Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 162 (2016) 254 – 263

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference on Efficient & Sustainable Water Systems Management toward Worth Living Development, 2nd EWaS 2016

Stochastic monthly rainfall time series analysis, modeling and forecasting in Kavala city, Greece, North-Eastern Mediterranean basin

Thomas Papalaskaris[a,*], Theologos Panagiotidis[b], Athanasios Pantrakis[c]

[a]Department of Civil Engineering-Democritus University of Thrace, Xanthi 67100,Greece
[b]Eastern Macedonia & Thrace Institute of Technology, Kavala 65404, Greece
[c]Deputy Head of Department of Technical Support & Natural Resources-Decentralized Administration of Macedonia & Thrace, Kavala 65404, Greece

Abstract

Rainfall 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 in meteorology (e.g. precipitation, 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 monthly rainfall in Kavala city, Kavala Prefecture, Region of Eastern Macedonia-Thrace, North-Eastern Greece, North-Eastern Mediterranean Basin, modeling past rainfall time series components structure and predicting future quantities in accordance to the past. The model which is mostly fit to both describe the past rainfall data and thus generate the most reliable future forecasts is selected rated by means of both the AIC- and BIC- (SBC-) model evaluation criteria. The conclusions of this research will provide local authorities (e.g. 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.) to develop strategic plans, policies and appropriate use of available water resources in Kavala city district.

Peer-review under responsibility of the organizing committee of the EWaS2 International Conference on Efficient & Sustainable Water Systems Management toward Worth Living Development

* Corresponding author. Tel.: +30-6977-507545; fax: +30-2510-250917.
E-mail address: tpapalas@civil.duth.gr

Peer-review under responsibility of the organizing committee of the EWaS2 International Conference on Efficient & Sustainable Water Systems Management toward Worth Living Development
doi:10.1016/j.proeng.2016.11.054

*Keywords:*Rainfall time series forecasting; auto regressive moving average models; trend; seasonality; SARIMA models.

1. Introduction

Rainfall can be definitely considered as a non-linear natural process raising significant difficulties especially while trying to forecast future values. Climate changes also provoke unfavorable conditions, worsening the problem causing the rainfall patterns to shift, sometimes quickly, from one state to another. Agricultural policy, crop engineering, tourism perspectives, flood protection civil works scheduling, rainwater harvesting strategy, urban water manipulation strategy, water storage reservoir capacity design and a great number of other pluralistic activities related to water resources management are strongly influenced by both short-term as well as long-term rainfall predictions and forecasts.A variety of statistical procedures are often employed to forecast rainfall amounts. One of the mostly widespread methods for time series data analysis is that elaborated within the general context of stochastic hydrology [1], also bearing the name of ARIMA (Autoregressive Integrated Moving Average). ARIMA modeling technique has been applied worldwide on a great 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 modeling [2].

Nomenclature

AR	Autoregressive models
MA	Moving Average models
ARMA	Autoregressive Moving Average models
ARIMA	Autoregressive Integrated Moving Average models
SARIMA	Seasonal Autoregressive Integrated Moving Average models
c	constant
e_t	white noise
$\phi_1,...,\phi_p$	Autoregressive modelsparameters
$\theta_1,...,\theta_q$	Moving Average models parameters
p,d,q	Non-seasonal part of Seasonal Autoregressive Integrated Moving Average models
P,D,Q,	seasonal part of Seasonal Autoregressive Integrated Moving Average models
ACF	Autocorrelation Function
PACF	PartialAutocorrelation Function
ADF	Augmented Dickey-Fuller Unit Root Test

2. Methodology

2.1. Study area

With the view to investigate rainfall patterns in Kavala city area, a seaside city, facing the Aegean Sea to the South, and surrounded by the Lekani mountain series ramifications to the North and East and the Paggaion Mountain branches to the West, total monthly recorded observations from the only one available, operating meteorological station, located at 40°56’25” N and 24°24’01” E, were analyzed, covering a time interval period of 10 years as depicted in Figure 1.

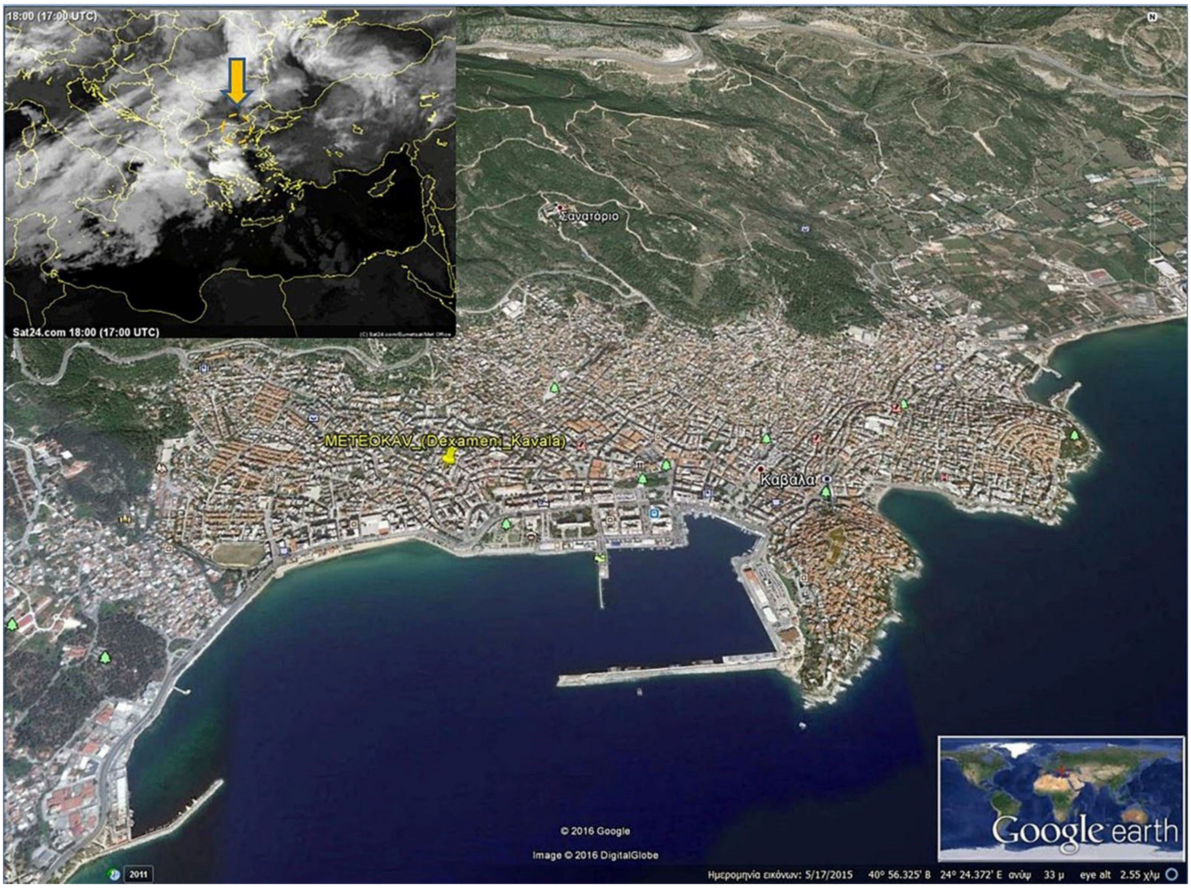

Fig. 1. Map of the district of Kavala city, Greece. Created using Google Earth and miscellaneous mapping (http://www.meteokav.gr/).

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 [3]:

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. Types 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 [4]. Thus an autoregressive model of order p can be written as,

$$y_t = c + \phi_1 \times y_{t-1} + \phi_2 \times y_{t-2} + \ldots + \phi_p \times y_{t-p} + e_t \tag{1}$$

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 [5],

$$y_t = c + \theta_1 \times y_{t-1} + \theta_2 \times y_{t-2} + \ldots + \theta_q \times y_{t-q} + e_t \tag{2}$$

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 [6]:

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

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

The primary aim of the present investigation is to find the potential patterns of flood and drought cycles occurred within the last decade in the city of Kavala, western part of Region of Eastern Macedonia and Thrace, North-Eastern Greece, North-Eastern Mediterranean Basin by employing mathematical and statistical models as well as by means of time series analysis techniques. SPSS, JMP and Stata software packages were employed to build and estimate the model parameters of the seasonal ARIMA (SARIMA) $[(0,0,0)X(0,1,1)_{12}]$ and the seasonal ARIMA (SARIMA) $[(0,0,0)X(1,1,1)_{12}]$ models of total recorded monthly rainfall in Kavala city, Greece.

3.1. Primordial data analysis

Primordial time series analysis was carried out dealing with total monthly rainfall data recorded from 2006 to 2014 employing the Box-Jenkins ARIMA model-building procedure. In the beginning a fairly simple time series chart was plotted using the unprocessed recorded data (total monthly rainfall data versus time/months) with the view to trace, by first sight, whether the mean or the variance of the time series either exhibits constancy over time or shifts as the time passes by, yielding the following Figure 2 [3, 7].

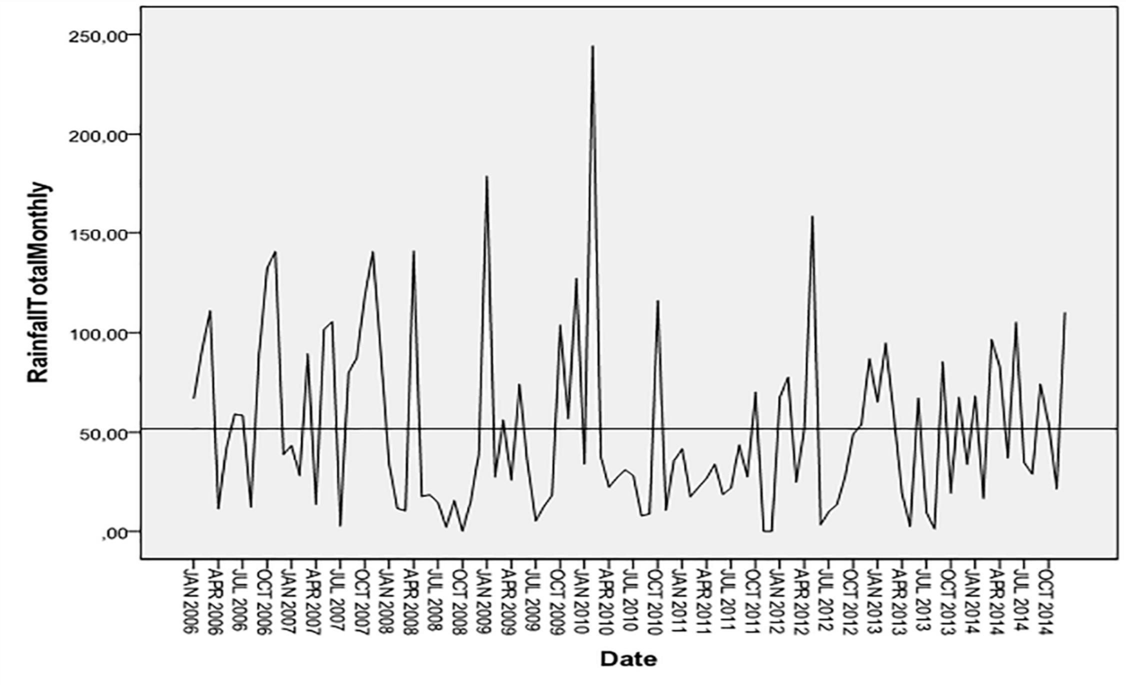

Fig. 2. Time series plot.

By first sight, the chart evidences that the time series seems to be stationary. 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. I selected to specify 72 autocorrelation lags concerning the total monthly rainfall 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 (108/4=27 regarding this particular case study) whilst their complementary suggestion that we would need at least 50 observations is also fulfilled since 108>50. The ACF and PACF charts examination shows evidences that seasonality trend exists which has to be eliminated with the view to achieve stationarity of the time 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 usual practical rule 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 [3,7]. The ACF plot reveals alternative positive and negative values, decaying to zero, suggesting the use of an autoregressive model [8]. In Figure 3, there are five charted autocorrelation values, at lags 13, 27, 42, 57 and 70, 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.

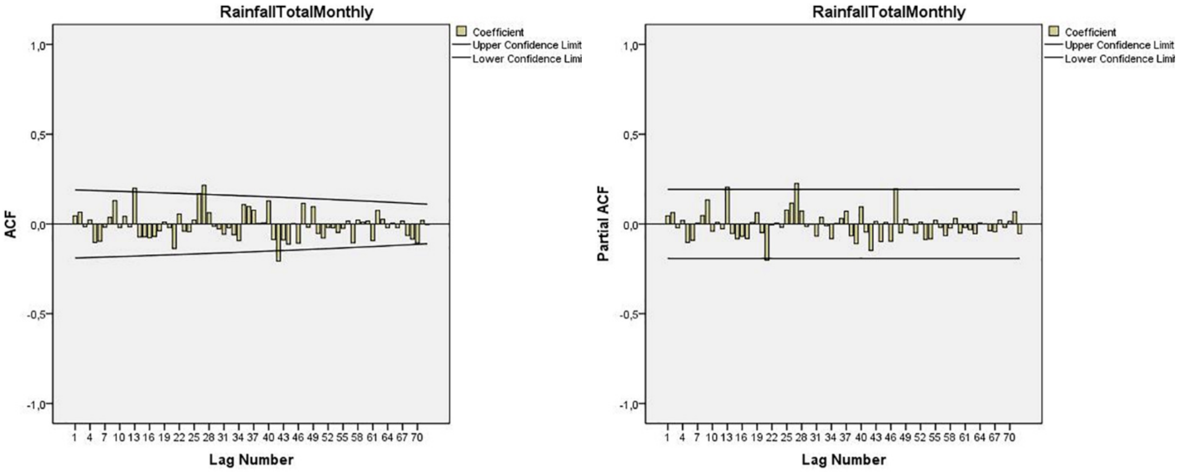

Fig. 3. ACF and PACF charts of the raw data.

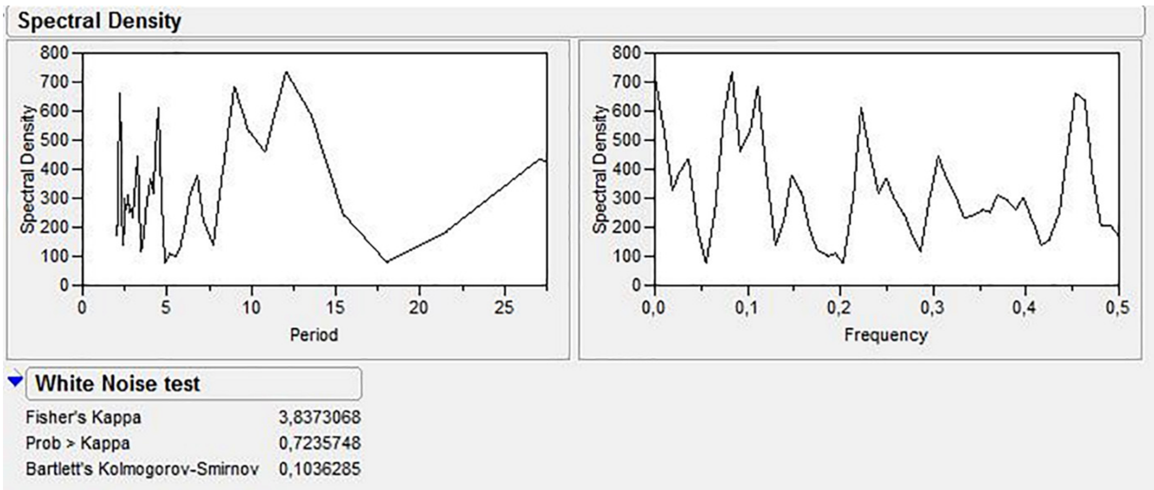

Fig. 4. Spectral density charts.

Visually inspecting, an anticipated conventional 6-month seasonal pattern, following a cyclical pattern between the summer period during which the recorded rainfall levels are very low or zero, and the winter period during which the rainfall reaches its highest levels, cannot be definitely either traced or recognized. On the contrary, we undoubtedly delineate that as an approximate 14-month seasonal pattern which might suggest a regional rainfall behavior, influenced from some additional local meteorological parameters, especially associated with the Mediterranean Basin which, in turn, contributes to that particular total monthly rainfall pattern appearance. 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 12.50-months time intervals, instead of 6-months intervals [7]. 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 total monthly recorded rainfall data. This pattern can be easily identified in the following Figure 4 (left chart), focusing on the high peak at close the period 12.

3.2. Data stationarity check

The Augmented Dickey-Fuller Unit Root Test, expressed within three different models, was employed and applied on the entire total monthly recorded rainfall 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 -10.284 is lower in value than critical values, -3.504, -2.889 and -2.579 all at 1%, 5% and 10% correspondingly; moreover, the regression coefficient L1 is negative in value (L1=-0.9548222), hence, we can accept the model as a valid one. In the second model (trend and intercept), the test statistic value -10.294 is lower in value than critical values, -4.034, -3.447 and -3.147 all at 1%, 5% and 10% in consequence; in addition, the regression coefficient L1 is once again negative in value (L1=-0.9592294), 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 -5.409 is lower in value than critical values, -2.597, -1.950 and -1.611 all at 1%, 5% and 10% successively; furthermore, the regression coefficient L1 is once again for this last model negative in value (L1=-0.3932618), 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 (total monthly recorded rainfall) is stationary and does not have a unit root [3].

```
                   delta:  1 day

. dfuller Y_TotalDailyRecordedRainfall, regress lags(0)

Dickey-Fuller test for unit root                  Number of obs   =      3651

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -49.566            -3.430            -2.860            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

----------------------------------------------------------------------------------------------
D.                           |
Y_TotalDailyRecordedRainfall |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------------------+----------------------------------------------------------------
Y_TotalDailyRecordedRainfall |
                         L1. |   -.80475   .0162358   -49.57   0.000    -.8365821   -.7729179
                             |
                       _cons |  1.383659     .10476    13.21   0.000     1.178265    1.589052
----------------------------------------------------------------------------------------------

.
```

Fig. 5. ADF test applied on the raw data (intercept only).

3.3. Data differencing (Seasonality)

Although the ADF test applied on the total 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 [3, 7].

3.4. Model identification

As soon as several tests concerning the total monthly recorded rainfall have been completed a few candidate models were considered that best meet the criteria and we finally concluded that the seasonal ARIMA (SARIMA) $[(0,0,0)X(0,1,1)_{12}]$ is considered the most suitable one appearing to have the least Normalized B.I.C. (7.976), M.A.P.E. (183.327), M.A.E. (36.811), Akaike's Information Criterion A.I.C. (729.981529), Schwarz's Bayesian Criterion S.B.C. (735.110226) [3, 7].

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 extend beyond the two continuous lines, plotted above and below the zero mean which imply the approximate 95% confidence limits [2, 3].

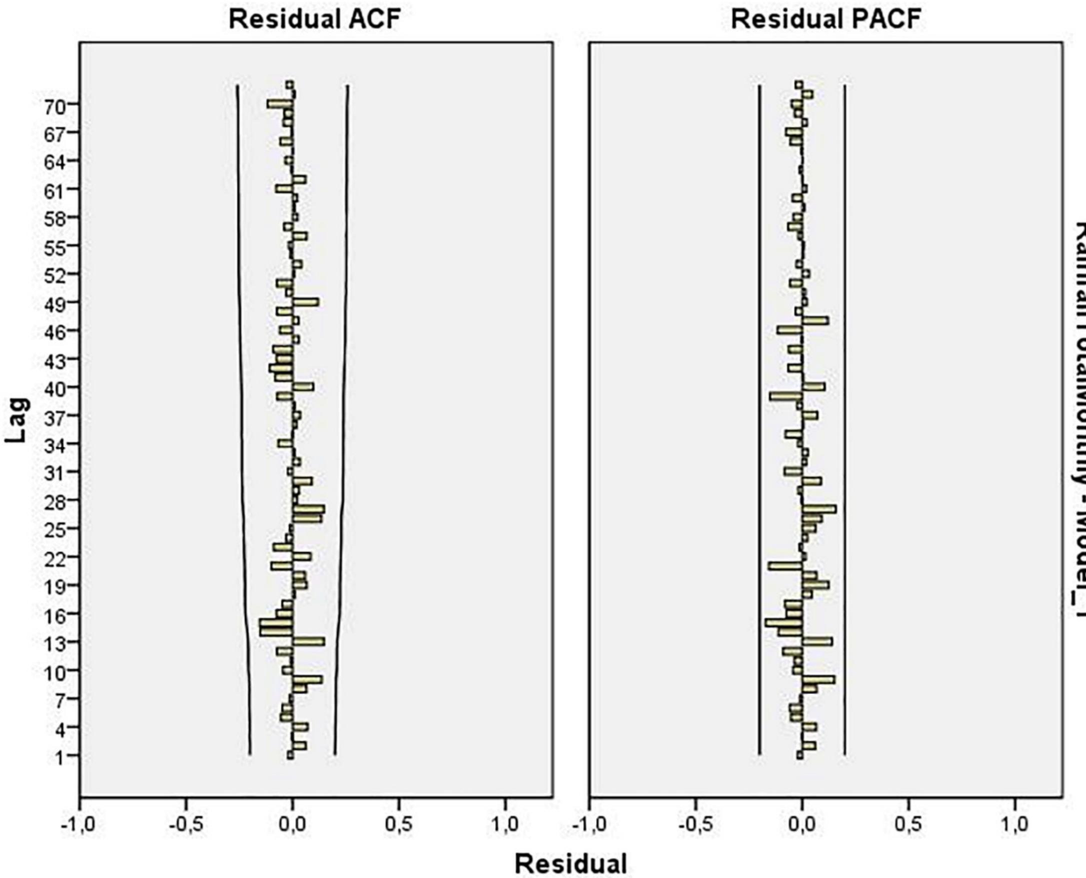

Fig. 6. ACF and PACF charts of SARIMA $[(0,0,0)X(0,1,1)_{12}]$ model Residuals.

3.6. Forecasting future values

Below in the Table 1 are depicted the forecasted values :

Table 1. Forecasted values of SARIMA [(0,0,0)X(0,1,1)$_{12}$] model.

SARIMA Model	Jan 2015	Feb 2015	Mar 2015	Apr 2015	May 2015	Jun 2015	Jul 2015	Aug 2015	Sep 2015	Oct 2015	Nov 2015	Dec 2015
Forecast	48.55	49.50	38.18	25.67	36.84	30.96	2.33	4.27	29.97	55.60	38.16	44.05

The observed, fit and forecast values as well as the upper and low confidence levels with reference to the finally selected SARIMA [(0,0,0)X(0,1,1)$_{12}$] model, are all summarized within the chart illustrated by Figure 7 [2, 3, 7].

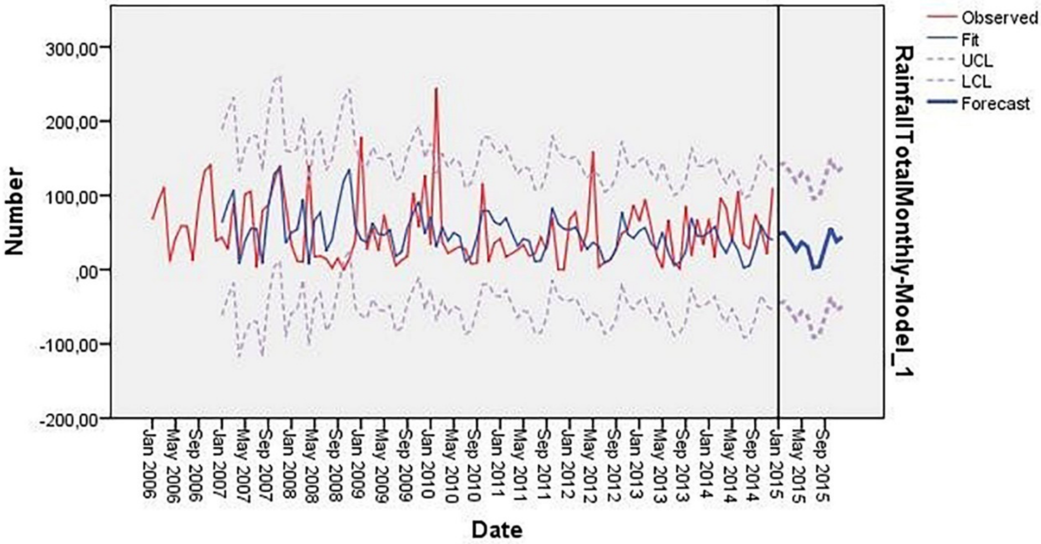

Fig. 7. Forecast chart of SARIMA [(0,0,0)X(0,1,1)$_{12}$] model.

4. Conclusions

After having examined a group of candidate SARIMA models we concluded that SARIMA [(0,0,0)X(0,1,1)$_{12}$] model best fit the total recorded monthly rainfall data of Kavala city, Greece, for the period 2006-2014.

Acknowledgements

The authors wishes to express that they feel grateful to Prof. Vlassios Hrissanthou, Department of Hydraulic Structures, Department of Civil Engineering, Faculty of Engineering, Democritus University of Thrace, for the valuable scientific knowledge within the general context of Stochastic Hydrology, he kindly transferred through the module, named, Stochastic Model Building in Hydrology taught as a part of the MSc. In Hydraulic Engineering, performed by the above mentioned department, as well as Mr. Giannis D. Zacharis, the meteorological station maintenance responsible and operator for broadcasting the real-time rainfall data through the free of charge, open access websites, named successively http://www.meteokav.gr, and https://www.facebook.com/Meteokav/?pnref=lhc.

References

[1] Box & Jenkins, 1976. *Time Series Analysis: Forecasting and Control*. Revised ed., Hoden-Day, San Francisco..

[2] S.H.Bari,M.T.Rahman,M.M. Hussain,S. Ray, Forecasting Monthly Precipitation in Sylhet City Using ARIMA Model,Civil Environ. Res.7:1 (2015) 69 – 77.

[3] A.R.Abdul-Aziz, M. Anokye, A. Kwame, L. Munyakazi, N.N.N. Nsowah-Nuamah, Modeling and Forecasting Rainfall Pattern in Ghana as a Seasonal Arima Process:The Case of Ashanti Region,Int. J. Human. Soc. Sci.3:3 (2013) 224 – 233.

[4] https://www.otexts.org/fpp/8/3 (accessed January 17, 2016).

[5] https://www.otexts.org/fpp/8/4 (accessed January 17, 2016).

[6] https://www.otexts.org/fpp/8/9 (accessed January 17, 2016).

[7] http://userwww.sfsu.edu/efc/classes/biol710/timeseries/timeseries1.htm (accessed January 17, 2016).

[8] http://www.statsref.com/HTML/index.html?arima.html (accessed January 17, 2016).al recorded monthly rainfall data of Kavala city, Greece, for the period 2006-2014.