

# The Use of Stochastic Models for Short-Term Prediction of Water Parameters of the Thesaurus Dam, River Nestos, Greece <sup>†</sup>

Antonis Sentas <sup>1</sup>, Lina Karamoutsou <sup>1</sup>, Nikos Charizopoulos <sup>2</sup>, Thomas Psilovikos <sup>3</sup>,  
Aris Psilovikos <sup>1,\*</sup> and Athanasios Loukas <sup>4</sup>

<sup>1</sup> Department of Ichthyology and Aquatic Environment, School of Agricultural Sciences, University of Thessaly, Fytoko st., 38446 N. Ionia Magnisias, Greece; antsentas@yahoo.gr (A.S.); lkaramoytsoy@gmail.com (L.K.)

<sup>2</sup> Laboratory of Minearology-Geology, Agricultural University of Athens, Iera Odos 75, 118 55 Athens, Greece; nchariz@gmail.com

<sup>3</sup> Department of Forest and Water engineering, Laboratory of Mechanical Science and Topography, School of Forestry & Natural Environment, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece; tvikos@ad.auth.gr

<sup>4</sup> Department of Civil Engineering, School of Engineering, University of Thessaly, Pedion Areos, 38221 Volos, Greece; aloukas@uth.gr

\* Correspondence: psiloviko@uth.gr; Tel.: +30-242-109-3154

† Presented at the 3rd EWaS International Conference on “Insights on the Water-Energy-Food Nexus”, Lefkada Island, Greece, 27–30 June 2018.

Published: 30 July 2018

**Abstract:** The scope of this paper is to evaluate the short-term predictive capacity of the stochastic models ARIMA, Transfer Function (TF) and Artificial Neural Networks for water parameters, specifically for 1, 2 and 3 steps forward ( $m = 1, 2$  and  $3$ ). The comparison of statistical parameters indicated that ARIMA models could be proposed as short-term prediction models. In some cases that TF models resulted in better predictions, the difference with ARIMA was minimal and since the latter are simpler in their construction, they are proposed for short-term prediction. Artificial Neural Networks didn't show a good short-term predictive capacity in comparison with the aforementioned models.

**Keywords:** stochastic models; ARIMA; Transfer Function (TF); prediction; Artificial Neural Networks (ANN); monitoring data; dissolved oxygen; water temperature; Thesaurus dam; River Nestos; Greece

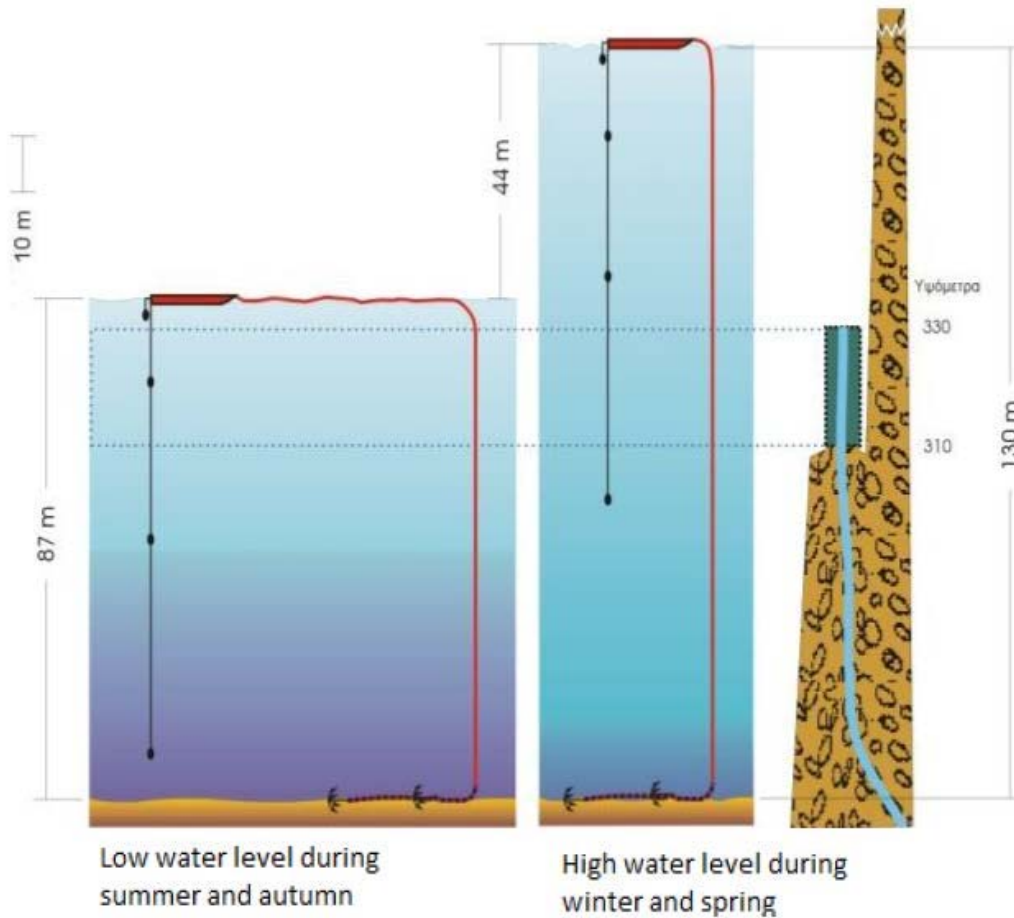
---

## 1. Introduction

The central part of River Nestos flow is one of the areas with the most intensive anthropogenic interference because of the construction and operation of Thesaurus and Platanovrissi dams. The purposes of the these dams are (a) hydroelectric power for the increased needs in the areas of Eastern Macedonia and Thrace; (b) anti-flooding protection of villages, cities, agricultural and touristic activities in Nestos downstream; (c) irrigation of the fertile deltaic plain and (d) ecological protection and preservation in Nestos Delta, which is one of the 11 Hellenic areas that is included in the Ramsar's Convention [1,2].

In the deeper section of the reservoir basin, upstream of the Thesaurus dam, a floating telemetric station (Figure 1) was anchored to record changes in water temperature ( $T_w$ ) and dissolved oxygen (DO) at four different depths (1, 20, 40 and 70 m). For these parameters, the daily

records of the years 2004 to 2007 were analyzed and evaluated [3–6]. This research was completed by the end of the year 2007.



**Figure 1.** Thesaurus floating station and measurement depths during the year [7].

This study aims to evaluate the predictive capability of ARIMA, Transfer Function (TF) and Artificial Neural Networks (ANN) models for short-term prediction and more specifically for 1, 2 and 3 steps forward ( $m = 1, 2$  and  $3$ ).

## 2. Methodology

An adequate number of continuous measurements (1440 on a daily basis) of the DO and  $T_w$  parameters, were used for the period from 19 January 2004 to 28 December 2007 and corresponded to four different depths (1, 20, 40 and 70 m). In order to evaluate the predictability of the models, the data set was split into two subsets: (a) the historical period with 1420 measurements and (b) the forecasting period with the last 20 measurements. The former was used to select the best adapted model in the data, while the latter was used in order to forecast the  $m$  step forward ( $\hat{y}_t(m)$ ) [7–9].

If  $m = k$  then the prediction occurs for “ $k$  steps forward”; this means the prediction of the observation at time  $t + k$ , given the observations up to time  $t$ , which is called start time. Four statistical parameters, MSE, RMSE, MAPE and NSC, were used in order to assess the forecast. Finally, 95% confidence intervals were calculated for all the models selected [10].

The linear and nonlinear stochastic models which were tested are: (a) ARIMA, (b) Transfer Function and (c) Neural Network models. Their compilation, as well as the prediction results for  $m = 1$ , are discussed in detail in Sentas et al., 2016 [10].

The 20-day forecast was considered sufficient as the Thesaurus reservoir was first operated in 2000. Moreover, since the end of 1997, when its construction was completed, its water filling has begun. From the end of 1997 to the early 2000 some intense phenomena of water’s stationarity were

observed, accompanied by anoxia, with reductive processes to dominate, by hydrogen sulfide production in the hypolimnion and by the development of sulfur bacteria [11].

By the time the reservoir started operating and the water being discharged for various uses (hydroelectric production, irrigation, flood protection, ecological), from the beginning of the year 2000 until nowadays, is still trying to restore the conditions of equilibrium between the natural environment and human interventions.

Thus, a period of four years (early 2004–late 2007) would be necessary for monitoring the hydro-system of Nestos-Thesaurus complex's characteristics, throughout the water column and at various depths. This is because slight changes occur in those characteristics (e.g., the temperature in hypolimnion). In any other case of an existing reservoir, it would take a maximum of 2 years for complete monitoring and reliable results' extraction. Nevertheless, in the case of the Thesaurus, it was considered that a database of 4 consecutive years is required.

Therefore, the forecasting was chosen to be performed at the end of the four-year period (December 2007), with no interim periods e.g., the spring of the same or the previous year, otherwise it would not exactly correspond to the interpretation of the physical problem.

### 3. Results

To test their predictive capacity, the following stochastic models were constructed and tested: (1) ARIMA, (2) Transfer Function and (3) Neural Network models. For each depth, an ARIMA model, a neural network model and a TF model were adapted to the dissolved oxygen time series. In total, four ARIMAs, four neural network models, and four TF models with a time-out of dissolved oxygen at each depth and input variable were adapted to the water depth of the corresponding depth. Using all of the above models, provision was made for  $m = 1$  step forward [12–15].

#### *ARIMA Models*

For each depth, an ARIMA model, a neural network model and a TF model were adapted to the dissolved oxygen time series [12]. In total, four ARIMAs, four neural network models and four TF models with output variable the dissolved oxygen at each depth and input variable the temperature of the corresponding water depth, were adapted. By using all the aforementioned models, a short-term prediction was attempted for  $m = 1, 2$  and  $3$  steps forward [16–18].

Details for the construction of the above models, as well as the prediction results for  $m = 1$ , are detailed in the paper Sentas et al., 2016 [10]. In the present paper, the results for  $2$  and  $3$  steps forward ( $m = 2$  and  $3$ ) are reported. In parallel, the results for  $m = 1$  are recorded in order to ensure better comprehension of short-term forecasting.

At  $1\text{ m}$  water depth the results are presented in Table 1.  $\hat{y}_i(m)$ , represents the forecast at the time  $t + m$  of a time series, in which observations are known up to time  $t$ . Similarly the results for depths  $20, 40$  and  $70$  meters are given in Tables 2–4, respectively.

The values listed in Table 1 lead to the conclusion that for both  $m = 2$  and  $m = 3$ , the TF model shows slightly better values for the statistical parameters than ARIMA and better than ANN, except in the case of the parameter NSC according to which the ANN model precedes.

At water depth of  $20\text{ m}$ , according to Table 2, for  $m = 1, 2$  and  $3$ , the ARIMA model results in better predictions with slight differences from the TF models.

At  $40\text{ m}$  water depth, according to Table 3, in all cases ( $m = 1, 2$  and  $3$ ) the TF models show better statistics. Also, the difference in the statistical parameters for steps  $1, 2$  and  $3$  between ARIMA and TF models is small.

At  $70\text{ m}$  water depth, according to Table 4, for prediction of  $1, 2$  and  $3$  steps forward ( $m = 1, 2$  and  $3$ ) the ARIMA models seem to precede in comparison to the others.

**Table 1.** Statistical parameters for comparison observed and predicted values using ARIMA, TF and ANN models and  $m = 1, 2$  and  $3$  (depth 1 m).

Thesaurus Station—DO Parameter—Depth 1 m					
$\hat{y}_t(m)$	Model	MSE	RMSE	MAPE	NSC
$m = 1$	ARIMA	<b>0.442</b>	<b>0.665</b>	<b>7.947</b>	<b>0.614</b>
	TF	0.452	0.672	7.974	0.605
	ANN	0.523	0.723	8.934	0.544
$m = 2$	ARIMA	0.852	0.923	11.158	0.256
	TF	<b>0.816</b>	<b>0.903</b>	<b>10.589</b>	0.196
	ANN	1.010	1.005	12.002	<b>0.599</b>
$m = 3$	ARIMA	0.960	0.979	11.713	0.163
	TF	<b>0.938</b>	<b>0.968</b>	<b>11.562</b>	0.532
	ANN	1.182	1.087	12.182	<b>0.545</b>

**Table 2.** Statistical parameters for comparison of observed and predicted values using ARIMA, TF and ANN models and  $m = 1, 2$  and  $3$  (20 m depth).

Thesaurus Station—DO Parameter—Depth 20 m					
$\hat{y}_t(m)$	Model	MSE	RMSE	MAPE	NSC
$m = 1$	ARIMA	<b>0.0077</b>	<b>0.0877</b>	<b>1.0552</b>	<b>0.7992</b>
	TF	0.0079	0.0891	1.0752	0.7931
	ANN	0.0434	0.2084	2.6721	-0.1303
$m = 2$	ARIMA	<b>0.0154</b>	<b>0.1241</b>	<b>1.505</b>	<b>0.9913</b>
	TF	0.0163	0.1278	1.6945	0.5751
	ANN	0.0570	0.2387	3.0021	0.0191
$m = 3$	ARIMA	<b>0.0215</b>	<b>0.1466</b>	<b>1.9787</b>	<b>0.9881</b>
	TF	0.0228	0.1511	2.0194	0.4061
	ANN	0.0882	0.2969	3.2124	0.0072

**Table 3.** Statistical parameters for comparison of observed and predicted values using ARIMA, TF and ANN models and  $m = 1, 2$  and  $3$  (40 m depth).

Thesaurus Station—DO Parameter—Depth 40 m					
$\hat{y}_t(m)$	Model	MSE	RMSE	MAPE	NSC
$m = 1$	ARIMA	0.0167	0.1293	1.5882	0.9167
	TF	<b>0.0131</b>	<b>0.1144</b>	<b>1.4791</b>	<b>0.9341</b>
	ANN	0.0170	0.1305	1.6781	0.9154
$m = 2$	ARIMA	0.0545	0.2334	2.5415	<b>0.9245</b>
	TF	<b>0.0301</b>	<b>0.1732</b>	<b>2.0954</b>	0.7155
	ANN	0.0510	0.2258	2.7102	0.6112
$m = 3$	ARIMA	0.0741	0.2722	3.6787	<b>0.9611</b>
	TF	<b>0.0515</b>	<b>0.2269</b>	<b>3.0231</b>	0.5930
	ANN	0.0893	0.2988	3.7487	0.4327

**Table 4.** Statistical parameters for comparison of observed and predicted values using ARIMA, TF and ANN models and  $m = 1, 2$  and  $3$  (70 m depth).

Thesaurus Station—DO Parameter—Depth 70 m					
$\hat{y}_t(m)$	Model	MSE	RMSE	MAPE	NSC
$m = 1$	ARIMA	<b>0.1376</b>	<b>0.3709</b>	<b>4.4140</b>	<b>-0.170</b>
	TF	0.1413	0.3759	4.3697	-0.202
	ANN	0.1913	0.4375	5.860	-0.628
$m = 2$	ARIMA	<b>0.1470</b>	<b>0.3834</b>	<b>5.1874</b>	-0.1809
	TF	0.1725	0.4153	5.4653	<b>0.0325</b>
	ANN	0.2130	0.4615	5.9012	-0.1109
$m = 3$	ARIMA	<b>0.1760</b>	<b>0.4195</b>	<b>6.2807</b>	-0.1771
	TF	0.2010	0.4484	6.7227	<b>0.0298</b>
	ANN	0.2278	0.4772	7.0136	-0.0208

### 3. Conclusions

This paper is focused on the comparative evaluation of the short-term predictive capacity of the stochastic models ARIMA, Transfer Function (TF) and Artificial Neural Networks. An adequate sample of four years daily monitoring program (19 January 2004–28 December 2007), of approximately 1440 records of Dissolved Oxygen (DO) and Water Temperature (Tw) parameters in Lake Thesaurus, were used in Monitoring Stations in four different depths (1, 20, 40 and 70 m).

In order to evaluate the predictability of the models, the data set was split into two subsets: (a) the historical period with 1420 measurements and (b) the forecasting period with the last 20 measurements. The former was used to select the best adapted model in the data, while the latter was used in order to forecast the  $m$  step forward ( $\hat{y}_t(m)$ ) [7–9].

For the short-term prediction of  $m = 1, 2$  and  $3$  steps forward, the performance of ARIMA models and—in some cases—TF models, is better with slight difference. This is due to the fact that as the distance between the starting point and the forecast increases ( $m$  increases), the contribution of the parameter observed values to the estimation of the forecast is diminished. ARIMA models could be proposed as short-term prediction models, although in some cases TF models resulted in better predictions, because (a) the difference with ARIMA models was not statistically significant and (b) the latter are simpler in their construction. On the other hand, Artificial Neural Networks didn't show a good short-term predictive capacity in comparison with the aforementioned models.

In order to predict more precisely the  $\hat{y}_t(m)$  values, the estimated values of  $y_t$  are used and not the observed ones. Therefore, the contribution of the observed values of water temperature at the corresponding depths, becomes more significant to the estimation of the forecasting values and the TF models result in better predictions than ARIMA.

To summarize, short-term forecasts of up to four days could be performed by both ARIMA and TF models. Nevertheless, ARIMA models are proposed due to their simplicity both regarding their compilation and use [6,7,17,18].

Further research can take place in three aspects: (a) applying forecasting for long-term prediction ( $m = 10, 15, \dots$ ); (b) comparing these models with other forecasting models and (c) applying these models in other deep water bodies and assess the comparison between them. The statistical analysis and stochastic modelling are very helpful tools to the assessment of water quality and quantity issues and could set the foundations for the Integrated Management of both aquatic ecosystems and water resources.

### References

1. Albanakis, K.; Mitrakas, M.; Moustaka-Gouni, M.; Psilovikos, A. Determination of the environmental parameters that influence sulphide formation in the newly formed Thesaurus reservoir, in Nestos River, Greece. *Fresenius Environ. Bull.* **2001**, *10*, 566–571.
2. Boskidis, I.; Gikas, G.D.; Sylaios, G.K.; Tsihrizis, V.A. Hydrologic and Water Quality Modelling of Lower Nestos River Basin. *Water Resour. Manag.* **2012**, *1*, 29.
3. Apostolou, A.; Koutrakis, M.; Pehlivanov, L.; Vassilev, M.; Stefanov, T.; Velkov, B. Notes of the Fish Fauna Composition of Mesta (Nestos) River in Regard to Management and Conservation. *Acta Zool. Bulg.* **2012**, *62*, 271–276.
4. Psilovikos, A.; Sentas, A.; Sahanidis, C.; Laopoulos, T. R.E.M.O.S. network in lignite mines in Western Macedonia, Greece. Assessment of the results for the year 2008. In Proceedings of the 2nd International Conference on Environmental Management, Engineering, Planning and Economics (CEMEPE), Mykonos, Greece, 21–26 June 2009; Volume II, pp. 731–736.
5. Psilovikos, A.; Sentas, A.; Sahanidis, C.; Laopoulos, T. Simulation and Trend Analysis of R.E.M.O.S. Water Quality and Quantity Data in Lignite Mines of Western Macedonia–Greece, for the Years 2007–2009. *Desalinat. Water Treat.* **2011**, *33*, 44–52.
6. Sentas, A.; Psilovikos, A. Comparison of ARIMA and Transfer Function (TF) Models in Water Temperature Simulation in Dam–Lake Thesaurus, Eastern Macedonia, Greece. In Proceedings of the International Symposium: Environmental Hydraulics, Athens, Greece, 23–25 June 2010; pp. 929–934.

7. Sentas, A.; Psilovikos, A. Dissolved oxygen assessment in Dam-Lake Thesaurus using stochastic modeling. In Proceedings of the International Conference: Protection and Restoration of the Environment, Thessaloniki, Greece, 3–6 July 2012; pp. 1573–1582.
8. Anastassopoulos, K. Treatment of big landslides affecting construction of Thissavros hydroelectric project. In Proceedings of the Hydro 2006, Maximizing the Benefits of Hydropower, Porto Carras, Greece, 25–27 September 2006.
9. Antonopoulos, V.; Papamichail, D.; Mitsiou, K. Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece. *Hydrol. Earth Syst. Sci.* **2001**, *5*, 679–691.
10. Sentas, A.; Psilovikos, A.; Psilovikos, T.; Matzafleri, N. Comparison of the performance of stochastic models in forecasting daily dissolved oxygen data in dam-Lake Thesaurus. *Desalin. Water Treat.* **2016**, *57*, 11660–11674.
11. Moustaka–Gouni, M.; Albanakis, K.; Mitrakas, M.; Psilovikos, A. Planktic autotrophs and environmental conditions in the newly-formed hydroelectric Thesaurus reservoir, Greece. *Arch. Hydrobiol.* **2000**, *149*, 507–526.
12. Ahmad, S.; Khan, I.; Parida, B. Performance of stochastic approaches for forecasting river water quality. *Water Res.* **2001**, *35*, 4261–4266.
13. Ljung, G.M.; Box, G.E. On a measure of lack of fit in time series models. *Biometrika* **1978**, *65*, 297–303.
14. Akaike, H. A Bayesian analysis of the minimum AIC procedure. *Ann. Inst. Stat. Math.* **1978**, *30*, 9–14.
15. Box, G.; Jenkins, G.; Reinsel, G. *Time Series Analysis, Forecasting and Control*, 3rd ed.; Prentice-Hall Inc.: Upper Saddle River, NJ, USA, 1994; p. 598.
16. Dorffner, G. *Neural Networks for Time Series Processing*; Neural Network World: Prague, Czech Republic; 1996; pp. 447–468.
17. Diamantopoulou, M.; Antonopoulos, V.; Papamichail, D. Cascade correlation artificial neural networks for estimating missing monthly values of water quality parameters in rivers. *Water Resour. Manag.* **2007**, *21*, 649–662.
18. Faruk, D.O. A hybrid neural network and ARIMA model for water quality time series prediction. *Eng. Appl. Artif. Intell.* **2010**, *23*, 586–594.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).