum values of 3 were set for each  $p, q, P$  and  $Q$  in SARIMA model combinations. Therefore, based on the given conditions more than 23 SARIMA models were identified further evaluation.

### PARAMETER ESTIMATION

**Fit statistics/Performance measures.** The summary of the fit statistics for out of sample forecast from October 2016 to September 2021 is shown in Table 2 for the first four best-fit SARIMA models out of so many models evaluated by GRETL software. Based on minimum performance measures, mainly by the minimum value of  $AI$  and  $HQ$  statistics, SARIMA (3, 0, 2) (3, 1, 3)<sub>12</sub> has performed best in forecasting mean monthly flows of Waterval River.

**Table 2.** Performance values of the first four best models

Parameter	SARIMA			
	(3,0,3) (3,1,3)	(3,0,2) (3,1,3)	3,0,1) (3,1,3)	(3,0,0) (3,1,3)
$AI$	2100.168	2098.172	2101.418	2104.832
$SB$	2157.775	2151.348	2150.163	2149.146
$HQ$	2122.559	2118.840	2120.364	2122.056
$RMSE$	1.342	1.342	1.444	1.332
$MAE$	0.956	0.956	1.094	0.946

Explanations:  $AI$  = Akaike information,  $SB$  = Schwarz Bayes criterion,  $HQ$  = Hannan–Quinn information criterion,  $RMSE$  = root mean square error,  $MAE$  = mean absolute error.

Source: own study.

**Parameters of selected models.** Model parameter values (Tab. 3) which were optimized using conditional maximum likelihood method were all within the range of minimum and maximum significance intervals.

**Table 3.** Parameter estimates for SARIMA (3, 0, 2) (3, 1, 3) model

Parameter	$\varphi_1$	$\varphi_2$	$\varphi_3$	$\theta_1$	$\theta_2$	$\theta_3$	$\Phi_1$	$\Phi_2$	$\Theta_1$	$\Theta_2$	$\Theta_3$
Value	-0.032	0.007	0.295	-1.177	-0.841	0.017	0.644	0.380	0.338	-0.213	-0.897

Explanations:  $\varphi_1, \varphi_2, \varphi_3$  = autoregressive parameters of non-seasonal components,  $\theta_1, \theta_2, \theta_3$  = moving average parameters of non-seasonal components,  $\Phi_1, \Phi_2$  = autoregressive parameters of seasonal components,  $\Theta_1, \Theta_2, \Theta_3$  = moving average parameters of seasonal components.

Source: own study.

### DIAGNOSTIC CHECKING

#### Autocorrelation test

From the plots of residual ACF and PACF (Fig. 4), all residual values at various lags were settled within tolerance interval at 95% confidence limits. This means each residual is very small relative to its standard error and shows the existence of no significant correlation between residuals. As a result, the errors obtained from the model are white noise or independent.

This graphical method of white noise test was also supported by Ljung–Box  $Q'$  and Durbin–Watson autocorrelation tests (Tab. 4). Ljung–Box  $Q'$  test for autocorrelation up to the maximum lag order of 36 was 29.1 at the  $p$ -value of 0.26. Similarly, the  $p$ -value of Durbin–Watson test for autocorrelation was 0.328. The larger  $p$ -values in these tests in comparison to significance level ( $\alpha = 0.05$ ) prove the absence of autocorrelations among residuals. Thus, the selected model is appropriate because of its capability of removing the residuals dependency from the data.

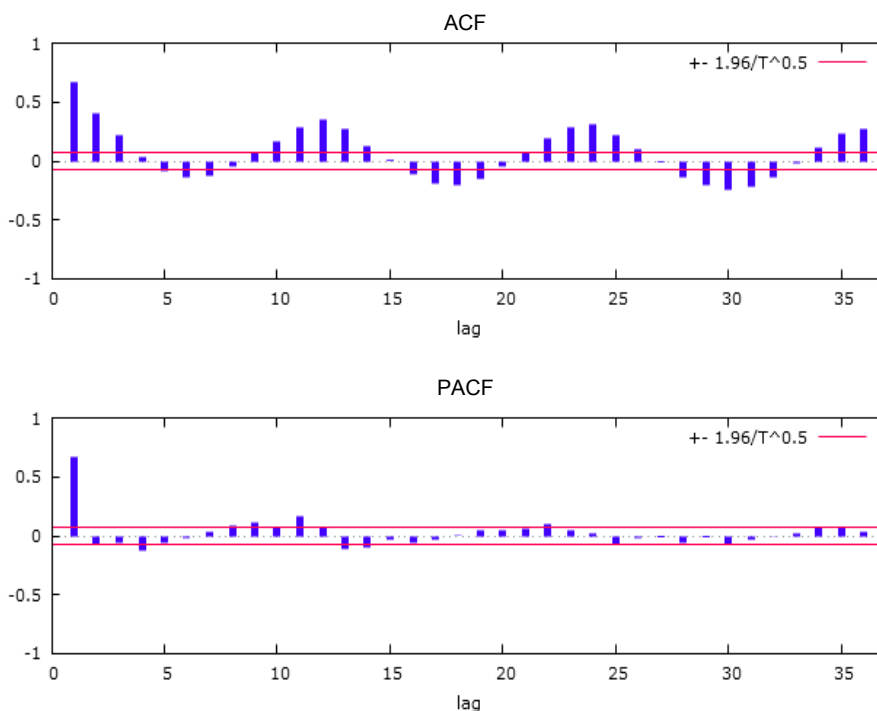


Fig. 2. Autocorrelation function (ACF) and partial autocorrelation functions (PACF) for historical Waterval River flows; source: own study

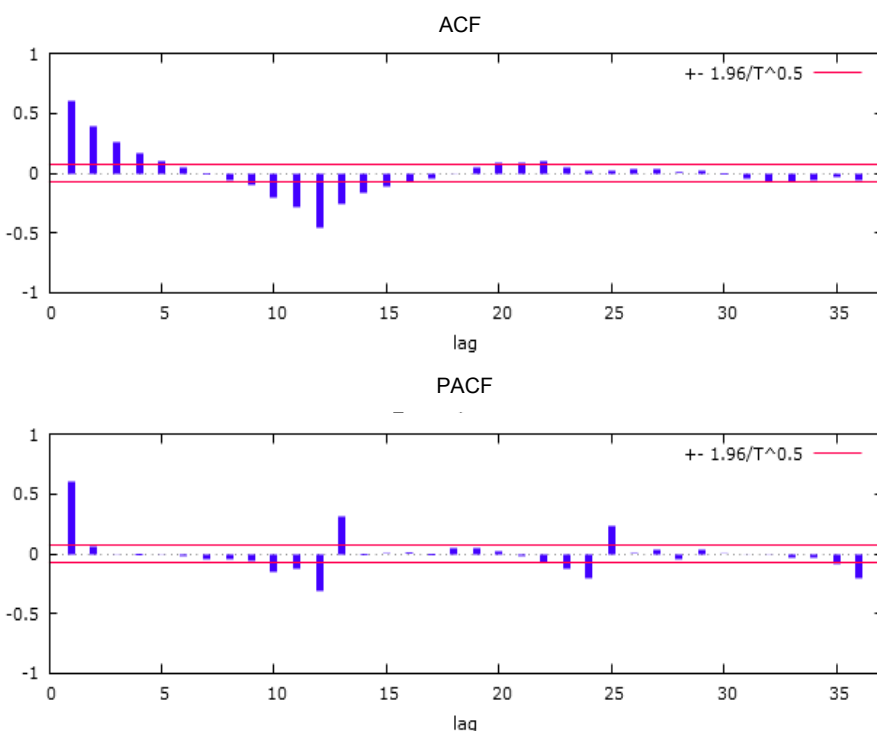


Fig. 3. Seasonally differenced autocorrelation function (ACF) and partial autocorrelation functions (PACF) for monthly Waterval River flows; source: own study

**HETEROSCEDASTICITY TESTS**

The results of heteroscedasticity test of residuals are shown in Table 4. The computed  $p$ -values of 0.804 in Breusch–Pagan’s test and 0.963 in White’s tests, were all greater than the significance level ( $\alpha = 0.05$ ). These prove that residuals are not heteroscedastic but

homoscedastic. This means the residuals from a linear SARIMA model have a variance that is dependent of the observations. Therefore, based on heteroscedastic tests, the selected model is adequate for Waterval River flow forecasting. Jarque–Bera normality test value of residuals is 4.86452 with  $p$ -value 0.0878. this proves that the residuals are normally distributed at  $\alpha = 0.05$ .

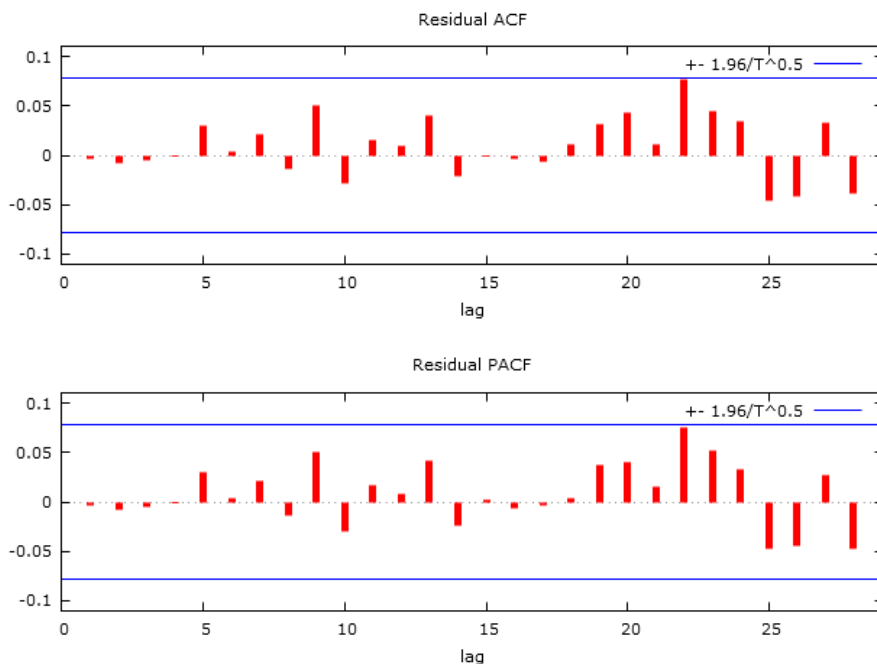


Fig. 4. Autocorrelation function (ACF) and partial autocorrelation functions (PACF) of residuals; source: own study

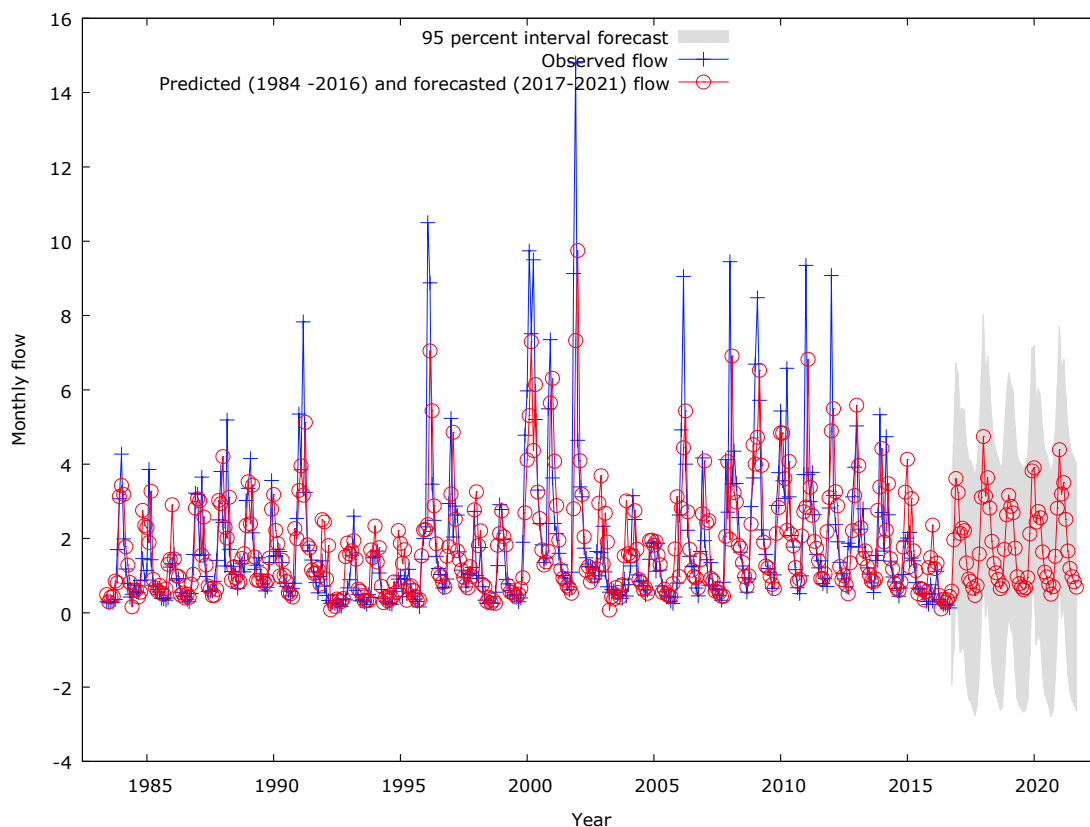


Fig. 5. Observed, predicted and forecasted mean monthly flows; source: own study

Table 4. Autocorrelation and heteroscedasticity tests

Parameter	Autocorrelation test		Heteroscedasticity test	
	Ljung-Box $Q'$	Durbin-Watson	Breusch-Pagan	White
$p$	0.26	0.328	0.804	0.963
$\alpha$	0.05	0.05	0.05	0.05

Source: own study.

SARIMA FORECASTING

A historical, validation and prediction (forecast) graph for monthly Waterval River flows time series using SARIMA (3, 0, 2) (3, 1, 3)<sub>12</sub> model is shown in Figure 5. In addition to its best performance, the pattern of the out of sample forecast graph is similar with

that of the validation and historical graph but unable to capture very high flows. In general, the selected model is appropriate for Waterval River forecasting.

From SARIMA (3, 0, 2) × (3, 1, 3)<sub>12</sub> model, the order of  $p$  is 3 means that the current time series ( $y_t$ ) is reliant on its preceding data  $y_{t-1}$ ,  $y_{t-2}$ , and  $y_{t-3}$ . The order  $P$  is 3 means that  $y_t$  is reliant on its preceding years' data of  $y_{t-12}$ ,  $y_{t-24}$ , and  $y_{t-36}$ . The order  $q$  is 2 means  $y_t$  is reliant on its preceding random shock  $\varepsilon_{t-1}$  and  $\varepsilon_{t-2}$ ; and the order  $Q$  is 3 means that  $y_t$  is reliant on its preceding random shocks of  $\varepsilon_{t-12}$ ,  $\varepsilon_{t-24}$  and  $\varepsilon_{t-36}$ . As non-seasonal component of this model was stationary, no non-seasonal differencing was used ( $d = 0$ ). However, seasonal differencing was used only once ( $D = 1$ ) to remove seasonality in the model.

## CONCLUSIONS

To aid planning and management of Waterval River system and for an optimal real-time adaptive operation of Vaal Dam reservoir, current and future Waterval River flows information is vital. To this end, Waterval River flow was modeled and forecasted using GRETl statistical software. Out of several models identified and evaluated, SARIMA (3,0,2)×(3,1,3)<sub>12</sub> model was selected for Waterval River flow forecasting for its minimum values of Akaike Information and Hannan–Quinn criteria. In addition to its best performance, the pattern of the out of sample forecast mean monthly flows graph is similar with that of the observed mean monthly flows graph. Consequently, the selected model is appropriate for Waterval River forecasting.

Thus, the forecasted mean monthly flows information and the developed model help water resource managers and decision makers for development and management of Waterval River, and optimal operation of Vaal Dam reservoir, in the Olifant basin. Comparison of forecast performance of SARIMA models with that of computational intelligent forecasting techniques was recommended for future study.

## Acknowledgements

Debre Markos University in providing study leave, the University of Johannesburg for providing Postdoctoral fellowship and Department of Water and Sanitation of South Africa for recording and making hydrological data freely accessible, are greatly acknowledged.

## REFERENCES

- ADHIKARY S.K., RAHMAN M.D.M., GUPTA A.D. 2012. A stochastic modelling technique for predicting groundwater table fluctuations with Time Series Analysis. *International Journal of Applied Sciences and Engineering Research*. Vol. 1. Iss. 2 p. 238–249.
- ADNAN R.M., YUAN X., KISI O., ANAM R. 2017. Improving accuracy of river flow forecasting using LSSVR with gravitational search algorithm. *Advances in Meteorology*. Vol. 2017. Art. ID 2391621 pp. 23. DOI 10.1155/2017/2391621.
- AKAIKE H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. Vol. 19. Iss. 6 p. 716–723.
- ALI S.M. 2013. Time series analysis of Baghdad rainfall using ARIMA method. *Iraqi Journal of Science*. Vol. 54. Iss. 4 p. 1136–1142.
- BARZEGAR R., MOGHADDAM A.A., ADAMOWSKI J., OZGA-ZIELINSKI B. 2017. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. *Stochastic Environmental Research and Risk Assessment* p. 1–15.
- BELAYNEH A., ADAMOWSKI J. 2013. Drought forecasting using new machine learning methods/Prognozowanie suszy z wykorzystaniem automatycznych samouczących się metod. *Journal of Water and Land Development*. No. 18 p. 3–12.
- BOX G.E.P., JENKINS G.M., REINSEL G.C. 2008. *Time series analysis, forecasting and control*. 4<sup>th</sup> ed. New Jersey. John Wiley and Sons, Inc. ISBN 978-0470272848 pp. 784.
- BROCKWELL P.J., DAVIS R.A. 2002. *Introduction to time series and forecasting*. 2<sup>nd</sup> ed. New York. Springer-Verlag. ISBN 978-1475777505 pp. 437.
- CHANG X., GAO M., WANG Y., HOU X. 2012. Seasonal autoregressive integrated moving average model for precipitation of time series. *Journal of Mathematics and Statistics*. Vol. 8. Iss. 4 p. 500–505.
- COTTRELL A., LUCCHETTI R.J. 2017. *Gretl user's guide*. Gnu Regression, Econometrics and Time-series Library. [online]. [Access date 11.03.2017]. Available at: <http://ricardo.ecn.wfu.edu/pub/gretl/manual/en/gretl-guide-a4.pdf>
- CRYER J.D., CHAN K.S. 2008. *Time series analysis: With applications in R*. New York. Springer-Verlag. ISBN 978-0-387-75958-6 pp. 491
- DICKEY D.A., FULLER W.A. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*. Vol. 74. Iss. 366 p. 427–431.
- DURBIN J., WATSON G.S. 1951. Testing for serial correlation in least squares regression, II. *Biometrika*. Vol. 38. Iss. 1–2 p. 159–179.
- GAUTAM R., SINHA A.K. 2016. Time series analysis of reference crop evapotranspiration for Bokaro District, Jharkhand, India. *Journal of Water and Land Development*. No. 30 p. 51–56. DOI 10.1515/jwld-2016-0021.
- GERRETSADIKAN A., SHARMA M.K. 2011. Modelling and forecasting of rainfall data of Mekele for Tigray region (Ethiopia). *Statistics and Applications*. Vol. 9. Iss. 1–2 p. 31–53.
- HAMILTON J.D. 1994. *Time series analysis*. Princeton. Princeton University Press. ISBN 978-0691042893 pp. 799.
- HANNAN E.J., QUINN B.G. 1979. The determination of the order of an autoregression. *Journal of the Royal Statistical Society*. Vol. 41 p. 190–195.
- JANHABI M., JHA R. 2013. Time-series analysis of monthly rainfall data for the Mahanadi River Basin, India. *Sciences in Cold and Arid Regions*. Vol. 5. Iss. 1 p. 73–84.
- KENDALL M. 1975. *Multivariate analysis*. London. Charles Griffin & Company. ISBN 978-0852642344 pp. 218.
- KOMORNÍKOVÁ M.A., SZOLGAY J., SVETLÍKOVÁ D.A., SZÖREOVÁ D., JURČÁK S. 2008. A hybrid modelling framework for forecasting monthly reservoir inflows. *Journal of Hydrology and Hydromechanics*. Vol. 1. Iss. 3 p. 145–62.

- KWIATKOWSKI D., PHILLIPS P.C.B., SCHMIDT P., SHIN Y. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*. Vol. 54 p. 159–178.
- LJUNG G.M., BOX G.E.P. 1978. On a measure of a lack of fit in time series models. *Biometrika*. Vol. 65. Iss. 2 p. 297–303.
- MOEENI H., HOSSIEEN B., FATEMI S.E. 2017. Stochastic model stationarization by eliminating the periodic term and its effect on time series prediction. *Journal of Hydrology*. Vol. 547 p. 348–364.
- NAILL P.E., MOMANI M. 2009. Time series analysis model for rainfall data in Jordan: A case study for using time series analysis. *American Journal of Environmental Sciences*. Vol. 5 p. 599–600.
- OSARUMWENSE O.I. 2013. Applicability of Box Jenkins SARIMA model in rainfall forecasting: A case study of Port-Harcourt south Nigeria. *Canadian Journal on Computing in Mathematics, Natural Sciences, Engineering and Medicine*. Vol. 4. Iss. 1 p. 1–4.
- PAPALASKARIS T., THEOLOGOS P., PANTRAKIS A. 2016. Stochastic monthly rainfall time series analysis, modeling and forecasting in Kavala City, Greece, North-Eastern Mediterranean Basin. *International Conference on Efficient and Sustainable Water Systems Management toward worth Living Development, 2<sup>nd</sup> EWaS 2016*. *Procedia Engineering*. Vol. 162 p. 254–263.
- PEKÁROVÁ P., ONDERKA M., PEKÁR J., RONČÁK P., MIKLÁNEK P. 2009. Prediction of water quality in the Danube River under extreme hydrological and temperature conditions. *Journal of Hydrology and Hydromechanics*. Vol. 1. Iss. 57 p. 3–15.
- PHILLIPS P.C.B., PERRON P. 1988. Testing for a unit root in time series regression. *Biometrika*. Vol. 75. Iss. 2 p. 335–346.
- SALAS J.D., DELLEUR J.W., YEVJEVICH V., LANE W.L. 1980. *Applied modelling of hydrologic time series*. Littleton, Colorado. Water Resource Publications pp. 484.
- SCHWARZ G.E. 1978. Estimating the dimension of a model. *Annals of Statistics*. Vol. 6. Iss. 2 p. 461–464.
- SINGH M., SINGH R., SHINDE V. 2012. Application of software packages for monthly stream flow forecasting of Kangsabati River in India. *International Journal of Computer Applications*. Vol. 20. Iss. 3 p. 7–14.
- TARIQ M.M., ABBASABD A.I. 2016. Time series analysis of Nyala rainfall using ARIMA method. *SUST Journal of Engineering and Computer Science*. Vol. 17. Iss. 1 p. 5–11.
- TEYMOURI M., FATHZADEH A. 2015. Stochastic modeling of monthly river flow forecasting (Case study: Atrak River Basin, Iran). *Journal of Selçuk University Natural and Applied Science*. Vol. 4. Iss. 2 p. 38–48.
- TIWARI M., ADAMOWSKI J., ADAMOWSKI K. 2016. Water demand forecasting using extreme learning machines. *Journal of Water and Land Development*. No. 28 p. 37–52. DOI 10.1515/jwld-2016-0004.
- VAHDAT S.F., SARRAF A., SHAMSNIA A., SHAHIDI N. 2011. Prediction of monthly mean Inflow to Dez Dam reservoir using time series models (Box-Jenkins). *International Conference on Environment and Industrial Innovation IPCBEE*. 12 (2001). Singapore. IACSIT Press p. 162–166.
- VALIPOUR M. 2015. Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications*. Vol. 22. Iss. 3 p. 592–598.
- WANG J., DU Y.H., ZHANG X.T. 2008. *Theory and application with seasonal time series*. 1<sup>st</sup> ed. Nankai. Nankai University Press.
- YADAV B., CH S., MATHUR S., ADAMOWSKI J. 2017. Assessing the suitability of extreme learning machines (ELM) for groundwater level prediction. *Journal of Water and Land Development*. No. 32 p. 103–112. DOI 10.1515/jwld-2017-0012.
- YANG T., ASANJAN A.A., WELLES E., GAO X., SOROOSHIAN S., LIU X. 2017. Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. *Water Resources Research*. Vol. 53. Iss. 4 p. 2786–2812.

---

## Kassahun Birhanu TADESSE, Megersa Olumana DINKA

### Zastosowanie modelu SARIMA do prognozowania miesięcznych przepływów rzeki Waterval w Południowej Afryce

#### STRESZCZENIE

Znajomość przyszłego przepływu wody w rzece jest istotna dla rozwoju i zarządzania w systemie rzeczny. W badaniach prezentowanych w niniejszym artykule prognozowano przepływ w rzece Waterval w Republice Południowej Afryki, używając modelu SARIMA i programu statystycznego GRETL. Do modelowania i budowania prognoz wykorzystano średnie miesięczne przepływy z lat 1960–2016. Różne pierwiastki jednostkowe i analiza trendu Manna–Kendalla dowiodły stacjonarności obserwowanych szeregów czasowych przepływu. Na podstawie sezonowo zróżnicowanych charakterystyk korelogramu oceniono różne modele SARIMA zoptymalizowano ich parametry i wykonano diagnostyczne sprawdzenie prognoz za pomocą białego szumu i testów heteroscedastyczności. Na podstawie minimum *AI* i kryteriów Hannana–Quinna (*HQ*), wybrano model SARIMA (3, 0, 2) × (3, 1, 3)<sub>12</sub> do prognozowania przepływu w rzece Waterval. W dalszych badaniach proponuje się porównanie prognozowania za pomocą modeli SARIMA i technik komputerowych.

**Słowa kluczowe:** analiza trendu, biały szum, heteroscedastyczność, ocena, test stacjonarności