

SEASONAL STOCHASTIC SIMULATION AND TIMESERIES MODELLING - ANALYSIS OF A KARSTIC SPRING IN CENTRAL MACEDONIA, GREECE

Lappas I.¹ and Lazaridou M.²

¹*Institute of Geology and Mineral Exploration, Department of Hydrogeology, 1st Spirou Loui St., Olympic Village, 13677 Acharnae, Athens, Greece*

²*Institute of Geology and Mineral Exploration, Branch of Thessaloniki, 1st Fragon St., 54626, Greece*

Abstract

The objective of this paper is to find an appropriate Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model for fitting the monthly discharge of a karstic spring located at the North of the city of Serres (Agios Ioannis, Mount Menikio) by considering the minimum of Akaike Information Criterion (AIC). Box-Jenkins methodology applies models to find the best fit of a timeseries to past values of this timeseries, in order to make forecasting and consists of a four-step iterative procedure: identification, estimation, diagnostic check and forecasting. Timeseries analysis and forecasting of hydrological parameters such as spring discharge may be useful in decision making and optimum water resources usage. In this study, monthly discharge measurements are analysed. Initial data are firstly transformed to normal and stationary using differencing methods. Autocorrelation and Partial Autocorrelation functions are calculated to determine the order of Autoregressive and Moving Average parameters and residuals are then checked to show the "white noise". The spring discharge data are forecasted based on the selected model up to 2008 and are then compared with measured values. The timeseries model SARIMA (2,1,1)(1,0,1)₁₂ could be used in monthly discharge forecasting at a short time (upcoming one year) with a simple and explicit model structure in order to help decision makers to establish priorities in terms of water demand management. Finally, the correlation coefficient between the observed and fitted data is essentially high, while the absolute and relative errors are significantly low.

Keywords: SARIMA model, Agios Ioannis, Box - Jenkins methodology, spring discharge forecasting, transformation, Autocorrelation Function (ACF).

Περίληψη

Ο αντικειμενικός στόχος της εργασίας είναι η εύρεση του κατάλληλου εποχικού μοντέλου SARIMA για την καλύτερη προσαρμογή των μηνιαίων παροχών της καρστικής πηγής, που βρίσκεται στα Βόρεια της πόλης των Σερρών (Άγιος Ιωάννης, Όρος Μενοίκιο), λαμβάνοντας υπόψη τη χαμηλότερη τιμή του κριτηρίου Akaike. Η μεθοδολογία των Box-Jenkins εφαρμόζεται ώστε να βρεθεί το κατάλληλο μοντέλο, που να προβλέπει μελλοντικές χρονοσειρές σύμφωνα με παλαιότερα δεδομένα ακολουθώντας μια επαναληπτική διαδικασία τεσσάρων (4) βημάτων: αναγνώριση, εκτίμηση, έλεγχος διάγνωσης και πρόβλεψη. Η ανάλυση χρονοσειρών και η πρόβλεψη υδρολογικών παραμέτρων, όπως η παροχή πηγών μπορεί να αποβούν σημαντικά χρήσιμες στη λήψη

αποφάσεων και τη βέλτιστη χρήση των υδατικών πόρων. Στη μελέτη αυτή, μηνιαίες μετρήσεις παροχών πηγής αναλύονται. Τα αρχικά δεδομένα κανονικοποιούνται και μετατρέπονται σε στάσιμα, χρησιμοποιώντας μεθόδους διαφόρισης. Οι συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης υπολογίζονται για τον προσδιορισμό της τάξης των παραμέτρων της Αυτοπαλινδρόμησης και του Κινούμενου Μέσου και κατόπιν ελέγχονται τα υπόλοιπα για την ύπαρξη “λευκού θορύβου”. Οι παροχές πηγών προβλέπονται, με βάση το επιλεγμένο μοντέλο, έως το 2008 και στη συνέχεια συγκρίνονται με ήδη μετρηθείσες τιμές. Το εποχικό μοντέλο SARIMA(2,1,1)(1,0,1)₁₂ μπορεί να χρησιμοποιηθεί για μηνιαία πρόβλεψη των παροχών πηγής για ένα σύντομο χρονικό διάστημα (έως ένα έτος), ούτως ώστε να τεθούν προτεραιότητες στη διαχείριση των υδατικών αναγκών. Τέλος, ο συντελεστής συσχέτισης μεταξύ των παρατηρημένων και των υπολογισμένων τιμών είναι σημαντικά υψηλός, ενώ τα απόλυτα και τα σχετικά σφάλματα σημαντικά χαμηλά.

Λέξεις κλειδιά: Μοντέλο SARIMA, Άγιος Ιωάννης, μεθοδολογία Box - Jenkins, πρόγνωση παροχών πηγής, μετασχηματισμός, συνάρτηση αυτοσυσχέτισης (ACF).

1. Introduction

Spring discharge forecasting is significant to water resources management and exploitation since groundwater from karstic limestones is of great importance to local and/or regional hydroeconomy due to high water demands. Timeseries modelling and analysis is one of the most widely used predicting methods having three modelling stages; identification, estimation and diagnostic check. Therefore, there has been considerable interest in stochastically modelling hydrological parameters. Hydrologically speaking a stochastic variable such as spring discharge is consisted of a deterministic and stochastic time-dependent part, namely white noise (Koutsoyiannis, 2000, 2008). 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. The existence of the deterministic part means that the timeseries are non-stationary and should be therefore stationary. Box - Jenkins models are significantly used because of the simple structure and the relatively small number of parameters to stationary and non-stationary procedures. 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 simplifying assumptions which cannot approach the physical – natural conditions (Ripley, 1987).

1.1. Site Location – Geomorphology

Menikio karstic aquifer belongs to the Water District of Central Macedonia (Figure 1). It is consisted of marbles surrounded by Serres and Drama basins and located at the Northeast of the city of Serres. The mountainous range has a Northeast to Southwest direction while the highlands are situated at Mantili (1903m), Thamnopoti (1952m), Mavromata (1963m), Kallipoli (1203m) and Polykorfo (1258m) sites. The highlands' characteristic geomorphological aspect is that they do not constitute an individual top of a mountain (as usually happens in other ranges) but instead smooth and almost flat surfaces with minor uplifts (relief with gentle slopes, which become steep towards the western outskirts). The karstic aquifer covers an area of approximately 289km² with the major and minor axis of 25.4Km and 16Km respectively. The mean elevation reaches 850m ranging between 100m and approximately 2000m a.s.l. The dendritic drainage network is mainly composed of intermittent rivers flowing in a NNW-SSE direction. The mountainous area show a sparse drainage network, developed due to the presence of permeable formations (intensively tectonized marbles), while the valley areas in the southwest form a relatively dense hydrographical network because of the less permeable alluvial deposits (Figure 1).

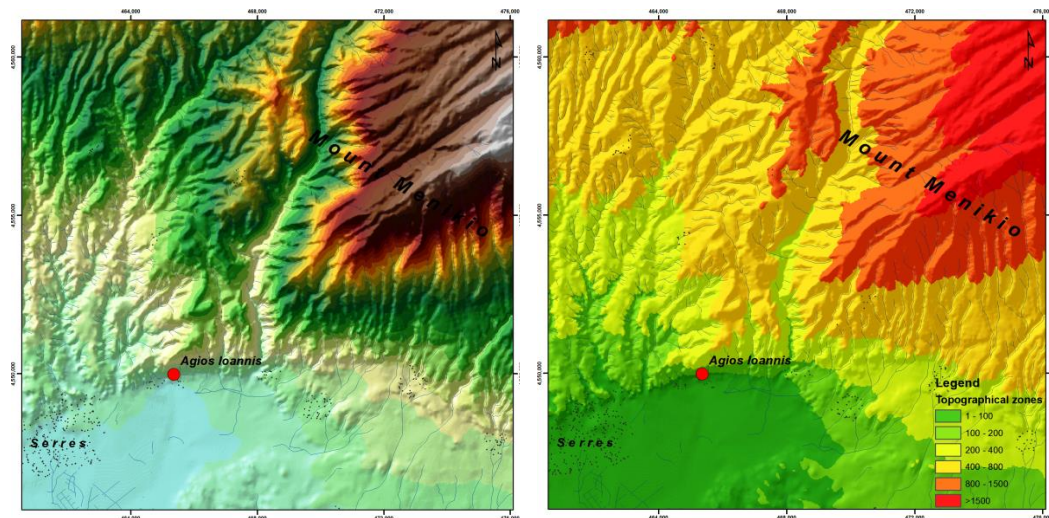


Figure 1 – (left) Geomorphological and (right) topographical zones maps.

1.2. Geological – Hydrogeological Setting

The mountainous range of Menikio geotectonically belongs to Rhodope zone forming an extensive karstic aquifer and is bounded by Serbo-Macedonian zone and Ring-Rhodope one (Lazaridou et al., 2001a). The Rhodope zone's metamorphic rocks occupy almost the entire area of mount Menikio and are consisted of three lithostratigraphic units such as gneiss (lower part of Rhodope mass), marbles (intensively tectonized) and lithological interchanges (schists, marbles and gneiss) (Figure 2).

The Menikio karstic aquifer is developed under the last two geological formations and characterized by several permeable lithologies either independent or with hydraulic communication between them (Figure 2). These individuals less extensive karstic aquifers are commonly associated with springs at different altitudes based on geological, stratigraphic and local conditions (tectonics, geomorphology etc.). The exact orientation of those aquifers has not been done yet due to the complex geological structure (neither a single karstic aquifer nor known karst base depth) and their unknown hydraulic behavior with adjacent aquifers (Lazaridou et al., 2001a, b). The groundwater flow and direction initially recharges the epikarstic water beds and then follows the deep structure of karst according to the already formed tectonic lines of faulting zones, the rock karstification etc.

Both springs location and karstic aquifers' delineation are defined by the non-permeable intercalations and is believed that karstic groundwater recharges the springs of mount Menikio as well as the basin's porous formations. Within the Menikio karstic aquifer two main permeable beds are discerned; one in the regional area of Eptamila site (hydraulic head of +67m a.s.l.) where Agios Ioannis spring is the main surface discharge point and the other in the wider area of Emmanouil Papa with hydraulic head of +122m a.s.l. Taking into account the pumping tests of several boreholes around the study area, the pumping rates were of the order of 60 - 130m³/h with drawdown ranges from few centimeters to 10m. Moreover, transmissivity values were estimated to be 10⁻¹ - 10⁻³m²/sec (Lazaridou, 2010).

Agios Ioannis spring. Located at the Southwest of mount Menikio, approximately 2km at the east of the city of Serres, at an altitude of ~60m, considered as contact - overflow spring (Figure 3). Based on the available data the discharge rate ranges from 0.44 to 0.79m³/sec with mean value of 0.61m³/sec. The mean yearly runoff is of the order of 19x10⁶m³ with a relative uniformity within a year (spring; 26.3%, winter; 25.2%, summer; 23.4%, fall; 25.1%).

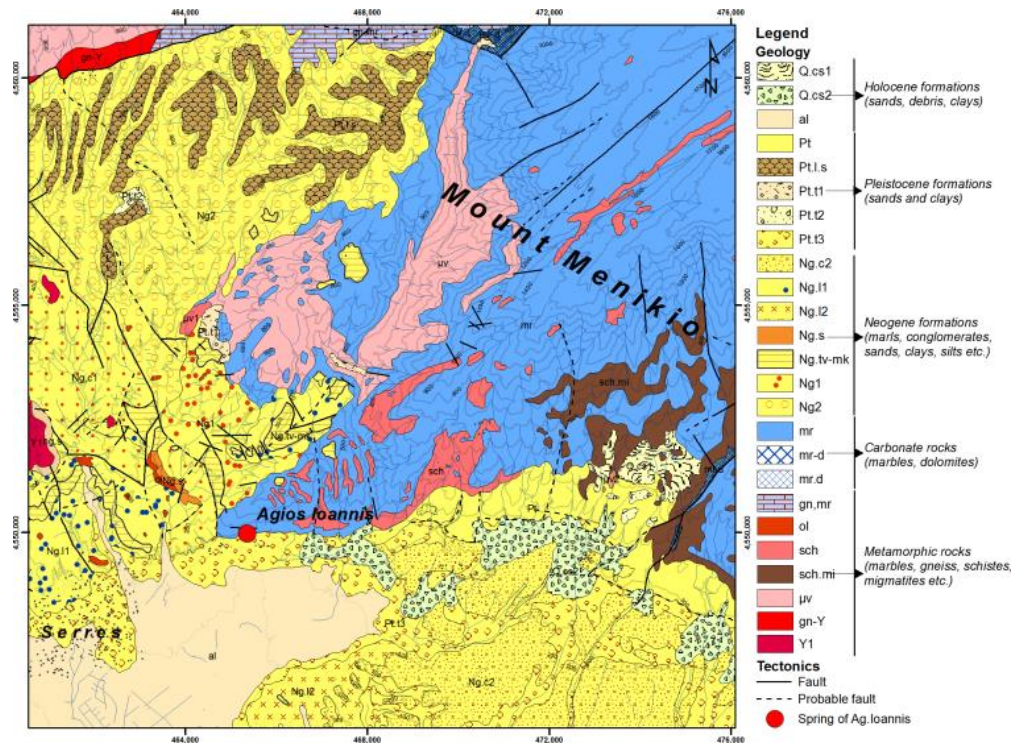


Figure 2 – Geological map of the study area (Xydas et al., 1985).

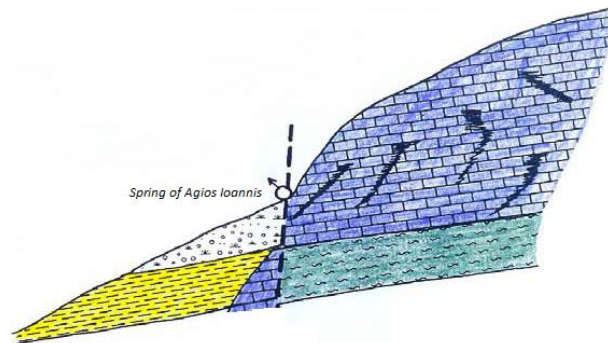


Figure 3 – Schematic cross-section of Agios Ioannis karstic spring.

Depletion curve – Dynamic spring groundwater resources. According to Agios Ioannis depletion curve, the groundwater resources volume is approximately $22 \times 10^6 \text{m}^3$ derived by the completion of the Maillet equation (Lazaridou et al., 2001a, b). Also, the coefficient of depletion is of the order of $2 \times 10^{-3} (\text{d}^{-1})$ representing the resources evacuation rate and showing the karstic aquifer's extensive hydrocapacity.

2. Applied Methodology

The most important stage in the process of modelling is to check for the series stationarity. Stationarity could be achieved mostly by differencing the series using data transformation methods such as Box-Cox, logarithmic and square root. There are also periodical tendencies named seasonality or periodicity. Finally, residual scores determine whether there are still patterns in the

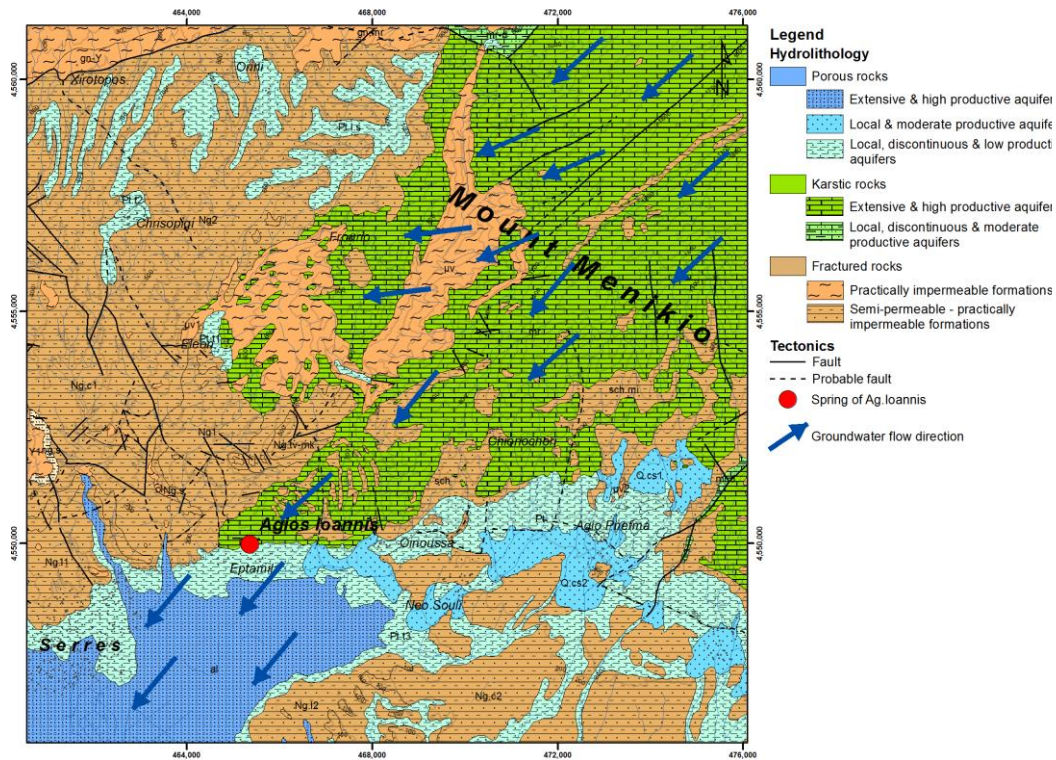


Figure 4 – Hydrogeological map of the study area.

data that are not accounted for. Residual scores are the differences between the scores predicted by the model and the actual scores for the series. After the aforementioned stages an equation is derived to adequately predict the future of the process called forecasting (Manakos *et al.*, 2004). Forecasting refers to the process of predicting future observations from a known timeseries which is the main objective in non-experimental use of the analysis.

In this paper, Agios Ioannis spring discharge is modeled by SARIMA model for the period from 1977 to 2001. Two divisions have been made, one calibration period from 1977 to 2001 to best fit the most suitable model and one validation period from 2003 to 2006. The derived model is then used for predicting discharge for the next two years until 2008. A preliminary timeseries analysis of the monthly discharge data has to be implemented to explore extreme values, homogeneity, step and trends. Box and Jenkins (1976) proposed a general form of SARIMA(p,d,q)(P,D,Q)₁₂ for non-stationary timeseries, in which p, q are autoregressive and moving average parameters and P, Q are seasonal autoregressive and moving average parameters, respectively (Box *et al.*, 1964; Papamichail, 1993; Manakos *et al.*, 2004, 2009). Moreover, these models show deterministic and stochastic trends and can explain both stochastic and seasonal variations. The general form of Seasonal ARIMA(p,d,q)(P,D,Q)_s is as follows:

Equation 1 - Formula for SARIMA(p,d,q)(P,D,Q)_s

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

where φ_p , Φ_P the autoregressive processes and θ_q , Θ_Q define moving average processes that number of p, P, q and Q are estimated from Autocorrelation and 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; identification, estimation and validation model. The

identification process is based on the behavior of autocorrelation and partial autocorrelation functions, efficient parameters' estimation can be obtained after identifying the autocorrelation and partial autocorrelation functions and finally, validation process is based on the goodness of fit. (Ripley, 1987).

3. Model Design and Timeseries Modelling

Agios Ioannis spring discharge data are derived from the integrated hydrogeological study made during the 3rd Community Support Framework concerning monthly values for the time period 1977 to 2008. In Figure 5, the observed discharge data are examined for the existence or not of trend and stationarity.

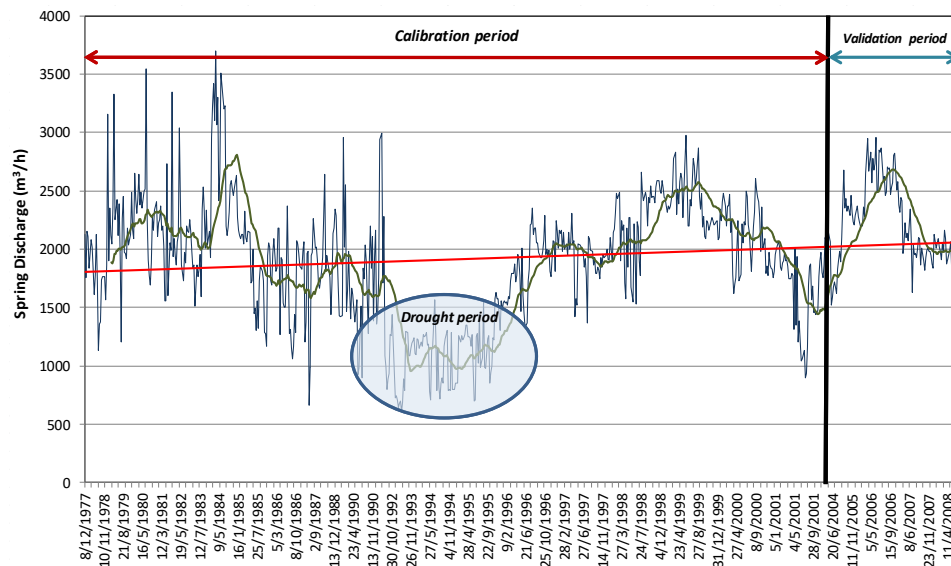


Figure 5 - Agios Ioannis monthly discharge data for the time period 1977 - 2008. The red thick line shows the trend while the green one the 24-month moving average of the observed timeseries. Both calibration and validation periods are also shown.

First of all, there was no need to transform the timeseries into normal and stationary using differencing method (Figure 6) because after testing the normality of the data series by normal Q - Q plot and histogram, showed that the data do have normal distribution, therefore no seasonal order differencing ($D=0$) is used. Moreover, the above data seem to have trend, that is, $d=1$ (seasonal trend), the timeseries is not of "white noise" and the seasonal length equals to 12 months. For stationarity checking the autocorrelation and partial autocorrelation functions are calculated for the estimation of p , q , P and Q components of the Autoregressive (AR) and Moving average (MA) model respectively. Finally, in order to find the model that fits the best, successive trials are made so that the predicted values to be as close as those of the observed ones (Figure 7).

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. Some of them are based on statistics (residuals sum up) such as AIC (Akaike Information Criterion) and SBC (Schwartz-Bayesian Criterion) and others on the forecasting error, like the Mean Percent Error (MPE), the Mean Square Error (MSE), the Mean Absolute Error (MAE) and the Mean Absolute Percent Error (MAPE) (Salas, 1992). The most suitable model is the one that the above statistics are the least. The Akaike information criterion gives the measure of the relative goodness of fit of a statistical model (Papamichail, 1993; Manakos *et al.*, 2004, 2009). The preferred model is the one with the minimum AIC value. Consequently,

AIC not only proves goodness of fit, but also discourages overfitting, that is, an increasing function of the number of estimated parameters.

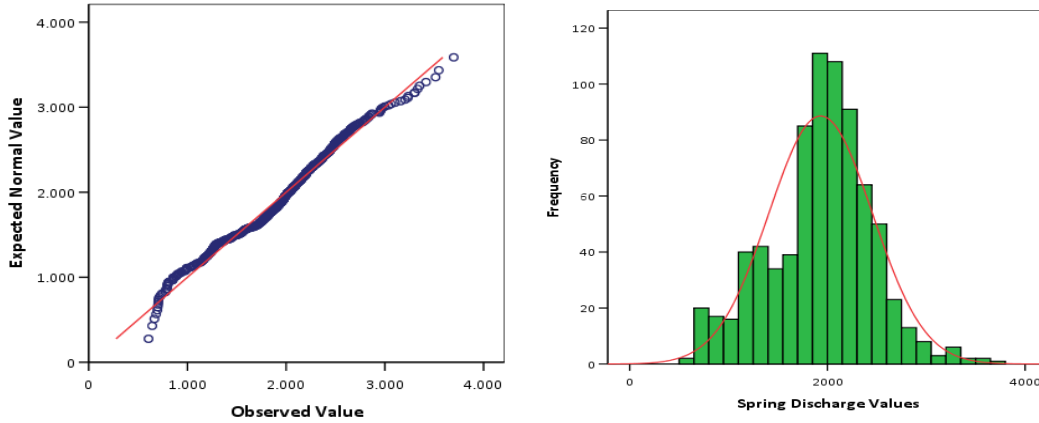


Figure 6 - Normal Q-Q plot ($p < 0.005$) and histogram with normal distribution curve showing the normality of the monthly spring discharge values.

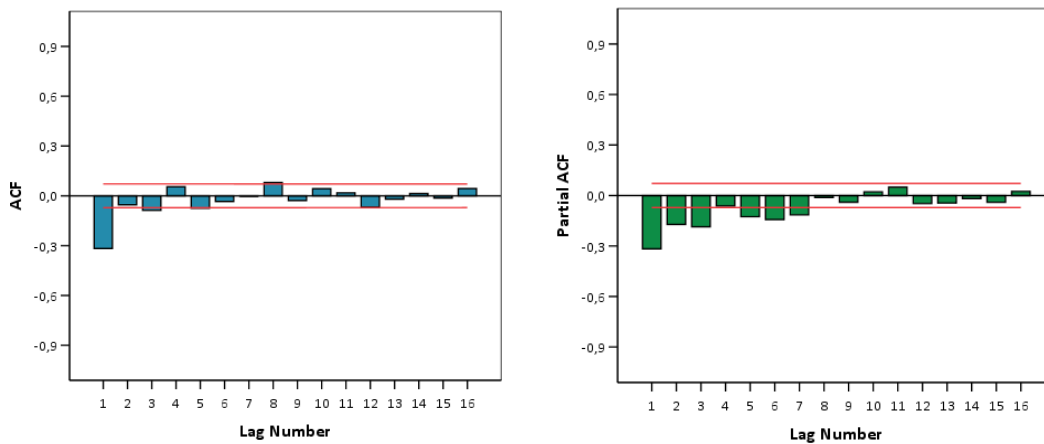


Figure 7 - Autocorrelation (ACF) and Partial Autocorrelation functions (PACF) of the discharge timeseries for the time period 1977 – 2001. The red thick lines show the confidence limits 95%.

Autocorrelation function (ACF) refers to the way the observations in a timeseries are related to each other while partial autocorrelation is used to measure the degree of association when the effects at other time lags are removed. 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. The time period from 1977 to 2001 is used for stochastic model calibration and then is validated during the period 2003 - 2006 for the confirmation of the model's parameters. The best fit and most suitable seasonal stochastic model satisfying the majority of the aforementioned criteria is SARIMA(2,1,1)(1,0,1)₁₂ and obviously can be used for monthly spring discharge estimation and forecasting. The residuals check takes place so as to prove if there is white noise taking into consideration the autocorrelation (ACF) and partial autocorrelation (PACF) functions and it seems that the residuals are of white noise which means uncorrelated. Additionally, (Figures 8, 9) shows the very good simulation and correlation of the above model on the measured discharge data.

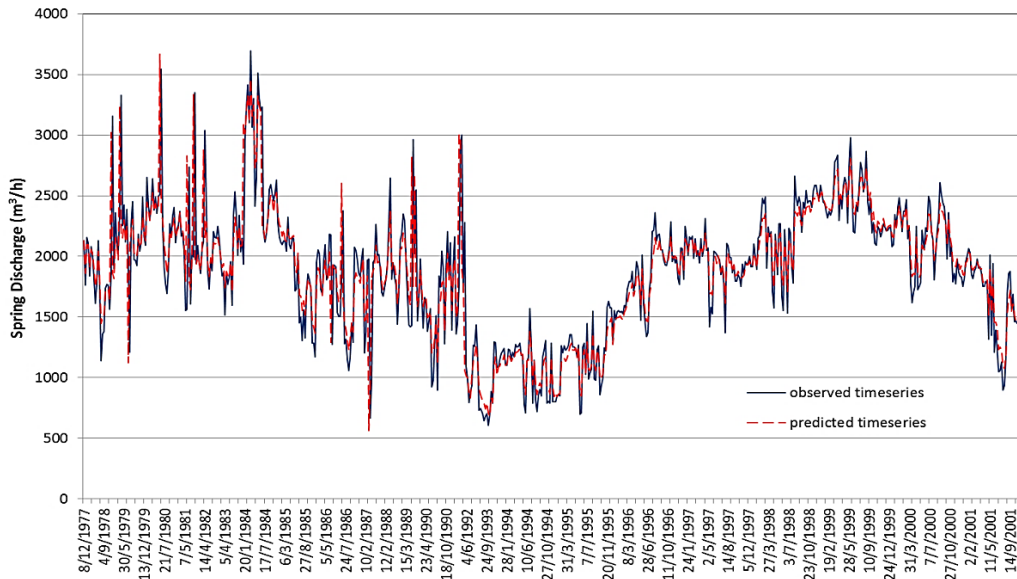


Figure 8 - SARIMA(2,1,1)(1,0,1)₁₂ model simulation at the observed timeseries.

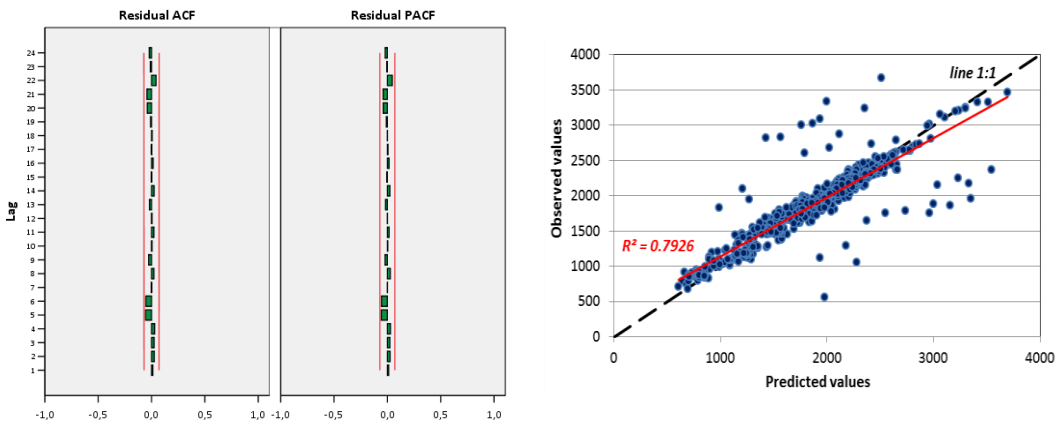


Figure 9 - (Left) Autocorrelation (ACF) and Partial Autocorrelation functions (PACF) of the residuals for the time period 1977 – 2001 (The red lines show the confidence limits 95%). (Right) SARIMA model generated series versus the observed monthly one ($R^2=0.7926$).

4. Spring Discharge Data Prediction

SARIMA model is applied as long as the above discharge timeseries can provide safe predictions and create, if possible, reliable future time period. As far as the model procedure is concerned the timeseries of months (1977 - 2008) is divided into two parts, the one of period month 1977 - 2001) and the other of period month of 2003 - 2006). SARIMA model is applied to the first time period of months trying to simulate the observed discharge values and then to the next already known months to validate and confirm the model. Re-estimation of model parameters is required for the last period month. The re-assessed model components slightly differ from the initial ones showing that the spring discharge data that are derived by the application of SARIMA(2,1,1)(1,0,1)₁₂ model may be used for a very satisfactory forecasting. Hence, it is shown that SARIMA(2,1,1)(1,0,1)₁₂ model being capable of adequately simulating the timeseries data, can be used to reliably predict Agios Ioannis discharge data. The analytical, mathematical expression of the above model is of the form:

Equation 2 – The SARIMA model analytical mathematical expression

$$(1-\phi_1B)(1-\phi_2B)(1-B^{12})Z_t=e_t \text{ or } Z_t=\phi_1(Z_{t-1}-Z_{t-13})\phi_2(Z_{t-2}-Z_{t-14})+Z_{t-12}+e_t$$

with $\phi_1=0.459$ and $\phi_2=0.048$. The very good fit of the model (Figure 10) is proved to be useful as far as the rational and sustainable water resources management and planning are concerned. The stochastic Seasonal ARIMA model seems to sufficiently describe the spring discharge and get acquainted satisfactorily to natural conditions while they are able to provide short term predictions - forecasting contributing to the sustainable exploitation of water resources. Finally, the seasonal ARIMA model can provide synthetic timeseries which may be used for the water resources planning proving its usefulness.

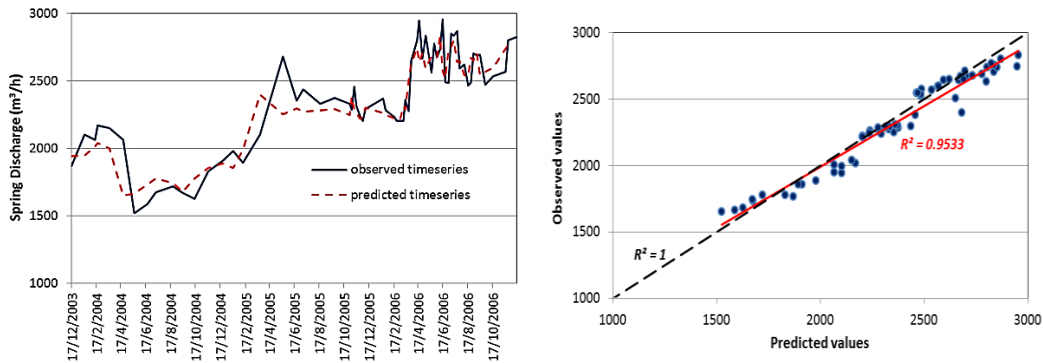


Figure 10 - SARIMA(2,1,1)(1,0,1)₁₂ model validation at the observed timeseries for the time period of 2003 – 2006 (R²=0.9533).

Table 1 - Monthly spring discharge predicted data for the reference years 2007 - 2008.

Month / Year	Estimated values (mm)	Observed values (mm)	Month / Year	Estimated values (mm)	Observed values (mm)
Jan-07	2601.4	2577.6	Jan-08	2039.6	2124.0
Feb-07	2405.5	2314.8	Feb-08	2016.8	2016.0
Mar-07	2488.8	2444.9	Mar-08	2021.6	2052.0
Apr-07	2472.0	2412.4	Apr-08	1952.0	1908.0
May-07	2300.5	2223.7	May-08	2026.5	2066.4
Jun-07	2154.4	2070.0	Jun-08	1956.8	1950.6
Jul-07	1864.2	1630.8	Jul-08	2021.5	2037.1
Aug-07	2033.1	1959.1	Aug-08	1973.0	1985.5
Sep-07	2013.6	2001.6	Sep-08	1950.8	1924.0
Oct-07	1971.5	1900.8	Oct-08	1864.6	1821.6
Nov-07	1987.4	1990.8	Nov-08	2038.0	2105.7
Dec-07	1953.6	1944.0	Dec-08	2127.9	2189.6

5. Conclusions

According to all mentioned above seasonal stochastic simulation is a method which can be easily applied since the major advantage is the flexibility to perform in complex systems describing them

reliably, without simplistic and robust assumptions and when analytical methods are either not applicable or require oversimplifying assumptions for the system. Nevertheless, it has to be pointed out that stochastic modelling is an approximate procedure and its accuracy depends on the sample size (the bigger the sample size the more reliable results). Finally, SARIMA models are capable to show deterministic and stochastic trends and efficiently simulate the stochastic and seasonal variations.

Discharge data are statistically and stochastically studied over Agios Ioannis karstic spring at the Northeast of the city of Serres. Using seasonal stochastic decomposition method, monthly mean discharge values are forecasted with high accuracy and reliability since average of observed and predicted data are proved almost the same. This is confirmed by the correlation coefficient between measured and predicted timeseries which is about 0.80. Moreover, SARIMA (2,1,1)(1,0,1)₁₂ model seems to be the most appropriate in simulating the observed timeseries and is capable to forecast future data for at least two years. Also, the goodness of fit as well as its easy adaptation to natural - physical conditions can be regarded a useful tool to the sustainable water resources management and exploitation.

6. Acknowledgements

The authors wish to express their thanks to all those hydrogeologists and technical staff of the Institute of Geology and Mineral Exploration (branch of Thessaloniki) for the unconditional help offered to all phases of the groundwater field sampling work.

7. References

- Box, G. and Cox, D., 1964. An Analysis of Transformations, *J. R. Stat. Soc.*, Ser. B, 26, 211-252.
- Box, G. and Jenkins, G., 1976. Time Series Analysis, Forecasting and Control, Holden - Day, San Francisco, California, U.S.A.
- Koutsoyiannis, D., 2000. A generalized mathematical framework for stochastic simulation and forecast of hydrologic timeseries, *Wat. Resour. Re.*, 36(6), 1519-1534.
- Koutsoyiannis, D. 2008. Stochastic Methods in Water Resources, National & Technical University of Athens.
- Lazaridou, M. and Papadopoulos, K., 2001a. Hydrogeological research of carbonate rocks of mounts Menikio and Agkistro, *2nd Community Support Framework*, I.G.M.E. Branch of Thessaloniki.
- Lazaridou, M. and Polizonis, E., 2001b. Quality monitoring and water resources control of Central Macedonia, Prefecture of Serres, *2nd Community Support Framework*, I.G.M.E. Branch of Thessaloniki.
- Lazaridou, M., 2010. Basins water balance. Quality parameters monitoring and protection measures of the groundwater in Central Macedonia, *3rd Community Support Framework*, I.G.M.E. Branch of Thessaloniki.
- Manakos, A. and 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.
- Manakos, A. and 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.
- Papamichail, D., 1993. Seasonal ARIMA Modelling of Acheloos River Monthly Discharge. *Proceedings of the 2nd Hydrogeological Congress*, Patra.
- Ripley, B.D., 1987. Stochastic Simulation, Wiley, New York.
- Salas, J.D., 1992. Analysis and Modelling of Hydrologic Timeseries, Maidment, D.R., ed., *Handbook of Hydrology*, New York.
- Xydas, S. and Staikopoulos, G., 1985. Serres sheet geological map, scale 1:50.000. I.G.M.E.