International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Statistical Analysis of Precipitation and Cross-Comparison of Stochastic and Artificial Neural Network Models for a Short-Term Rainfall Monthly Forecast in Atalanti Gauge Station (Central – Eastern Greece)

Lappas Ioannis
Dr. Hydrogeologist, General Secretariat for Natural Environment and Water, Department of Protection and Management of Water Environment, Division of Surface and Ground Waters, Amaliados 17 Str., Ambelokipi-Athens, P.C. 11523,
i.lappas @ prv.ypeka.gr, tel.+30 2131515416

Abstract— *The objective of this paper aims at modeling the precipitation data from Atalanti rain gauge station in Central-Eastern Greece through an Artificial Neural Network (ANN) with multi-layer perceptron and a Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model which can simulate systems characterized by complicated physical processes. First, a thorough statistical analysis of the rainfall timeseries and its residuals as well was carried out so as to figure out its main statistics before proceeding to further analysis. The use of stochastic methods, introduced by Box and Jenkins, has found wide application for fitting and forecasting the monthly rainfall timeseries which may be useful in decision making as well as risk management and water resources usage optimization. Also, the ANN model has the ability of identifying non-linear relationships between inputs and output datasets, therefore, a suitable tool for assessing various hydrological impacts. This essay illustrated the application of a feed-forward back propagation learning process with various algorithms with performance of multi-layer perceptors (MLP) to reproduce a given data file. During timeseries preparation, raw data were initially transformed to normal and stationary using differencing methods and autocorrelation plots of residuals were made to check the uncorrelated "white noise". The objective is to find an appropriate model for best fitting the precipitation and to predict future series quantities according to the past. In this study, the monthly precipitation data of Atalanti meteorological station were investigated based on a 60-year period (1955-2014). This period was separated into training (70%, from 1955 to 1996), testing (20%, from 1997 to 2008) and validation one (10%, from 2009 to 2014) to find the best fit of the timeseries to past already measured values. The best performance between observed and modeled results based on the model evaluation criteria, namely, R^2, MSE, RMSE, model efficiency (E), MAE and % Bias, was achieved by the stochastic SARIMA $(1,0,0)(0,1,0)_{12}$ model. Nevertheless, it has been proved that both stochastic and ANN models are useful tools for forecasting various hydrological processes and could be used in monthly rainfall prediction at a short time (one or two years) in order to help decision makers to establish priorities in terms of water demand management.*

Keywords— *Box – Jenkins models; timeseries forecasting; residuals; multi-layer perceptors (MLP); feed-forward back propagation; model evaluation criteria.*

I. Introduction

Precipitation prediction is the most important issue to water resources management (e.g., irrigation), planning, exploitation and to local as well as regional hydroeconomy because of high water demands (Amiri, 2004; Karel, 2011). Timeseries modelling and analysis is one of the most widely used predicting methods having three modelling stages; identification, estimation and diagnostic check for medium to long-term forecasting. A stochastic, hydrological variable, such as rainfall, is consisted of a deterministic N_t which is composed by trend, periodicity and persistence and a stochastic Z_t time-dependent part, namely, white noise (Koutsoyiannis, 2000, 2008; Mimikou, 1994; Papoulis, 2002). The utmost purpose of timeseries modelling is to find a best fit to a dataset that can be defined by a model used for forecasting. Box-Jenkins (1976) models are significantly used because of the simplified mathematical structure and the relatively small number of parameters used to both stationary and non-stationary procedures, therefore there has been considerable interest in stochastically modelling weather parameters data. Finally, the great usefulness of these models is for the analysis and prediction of the timeseries as well as to study and analyze complex cases when other methods are either not applicable or require extremely complex equations which are not able to approach the physical conditions occurred in nature (Ripley, 1987; Salas, 1992).

On the other hand, Artificial Neural Networks (ANN), as a flexible mathematical structure, are developed to predict the hydrological parameters having the ability of giving solutions to complex physical processes' problems with non-linear relationships since they do not require the a priori knowledge of a mathematical form used to describe an explicit description of the complex nature which, by default, exhibit extreme variability (Raman & Sunilkumar, 1995; Sajikumar & Thandaveswara, 1999). The neural networks need to be trained to perform a particular function by adjusting the values of the connections (weights) between elements. The weights are adjusted based on a comparison of ANN output and the target until they satisfactorily match. The model is considered to behave satisfactorily if its performance during the testing period is similar to that during the training one. The output timeseries is the response of the system and reflects the system processes. The most commonly used ANN method for modelling rainfall processes are the multi-layer feed forward

Lappas Ioannis, "Statistical Analysis of Precipitation and Cross-Comparison of Stochastic and Artificial Neural Network Models for a Short-Term Rainfall Monthly Forecast in Atalanti Gauge Station (Central – Eastern Greece)," *International Research Journal of Advanced Engineering and Science*, Volume 5, Issue 2, pp. 49-58, 2020.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

perceptron (MLP).

Before proceeding to the aforementioned modelling, a statistical analysis of the rainfall timeseries has to be performed so as to determine those months with the greatest rainfall values, check the normality, the trend and the persistence as well as the outliers to figure out the exceedance probability of certain values, etc. To do so several diagrams were drawn, probability functions were estimated, the extreme values, the trend line and many other statistical parameters were detected and calculated on a yearly and monthly basis.

II. Materials and Methods

A. *Statistical Approach of Rainfall Timeseries*

The rainfall data of Atalanti meteorological station are derived from the Ministry of Environment and Energy concerning monthly values for the time period 1955-2014. As illustrated in Fig.1, the historical precipitation timeseries of the aforementioned rain gauge is displayed. A preliminary exploratory analysis of the yearly as well as monthly rainfall data is firstly implemented to detect extreme values, homogeneity, step and trends (Koutsoyiannis, 1997; Mimikou, 1994; Manakos, 1999). The mean annual precipitation of the selected meteorological station is about 573.3mm with the 75% of the total rainfall occurring in the wet season from October to April. The minimum and maximum rainfall events took place in 2000 (272.4mm) and 1994 (842.3mm), respectively. Also, the trend line (blach thick line) shows that there is no evident increase or decrease in precipitation during the 60-year data period. The below bar charts show that the most wet and humid month of the year is December followed by November and January. Instead, the months with the lowest rainfall are July, August and June as expected since the rainfall station belongs to C_{sa} climate type (typical Mediterranean type).

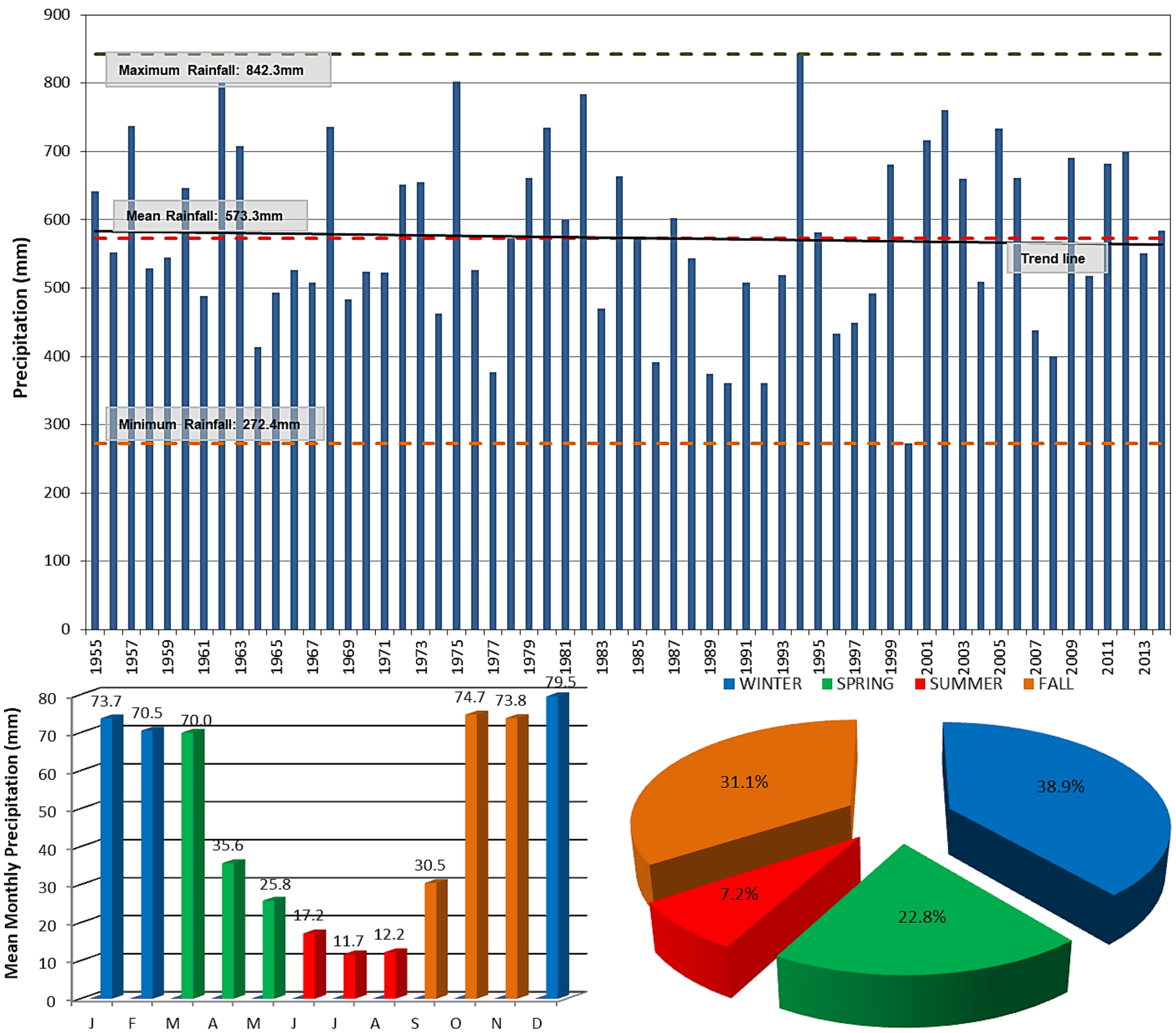

Fig. 1: Atalanti station's yearly rainfall distribution for the time period 1955-2014. The black thick line shows the over-annual trend while the dashed ones represent the mean, maximum amd minimum rainfall years. Seasonal precipitation is also displayed.

Also, it can be shown that the major rainfall events start in October, reach their peak in December and remain at high levels until March (Lappas, 2018).

Moreover, in pie chart the seasonal rainfall distribution is illustrated. As seen, winter accumulates precipitation at a rate of 38.9%, followed by 31.1% during fall, spring rate by 22.8%, while during summer season rainfall takes place only by 7.2%. Evaluating the two basic periods of a hydrological year (wet and dry) it seems that 77.6% of annual precipitation corresponds to wet period and only 22.4% to the dry one.

B. Applied Methodology

In this paper, precipitation based on Atalanti meteorological station was modeled by two methodologies; SARIMA (Seasonal AutoRegressive Integrated Moving Average) and Artificial Neural Network (ANN) model for the period from 1955 to 2014. The historical timeseries was divided into three periods, one from 1955 to 1996 (calibration period) which was used for choosing the most suitable model after the application of Box-Cox (1964) transformation to make the timeseries stationary, the other from 1997 to 2008 to test the model and the other one from 2009 to 2014 to validate it.

Seasonal AutoRegressive Integrated Moving Average (SARIMA)

Box and Jenkins (1976) proposed a general form of Seasonal ARIMA $(p,d,q)(P,D,Q)_{12}$ for non-stationary timeseries, in which p and q are autoregressive and moving average parameters and P and Q are seasonal autoregressive and moving average parameters, respectively (Khazavi et al, 2012; Lappas, 2018; Manakos & Dimopoulos, 2004; Manakos & Georgiou, 2009; Papamichail, 1989, 1993; Paschalis, 2009). The general form of Seasonal ARIMA $(p,d,q)(P,D,Q)_s$ is as follows:

$$\varphi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_Q(B^s)a_t$$

where,
φ_p and Φ_P define autoregressive processes,
θ_q and Θ_Q define moving average processes,
p, *P*, *q* and *Q* are estimated from ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) of the series,
D and *d* show the order of seasonal and non-seasonal differencing used to make the series stationary,
B is the backward operator,
s is the period of the season and
a_t is the white noise.

SARIMA modelling includes three steps (Koutsoyiannis, 2000; Manakos & Georgiou, 2009; Papamichail, 1989, 1993; Paschalis, 2009), namely, identification, estimation and validation process used for modelling the patterns in the raw data. First step; the identification model is based on the behavior of autocorrelation and partial autocorrelation functions where the values of *p* and *q* are estimated. Second step; efficient estimate of the parameters can be obtained only after identifying the autocorrelation and partial autocorrelation functions. Third step; validation model lies on the goodness of fit test that verifies the validity of the model. In this stage, the residuals (the differences between the scores predicted by the model and the actual scores for the series) of the model are usually considered to be time-independent and normally distributed over time (Ripley, 1987, Abudu et al., 2010). The foremost step in the process of modelling is to check for the series' stationarity, as the estimation procedures are available only for stationary series. If the model is found to be non-stationary, stationarity could be achieved mostly by differencing the series using data transformation methods such as Box-Cox, logarithmic and square root. Residual scores are examined to determine if there are still patterns in the data that are not accounted for, that is, checking the randomness. Identifying and modelling the patterns in the data are sufficient to produce an equation, which is then used to forecast (Manakos & Dimopoulos, 2004).

Artificial Neural Network (ANN)

The basic structure of ANN usually consists of three layers; the input, where the data are introduced to the network, the hidden, where data are processed and the output layer, where the results of given outputs are produced (Fig.2).

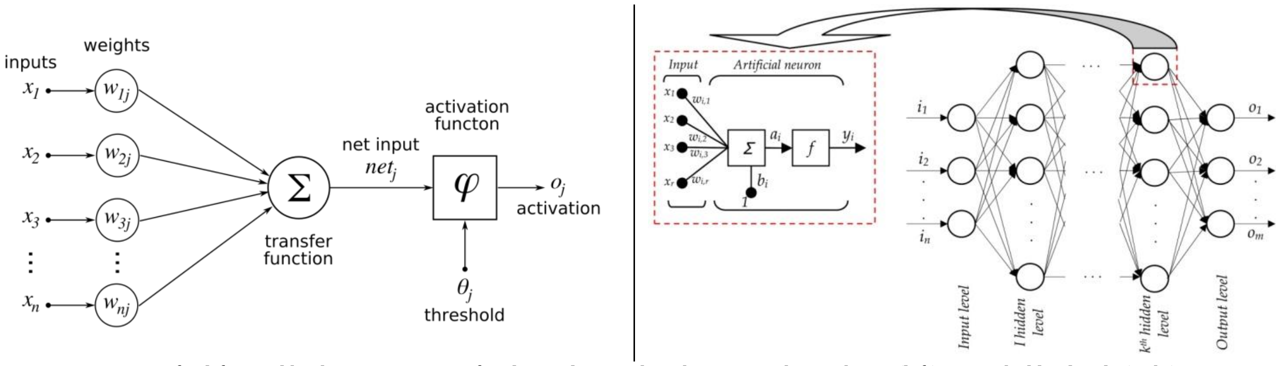

Fig. 2: A feed-forward back-propagation artificial neural network architecture either with one (left) or two hidden levels (right).

The neurons basically consist of inputs which are multiplied by weights and then computed by a mathematical function which determines the activation of the neuron. Another function computes the output of the artificial neuron. The incoming data are processed by non-linear transfer functions at hidden and output layers to get the output. The transfer function transforms weighted input to output. A common transfer function is the sigmoid one which is an Sshaped graph. In hydrology, multiple-layer perceptron (MLP) networks, a single hidden layer with a sigmoid transfer function and an output layer with a linear transfer function, are preferred for their simplicity and effectiveness. The primary goal in training an ANN model is to minimize the error at the output layer by searching for a set of connection strengths that cause the ANN to produce outputs that are equal to or closer to the targets through a back-propagation training algorithm. The back-propagation algorithm uses supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute and then the error (difference between observed and modeled results) is calculated (Barros & Kuligowski, 1998; Dawson & Wilby, 2001; Haykin, 1998). After training, the neural network can simulate the model. The final step involves testing the adequacy of the selected model. A basic neural model can be characterized by the functional descriptions of the connection network and the network activation as follows:

$$S_j = \sum_{i=0}^{n} w_{ij} \times x_i$$

Each node j receives incoming signals from every node i in the previous layer. Associated with each incoming signal x_i is a weight w_{ji}. The effective incoming signal S_j to node j is the weighted sum of all the incoming signals. The effective incoming signal, S_j, is passed through a non-linear activation function (transfer function) to produce the outgoing signal of the node. The transfer function, which is a sigmoid one in this paper, is given as shown below:

$$Y_j = f(S_j) = \left(\frac{1}{1 + \exp(-S_j)} \right)$$

where, Y_j is bounded between 0 and 1.

a. Models' Evaluation Criteria

The comparison between simulated results and the observed data was evaluated statistically. In order to select the appropriate model in timeseries modelling there are several criteria which may be used for representing a given set of data. With contribution of Minitab statistical software successive trials were performed for the best fit model, so that the predicted values of the stochastic and ANN model to be as close as those of the observed ones. Model evaluation criteria were used either based on statistics summarized from residuals or on the forecasting error (Abudu et al., 2010). For the first methods one can mention AIC (Akaike Information Criterion) given by the equation:

$$AIC = -2logL + 2m$$

where,
$m = p + q + P + Q$ and
L is the likelihood function.

The Akaike information criterion is a measure of the relative goodness of fit of a statistical model (Papamichail, 1993, Manakos, 1999; Shamsnia et al., 2011). Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Therefore, AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters (overfitting).

Another criterion is SBC (Schwartz-Bayesian Criterion) given by the equation:

$$SBC = log\sigma^2 + (mlogn)/n$$

where,
σ is the standard deviation of the population.
and the integer number n given by the equation:

$$n=p+P+q+Q+pP+qQ+1$$

The model, in which the above statistics were the least, was chosen as the appropriate one.

The most important measures for evaluating model performance used in this paper taking into account the forecasting error (Barros & Kuligowski, 1998; De Vos, 2003; Meher, 2014; Salas, 1992) were the coefficient of correlation-*r*, the Mean Squared Error-*MSE*, the Root Mean Squared Error-*RMSE*, the Nash-Sutcliffe (1970) coefficient of efficiency-*E*, the Mean Absolute Error-*MAE* and the percent (%) Bias. Only when the performance of the model is satisfactory, both in calibration and testing-validation periods can the model be used with confidence. The coefficient of determination-R^2 is obtained by performing a linear regression between the predicted values and the observed. These criteria were employed to measure the models' goodness-of-fit and used to test the model efficiency in all phases (calibration-training, testing and validation). The equation of correlation coefficient is as follows:

$$r = \frac{\sum_{i=1}^{n}(R_{obs} - \overline{R_{obs}})(R_{sim} - \overline{R_{sim}})}{\sqrt{\sum_{i=1}^{n}(R_{obs} - \overline{R_{obs}})^2 \sum_{i=1}^{n}(R_{sim} - \overline{R_{sim}})^2}}$$

The difference between the simulated output and observed one was measured by the Mean Squared Error (MSE) function as:

$$MSE = \frac{1}{n} \sum_{t=1}^{n} \left(R_{sim} - R_{obs} \right)^2$$

which measures the fit of the simulated rainfall (R_{sim}) to the observed one (R_{obs}). The objective is to minimize the sum of the squared errors. The smaller MSE the better model fit would be.

Both RMSE and MAE are desirable when the evaluated output data are smooth or continuous and are given by the following equations:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(R_{sim} - R_{obs} \right)^2}$$

$$MAE = \frac{1}{n} \sum_{t=1}^{n} \left| R_{sim} - R_{obs} \right|$$

Nash-Sutcliffe efficiency can range from $-\infty$ to 1. An efficiency of 1 *(E = 1)* corresponds to a perfect match of modelled to the observed data. Essentially, the closer the model efficiency is to 1, the more accurate the model is.

$$E = 1 - \frac{\sum_{i=1}^{n}(R_{sim} - R_{obs})^2}{\sum_{i=1}^{n}(R_{obs} - \overline{R_{obs}})^2}$$

Lappas Ioannis, "Statistical Analysis of Precipitation and Cross-Comparison of Stochastic and Artificial Neural Network Models for a Short-Term Rainfall Monthly Forecast in Atalanti Gauge Station (Central – Eastern Greece)," *International Research Journal of Advanced Engineering and Science*, Volume 5, Issue 2, pp. 49-58, 2020.

Different statistics represent the *goodness-of-fit* of the calibrated monthly precipitation to the observed ones like the percent (%) Bias computed by the following equation:

$$\%Bias = 100 \cdot \left(\overline{R}_{sim} - \overline{R}_{obs} \right) / \overline{R}_{obs}$$

where, $\overline{R}_{sim}$ represents the mean of the model generated rainfall and $\overline{R}_{obs}$ represents the mean of the observed rainfall.

III. Results and Discussion

A. SARIMA Model Design and Structure

As shown below (Fig.3), the historical precipitation timeseries examined for the existence or not of trend and stationarity was displayed looking also for trends, if any, as well as moving averages.

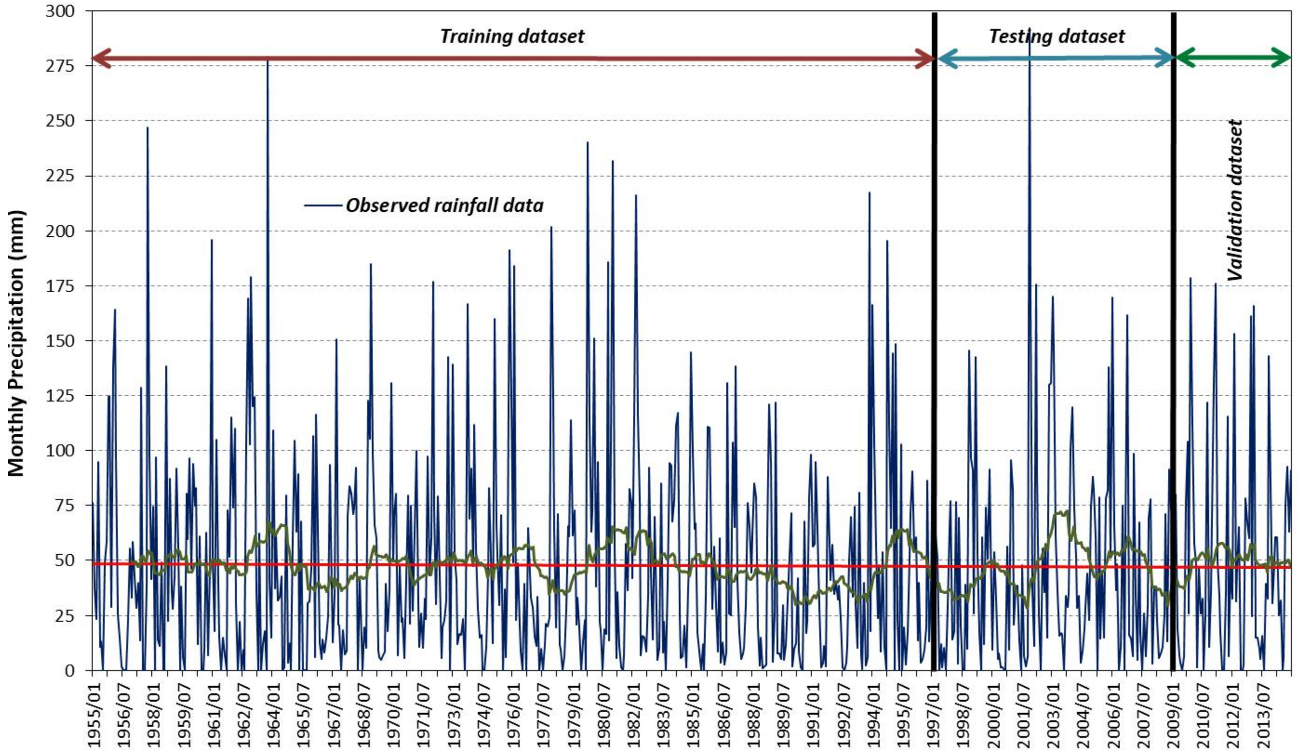

Fig. 3: Atalanti station's monthly rainfall data for the time period 1955-2014. The red thick line shows the trend while the green one the 24-month moving average of the observed timeseries. Also, calibration (training), testing and validation periods are also shown.

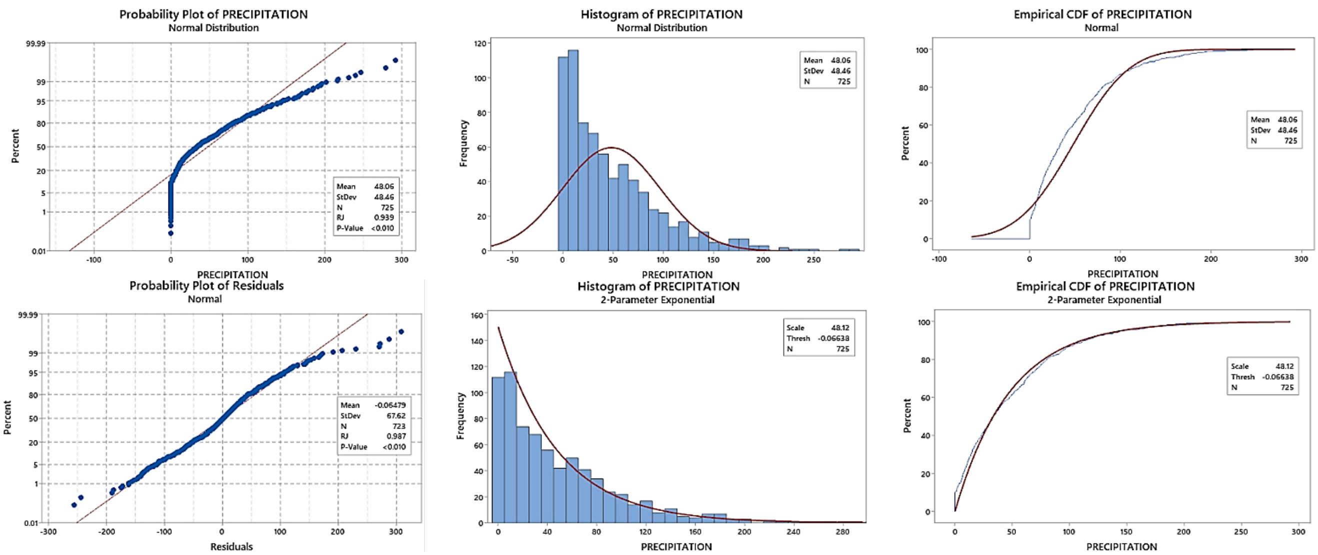

Fig. 4: Probability plot, histogram and empirical CDF (Cumulative Distribution Function) of precipitation showing the non-normality of the monthly rainfall timeseries (up). Two-parameter Exponential distribution seems to fit the best after Box-Cox transformation and Seasonal Differencing (D=1) method (down).

In preparation of the time series to use the Box-Jenkins forecasting model (SARIMA), the timeseries was initially transformed to normal and stationary using differencing method. Normality of the monthly mean rainfall series was tested by normal probability paper, histogram of series and Kolmogorov-Smirnof, Ryan-Joiner (similar to Shapiro-Wilk) and chi square (χ^2) test (Fig.4, Fig.5). The implementation of those tests showed that the historical rainfall timeseries did not have normal distribution and in order to obtain a series which had approximately a normal distribution, seasonal first order differencing (D=1) was performed. Furthermore, as illustrated above, the monthly rainfall data did not have any trend, that is, d=0 (non – seasonal trend), the timeseries was not of "white noise" and that the seasonal length equals to 12 months (1 year). However, for stationarity checking the Box-Cox transformation was used and then the autocorrelation and partial autocorrelation functions were calculated for the estimation of *p* and *q* components of the Autoregressive (AR) and Moving average (MA) model, respectively.

Autocorrelation refers to the way the observations in a timeseries are related to each other and is measured by the simple correlation between current observation (Y_t) and observation from p periods before the current one (Y_{t-p}). Partial autocorrelations are used to measure the degree of association between y_t and y_{t-p} when the y-effects at other time lags are removed. Theoritical ACFs and PACFs (autocorrelations versus lags) are available for the various models chosen for various values of orders of autoregressive and moving average components i.e. p and q (Fig.5). Comparing the correlograms (plot of sample ACFs/PACFs versus lags) obtained from the given measured data with the theoretical ones, one may find a reasonably good match and choose one or more SARIMA models.

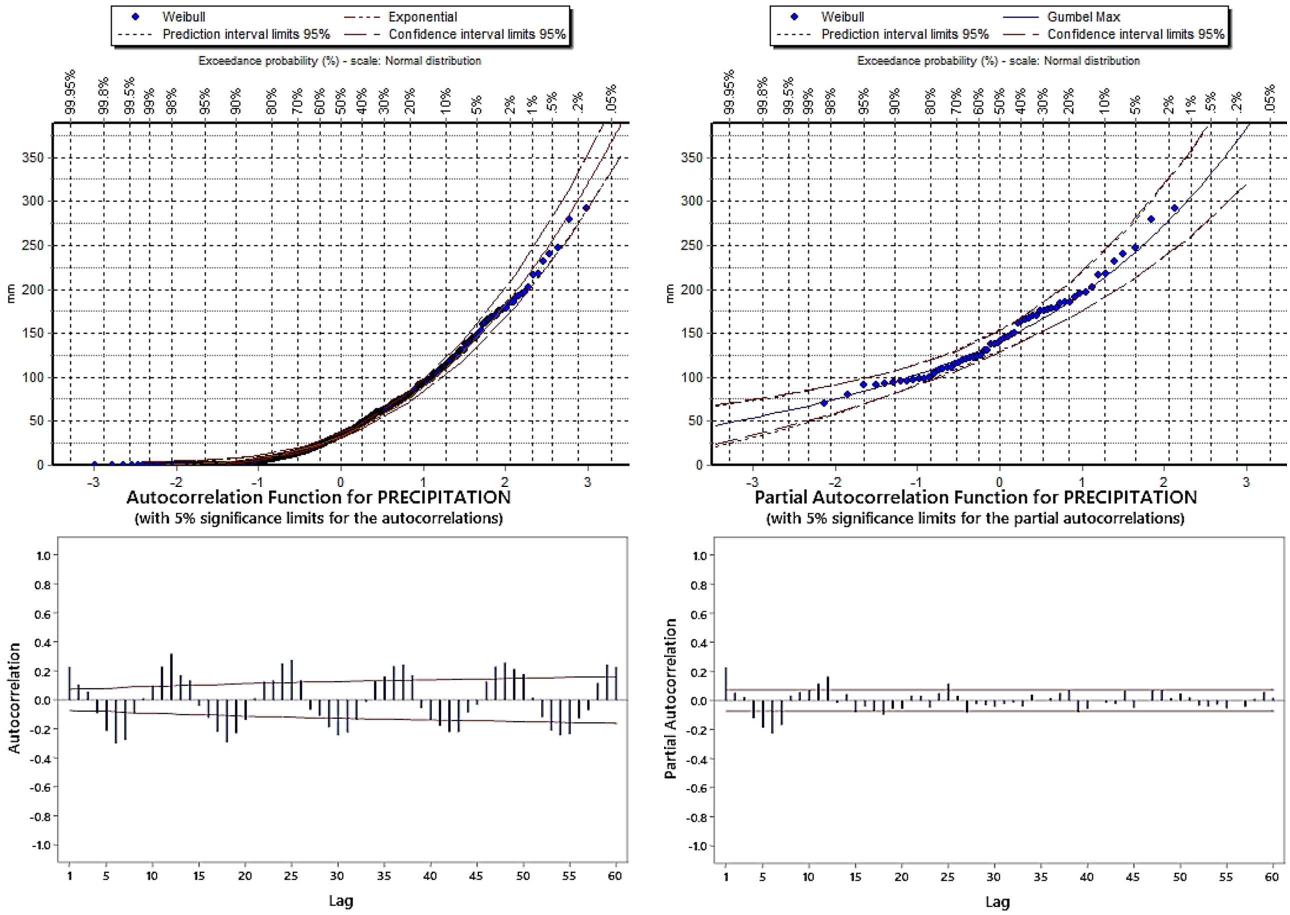

Fig. 5: Exponential and Gumbel distribution curves for the monthly and maximum rainfall timeseries, respectively (up). Autocorrelation (ACF) and Partial Autocorrelation functions (PACF) of the Atalanti station's Box-Cox transformed rainfall timeseries for the time period 1955-2014 (down).

According to this procedure the timeseries of 60 years (1955-2014) was divided into three parts, the one of 42 years (1955-1996), the second of 12 years (1997-2008) and the last one of 6 years (2009-2014) so as the model's parameters to be confirmed. The model's components that were re-assessed slightly differ from those of the first part (1955-1996). The best fit and most suitable seasonal stochastic model satisfying the most of the above model criteria was SARIMA $(1,0,0)(0,1,0)_{12}$ which was then checked. The following figures show the 12-year and 6-year rainfall data predictions that came up with the application of the SARIMA $(1,0,0)(0,1,0)_{12}$ model proving the very good fitting, showing also that the data derived by the application of above stochastic model may be used for a very satisfactory and reliable forecasting. The analytical, mathematical expression of the above model was of the form:

Lappas Ioannis, "Statistical Analysis of Precipitation and Cross-Comparison of Stochastic and Artificial Neural Network Models for a Short-Term Rainfall Monthly Forecast in Atalanti Gauge Station (Central – Eastern Greece)," *International Research Journal of Advanced Engineering and Science*, Volume 5, Issue 2, pp. 49-58, 2020.

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

$$(1-\varphi_1 B)(1-B^{12})Z_t=e_t \text{ or } Z_t=\varphi_1(Z_{t-1}-Z_{t-12})+Z_{t-12}+e_t$$

with $\varphi_1=0.0344$ (standard error (SE) = 0.074, T-value = 4.77 and p-value = 0.0001<0.05).

Model's residuals were also used for validation. The residuals' checking procedure took place to prove whether or not they were of "white noise" and this was achieved by the autocorrelation (ACF) and partial autocorrelation (PACF) functions. In Fig.6 it seems that most of the residuals were uncorrelated. However, due to a slight data remaining trend, some of the residuals exceeded the confidence limits 95%. It has to be mentioned that all the R^2 stages were similar and so slose to line 1:1. So the best fit seasonal stochastic model was SARIMA $(1,0,0)(0,1,0)_{12}$ which was the one that could be used for monthly rainfall values estimation and prediction of Atalanti meteorological station (Table 1). In addition, Fig.6 shows the very good simulation of the above model on the observed precipitation timeseries. Finally, SARIMA model is applied as long as the above rainfall timeseries can provide safe predictions and create, if possible, reliable future time period.

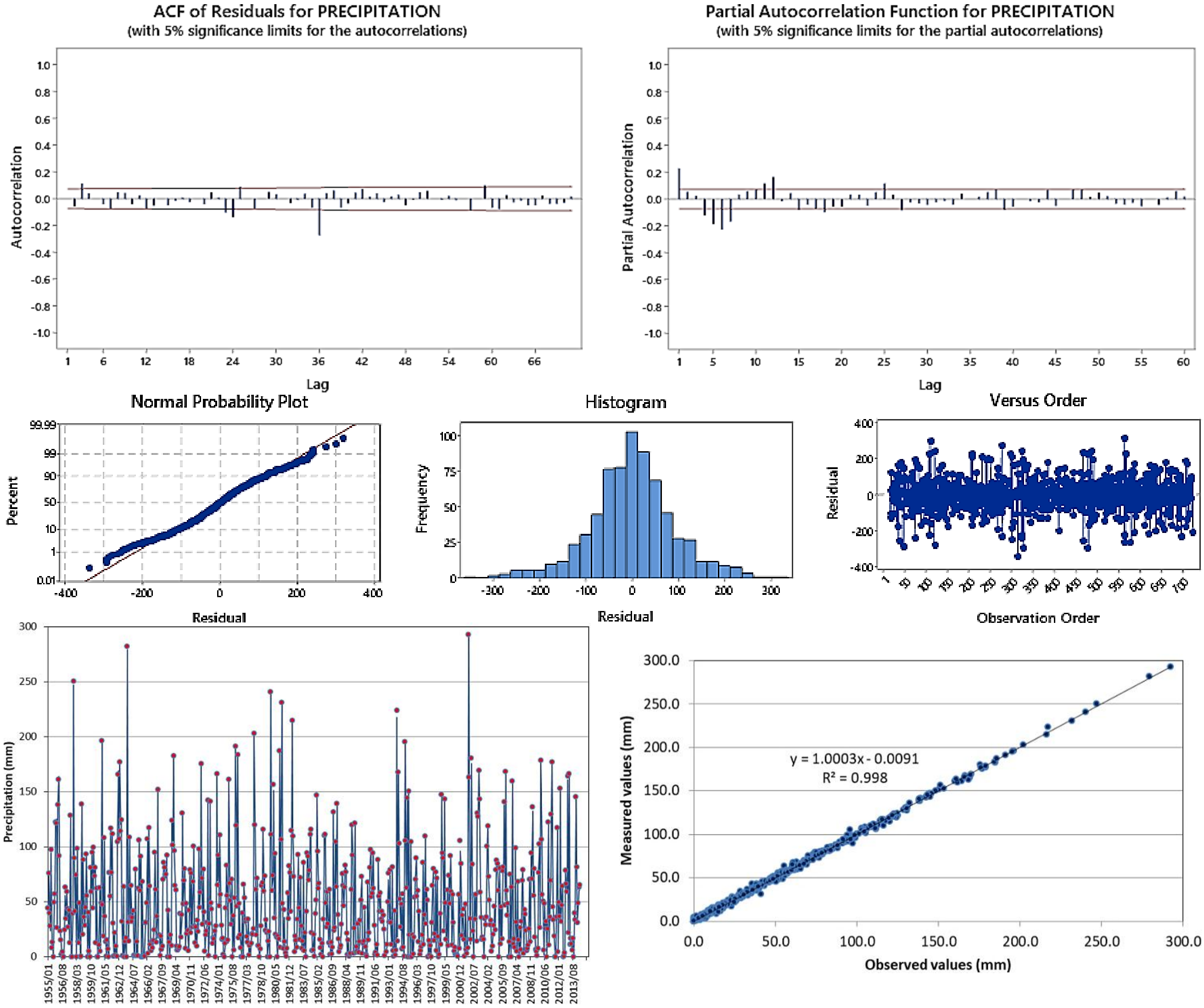

Fig. 6: Autocorrelation (ACF) and Partial Autocorrelation functions (PACF) of the residuals for the time period 1955-2014 (up). Residual plots of precipitation showing its normality and randomness (middle). Seasonal ARIMA $(1,0,0)(0,1,0)_{12}$ model generated series (red dots) versus the observed monthly rainfall timeseries with R^2=0.99 (down).

TABLE 1: Model parameters comparison and evaluation of the monthly rainfall timeseries (1955-2014) at the Atalanti meteorological station. The best fit SARIMA model is represented by bold and *italics*.

SARIMA model	n	AIC	SBC	SARIMA model	n	AIC	SBC	SARIMA model	n	AIC	SBC
$(2,1,0)(1,1,1)_{12}$	7	26.27	3.72	$(1,1,0)(1,1,1)_{12}$	5	19.56	3.71	$(0,1,0)(0,1,1)_{12}$	4	38.17	3.75

B. ANN Model Design and Structure

The first and the most critical step in developing an effective ANN model is input and output definition as well as the selection of the input variables that have the most significant impact on model performance since a good subset of input variables can substantially improve model performance. The determination as to whether a parameter input is significant or not is dependent on the error of a trained

International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

model, which is not only a function of the inputs, but also model structure and calibration. Consequently, a separate validation set is needed to ensure that the model can generalize within the range of the data used for calibration (Aksoy & Dahamsheh, 2009, Nastos et al., 2013, Tolika et al., 2007). The used data as model inputs included constant monthly recorded precipitation, air monthly temperature (average), evapotranspiration (potential and actual) as well as rainfall data during the periods t-1, t-3, t-12 and t-24 in relation with monthly observed precipitation. Output layer (target) was the predicted rainfall. Before applying the MLP method, the input data were normalized and standardized to fall in the range [0,1] and ensure that each variable is treated equally in the model in order to obtain optimal results. Prior to any data pre-processing was carried out, the whole dataset was divided into subsets i.e. training, testing and validation. Data pre-processing is necessary to ensure all variables receive equal attention during the training process speeding up also the learning process (Maier & Dandy, 2000). Hence, the available observed and calculated data (potential evapotranspiration through Thornthwaite equations and actual evapotranspiration through Thornthwaite-Mather model) were separated as 70% for training (1955-1996), 20% for testing (1997-2008) and 10% for validation (2009-2014). Obviously, the statistical properties of the various data subsets (e.g., training, testing and validation) need to be similar (similar statistical properties) to ensure that each subset represents the same statistical population (Fig.7).

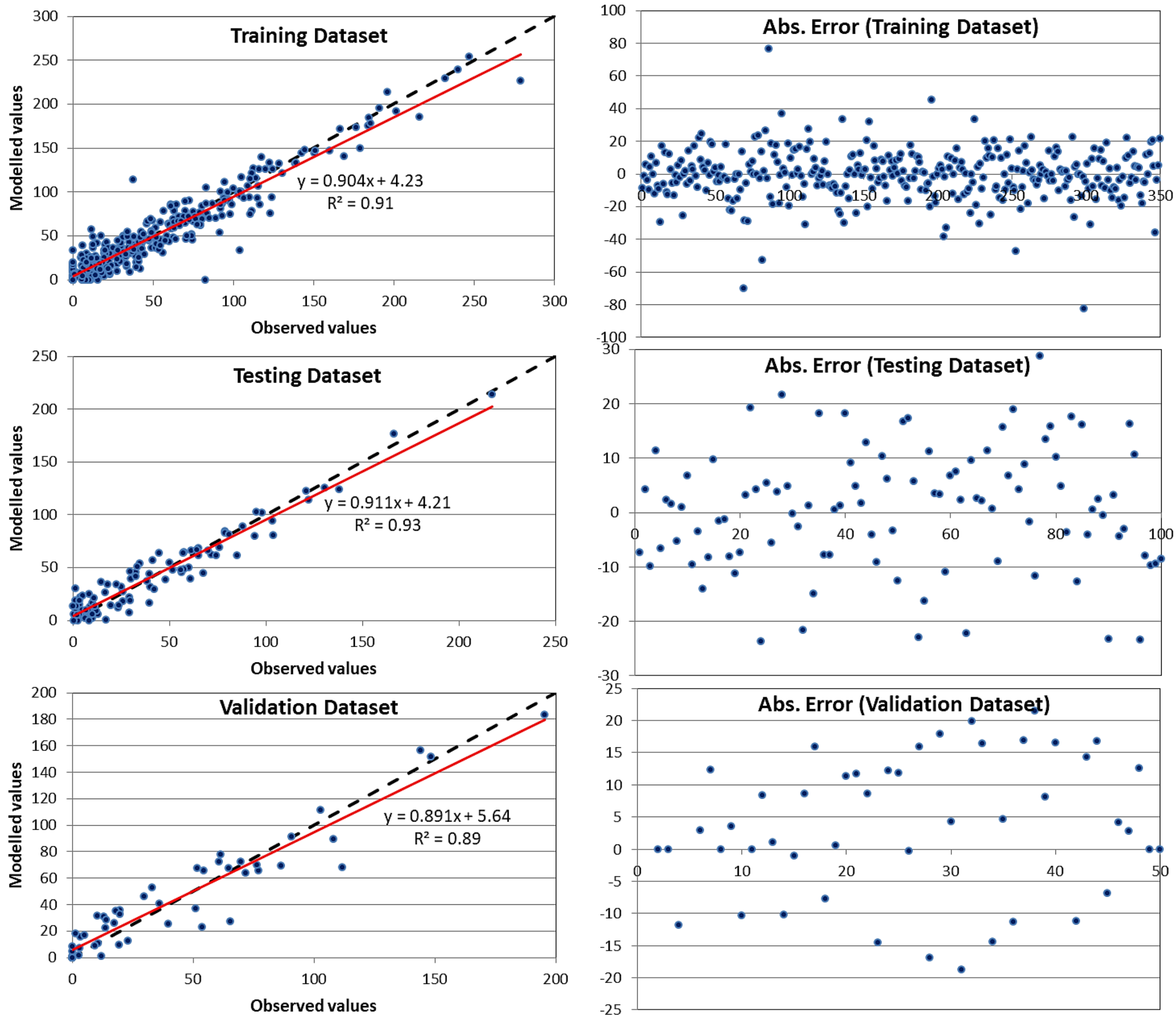

Fig. 7: Training, testing and validation dataset comparison between observed and modelled values.

During the data pre-processing the selection of the network parameters such as the transfer function, learning algorithm, etc. has to be made (Daliakopoulos et al., 2005, Tokar & Johnson, 1999). Once the ANN model structure is defined, data are collected and fed to the model. The approach used to ANN training was the supervised training algorithms. The method most commonly used for finding the optimum weight combination of feed-forward MLP neural networks is the back-propagation algorithm. During the training process, one or several of the network parameters change to improve the network's performance. This process continues until weights converge to the desired error level or to a given acceptable level. After training is complete, testing and validation are the final steps in the development process (Dibike & Solomatine, 2001; Minns & Hall, 1996; Mutly et al., 2008; Smith & Eli, 1995). The trained results are compared to observed results and the trained model is assumed to be successful if the model gives good results for the given test set. Also, the testing set is used to reduce the overtraining of the network and to determine the optimum number of hidden layer nodes and the optimum values of the internal variables. Training stops when the error of the testing set starts to increase. The latter of the three datasets is used to validate the performance of a trained and tested ANN dealing with comparing the results of the developed model to observed ones. The main purpose of the model validation stage is to ensure that the model has the ability to generate forecasted data. As illustrated in Table 2, the evaluation criteria of all the ANN model phases were proved to be quite satisfactory, meaning that the selected model can be used for a short-term rainfall prediction.

TABLE 2: Model evaluation criteria.

Evaluation Criteria	Training Period (1955-1996)	Testing Period (1997-2008)	Validation Period (2009-2014)
r	0.954	0.964	0.943
R^2	0.91	0.93	0.89
MSE	296.2	151.9	176.1
RMSE	17.2	12.3	13.4
MAE	9.83	7.83	8.22
E	0.853	0.875	0.866
%Bias	-4.23	5.17	-3.51

IV. CONCLUSIONS

Both stochastic and artificial neural network simulation are powerful methods, quite easily applicable and flexible. Their main advantage is the capability to perform in complex, synthetic systems describing them faithfully without simplistic assumptions. However, it is an approximate procedure and the accuracy of its results depends on the sample size. These simulations become powerful tools when a complex, natural system is to be studied and analytical (or other numerical) methods are not applicable or are very difficult requiring extremely complex equations and robust assumptions. Rainfall timeseries data of 60 years (1955-2014) were statistically described and studied over the Central – Eastern Greece at the Atalanti gauge station. Using seasonal model method, monthly mean rainfall values were simulated with high accuracy which was confirmed by the evaluation criteria (AIC, BSC, n) applied to training, testing and validating dataset. The SARIMA $(1,0,0)(0,1,0)_{12}$ model seemed to be the most suitable in simulating the observed timeseries and forecasting the rainfall data at Atalanti meteorological station. Moreover, the seasonal ARIMA model can provide synthetic timeseries which may be used for the water resources planning proving its usefulness. On the other hand, ANN models provide a trustworthy, if adequately trained, mathematical and statistical method for data analysis. The structure of the paper's model consisted of a supervised multi-layer perceptron with feed-forward back propagation approach (sigmoid function). The used input data in ANN were separated into three categories; traing (70%), testing (20%) and validation (10%) period. Based on statistical analysis, the results during all the stages in terms of r, R^2, MSE, RMSE, MAE, E and %bias had satisfactory similar values during all the model's stages. Therefore, ANN was able to model the monthly rainfall data with fairly good accuracy when proper input variables are included. Nevertheless, comparing the two models it seemed that the applied stochastic model much better simulated the observed rainfall timeseries in relation with the applied ANN model since the SARIMA statistical indices were much better than the ANN's. The fairly good fit of the aforementioned models as well as their relatively easy adaptation to physical conditions can be considered as a tool for the rational and sustainable water resources management and exploitation since many mathematical models fail to simulate the complex behaviour of most hydrological problems (non-linear relationships).

ACKNOWLEDGMENT

The author wish to express his thanks to the Ministry of Environment and Energy for the Atalanti meteorological station's precipitation dataset.

Conflict of Interests
No potential conflict of interest was reported by the authors.

REFERENCES

[1] Abudu, S., Cui, C. L., King, J. P., Abudukadeer. K. 2010. Comparison of Performance of Statistical Models in Forecasting Monthly Streamflow of Kizil River, China. Water Science and Engineering, 3(3), 269-281 pp.

[2] Aksoy, H., Dahamsheh, A. 2009. Artificial neural network models for forecasting monthly precipitation in Jordan. Stochastic Environmental Research and Risk Assessment. Vol. 23, No 7, pp.917-931.

[3] Amiri, A. 2004. Studying the Probability Climate Change and Fitting a Suitable ARIMA Model on Temperature and Rain Data of Gilan Province, IRIMO, Gilan Met. Vol. 5–6: 78-125 pp.

[4] Barros, A., Kuligowski, R. 1998. Experiments in short-term precipitation forecasting using artificial neural networks. Stochastic and statistical methods on hydrology and environmental engineering time series analysis in hydrology and environmental engineering, Dordrecht, The Nederlands, pp.229-242.

[5] Box, G., Cox, D. 1964. An Analysis of Transformations. J. R. Stat. Soc., Ser. B, 26: 211-252 pp.

[6] Box, G., Jenkins, G. 1976. Time Series Analysis, Forecasting and Control, Holden – Day, San Francisco, California, U.S.A.

[7] Daliakopoulos, N., Coulibaly, P., Tsanis, K. 2005. Groundwater level forecasting using artificial neural networks. Journal of Hydrology, 309, pp.229-240.

[8] Dawson, C., Wilby, R. 2001. Hydrological modeling using artificial neural networks, Prog. Phys. Geog., Vol 25, No 1, pp.80-108.

[9] De Vos, J. 2003. Rainfall-runoff modelling using artificial neural networks. Master's thesis, Delft University of Technology.

[10] Dibike, B., Solomatine, P. 2001. River flow forecasting using artificial neural networks. Phys. Chem. Earth (B), 26, 1, pp.1-7.

[11] Haykin, S. 1998. Neural Networks – A Comprehensive Foundation. Second Edition. Prentice Hall, Englewood Cliffs, NJ.

[12] Karel, H. 2011. SARIMA Models for Temperature and Precipitation Timeseries in the Czech Republic for the Period 1961–2008. Journal of Applied Mathematics, vol.IV.

[13] Khajavi, E., Behzadi, J., Nezami, M., Ghodrati, A., Dadashi, M. A. 2012. Modelling ARIMA of Air Temperature of the Southern Caspian

Sea Coasts. International Research Journal of Applied and Basic Sciences Vol., 3 (6), 1279-1287 pp.
[14] Koutsoyiannis D. 1997. Statistical Hydrology. NTUA, Department of Water Sources, Athens.
[15] Koutsoyiannis, D. 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic timeseries. Wat. Resour. Re. 36(6), 1519-1534 pp.
[16] Koutsoyiannis, D. 2008. Stochastic Methods in Water Resources, National & Technical University of Athens.
[17] Lappas, I. 2018. Applied hydrogeological research in coastal aquifers. Case study of the coastal part of Atalanti region, Prefecture of Fthiotida. PhD Thesis Dissertation, School of Mining and Metallurgical Engineering, National and Technical University of Athens, p.487.
[18] Maier, H., Dandy, G. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling & Software, 15, pp.101-124.
[19] Manakos, A. 1999. Hydrogeological Behaviour and Stochastic Simulation of Krania Elassona Karstic Aquifer in Thessaly. PhD Thesis, Aristotle University of Thessaloniki, 214 pp.
[20] Manakos, A., Dimopoulos, G. 2004. Contribution of Stochastic Models to the Sustainable Water Management. The example of Krania Elassona Karstic Aquifer in Thessaly. Proceedings of the 10th International Conference of Greek Geological Society, Thessaloniki.
[21] Manakos, A., Georgiou, P. 2009. Timeseries modelling of groundwater head Using Seasonal Stochastic Models SARIMA. Proceedings of the Common Conference of the 11th Hellenic Hydrotechnical Society and of the 7th Conference of the Hellenic Committee of Water Management, Volos.
[22] Meher, J. 2014. Rainfall and runoff estimation using hydrological models and ANN techniques. PhD, National Institute of Technology, Rourkela, India, p.218.
[23] Mimikou, M. 1994. Technology of Water Resources Systems, NTUA, Athens.
[24] Minns, A. and Hall, M. (1996). Artificial neural networks as rainfall-runoff models. Hydrologic Sci. 41(3), pp.399-417.
[25] Mutlu, E., Chaubey, I. Hexmoor, H., Bajwa, S. 2008. Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed. Hydrol. Process.
[26] Nash, E., Sutcliffe, V. 1970. River flow forecasting through conceptual models: 1. A discussion of principles. Journal of Hydrology, 10(3), pp.282-290.
[27] Nastos, P., Moustris, K., Larissi, I., Paliatsos, A. 2013. Rain intensity forecast using Artificial Neural Networks in Athens, Greece. Atmospheric Research, 119, pp.153–160.
[28] Papamichail, D. 1989. Investigation and Comparison of Functional Models with Stochastic Procedures for the Solution of Hydrological Problems. PhD Thesis, Thessaloniki.
[29] Papamichail, D. 1993. Seasonal ARIMA Modelling of Acheloos River Monthly Discharge. Proceedings of the 2nd Hydrogeological Congress, Patra.
[30] Papoulis, A. 2002. Probality, Random Variables and Stochastic Processes, Fourth Edition, Mc-Graw Hill.
[31] Paschalis, A. 2009. Stochastic Simulation of the Spatial Structure of Rain. Thesis, NTUA, Department of Water Resources and Environment, Athens.
[32] Raman, R., Sunilkumar, N. 1995. Multi-variate modeling of water resources time series using artificial neural networks. Hydrological Sci. 40, pp.145-163.
[33] Ripley, B. D. 1987. Stochastic Simulation, Wiley, New York.
[34] Sajikumar, N., Thandaveswara, B. 1999. A non-linear rainfall runoff model using an artificial neural network. Journal of Hydrology, 216, pp.32-35.
[35] Salas, J. D. 1992. Analysis and Modelling of Hydrologic Timeseries. Maidment, D. R., ed., Handbook of Hydrology, New York.
[36] Shamsnia, S. A., Shahidi, N., Liaghat, A., Sarraf, A., Vahdat, S. F. 2011. Modelling of Weather Parameters Using Stochastic Methods (ARIMA Model) (Case Study: Abadeh Region, Iran). International Conference on Environment and Industrial Innovation, IPCBEE, vol.12, Singapore.
[37] Smith, J., Eli, R. 1995. Neural-network models of rainfall-runoff process. J. Water Resour. Plng. and Mgmt., pp.499-508.
[38] Tokar, A., Johnson, P. 1999. Rainfall-runoff modeling using artificial neural networks. Journal of Hydrologic Engineering, 4(3), pp.232-239.
[39] Tolika, K., Maheras, P., Vafiadis, M., Flocas, H., Arseni–Papadimitriou, A. 2007. Simulation of seasonal precipitation and rain days over Greece: a statistical downscaling technique based on artificial neural networks (ANNs). International Journal of Climatology, Vol 27, pp.861-881.