European Water 57: 93-100, 2017.

Time series analysis of water characteristics of streams in Eastern Macedonia – Thrace, Greece

T. Papalaskaris[1*] and G. Kampas[2]
[1] *Department of Civil Engineering, Democritus University of Thrace, Kimmeria Campus, 67100 Xanthi, Greece*
[2] *Deputy Head of Water Directorate of Eastern Macedonia - Thrace, 65404 Kavala, Greece; e-mail: dy-amt@damt.gov.gr*
[*] *e-mail: tpapalas@civil.duth.gr*

Abstract: Water is one of the most important sources of water on earth supporting the existence of the majority of living organisms. Time series analysis, modeling and forecasting constitutes a tool of paramount importance with reference to a wide range of scientific purposes both in meteorology and hydrology (e.g. precipitation, water quality, humidity, temperature, solar radiation, floods and draughts). The present research applies the Box-Jenkins approach, employing SARIMA (Seasonal Autorregressive Integrated Moving Average) model to perform short term forecasts of maximum monthly water temperatures Nestos River at the junction with its trans-boundary tributary, Despatis River, at the specific location called Nestos River – Gefira Despati, situated within the range of Paranesti area, Drama Prefecture, Region of Eastern Macedonia-Thrace, North-Eastern Greece, North-Eastern Mediterranean Basin, modeling past maximum monthly water temperatures time series components structure and predicting future quantities in accordance to the past ones. The model which is mostly fit to both describe the past water quality data and thus generate the most reliable future forecasts is selected and rated by means of the (BIC-), (MAPE), (MAE) and (RMSE) model evaluation criteria. The conclusions of this research will provide local authorities (e.g. Water Directorate of Eastern Macedonia – Thrace, General Secretariat for Civil Protection, European Center for Forest Fires, Deputy Governor of Agricultural Economy, daily fire risk maps designers, hydraulic, irrigation and environmental engineers, city inhabitants, farmers etc., biologists, fish scientists etc.) valuable data in order to develop strategic plans, policies and appropriate use of available water resources within the Region of Eastern Macedonia-Thrace.

Key words: Water temperature time series forecasting; auto regressive moving average models; trend; seasonality; SARIMA models

1. INTRODUCTION

Water temperature is undoubtedly world over considered as one of the most important water quality indicator. It is a significant water quality and environmental aspect because it essentially controls the range of kinds and types of aquatic life, determines the maximum dissolved oxygen density in the water, and affects the rate of chemical and biological reactions. It can be affected both by natural causes, (e.g. sun, precipitation, surface run-off, groundwater and upstream headwaters inputs, reciprocal transmissions of heat to the air and heat received or transmitted by evaporation or condensation) as well as by artificial causes (e.g. discharge of cooling water and/or heated industrial effluents, alterations to streams shading due to agriculture and water harvesting, urban development and climate change).Water temperature in streams and rivers changes over time much shorter time in comparison to those of lakes. Their seasonal and diurnal water temperatures are much more closely connected with the specific atmospheric temperature patterns, established within the surrounded area climate, in accordance with the size and origin of those natural watercourses than do those of lakes. A variety of statistical procedures are often employed to forecast rainfall amounts. In warm water streams the temperatures should not exceed 31.67 ^{o}C, whilst at the same time in cold water streams the temperatures should not exceed 20.00 ^{o}C, since often heat occurring during summer can provoke fish kills, especially those inhabiting stream ponds due to the fact that the more the stream water temperature increases, the less dissolved oxygen (crucial for aquatic life survival) is available into the stream water. One of the most common methods for time series data analysis is that which was developed within the general concept of stochastic hydrology, also known with the name of ARIMA (Autoregressive Integrated Moving

Average) (Box and Jenkins, 1970). ARIMA modeling procedure has been applied the world over on a significant number number of not only financial but additionally, hydrological time series data in order to forecast rainfall data, water reservoir inflow/outflow discharge patterns as well as river flows and river water characteristics modeling (Bari et al., 2015; Papalaskaris et al., 2016).

2. METHODOLOGY

2.1 Study area

With the view to investigate maximum monthly temperature patterns in Despatis river, a trans-boundary stream, originating in Bulgaria and discharging into Nestos River, daily recorded observations from the only one available, operating conventional gauging station (location name: "Nestos River – Despatis Bridge") which records dissolved oxygen, water conductivity, water temperature, turbidity, pH, O.R.P. (oxidation reduction potential) and water level, located at 41.404956° N and 24.100250° E (WGS84 coordinates), were analyzed, which started its operation by the 1st of April 2013 covering a time interval period of 3 years and 3 months (1st of April 2013 - 31st of July 2016) as depicted in Figure 1.

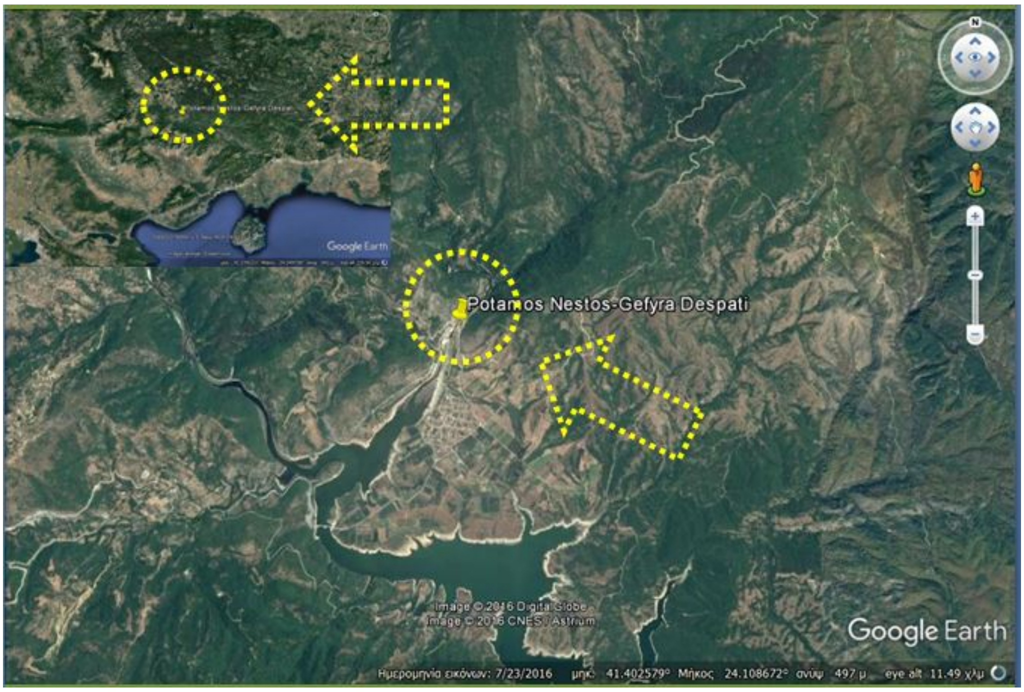

Figure 1. Nestos River- Despatis Bridge water quality gauging station, Paranesti area, Drama Prefecture, Greece (Source: Google Earth).

2.2 The Box-Jenkins model-building procedure

The statisticians Box and Jenkins (1970) developed a modeling method, primarily for financial time series analysis, dealing with stationary time series, and fitting either autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) or seasonal autoregressive integrated moving average (SARIMA) models with the view to discover the most appropriate match of a time series data to previous values of the same time series, with the view to perform future predictions and forecasts. The model-building procedure incorporates the following successive stages (Box and Jenkins, 1970; Abdul-Aziz et al., 2013).

2.2.1 Model identification and selection

Verifying the stationarity of the variables, tracing and locating seasonality if exists, within the time series data under investigation (situation which can be treated by seasonal differencing) and interpreting charts of autocorrelation and partial autocorrelation functions of the time series data under examination with the view to conclude which autoregressive or moving average constituent would be the most appropriate to take place in the model.

2.2.2 Model parameters estimation

Using calculation procedures in order to assay the most competent coefficients for the preferable ARIMA model, in most cases, by means of computation methods like either maximum likelihood estimation or least-squares estimation.

2.2.3 Model checking and forecasting

By examining whether the elaborated model complies with the requirements of a stationary univariate process, namely, the residuals should be independent between each other and exhibit constancy in terms of mean and variance along the entire length of the time series, as the time passes by; this can be carried out by plotting the mean and variance of residuals over time and executing a Ljung-Box test or/and charting autocorrelation and partial autocorrelation functions of the residuals as a means to verify whether the model we built best fit our time series data or not.

2.3 Type of ARIMA models

2.3.1 Autoregressive (AR) and Moving Average (MA) models

In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. The term autoregression indicates that it is a regression of the variable against itself (Tamura, 2004; Hyndman and Athanasopoulos, 2012; PennState Eberly College of Science, 2017). Thus an autoregressive model of order p can be written as,

$$y_t = c + \varphi_1 \times y_{t-1} + \varphi_2 \times y_{t-2} + \cdots + \varphi_p \times y_{t-p} + e_t \qquad (1)$$

where c is a constant and e_t is white noise. This procedure resembles a multiple regression but with lagged values of y_t as predictors. Reference is made to this type of model as an AR(p) model.

2.3.2 Autoregressive-Moving-Average (ARMA) models

Instead of using past values of the forecast variable in a regression process, a moving average model uses past forecast errors within a regression-resembling model building (Hyndman and Athanasopoulos, 2012),

$$y_t = c + e_t + \theta_1 \times y_{t-1} + \theta_2 \times y_{t-2} + \cdots + \theta_q \times y_{t-q} \qquad (2)$$

where e_t is white noise. We make reference to this type of model as an MA(q) model.

2.3.3 Seasonal (ARIMA) models

ARIMA models possess also the ability to model seasonal data of a great extent. A seasonal ARIMA model, or so- called SARIMA model, is generated by incorporating supplementary seasonal components in the ARIMA models we have already mentioned above, and it can be written in the following form (Hyndman and Athanasopoulos, 2012; Nau, 2017):

$$\text{ARIMA (p,d,q,) (P,D,Q,)}_{m} \tag{3}$$

with p=non-seasonal (AR) order, d=non-seasonal differencing, q=non-seasonal (MA) order, P=seasonal (AR) order, D=seasonal differencing, Q=seasonal (MA) order, and where m=number of periods per season or in other words, time interval of repeating seasonal behavior. The seasonal components of the model are written by means of uppercase letters, whilst, on the contrary, we write the non-seasonal components of the model via lowercase letters. The seasonal portion of the model is comprised of components greatly resembling the non-seasonal terms of the model, yet they include backshift operators of the seasonal period. The additional seasonal components are multiplied through a simple way with the non-seasonal components of the model.

3. RESULTS

Primordial time series analysis was carried out dealing with maximum monthly temperature data recorded from 01.04.2013 to 31.07.2016 employing the Box-Jenkins ARIMA model-building procedure.

3.1 Primordial data analysis

By definition, a time series is considered non-stationary, when its values mean and variance either do not exhibit stability over time in terms of mean and variance; The ACF and PACF charts inspection can also provide clues of trend and seasonality existence (Senter, 2008; Abdul-Aziz et al., 2013). We selected to specify 38 autocorrelation lags concerning the maximum monthly water temperature data although Box and Jenkins (1970) suggested that the estimated autocorrelations would be calculated for k=0,1,…k, was not larger than N/4 (40/4=10) series data. Seasonality often results to non-stationary time series owing to the fact that the average values along different parts of the time series occurring within the seasonal time interval differ from the average values along other segments of the same time series. A practical rule, usually employed, while interpreting an ACF chart is if there are designed autocorrelation bars that are larger in value than two standard errors apart from the zero mean, then they suggest proxy of autocorrelation which is statistically significant. The ACF plot reveals alternative positive and negative values, decaying to zero, suggesting the use of an autoregressive model (Smith, 2015). In Figure 3, there are three charted autocorrelation values, at lags 1, 2 and 3, that either extend more (or close to) than two standard errors from the mean, represented by the zero value whilst the two continues lines designed above and below the zero mean stand for the approximate 95% confidence limits.

Visually inspecting, an anticipated conventional 6-month seasonal pattern, following a cyclical pattern between the summer period during which the recorded water temperature values are at their highest levels, and the winter period during which the water temperature values reach its lowest levels, can be definitely both traced and recognized.

Furthermore it should be noted that the performance of the time series analysis in the spectral domain, and after having examined the spectral density graph, proved that the largest seasonal pattern takes place at approximately 12.00-months time intervals, instead of 6-months intervals (Senter, 2008). Hence, we conclude that charting the ACF plot the seasonality pattern is unfolded which couldn't be discovered by the simple linear equation describing the linear trend line associated with the original raw maximum monthly recorded water temperature data. This pattern can be easily identified in the following Figure 4, focusing on the high peak at close the frequency 0.1.

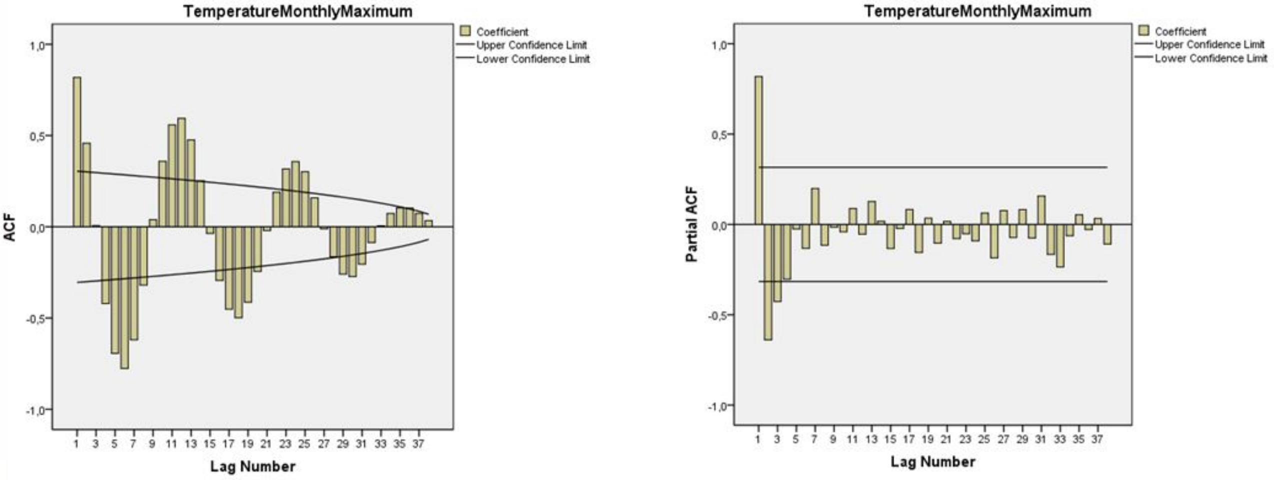

Figure 3. ACF and PACF charts of the raw data of the original observed maximum monthly water temperature, at Nestos River – Gefira Despati - Paranesti area station, Drama Prefecture, Greece.

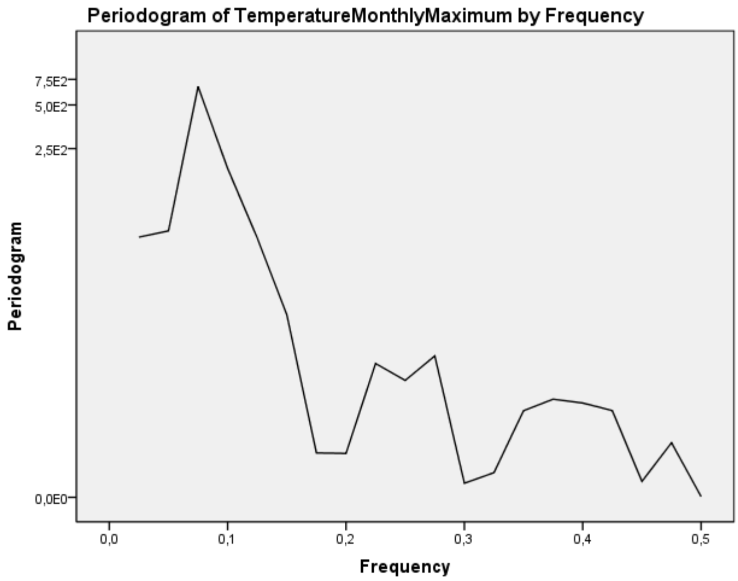

Figure 4. Spectral density chart of the raw data of the original observed maximum monthly water temperature, at Nestos River – Gefira Despati - Paranesti area station, Drama Prefecture, Greece.

3.2 Data stationarity check

The Augmented Dickey-Fuller Unit Root Test, expressed within three different models, was employed and applied on the entire maximum monthly recorded water temperature time series data with the view to investigate and verify whether the rainfall time series is stationary or not. The Figure 5, depicts the outcomes of the test: In the first model (intercept only), the test statistic value -5.140 is lower in value than critical values, -3.655, -2.961 and -2.613 all at 1%, 5% and 10% correspondingly; moreover, the regression coefficient L1 is negative in value (L1=-0.8365932), hence, we can accept the model as a valid one. In the second model (trend and intercept), the test statistic value -5.329 is lower in value than critical values, -4.251, -3.544 and -3.206 all at 1%, 5% and 10% in consequence; in addition, the regression coefficient L1 is once again negative in value

(L1=-0.8677031), consequently, the model can be definitely considered as a valid one. In the third and last model (neither trend nor intercept), the test statistic value -2.830 is lower in value than critical values, -2.638, -1.950 and -1.606 all at 1%, 5% and 10% successively; furthermore, the regression coefficient L1 is once again for this last model negative in value (L1=-0.3478649), therefore, the model can be definitely accepted as a valid one. Evidently, after having performed, examined and checked all three different Augmented Dickey-Fuller Unit Root Tests we concluded in all cases that we reject the null hypothesis H_0, and we accept the alternative hypothesis H_1, which declares that the variable Y (maximum monthly recorded water temperature) is stationary and does not have a unit root.

```
. dfuller Y, regress lags(0)

Dickey-Fuller test for unit root                   Number of obs   =        39

                               ---------- Interpolated Dickey-Fuller ----------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -5.140            -3.655            -2.961            -2.613
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
         D.Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           Y |
         L1. |  -.8365932   .1627485    -5.14   0.000    -1.166353   -.5068334
             |
       _cons |   46.90517   11.95831     3.92   0.000     22.67534    71.13501
------------------------------------------------------------------------------
```

Figure 5. ADF test applied on the raw data (intercept only) of the original observed maximum monthly water temperature, at Nestos River – Gefira Despati - Paranesti area station, Drama Prefecture, Greece.

3.3 Data differencing (seasonality)

Although the ADF test applied on the maximum water temperatures recorded monthly time series of the raw data revealed it is a stationary time series, after examining the ACF chart depicted in the Figure 3, we considered that the time series data should be seasonally differenced, (due to a few spikes cut the 95% confidence limits), by order D=1, in order to eliminate seasonality (Senter, 2008; Abdul-Aziz et al., 2013).

3.4 Model identification

As soon as several tests concerning the maximum monthly recorded water temperatures have been completed a few candidate models were considered that best meet the criteria and we finally concluded that the seasonal ARIMA (SARIMA) $[(0,1,0)X(0,1,0)_{12}]$ is considered the most suitable one appearing to have the highest R-squared (R^2) (0.853), whilst simultaneously the least Normalized B.I.C. (1.586), M.A.P.E. (17.077), M.A.E. (1.820), Root Mean Squared Error (R.M.S.E.) (2.079) (Senter, 2008; Abdul-Aziz et al., 2013) .

3.5 Model adequacy and diagnostic tests

By examining in Figure 6 the autocorrelation and partial autocorrelation charts of the residuals we verify that there aren't any bar peak values that significantly extend beyond the two continuous lines, plotted above and below the zero mean which imply the approximate 95% confidence limits (Abdul-Aziz et al., 2013; Bari et al., 2015).

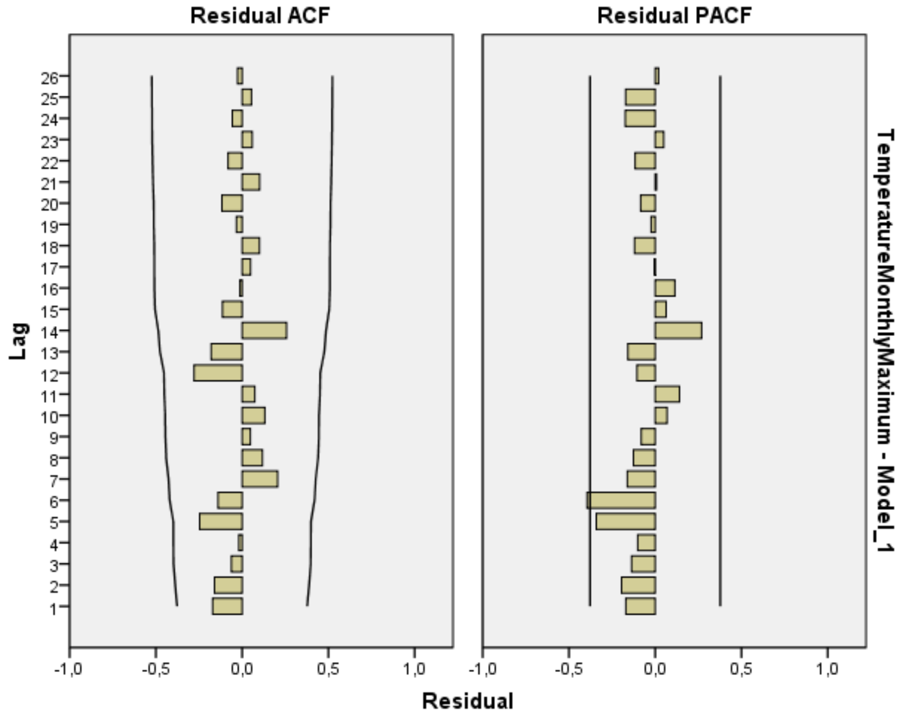

Figure 6. ACF and PACF charts of SARIMA [(0,1,0)X(0,1,0)$_{12}$] model Residuals, at Nestos River – Gefira Despati - Paranesti area station, Drama Prefecture, Greece.

3.6 Forecasting future values

In the Table 1 are depicted the forecasted values :

Table 1. Forecasted values of SARIMA [(0,1,0)X(0,1,0)$_{12}$] model, at Nestos River – Gefira Despati - Paranesti area station, Drama Prefecture, Greece.

SARIMA Model	Aug 2016	Sep 2016	Oct 2016	Nov 2016	Dec 2016	Jan 2017	Feb 2017	Mar 2017	Apr 2017	May 2017	Jun 2017	Jul 2017
Forecast	21.84	19.47	15.61	12.05	7.39	5.42	10.06	10.70	16.53	19.37	23.31	24.14

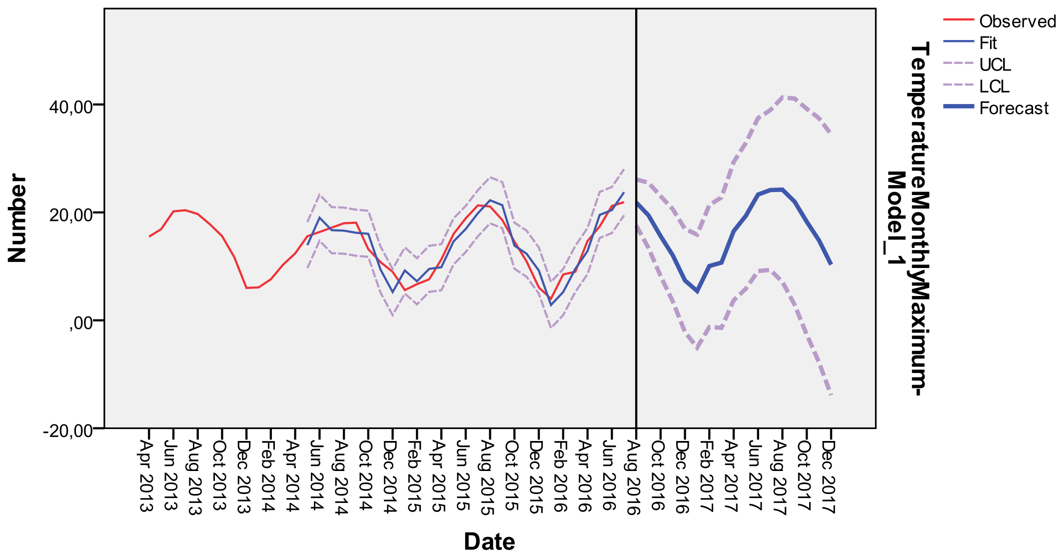

Figure 7. Forecast chart of SARIMA [(0,1,0)X(0,1,0)$_{12}$] model Residuals, at Nestos River – Gefira Despati - Paranesti area station, Drama Prefecture, Greece.

The observed, fit and forecast values as well as the upper and low confidence levels with reference to the finally selected SARIMA [(0,1,0)X(0,1,0)$_{12}$] model, are all summarized within the chart illustrated by Figure 7 (Senter, 2008; IBM Corporation, 2011; Abdul-Aziz et al., 2013; Bari et al., 2015).

4. CONCLUSIONS

The stream flow water maximum monthly temperature patterns, observed at Nestos River, Gefira Despati, Paranesti Region gauging station have been studied using the Box-Jenkins (SARIMA) model methodology. Adaili water temperature record spanning the period of April 2013 – July 2016 has been used to develop the most fit model. After having examined a group of candidate SARIMA models and having compared them based on several statistical criteria (e.g. R-squared, B.I.C., M.A.E., M.A.P.E., R.M.S.E.) we concluded that SARIMA [(0,1,0)X(0,1,0)$_{12}$] model best fits the maximum recorded monthly water temperature data of Nestos River – Gefira Despati, Paranesti area, Drama Prefecture, NE Greece, for the period 01.04.2013-31.07.2016. The study reveals that the Box-Jenkins (SARIMA) model methodology could be implemented as a suitable tool to simulate the maximum monthly water temperature at Nestos River, Gefira Despati, Paranesti Region gauging station for the up-coming months (up to July 2017). Although, undoubtedly, the operation time span can be considered relatively short in order to come up with secure conclusions, the continuous future gauging station operation will feed the modeling procedure with more raw data in an attempt to develop a more adequate model. The results achieved for water temperature forecasting will assist scientists and decision makers to develop policies concerning river pollution, fish farming, river fauna and flora sustainability etc.

REFERENCES

Abdul-Aziz A.R., Anokye M., Kwame A., Munyakazi L. and Nsowah-Nuamah N.N.N. (2013). Modeling and forecasting rainfall pattern in Ghana as a seasonal ARIMA process: The case of Ashanti Region. International Journal of Humanities and Social Science, 3 (3), 224-233.

Bari S.H., Rahman M.T., Hussain M.M. and Ray S. (2015). Forecasting monthly precipitation in Sylhet City using ARIMA model. Civil and Environmental Research, 7 (1), 69-77.

Box G.E.P. and Jenkins G. (1970) (Revised edition: 1976). Time Series Analysis: Forecasting & Control, Holden-Day Publishments, San Fransisco

Hyndman R.J., G. Athanasopoulos. (2012) 8.3: Autoregressive models. Available from: https://www.otexts.org/fpp/8/3 [Accessed 17th January 2016].

Hyndman R.J., G. Athanasopoulos. (2012) 8.4: Moving average models. Available from: https://www.otexts.org/fpp/8/4 [Accessed 17th January 2016].

Hyndman R.J., G. Athanasopoulos. (2012) 8.9: Seasonal ARIMA models. Available from: https://www.otexts.org/fpp/8/9 [Accessed 17th January 2016].

IBM Corporation (1989, 2011). IBM SPSS Forecasting 20, 95p

Nau R. (2017) ARIMA Models for time series forecasting. Available from: https://people.duke.edu/~rnau/seasarim.htm [Accessed 17th January 2016].

Papalaskaris T., Panagiotidis T. and Pantrakis A. (2016). Stochastic monthly rainfall time series analysis, modeling and forecasting in Kavala city, Greece, North-Eastern Mediterranean Basin. Proceedings 2nd International Conference on Efficient & Sustainable Water Systems Management toward Worth Living Development ("2nd E.W.a.S. 2016") (eds. V. Kanakoudis, G. Karatzas and E. Keramaris), Chania, Crete, Greece, 2016, Theme F: Hydrology, ID 064.

Papalaskaris T., Panagiotidis T. and Pantrakis A. (2016). Stochastic monthly rainfall time series analysis, modeling and forecasting in Kavala city, Greece, North-Eastern Mediterranean Basin. Procedia Engineering, 162, 254-263.

PennState Eberly College of Science. (2017) 2.1: Moving Average Models (MA models). Available from: https://onlinecourses.science.psu.edu/stat510/node/48 [Accessed 17th January 2016].

Senter A. (2008) Time series analysis. Available from: http://userwww.sfsu.edu/efc/classes/biol710/timeseries/timeseries1.htm [Accessed 17th January 2016].

Smith M. J. (2015) ARMA and ARIMA (Box-Jenkins) models. Available from: http://www.statsref.com/HTML/index.html?arima.html [Accessed 17th January 2016].

Tamura H. (2004) Chapter 6: Autoregression. Available from: http://faculty.washington.edu/htamura/qm500/king/Ch06.pdf [Accessed 17th January 2016].