Bioautomation, 2008, 9, 79-87 ISSN 1312 – 451X

Comparative Analysis of Two Models of the Strouma River Ecosystem

Tatiana Ilkova, Mitko Petrov*

Centre of Biomedical Engineering "Prof. Ivan Daskalov" – Bulgarian Academy of Sciences
105 Acad. G. Bonchev Str., Sofia 1113, Bulgaria
E-mail: tanja@clbme.bas.bg, mpetrov@clbme.bas.bg

* *Corresponding author*

Received: January 9, 2008 **Accepted: April 2, 2008**

Published: April 15, 2008

Abstract: *A modified method of regression analysis for modelling of the water quality of river ecosystems is offered. The method is distinguished from the conventional regression analysis of that the factors included in the regression dependence are time functions. Two type functions are tested: polynomial and periodical. The investigations show better results the periodical functions give. In addition, a model for analysis of river quality has been developed, which is a modified method of the time series analysis. The model has been applied for an assessment of water pollution of the Strouma river. An assessment for adequately of the obtained model of the statistical criteria – correlation coefficient, Fisher function and relative error is developed and it shows that the models are adequate and they can be used for modelling of the water pollution on these indexes of the Strouma river. The analysis of the river pollution shows that there is not a materially increase of the anthropogenic impact of the Strouma river in the Bulgarian part for the period from 2001 to 2004.*

Keywords*: Regression analysis, Times series analysis method, Water quality, Strouma river.*

Introduction

Modelling of the water pollution gives a chance to define the basic tendencies in the evolution of water quality for a period. The use of deterministic models for the water ecosystems functioning demands complex information about the river system, the catchment pollution and the seasonal and annual fluctuations of water quality for a period covering many years. The models for assessment of river ecosystems quality are based on the "black box" method, where the characteristics of the processes are not directly taken into consideration. The data are analysed using the inlet and outlet information. The developed models are of essential meaning for assessing the water quality of the river ecosystem [1, 2].

The analysis of the time series shows the availability of periodical and long-term changes of some indices for water pollution, which may be used in water quality management. On analyzing the time series the physical, chemical, biological and biochemical characteristics of the processes are not directly taken into consideration; an attempt however, was made for a description in the time of water quality in definite points as an integral result of water ecosystem functioning. The time series analysis for water quality is a part of the retrospective modelling of water ecosystem functioning carried out with deterministic methods. The integration of determination and statistical models for water quality is a necessary means of water ecosystem management. These models use information from environmental monitoring, and their realizations [6, 7].

The major goal of the present investigation is developing of models of the water pollution by a modified regression analysis (MRA) method and modified times series analysis (MTSA) method, and comparison between them at the end of Bulgarian part – the border with Greece.

Materials and methods

A MRA method for modelling of the water river quality

The polynomial models from second order are some the most used in the practice. They allow the investigation phenomenon to be described in comparison a wide area of change of the input variables. In a general type, they have the type [3, 5]:

$$\mathbf{y} = \mathbf{B}_0 + \sum_{i=1}^{m} \mathbf{B}_i x_i + \sum_{i=1}^{m} \mathbf{B}_{i,i} x_i^2 + \sum_{i=1}^{m} \sum_{j=i+1}^{m-1} \mathbf{B}_{i,j} x_i\, x_j \,, \tag{1}$$

where: $\mathbf{B}_0 = \mathbf{B}_0[b_0^1, b_0^2, ..., b_0^m\,; \mathbf{B}_{i,i} = \mathbf{B}_{i,i}[b_{1,1}^1, b_{2,2}^1,..., b_{m,m}^1; b_{1,1}^2, b_{2,2}^2,..., b_{m,m}^2; ..., b_{1,1}^m,..., b_{2,2}^m,..., b_{m,m}^m]$; $\mathbf{B}_{i,j} = \mathbf{B}_{i,j}[b_{1,2}^1, b_{1,3}^1, ..., b_{i,j}^1;\ b_{1,2}^2, b_{1,3}^2, ..., b_{i,j}^2; ..., b_{1,2}^m, b_{1,3}^m, ..., b_{i,j}^m]$, i = 1, …, m, j = i + 1, m–1; m is number of the investigations index; **y**–vector of the investigated indexes of the pollution at the end of the Bulgarian part, $\mathbf{y} = \mathbf{y}\left[y^1, y^2, ..., y^m\right]^T$; $\mathbf{x}_i$ – the same investigated indexes at the beginning of the Bulgarian part of the river.

In contrast to the conventional regression analysis, the factors participant in the regression dependence are the time functions. In the paper, the following models are investigated:

$$x_i(t) = \sum_{k=0}^{r} a_{i,k}\, t^k \tag{2}$$

or

$$x_i(t) = c_i \sin\left(\pi\left(t - p_i\right)/\, q_i\right) \tag{3}$$

where: r is the polynomial degree, $a_{i,k}$ – the coefficients in the polynomial functions (2) of each index; c_i, p_i and q_i – the coefficients on each index for the periodical functions.

The choice of a suitable function from (2) and (3) is made based on the correlation coefficient. The regression analysis requires some statistical verification of the experimental data to be made. The verification is based on the mathematical statistic methods, as far as the investigated indexes are shown chance quantities and chance processes. An algorithm and program is developed of the MRA method [5].

A MTSA method for modelling of the water river quality

For modelling of water pollution, a method of TSA is a determined component (trend) – $\mathbf{x}_T$ describing the regularity of the development of the examined phenomenon, periodical component – $\mathbf{x}_P$ and stochastic variable – ε_t [1]:

$$\mathbf{x} = \mathbf{x}_T + \mathbf{x}_P + \varepsilon_t \,, \tag{4}$$

where: **x** is the vector of the investigated indexes of the pollution.

The determined component (*trend*) – $\mathbf{x}_T$ is a polynomial of 1st to 3rd degrees and the periodical component – $\mathbf{x}_P$ is described by the order of Fourier.

In contrast to the conventional method of TSA the determined component is a polynomial with high degree:

$$\mathbf{x}_T = \sum_{j=0}^{r} a_j t^j , \qquad (5)$$

where: a_j are coefficients of polynomial, $j = 0, 1, 2, \ldots, r$; r – degree of the polynomial, $r \le 5$.

The main trend shows the main tendencies in the alteration of the studied indices, and it is a straight line:

$$\mathbf{x}_T^M = A_0 + A_1 t \qquad (6)$$

In contrast to the classical method for analyzing temporary series, where the Fourier series are used, the present research proposes the uses of the periodical functions from the type:

$$\mathbf{x}_P = \sum_{k=0}^{p} b_k \sin(2\pi t / c_k + d_k), \qquad (7)$$

where: p are number of the periodical functions; b_k, c_k and d_k – coefficients in the periodical functions, $k = 0, \ldots, p$.

The number of the periodical functions p in (7) and the polynomial degree (5) is determined on basis the statistical criteria experimental Fisher function – F_E and experimental correlation coefficient – R^2_E. Then the model (4) for analysis and prognosis has the following form ($\varepsilon_t = 0$) [4]:

$$\mathbf{x} = \sum_{j=0}^{r} a_j t^j + \sum_{k=0}^{p} b_k \sin\left(\frac{2\pi}{c_k} t + d_k\right) \qquad (8)$$

An algorithm and program is developed for the MTSA method [4].

The assessment for the models adequate is made on the basis of the following statistical criteria: Fisher function F, correlation coefficient R^2 and relative error S_L. The relative error is determined by the relation [2, 7]:

$$S_{L_j} = \sqrt{\frac{1}{(n-k)} \sum_{i=1}^{n} \left(1 - \frac{y_{i,j}}{x_{i,j}}\right)^2} , \qquad (9)$$

where: n are number experimental observations; k – number of the coefficients in the model; $x_{i,j}$ and $y_{i,j}$ – experimental and the simulated of the models; $j = 1, \ldots, m$.

Indexes value S_{L_j} is in the interval from 0 to 1.

An application of the MRA and MTSA methods for modelling of the Strouma river pollution in the Bulgarian part

The used information is received from the National System for Ecological Monitoring for the period from 2001 to 2004. The following indices of biogenic water pollution have been measured: *biochemical oxygen demand* (BOD), *Permanganate oxidation* (KMn_3O_4-oxidation), *ammonia nitrogen* ($N\text{-}NH_4$), *nitrate nitrogen* ($N\text{-}NO_3$) and *phosphate* (PO_4) at the Struma river near the town of Batanovtsi (*Point* 1 – input) and at the final point, which is the village of Marino Pole of the Bulgarian section (*Point* 2 – output) for modelling with MRA. For modelling with MTSA we examined the water pollutions of the Strouma river in point at the end of Bulgarian part – *Point 2* [4, 5].

The coefficients in the regression models from the type (1) are determined by a subprogram for multiple regression. The investigation of the indexes at the beginning (*Point* 1) shows better results gives the periodical functions from the type (3). The modelling of the MTSA the trend for investigated indexes shows the model from the third degree has the best statistical index. The comparison is made on the Fisher function [4]:

$$\mathbf{x} = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \sum_{k=0}^{3} b_k \sin\left(2\pi t / c_k + d_k\right) \quad (10)$$

At the pollution assessment it is useful the basic to be determined. It shows the tendencies in the river pollution for investigated indexes and it is a linear function on the time:

$$\mathbf{x} = A_0 + A_1 t \quad (11)$$

Comparative analysis of the models

The statistical indexes for the assessment for adequacy of MRA and MTSA models are shown in Table 1 [4, 5] and the results for both models are shown in Figs. 1-5.

Table 1. Statistical indexes

Index	MRA models			MTSA models		
	$R^2_{E_1}$	F_{E_1}	S_{L_1}	$R^2_{E_2}$	F_{E_2}	S_{L_2}
BOD	0.918	1.38	1.545	0.993	1.096	1.734
KMn_3O_4-oxidation	0.937	1.57	1.288	0.996	1.065	1.908
$N\text{-}NH_4$	0.934	1.95	1.361	0.996	1.051	3.080
$N\text{-}NO_3$	0.947	1.90	3.600	0.993	1.085	0.620
PO_4	0.881	1.29	2.056	0.997	1.038	0.752

In the Table 1 $R^2_{E_i}$ are experimental correlation coefficients, F_{E_i} – experimental Fisher function and S_{L_i} – relative error, i = 1, 2. The theoretical value of the correlation coefficient at a degree of freedom n = 33 is $R^2_T(33)$ = 0.349 and the theoretical value of the Fisher function is F_T = 2.119. On the experimental correlation coefficient and experimental Fisher function the MSTA models are vastly better than MRA models ($R^2_{E_2} > R^2_{E_1}$ and $F_{E_2} < F_{E_1}$). On the relative error, the MRA models for the first three indexes show better results $S_{L_1} < S_{L_2}$ and for the last two indexes, the MTSA models are better $S_{L_1} > S_{L_2}$.

Only of above mentioned, the models are compared with the following indexes: 1) measure for definiteness Q_j and 2) Fisher function – F_C. The measure for definiteness is determined by the relation [3]:

$$Q_j = \sqrt{1 - \frac{(n-k-1)s^2_{r_j}}{(n-1)s^2_{x_j}}}\,, \quad (12)$$

where: $s^2_{r_j} = \frac{1}{n-k}\sum_{i=1}^{n}\left(\eta_{i,j} - \eta^*_{i,j}\right)^2$; $\eta_{i,j} = \frac{x_{i,j} - \bar{x}_j}{s_{x_j}}$, $\eta^*_{i,j} = \frac{y_{i,j} - \bar{y}_j}{s_{y_j}}$;

$$s^2_{x_j} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{i,j} - \bar{x}_j\right)^2;\ s^2_{y_j} = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_{i,j} - \bar{y}_j\right)^2.$$

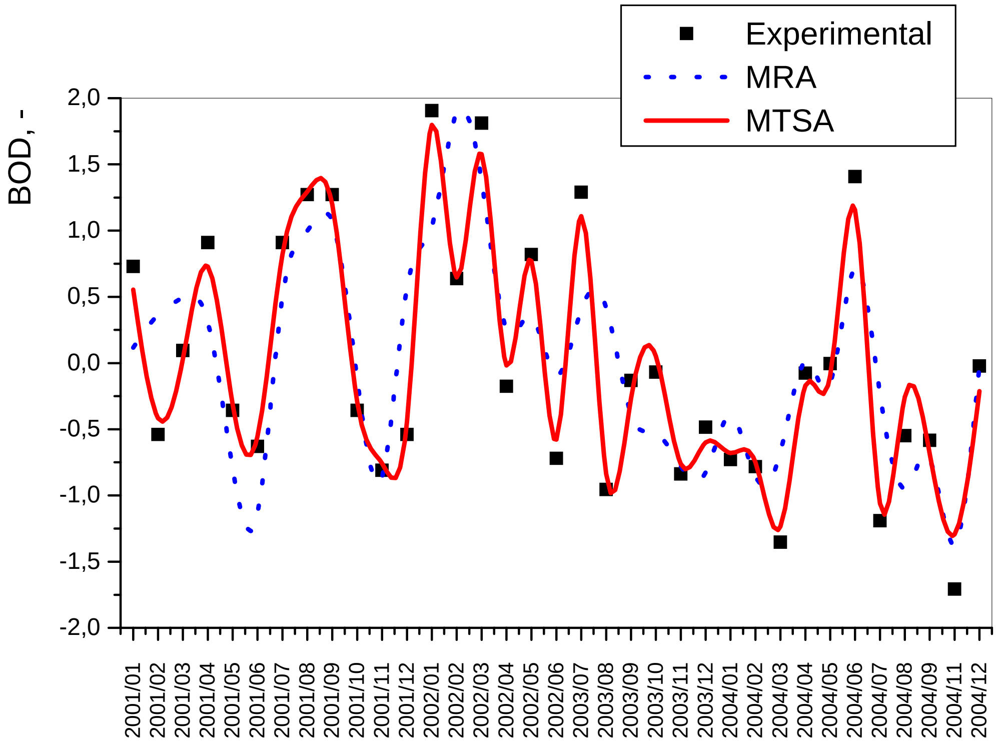

Fig. 1 MRA and MTSA models of the BOD

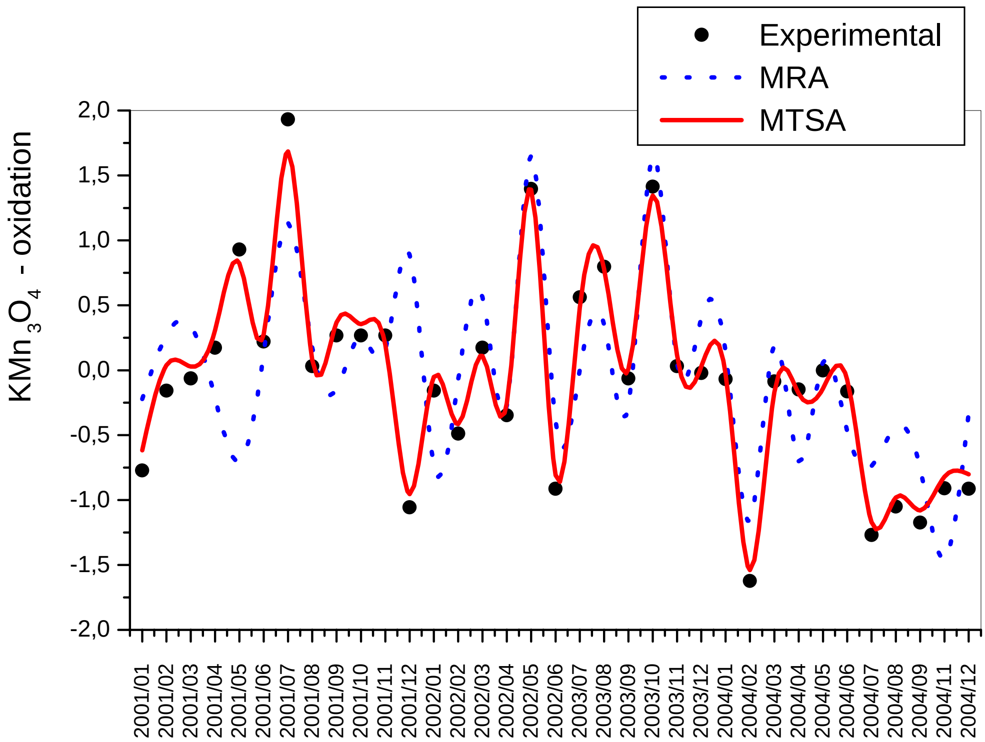

Fig. 2 MRA and MTSA models of the KMn_3O_4-oxidation

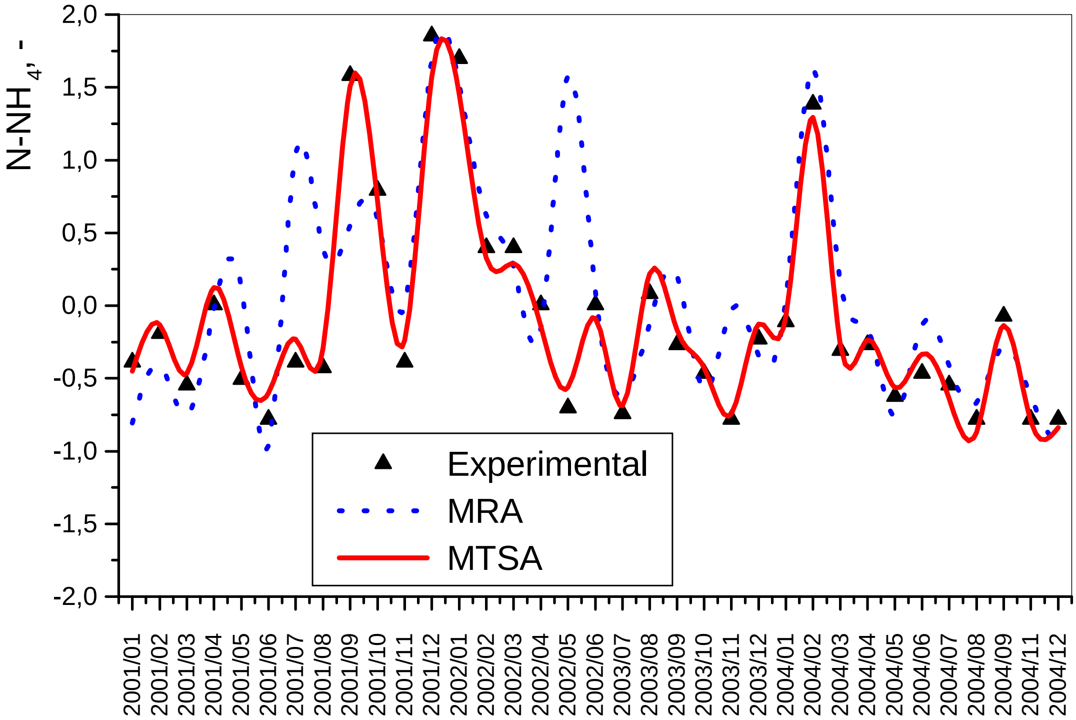

Fig. 3 MRA and MTSA models of the $N\text{-}NH_4$

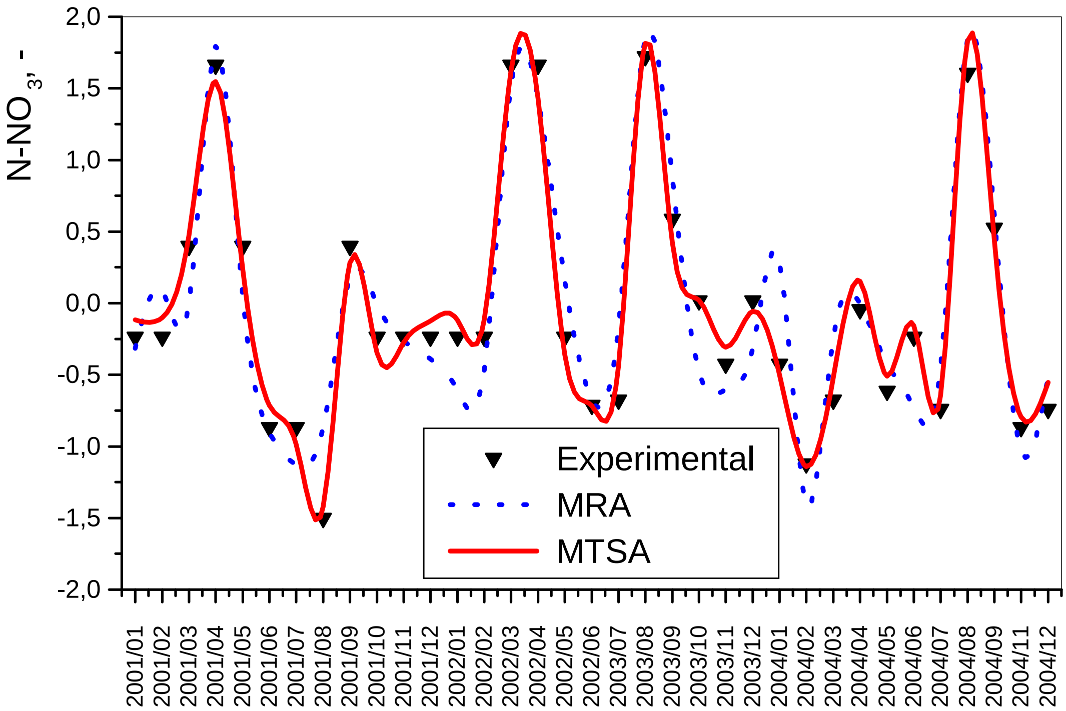

Fig. 4 MRA and MTSA models of the $N\text{-}NO_3$

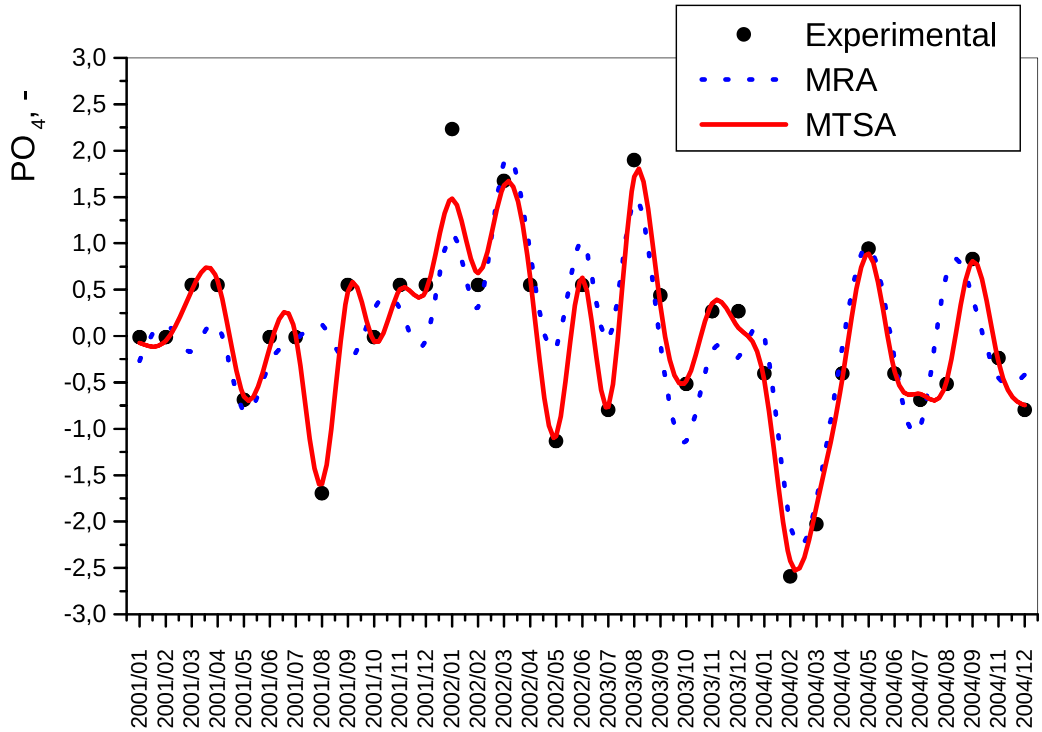

Fig. 5 MRA and MTSA models of the PO_4

As *the measure for definiteness* Q_j is nearer to one as the model describes better the experimental results.

The Fisher function F_C is determined by the relation [3]:

$$F_j = \frac{s^2_{1r_j}}{s^2_{2r_j}}, \tag{13}$$

where: $s^2_{1r_j}$ and $s^2_{2r_j}$ are the residuals standard deviation of the *model* 1 (MRA models) and *model* 2 (MTSA models), calculated by the equations:

$$s^2_{1r_j} = \frac{1}{n-m_1}\sum_{i=1}^{n}\left(x_{i,j} - y_{1i,j}\right)^2, \quad s^2_{2r_j} = \frac{1}{n-m_2}\sum_{i=1}^{n}\left(x_{i,j} - y_{2i,j}\right)^2 \tag{14}$$

In the numerator always is bigger from the two dispersions, i.e. $s^2_{1r_j} > s^2_{2r_j}$.

The received F_C is compared with the tabular values F_T for a given level of appreciable α and degree of freedom $\nu_1 = n - m_1$ и $\nu_2 = n - m_2$, where m_1 and m_2 is a number of the coefficients in the *model* 1 and 2.

If $F_C > F_T$, then we accept the *model* 2 (MTSA) leads to a materially decrease of the remainder dispersion and therefore the *model* 1 is better.

The value of the indexes (13) and (14) are shown in Table 2.

Table 2. Value of the indexes (13) and (14)

Index	BOD	KMn_3O_4-oxidation	N-NH_4	N-NO_3	PO_4
$Q_{MRA, j}$	0.785	0.752	0.839	0.945	0.813
$Q_{MTSA, j}$	0.970	0.938	0.925	0.979	0.967
$F_{C, j}$	9.020	4.980	4.723	3.516	7.220

The theoretical values of the Fisher function F_T at degrees of freedom $\nu_1 = 14$ and $\nu_2 = 19$ and level of considerable $\alpha = 0.05$ is $F_T(0.05, 14, 19) = 2.26$ [5].

From the presented results in Table 2 it is shown on the indexes Q_j and F_C, the MTSA models give better results than MRA models. Therefore, the MTSA models are more suitable for using at the investigation of the river ecosystems pollution. The MRA models can be used at analysis and assessment of water ecosystems pollution for a definite (final) time period.

Conclusions

1. The developed comparative analysis of the two models for modelling of the Strouma river pollution has shown the MTSA models are better than the MRA models. Another thoroughly priority of the times series analysis is it can be used also for pollution prognosis.
2. The MRA models are rather sensitive to the prerequisites for their using. The mistakes in determining of the input variables lead to considerable mistakes in the determination the regression coefficients, as some time it can be reached to change of the sign of the coefficients. The regression analysis decides actually the problem for approximation in the given area of the space of the input variable, determined by the experimental taken down point. Every experience an *extrapolation* outside this area to be made is *ungrounded*.
3. The obtained regression models can be used for analysis and modelling of the river ecosystems pollution for a definite time period and the models, based on the modified times series analysis could be profited as analysis and modelling as for prognosis of the river ecosystems pollution.

Acknowledgements

The investigations are partially supported by Fund for Scientific Investigations of the Ministry of Education and Science of Bulgaria by grant ES–YS–1501/2005.

References

1. Box G., M. Jenkins, G. Reinse (1994). Time Series Analysis: Forecasting & Control, 3rd ed., Prentice-Hall, New Jersey.
2. Crabree R., L. Cluckre, C. Forste, C. Crocket (1986). A Comparison of Two River Quality Models, Water Research, 20, 53-61.
3. Gatev G. (1973). Methods, Algorithms and Programs for Modelling and Optimization of the Manufacture, Technique, Sofia, (in Bulgarian).
4. Ilkova T., M. Petrov (2007). A Modified Times Series Analysis Method for Modelling and Prognosis of River Water Quality: Application of the Strouma River, Proceedings of the International Conference Automatic and Informatics, Sofia, Bulgaria, 3-6 October, I-17-I-20.
5. Ilkova T., M. Petrov (2007). An Application of Regression Analysis for Modelling of Water Quality of River Ecosystems, Ecology – Scientific Articles, 2, 273-290.

6. Straskraba M. (1982). Application of Predictive Mathematical Models of Reservoir Ecology and Water Quality, Canadian Water Resources Journal, 7, 283-318.
7. Topnikov V., V. Vavilin (1992). Sravnitelnaja Ocenka Matematicheskih Modelej Samoochishchenija Rek, Vodnie Ressursi, 1, 59-76, (in Russian).

Tatiana Ilkova, Ph.D.
E-mail: tanja@clbme.bas.bg

Tatiana Ilkova (born 1970) graduated Technical University - Sofia, specialty Engineering of Biotechnology in 1995. She has worked as a Research Associate in the Centre of Biomedical Engineering “Prof. Ivan Daskalov”, Bulgarian Academy of Sciences since 1995. She achieved Ph.D. degree in 2008. Her scientific interests are in the fields of bioprocesses systems, modelling and optimization of bioprocesses and modelling of ecological systems. She has published about 120 scientific publications with more than 20 known citations.

Mitko Petrov, Assoc. Prof., Ph.D.
E-mail: mpetrov@clbme.bas.bg

Mitko Petrov (born 1959) graduated Technical University - Sofia as a mechanical engineer in 1987. He has worked as a Research Associate in the Centre of Biomedical Engineering “Prof. Ivan Daskalov”, Bulgarian Academy of Sciences since 1988. He achieved Ph.D. degree in 2004. He has been an Associate Professor since 2007. His scientific interests are in the fields of modelling and optimization of apparatus of bioprocess systems and modelling of ecological systems. He has published about 120 scientific publications with more than 20 known citations.