Desalination and Water Treatment
www.deswater.com
doi: 10.1080/19443994.2015.1128984

57 (2016) 11660–11674
May

Taylor & Francis
Taylor & Francis Group

Comparison of the performance of stochastic models in forecasting daily dissolved oxygen data in dam-Lake Thesaurus

Antonis Sentas[a], Aris Psilovikos[a,*], Thomas Psilovikos[b], Niki Matzafleri[a]

[a]*Department of Ichthyology and Aquatic Environment, School of Agricultural Sciences, University of Thessaly, Fytoko st., N. Ionia Magnisias 38446, Greece, Tel./Fax: +30 2421093154; emails: antsentas@yahoo.gr (A. Sentas), psiloviko@uth.gr (A. Psilovikos), matzafleriniki@yahoo.gr (N. Matzafleri)*

[b]*General Directorate of Technical Services and Mechanization, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece, Tel./Fax: +30 6948060210; email: tvikos@ad.auth.gr*

Received 31 March 2015; Accepted 19 November 2015

ABSTRACT

This study presents the development and validation of three different stochastic models on the basis of (a) their efficiency to forecast and (b) their ability to utilize auxiliary environmental information. The three models are ARIMA models, transfer function (TF) models, and artificial neural networks. Four-year (2004–2007) daily measurements of dissolved oxygen at four different depths (1, 20, 40 and 70 m) of Thesaurus dam-lake in River Nestos, Eastern Macedonia, Greece, were used to obtain the best models for these time series. For the final selected models, four statistical criteria (mean square error (MSE), root-mean-square error (RMSE), MAPE, and NSC) were used to evaluate the accuracy of the forecast and to compare the forecasting ability for one step ahead of each approach. For 1- and 20-m depth, the best forecast is obtained by ARIMA models, while for the 40-m depth, TF models gives the best forecast. Finally for the 70-m depth, according to the MSE, RMSE, and NSC statistical criteria, ARIMA models are the best, while for the MAPE, TF models are the best. Further research could be carried out concerning on (a) the comparison of these models with other forecasting ones, (b) the application of forecasting for more than one step ahead ($m = 2, 3, \ldots$), and (c) the implementation of such models in other deep lakes and the assessment of the comparison between them.

Keywords: Stochastic modeling; ARIMA; Transfer function—TF; Artificial neural networks—ANNs; River Nestos; dam-Lake Thesaurus

1. Introduction

The decision-making regarding water resources management issues requires a careful modeling, forecasting, and analysis of water quality for different possible scenarios. Statistics and computer science have improved modeling approaches for discovering patterns found in water resources time series data. Much effort has been devoted over the past several decades to the development and improvement of time

*Corresponding author.

Presented at the 12th International Conference on Protection and Restoration of the Environment (PRE XII) 29 June–3 July 2014, Skiathos Island, Greece

series prediction models [1–3]. A quite significant number of scientific papers focus on the prediction of river water quality and quantity parameters in water bodies [4–6]. River water quality prediction is required for a proper management of the river basin [7–9].

The linear stochastic models including ARIMA, SARIMA, and transfer function (TF) models are the most widely used in forecasting water quantity and quality parameters at different timescales. They were built up following the three modeling stages: (a) model identification, (b) model estimation, and (c) model diagnostic checking. The ARIMA and SARIMA models (univariate models) are based on past information and do not take into account the effects of other parameters [10–12]. The TF models are more complicated than the multivariate models. To develop such models, the computing of the cross-correlation function (CCF) took place. Recently, the developed machine learning techniques, artificial neural networks (ANNs), have been used for prediction and forecasting in a number of water-related areas. Such models have the capability to locate complex nonlinear relationships between input and output parameters, without underlying assumptions of linearity or normality [5,13–15].

The conservation, protection, and management of the environment start to be taking into account only the last half of the twentieth century. This fact was due to the facing problems of the quantitative and qualitative degradation of aquatic ecosystems and water resources. The assessment of water status requires (a) systematic and long-term monitoring of specific water quantity and quality parameters and (b) statistical analysis of these data [16–18].

Nestos is a transboundary river, between Greece and Bulgaria. Since the beginning of the twentieth century, large-scale hydraulic works have been constructed along its course. Nestos presents scientific interest especially for issues of monitoring and management from both countries. This need began to increase during the late eighties [19–22]. Nowadays, there are monitoring programs, management strategies, and bilateral agreements from both EU countries (Greece and Bulgaria), under the aegis of WFD 2000/60.

Dam-lake Thesaurus is the study area of this work. It is the deepest lake in Greece (140 m) and is located in the intermountainous basin of Nestos. The dam began to operate in 1997, and it was focused on hydroelectric power, irrigation needs, and anti-flooding protection. Moustaka–Gouni et al. [23] and Psilovikos [24] study the horizontal distribution of phytoplankton species and also the frequency and distribution of sulfide bacteria in the water column, during the first years of dam operation (1997–1999). Albanakis et al. [25] in 2001 analyzed the measurements of six water-quality parameters and noted strong stratification of water column once a year. So, Thesaurus was found to be Monomictic Lake, having a permanent cold hypolimnion under anoxic conditions [25]. Systematic monitoring data confirmed the results above and consists the base for the sustainable management of Nestos [26–28].

The purpose of this study was to develop and to validate different stochastic models to describe the dissolved oxygen (DO) values at different depths of dam-lake Thesaurus in River Nestos.

2. Material and methods

2.1. Study case

Our study area is the dam-lake of Thesaurus in the Hellenic part of River Nestos, which is a transboundary surface water resource in the Balkan Peninsula (Fig. 1). Thesaurus is the deepest artificial lake in the Hellenic territory, and the rough geomorphology of the area is due to the steep relief of its catchment's area and the depth of the dam-lake [29–31]. A four-year monitoring program (19 January 2004–28 December 2007) based on daily measurements of DO and water temperature (Tw), in four different depths (1, 20, 40, and 70 m) of Thesaurus, was used to obtain the best-fitted models for the specific time series [27,28,32,33].

Three modeling and forecasting techniques were evaluated on the basis of (a) their efficiency to forecast and (b) their ability to utilize auxiliary environmental information: ARIMA, TF, and ANNs models.

2.2. ARIMA models

To fit an ARIMA model at the existing data, the three steps of Box and Jenkins method were followed: (a) model identification, (b) model estimation, and (c) model diagnostic checking [34].

At the first step, the type and order of the ARIMA model are identified from the observed time series. This step is purposed to determine the differencing or other proper transformations to obtain a stationary series, i.e. a series with a constant mean and a constant variance through time. Stationarity is a necessary condition in constructing an ARIMA model. The graphs of the sample autocorrelation (ACF) and partial autocorrelation functions help to obtain an initial empirical point of view for the type of the model. For the seasonal ARIMA model, the identification phase starts with the estimation of the seasonal model and

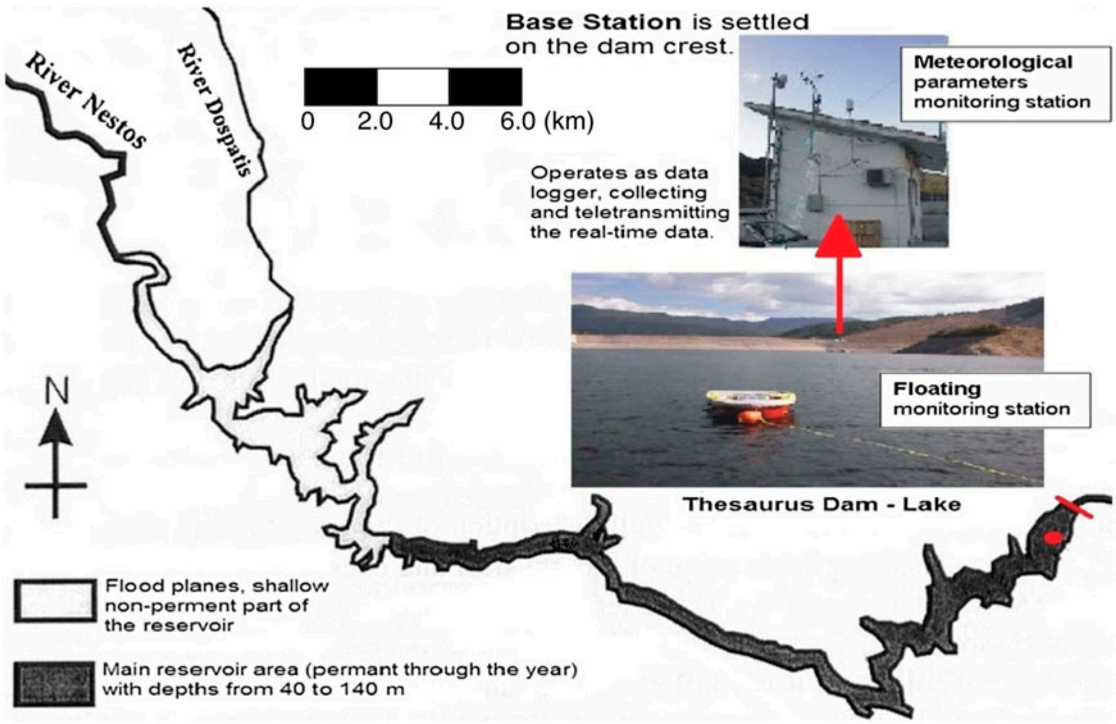

Fig. 1. The location map of Thesaurus dam-lake and the monitoring stations.

afterward with the analysis of its residuals. In a well-identified seasonal model, its residuals show the non-seasonal term.

The second step includes the estimation of the model parameters from the data. The parameters with significance greater than 0.05 are rejected. For the fitting of a model, the following criteria are used: (a) Akaike's Information Criterion [33] and (b) Schwartz's Bayesian Criterion [35].

Finally, in the third step, diagnostic checks are applied to determine whether or not the chosen model adequately represented the given set of data. The independence of the residuals are tested with the graph of the residuals autocorrelation function (RACF, $r_k(a)$) and the Q-statistic [34], which follows X^2 distribution with m–k degrees of freedom, where m is the number of RACF and k is the number of estimated model parameters. The Kolmogorov–Smirnov test [36] and the Q–Q plot are used for the residuals fitting to the normal distribution.

2.3. TF models

For the construction of the TF models, an appropriate ARMA model is applied to the input series and obtains the white noise series α_t. The output series are transformed using the above pre-whitening model to obtain another white series β_t. The *CCF* between α_t and β_t is proportional to the impulse response function (IRF). From the pattern of the IRF, we calculate the parameters b, r, s, and preliminary estimates $\hat{\omega}_j$ and $\hat{\delta}_j$ [34], which are necessary for the estimation of the TF model. For the new residuals, a traditional ARMA model is estimated as described above. After this identification, the total model is estimated.

2.4. Artificial neural networks

Using ANNs for forecasting, the modeling philosophy is similar to the traditional statistical approaches that are used. The unknown model parameters are adjusted, in order to obtain the best fit between the inputs and the corresponding outputs [5,8,13,14,37–39].

For time series forecasting, the inputs to the network consist of past samples of the time series and the output is the predicted value. The relationship between the inputs and the outputs is:

$$y_t = f(y_{t-1},\ y_{t-2},\ y_{t-3},\ldots,y_{t-p}) \tag{1}$$

where y_t is the observation at time t; p is the prediction order, and f the nonlinear TF.

To evaluate the accuracy of the models' forecasting ability, the data (1,440 observations) were divided in two subsets: (1) a main set (historical period) with 1,420 measurements and (2) a trial set (forecasting period) with the last 20 measurements. The first set was used for choosing the model with the best fit to the data, while the second was used to forecast one step ahead ($\hat{y}_t(1)$). The forecasting "one step ahead" means that the forecasting occurs at time $t + 1$ and the observations from time $t = 0$, until time t, which is the starting time.

2.5. Forecasting assessment

To evaluate the forecasting ability, four statistical criteria were used: (a) mean square error (MSE), (b) root-mean-square error (RMSE), (c) mean absolute percentage error (MAPE), and (d) Nash–Sutcliffe coefficient of efficiency (NSC).

(1) The MSE measures the average of the squares of the "errors". The MSE which is given by the following equation:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

(2) The RMSE is given by the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

(3) The mean absolute percentage error is given by the following equation:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100 \tag{4}$$

MAPE is another statistic, which measures the performance and forecasting accuracy of the models. The smaller values of the above criterions show that the estimated values are close to the real values, therefore, better performance and forecasting ability of the model.

(4) The Nash–Sutcliffe coefficient of efficiency is given by the following equation:

$$\mathrm{NSC} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y} - y_i)^2} \tag{5}$$

where $\bar{y}$ is the sample mean value [40].

The values of this statistic are less than one (1) and compare the mean value of a parameter with the estimators of the model parameters. According this statistic, the model with NSC near unit is the best [37].

Thesaurus dam was constructed during the year 1997. For approximately three years, it has been filled with water, before it starts to operate in the year 2000. During this period of three years, water stagnation phenomena and anoxia at the hypolimnion were observed [10,23]. Since the beginning of the 2000s, when Thesaurus starts to operate, the complicated hydrosystem of Nestos and Thesaurus is trying to restore the dynamic balance between the natural environment and the human intervention. In this study, a four-year monitoring period on a daily basis (historical period) is our database. It is an adequate period of time for the construction of the stochastic models and the extraction of the statistical results of water temperature (Tw) and DO. These parameters are monitored at four different depths in the water column (1, 20, 40, and 70 m). The forecasting period is chosen at the end of this period, and 20 steps (d) are considered sufficient.

3. Results and discussion

3.1. ARIMA models

For each one-time series, different trials were carried out to choose the best-fitted ARIMA forecasting model. The final ARIMA model was chosen, so that to satisfy all the statistical diagnostics with significant parameters. The symbolic and analytic equations of the final models for each depth, as well as the significance of Ljung–Box–Pierce statistic, are given in the following table (Table 1).

The graph of the monitored DO time series for 1, 20, 40, and 70 m, the estimated ARIMA time series, and the 95% confidence levels for each one of the depths are shown in Fig. 2. The time increment from 25 of November to 8 of December, 2007 is part of the historical and simulation period. On the other hand, the period from 9 to 28 of December is part of the validation and forecasting period. It was found that more than 20 d do not provide reliable forecasting results.

The scatter plots of the observed daily DO values and one day ahead forecasts of ARIMA models are given in Fig. 3. In the same figure, the correlation coefficient, the line of the best fitting (line 1:1), and the regression line are presented and give a graphical representation of the four models forecasting ability.

The values of the four statistical criteria (MSE, RMSE, MAPE, and NSC), which are used for the assessment of the fitted models, are given in Table 2.

Table 1
ARIMA models for DO in 1, 20, 40, and 70 m

DO (m)	Model	Analytic equations	Sig. $Q(18)$	
1	ARIMA(3, 1, 5)	$y_t = \dfrac{(1 - 0.5B^2 - 0.790B^3 + 0.366B^4 + 0.102B^5)e_t}{(1 - B)(1 + 0.397B - 0.245B^2 + 0.880B^3)}$	0.136	(6)
20	ARIMA(2, 1, 3)	$y_t = \dfrac{(1 - 0.722B - 0.498B^2 + 0.274B^3)e_t}{(1 - B)(1 - 0.428B + 0.534B^2)}$	0.250	(7)
40	ARIMA(2, 1, 13)	$y_t = \dfrac{(1 + 0.108B + 0.308B^2)e_t}{(1 - B)(1 - 0.339B^2 + 0.189B^3 + 0.082B^{13})}$	0.850	(8)
70	ARIMA(2, 1, 10)	$y_t = \dfrac{(1 - 0.563B^2)e_t}{(1 - B)(1 - 0.206B - 0.477B^2 + 0.252B^3 + 0.123B^{10})}$	0.585	(9)

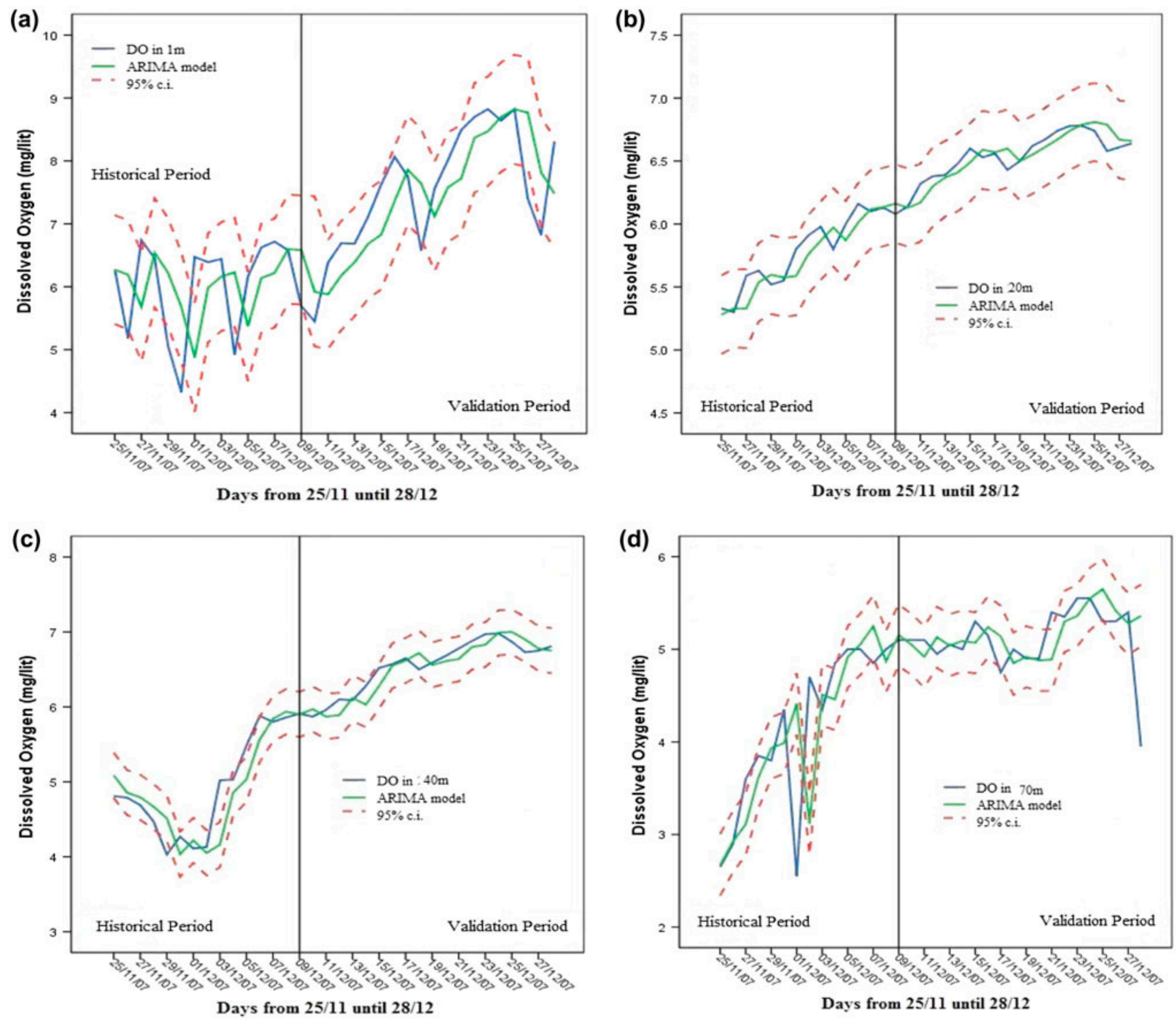

Fig. 2. Daily observed and predicted DO values with ARIMA, models for the depths (a) 1 m, (b) 20 m, (c) 40 m και, and (d) 70 m for the historical and validation period.

The ranking of the used models, beginning from the best to the worst one, according to MSE, RMSE, and MAPE criteria is: (a) model of 20 m, (b) model of 40 m, (c) model of 70 m, and finally (d) model of 1 m. The results are different only for the NSC criterion, where the ranking is (a) model of 40 m, (b) model of 20 m, (c) model of 1 m, and finally (d) model of 70 m with a negative value.

Table 2
Comparison of ARIMA models performance for testing data set

ARIMA	MSE	RMSE	MAPE	NSC
1 m depth	0.4425	0.6652	7.9472	0.614
20 m depth	0.0077	0.0877	1.0552	0.799
40 m depth	0.0167	0.1293	1.5882	0.916
70 m depth	0.1376	0.3709	4.4140	−0.170

3.2. TF models

To investigate a possible correlation between *DO* and Tw for each depth, the sample cross-correlation function (SCCF) is obtained from the ARIMA models and showed significant correlations (Fig. 4). This means that there is a direct correlation between the two time series. So, for the construction of the four TF models, the Tw values were used as input time series and the DO values as output time series, for each depth. The ARIMA models that were used for the pre-whitening of these two parameters are given in Table 3.

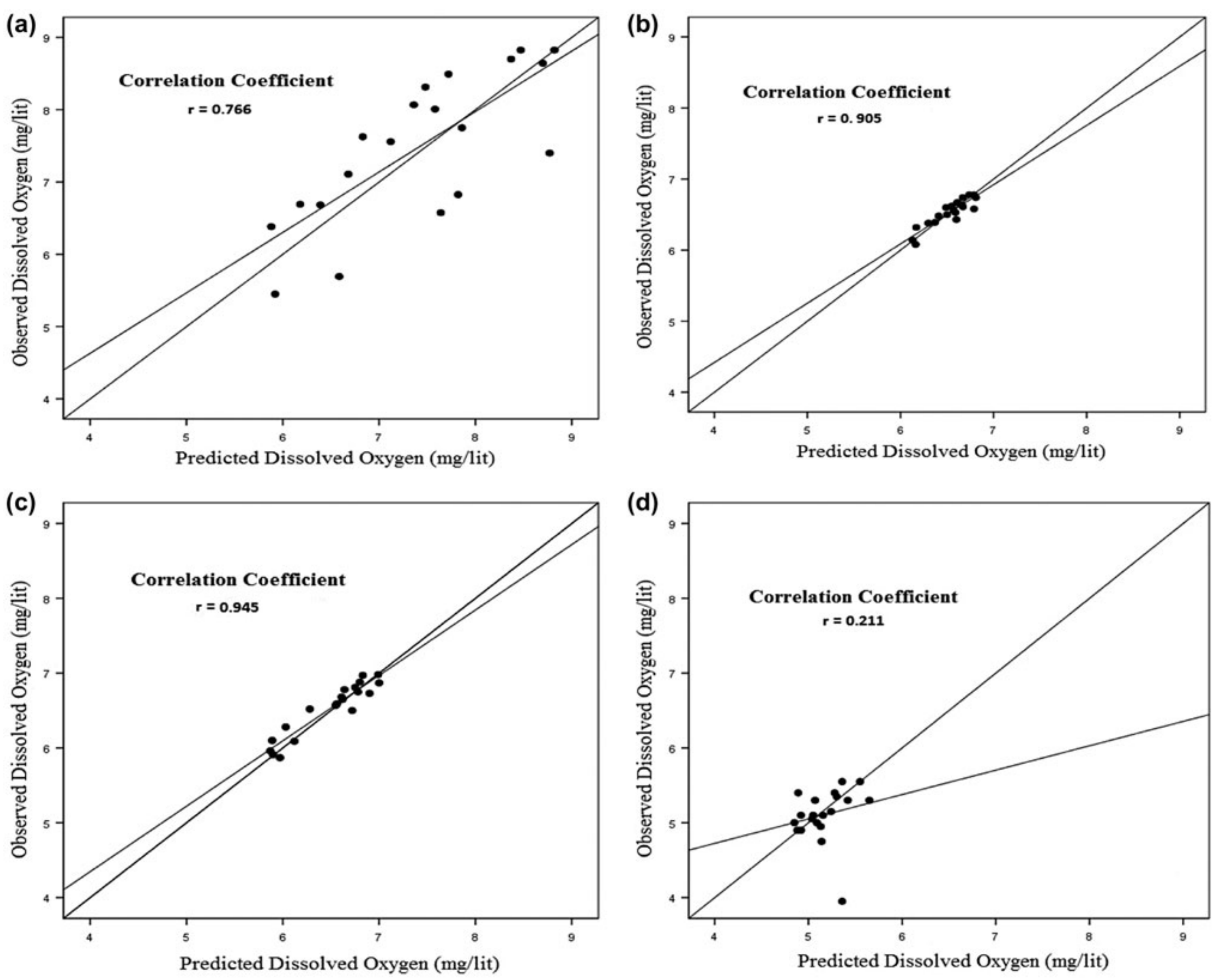

Fig. 3. Scatter plots of the observed and predicted values for the ARIMA models, (a) 1 m, (b) 20 m, (c) 40 m, and (d) 70 m.

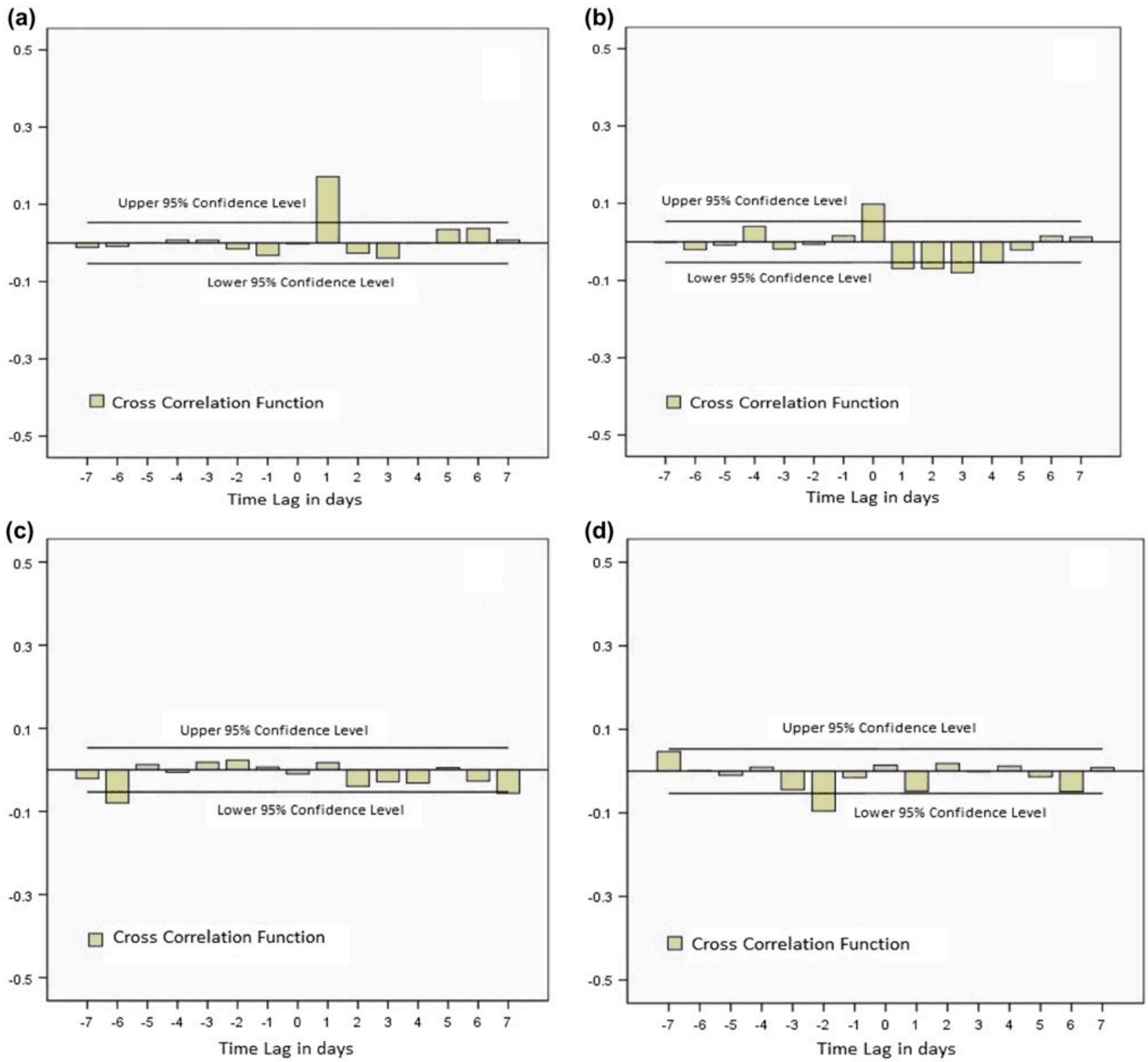

Fig. 4. The SCCF between DO and Tw for (a) 1 m, (b) 20 m, (c) 40 m, and (d) 70 m.

The analytic equations of the final TF models for each one depth are:

$$\text{TF-1}: \ y_t = 0.073 - 0.004x_t + \frac{(1 - 0.440B^2 - 0.755B^3 + 0.346B^4 + 0.107B^5)}{(1 - B)(1 - 0.465B + 0.809B^3)} e_t \tag{10}$$

Table 3
Pre-whitening ARIMA models

Depth	Dissolved oxygen	Water temperature
1 m	ARIMA(3, 1, 5)	ARIMA(1, 1, 13)
20 m	ARIMA(2, 1, 3)	ARIMA(1, 1, 2)
40 m	ARIMA(2, 1, 13)	ARIMA(1, 1, 2)
70 m	ARIMA(2, 1, 10)	ARIMA(1, 1, 2)

Table 4
The significance of $Q(18)$ statistic for TF models

	TF-1	TF-20	TF-40	TF-70
Significance of $Q(18)$ statistic	0.106	0.103	0.229	0.10

$$\text{TF-20}: \; y_t = 0.091x_t - 0.09x_{t-1} + \frac{(1 - 0.706B - 0.513B^2 + 0.276B^3)}{(1-B)(1 - 0.421B - 0.540B^2)} e_t \tag{11}$$

$$\text{TF-40}: \; y_t = \frac{0.048(B-1)}{1 - 0.856} x_{t-6} + \frac{(1 + 0.943B^2 - 0.09B^4 + 0.216B^5 - 0.067B^6 + 0.083B^7)}{(1-B)(1 - 0.128B + 0.983B^2)} e_t \tag{12}$$

$$\text{TF-70}: \; y_t = -0.113x_{t-2} + 0.113x_{t-3} + \frac{(1 - 1.089B + 0.23B^2 + 0.06B^4)}{(1-B)(1 - 0.895B)} e_t \tag{13}$$

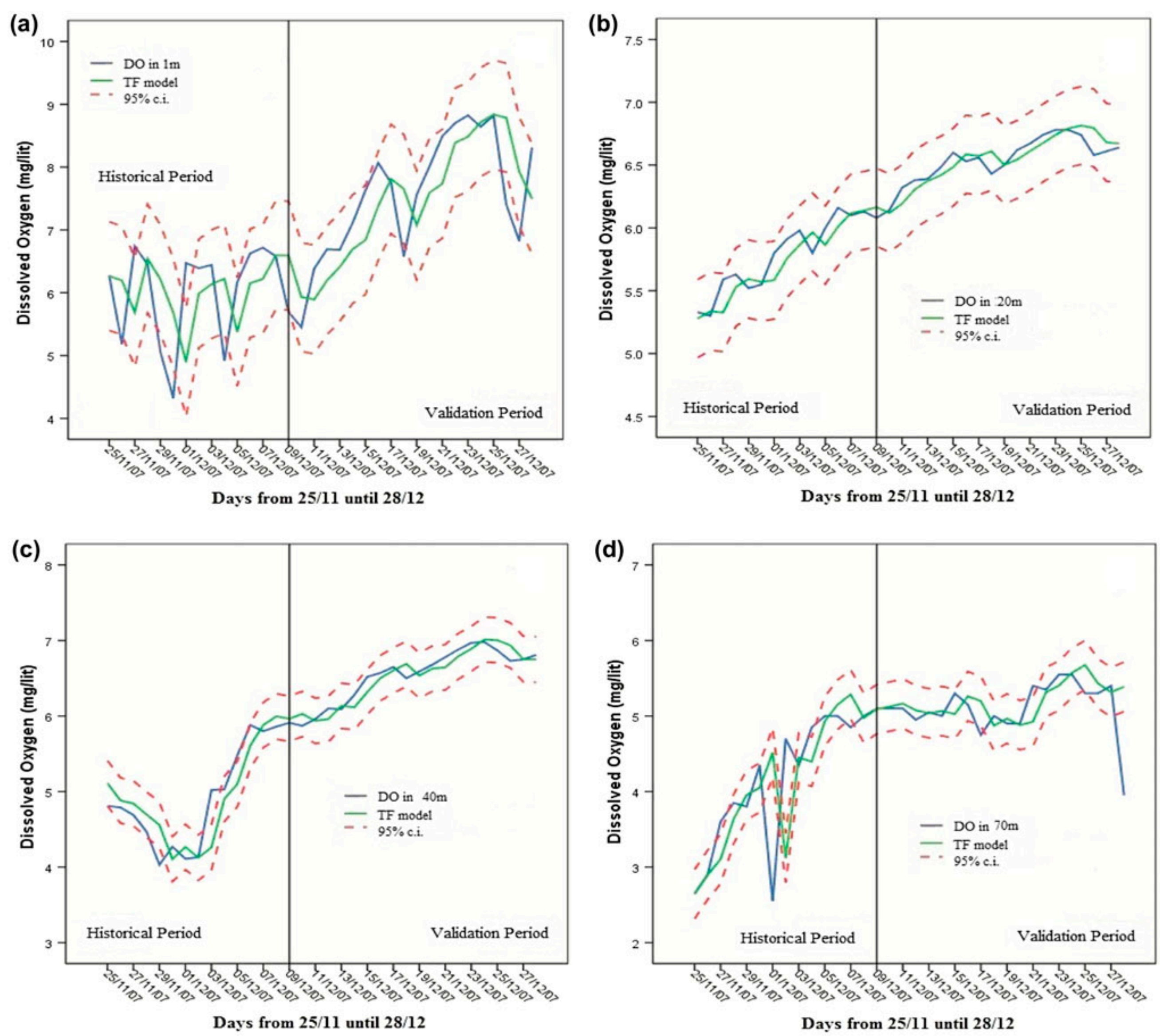

Fig. 5. Daily observed and predicted DO values with TF models for the depths (a) 1 m, (b) 20 m, (c) 40 m και, and (d) 70 m for the historical and validation periods.

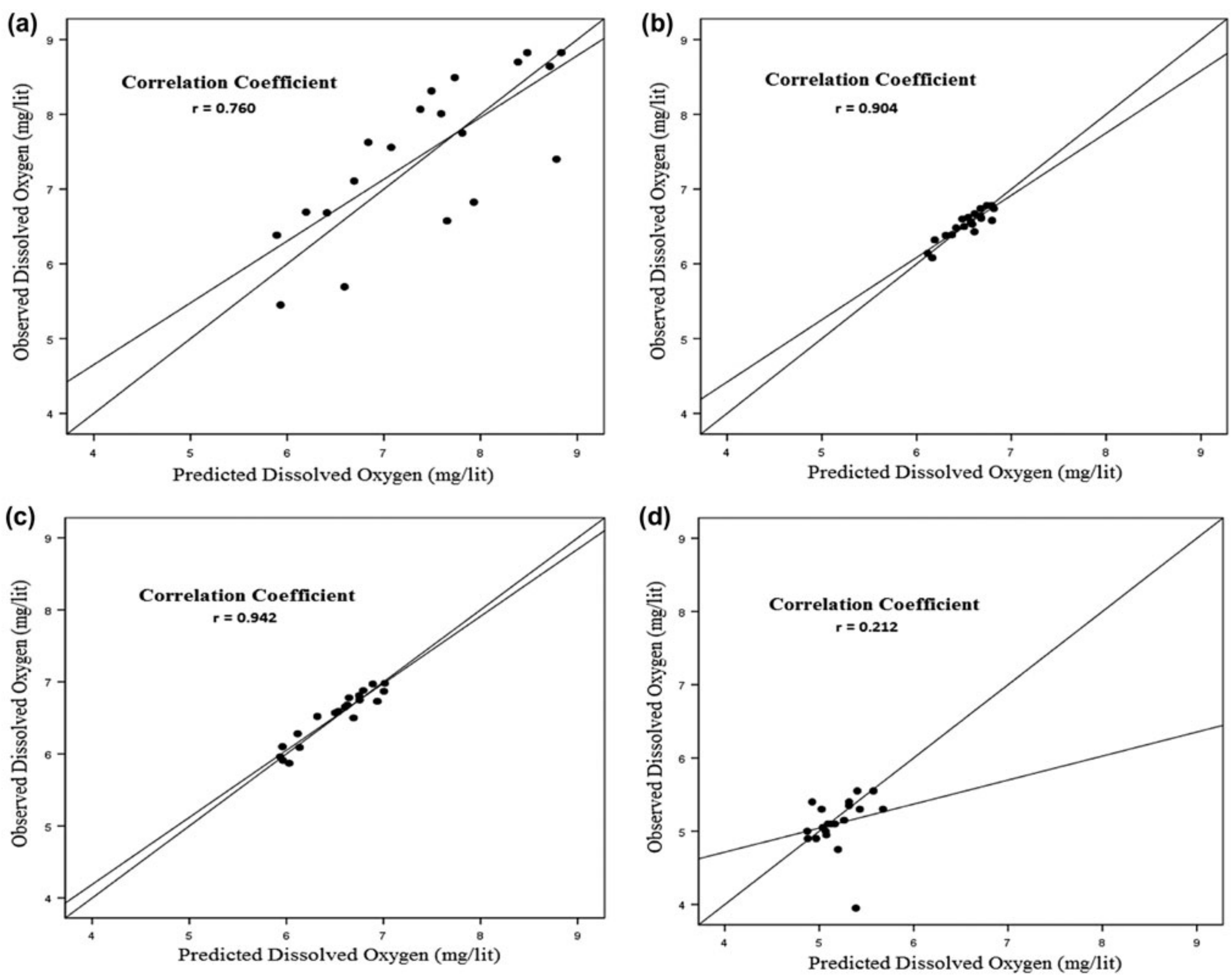

Fig. 6. Scatter plots of the observed and predicted values for the TF models, (a) 1 m, (b) 20 m, (c) 40 m, and (d) 70 m.

Table 5
Comparison of TF models performance for testing data set

Transfer function	MSE	RMSE	MAPE	NSC
1-m depth	0.4527	0.6728	7.9748	0.605
20-m depth	0.0079	0.0891	1.075	0.793
40-m depth	0.0131	0.1144	1.4791	0.934
70-m depth	0.1413	0.3759	4.3697	−0.202

For all the models above, the significance of the Ljung–Box–Pierce $Q(r)$ statistic, for $r = 18$ is greater than 0.05 and the models are adjusted. Table 4 shows the exact values.

The graph of the monitored DO time series for 1, 20, 40, and 70 m, the estimated TF time series, and the 95% confidence levels for each one of the depths are shown in Fig. 5.

The scatter plots of the observed daily DO values, the one day ahead forecasts of TF models and the correlation coefficient for the values in the four depths are given in Fig. 6. The values of the correlation coefficients are 0.760 for 1-m, 0.904 for 20-m, 0.949 for 40-m, and 0.212 for 70-m depth. These values are different from the ARIMA models' corresponding values at the third decimal digit, which means that the two models are almost equivalent.

Both the historical and the validation periods are the same as in ARIMA case study. The values of the statistical criteria (MSE, RMSE, MAPE, and NSC) are shown in Table 5. According to the MSE, RMSE, and MAPE criteria, the ranking of the used models, beginning from the best to the worst one, is (a)

Table 6
The statistics of ANNs models in the four depths

	Statistics of ANN	ANN-1	ANN-20	ANN-40	ANN-70
Training sample	Sum of squares error (SSE)	23.679	8.867	6.369	4.418
	Relative error (RE)	0.48	0.017	0.013	0.09
Testing sample	Sum of squares error (SSE)	7.082	2.335	1.416	0.968
	Relative error (RE)	0.027	0.013	0.007	0.005

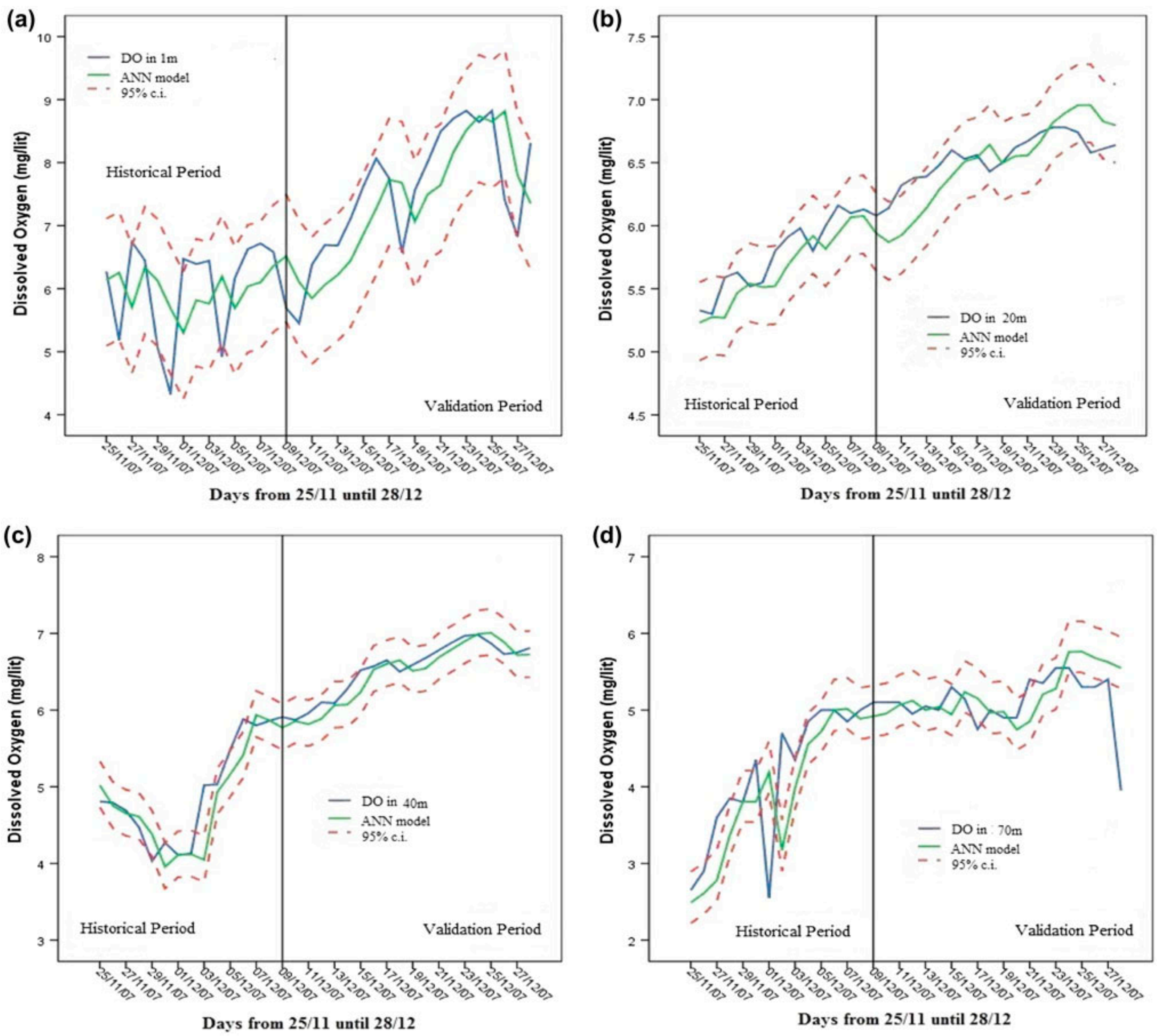

Fig. 7. Daily observed and predicted DO values with ANNs models for the depths (a) 1 m, (b) 20 m, (c) 40 m και, and (d) 70 m for the historical and validation periods.

Table 7
Comparison of ANNs models performance for testing data set

ANNs	MSE	RMSE	MAPE	NSC
1-m depth	0.5233	0.7233	8.934	0.544
20-m depth	0.0434	0.2084	2.672	−0.130
40-m depth	0.0170	0.1305	1.678	0.915
70-m depth	0.1913	0.4375	5.860	−0.628

model of 20 m (TF-20), (b) model of 40 m (TF-40), (c) model of 70 m (TF-70) and finally (d) model of 1 m (TF-1). The results are different only for the NSC criterion, where the ranking is: (a) model of 40 m (TF-40), (b) model of 20 m (TF-20), (c) model of 1 m (TF-1), and finally (d) model of 70 m (TF-70) with a negative value.

3.3. Artificial neural networks

As mentioned above, the 1,440 daily measurements are divided in two subsets (historical and forecasting periods). The 70% of the historical subset is defined as training sample, while the rest 30% as validation or testing sample. The most commonly used type of ANNs in hydrological modeling is the feed-forward multilayer perceptron, which was used in this study [9]. The structure of the final neural networks is as follows:

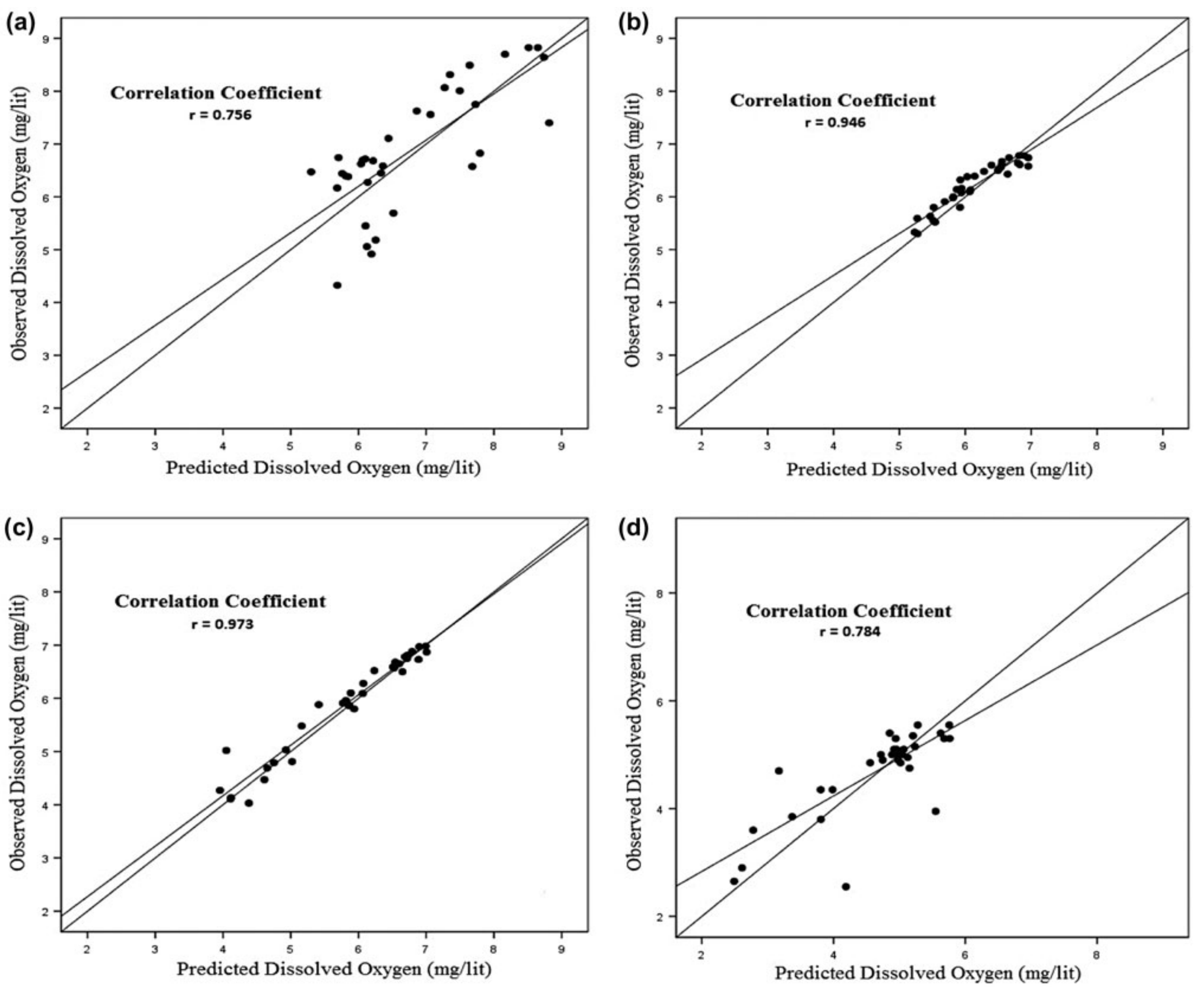

Fig. 8. Scatter plots of the observed and predicted value for the ANN models, (a) 1 m, (b) 20 m, (c) 40 m, and (d) 70 m.

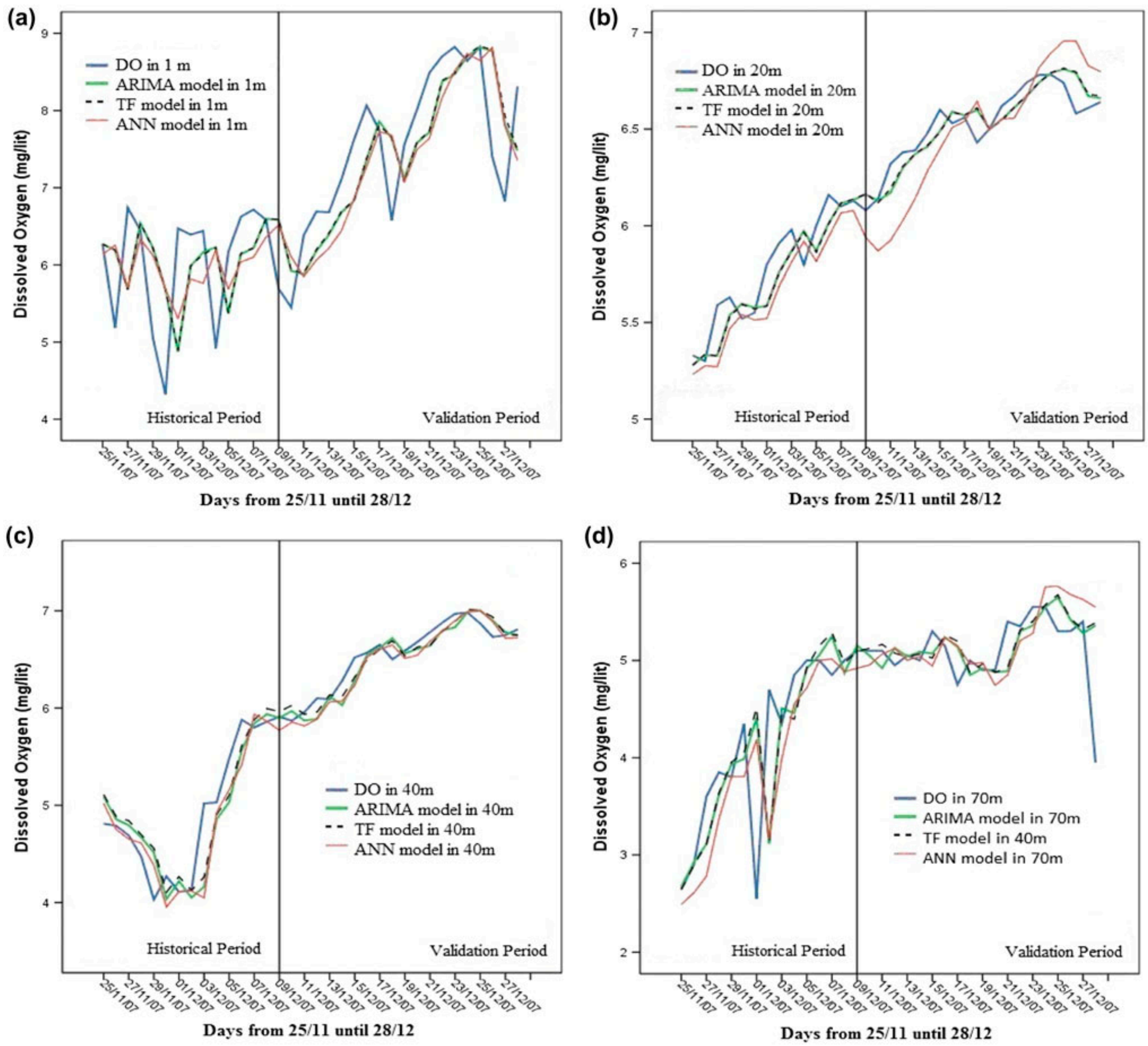

Fig. 9. Daily observed and predicted DO values with ARIMA, TF, and ANNs models for the depths (a) 1 m, (b) 20 m, (c) 40 m και, and (d) 70 m for the historical and validation periods.

(1) One hidden layer.
(2) The sigmoid function as the input activation.
(3) The identity function as the output activation.
(4) As the rule of convergence of the algorithm is the nonincrease of error after two steps, or the completion of maximum time of 15 min, or the completion of the maximum number of 2,000 epochs.
(5) 1,000 repeats for each model.
(6) Selection of the final model using the sum of square error (SSE) on the validation subtest.

Using the above methodology, four ANNs models were constructed based on the structure of the ARIMA models, at the four depths. In any case, as input time series past DO values were used resulting to the following ANNs:

$$\text{ANN-1}: \ y_t = f(y_{t-1},\ y_{t-2},\ y_{t-3},\ y_{t-4}) \tag{14}$$

$$\text{ANN-20}: \ y_t = f(y_{t-1},\ y_{t-2},\ y_{t-3}) \tag{15}$$

$$\text{ANN-40}: \ y_t = f(y_{t-1},\ y_{t-2}, y_{t-3}) \tag{16}$$

Table 8
Comparison of model performance for testing data set

Depth	Statistical measures	ARIMA	TF	ANNs
1 m	MSE	**0.4425**	0.4527	0.5233
	RMSE	**0.6652**	0.6728	0.7233
	MAPE	**7.9472**	7.9748	8.934
	NSC	**0.614**	0.605	0.544
20 m	MSE	**0.0077**	0.0079	0.0434
	RMSE	**0.0877**	0.0891	0.2084
	MAPE	**1.0560**	1.0750	2.672
	NSC	**0.799**	0.793	−0.130
40 m	MSE	0.0167	**0.0131**	0.0170
	RMSE	0.1293	**0.1144**	0.1305
	MAPE	1.5882	**1.4791**	1.678
	NSC	0.916	**0.934**	0.915
70 m	MSE	**0.1376**	0.1413	0.1913
	RMSE	**0.3709**	0.3759	0.4375
	MAPE	4.414	**4.3697**	5.860
	NSC	**−0.170**	−0.202	−0.628

Note: Bold values indicate the best fitted model in 1, 20, 40, and 70 m respectively.

$$\text{ANN-70}: \ y_t = f(y_{t-1},\ y_{t-2},\ y_{t-3}) \tag{17}$$

where ANN-*k* is the neural network that corresponds to the *k*-meters depth.

In Table 6, the statistical tests of the used ANNs models are shown, thus the sum of squares error—SSE and the relative error—RE of the residuals of the fitting models, in both the training and testing samples.

In Fig. 7, the graph of the monitored DO time series for 1, 20, 40, and 70 m, the estimated ANNs models and the 95% confidence levels for each one of the depths are shown. Both the historical and the validation periods are the same as in the previous ARIMA and TF case studies.

The statistical criteria for the ranking of the forecasting ability of ANNs models are given in Table 7. According to the values of the three statistical criteria (MSE, RMSE, and MAPE), the best forecast occurs with (a) ANN-40, (b) ANN-20, (c) ANN-70, and (d) ANN-1 models. Concerning on NSC criterion, the ranking of the models is (a) ANN-40, (b) ANN-1, (c) ANN-20, and ANN-70.

The values of the correlation coefficients, the scatter plots of the observed daily DO values, and one day ahead forecasts of ANN models of the four daily DO values are given in Fig. 8. The correlation coefficients values are 0.756 for 1 m, 0.946 for 20-m, 0.973 for 40-m, and 0.784 for 70-m depth. The above measures showed that the ANN models were having sufficiently better behavior compared with ARIMA and TF models.

3.4. Comparison of the models

To give a graphical representation of the comparison between ARIMA, TF, and ANNs models, the monitored and the predicted DO values are plotted with the three different approaches for each of the sensor depths of 1, 20, 40, and 70 m (Fig. 9).

Also, the one day ahead forecasting performances of all models for the forecasting phase are given in Table 8.

4. Conclusions

The comparison of the models' forecasting ability was the purpose of this study. The 1,440 DO daily monitoring values, at each one of the four depths (1, 20, 40, and 70 m) in Thesaurus dam-lake, were used for the application of three different models: ARIMA, TF, and ANN_S. From the statistical criteria analysis above, it became obvious that in most cases, the best forecast was obtained from the ARIMA models. The TF models followed with almost the same statistical values, while the ANNs had a smaller forecasting ability.

The conclusions (obtained from Table 8) are the following:

(1) *For 1-m depth*: the ARIMA model gives the best results for all statistics, with the TF model to come second with a slight difference at the second decimal.
(2) *For 20-m depth*: the situation is the same as in the 1-m depth, with the ANN to follow with a slight difference from the other two models.
(3) *For 40-m depth*: all statistics agree that the TF gives the best forecast, with the ARIMA to follow and the ANN to come last.
(4) *For 70-m depth*: according to the MSE, RMSE, and NSC statistics, the best forecasting ability belongs to the ARIMA, the TF follows and the ANN comes last. As for the MAPE statistic, best model is the TF, followed by the ARIMA and the ANN.

Further research can take place in three aspects: (a) comparing these models with other forecasting models, (b) applying forecasting for more than one steps ahead (m = 2, 3, …), and (c) applying these models in other deep water bodies and assess the comparison between them.

References

[1] M. Valipour, M.E. Banihabib, S.M.R. Behbahani, Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir, J. Hydrol. 476 (2013) 433–441.

[2] M. Valipour, Long-term runoff study using SARIMA and ARIMA models in the United States, Meteorol. Appl. 22(3) (2015) 592–598.

[3] A. Kurunc, K. Yurekli, O. Cevik, Performance of two stochastic approaches for forecasting water quality and stream flow data from Yesilirmak River, Turkey, Environ. Modell. Softw. 20 (2004) 1195–1200.

[4] A. Najah, A. Elshafie, O.A. Karim, O. Jaffer, Prediction of Johor River water quality parameters using Artificial Neural Networks, Eur. J. Sci. Res. 28(3) (2009) 422–435.

[5] M. Diamantopoulou, V. Antonopoulos, D. Papamichail, Cascade correlation artificial neural networks for estimating missing monthly values of water quality parameters in rivers, Water Resour. Manage. 21 (2007) 649–662.

[6] A. Sentas, S. Margoni, K. Albanakis, Statistical research and assessment of water quality parameters of "REMOS" telemetric monitoring station in River Nestos Delta for the year 2004 (in Greek), Proceedings of the 2nd International Conference: Small Scale Water & Waste Treatment, Skiathos, Greece, vol. 1, 2008, pp. 463–468.

[7] A. Sentas, A. Psilovikos, Comparison of ARIMA and Transfer Function (TF) models in water temperature simulation in Dam-Lake Thesaurus, Eastern Macedonia, Greece, Proceedings of the International Symposium: Environmental Hydraulics, Athens, Greece, vol. 2, 2010, pp. 929–934.

[8] S. Abudu, J. King, Z. Sheng, Comparison of the performance of statistical models in forecasting monthly total dissolved solids in the Rio grande1, JAWRA J. Am. Water Resour. Assoc. 48(1) (2012) 10–23.

[9] H. Wang, Y. Gao, Elman's Recurrent Neural Network applied to forecasting the quality of water diversion in the water source of Lake Taihu, Proceedings of the International Conference on Biology, Environment and Chemistry IPCBEE, Singapore, 2010.

[10] S. Ahmad, I. Khan, B. Parida, Performance of stochastic approaches for forecasting river water quality, Water Res. 35(18) (2001) 4261–4266.

[11] H.C. Guo, L. Liu, H. Huang, A stochastic water quality forecasting system for the Yiluo River, J. Environ. Inform. 1(2) (2003) 18–32.

[12] O.F. Durdu, A hybrid Neural Network and ARIMA model for water quality time series prediction, Eng. Appl. Artif. Intell. 23 (2010) 586–594.

[13] M.J. Diamantopoulou, V.Z. Antonopoulos, D.M. Papamichail, The use of Neural Network technique for the prediction of water quality parameters of Axios River in Northern Greece, Eur. Water 12(11) (2005) 55–62.

[14] M. Diamantopoulou, V. Antonopoulos, D. Papamichail, Cascade correlation artificial neural networks for estimating missing monthly values of water quality parameters in rivers, Water Resour. Manag. 21 (2007) 649–662.

[15] M. Kotopouli, Ar. Psilovikos, N. Gkitsakis, M.Sapountzis, G. Dimos, Time Series Analysis of rain amount at rainfall station Nestorio Kastoria using ANN, Hydrotech 18–19 (2009) 49–64 (in Greek).

[16] K.A. Mitsiou, V.Z. Antonopoulos, D.M. Papamichail, Statistical analysis of water quality parameters time series of Strymon River, Hydrotech 9 (1999) 59–74 (in Greek).

[17] A. Psilovikos, K. Albanakis, S. Margoni, Ar. Psilovikos, C. Makrygiorgos, Some data of the operation of "REMOS" system in Nestos Delta for the quantitative and qualitative monitoring of water resources after the Dam construction, Proceedings of the 3rd Hellenic Conference of Agriculture Engineering, Thessaloniki, Greece, 2003, pp. 257–264 (in Greek).

[18] Y. Mylopoulos, E. Kolokytha, E. Kampragou, D. Vagiona, A COMBINED Methodology for transboundary river basin management in Europe. Application in The Nestos–Mesta catchment area, Water Resour. Manage. 22(8) (2008) 1101–1112.

[19] A. Psilovikos, E. Vavliakis, T. Langalis, Natural and anthropogenic processes of the recent evolution of Nestos Delta, Bull. Geol Soc Greece 20 (1988) 313–324.

[20] I. Choleev, G. Baltakov, Basic features of the late Cenozoic evolution of the Mesta valley system on Bulgaria territory, Proceedings of the International Conference: Geographica Rhodopia, Sofia, Bulgaria, vol. 1, 1989, pp. 14–17.

[21] D. Argiropoulos, E. Papachristou, J. Ganoulis, Statistical assessment of water pollution in the Aegean rivers: the case of Nestos, 6th meeting of the Regional Agency for the Environment, Provence-Alpes-Cote d' Azur, France, 1994.

[22] D. Argiropoulos, J. Ganoulis, E. Papachristou, Water quality assessment of the Greek part of Nestos (Mesta) River. in: Transboundary Water Resources Management (Ed.), Springer-Verlag, Germany, Heidelberg, 1996, pp. 427–438.

[23] M. Moustaka-Gouni, K. Albanakis, M. Mitrakas, A. Psilovikos, Planktic autotrophs and environmental conditions in the newly-formed hydroelectric Thesaurus Reservoir, Greece Archiv. Hydrobiol. 149(3) (2000) 507–526.

[24] A. Psilovikos, Research of the anoxic problem at the artificial lake of Thesaurus, River Nestos, Technical Report, Research committee of AUTH 7730 2001, pp. 492.

[25] K. Albanakis, M. Mitrakas, M. Moustaka-Gouni, A. Psilovikos, Determination of the environmental parameters that influence sulphide formation in the newly formed Thesaurus reservoir, in Nestos River, Greece Fresen. Environ. Bull. 10(6) (2001) 566–571.

[26] K. Anastassopoulos, Treatment of big landslides affecting construction of Thesaurus hydroelectric project, Proceedings of Hydro 2006, Maximizing the Benefits of Hydropower, 2006, Porto Carras, Greece, Paper 6.05, 2006.

[27] A. Sentas, A. Psilovikos, Comparison of ARIMA and transfer function (TF) models in water temperature simulation in Dam–Lake Thesaurus, Eastern Macedonia, Greece, Proceedings of the International Symposium: Environmental Hydraulics, Athens, Greece, 2010, pp. 929–934.

[28] A. Sentas, A. Psilovikos, Dissolved oxygen assessment in Dam–Lake Thesaurus using stochastic modeling, Proceedings of the International Conference: Protection and Restoration of the Environment XI, Thessaloniki, Greece, 2012, pp. 1573–1582.
[29] Ar. Psilovikos, S.Margoni, Ant. Psilovikos, Simulation and trend analysis of the water quality monitoring daily data in Nestos River Delta. Contribution to the sustainable management and results for the years 2000–2002, Mar. Pollut. Bull. 116(1–3) (2006) 543–562.
[30] E. Nakova, F. Linnebank, B. Bredeweg, P. Salles, Y. Uzunov, The river Mesta case study: A qualitative model of dissolved oxygen in aquatic ecosystems, Ecol. Inform. 4(5–6) (2009) 339–357.
[31] A. Psilovikos, A. Sentas, Comparison and assessment of the monitoring data of two REMOS stations in Nestos and Pagoneri for the year 2004. The base for an integrated water management, Desalination 248(1–3) (2009) 1016–1028.
[32] A. Psilovikos, A. Sentas, C. Sahanidis, T. Laopoulos, Trend analysis and assessment of water quality and quantity monitoring data in lignite mines of Western Macedonia, Greece, Desalin. Water Treat. 33 (2011) 44–52.
[33] H. Akaike, A Bayesian analysis of the minimum AIC procedure, Ann. Inst. Stat. Math. 30 (1978) 9–14.
[34] G. Box, G. Jenkins, G. Reinsel, Time Series Analysis, Forecasting and Control, third ed., Prentice-Hall, London, 1994.
[35] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.
[36] J.D. Gibbons, Nonparametric Methods for Quantitative Analysis, third ed., American Sciences Press, Ohio, Columbus, 1997.
[37] S. Palani, S.Y. Liong, P. Tkalich, An ANN application for water quality forecasting, Mar. Pollut. Bull. 56 (2008) 1586–1597.
[38] S. Haykin, Hydrology and Earth System Sciences, third ed., Pearson Education, Hamilton, Ontario, Canada, 2009.
[39] V. Antonopoulos, D. Papamichail, K. Mitsiou, Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece, Hydrol. Earth Syst. Sci. 5(4) (2001) 679–692.
[40] H.R. Nash, J.V. Sutcliffe, River flow forecasting using stochastic models, Stoch. Environ. Res. Risk Assess. 19 (1970) 326–339.