Bioautomation, 2007, 8, Suppl. 1, 87 – 94 ISSN 1312 – 451X

An Algorithm for Modified Times Series Analysis Method for Modeling and Prognosis of the River Water Quality

Ilkova T.*, **Petrov M.**

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

**Corresponding author*

Summary: An algorithm and programs for modeling, analysis, and prognosis of river quality has been developed, which is a modified method of the time series analysis (TSA). The algorithm and program are used for modeling and prognosis of the river quality of Bulgarian river ecosystems.

Keywords: Times Series Analysis, Algorithm of Modified Times Series Analysis Method, River Water Quality.

1. INTRODUCTION

The Commission to the Council and the European Parliament on implementation of Council Directive 91/676/EEC concerning the protection of waters against pollution gives that almost all Member States, with the exception of Italy, Portugal and Greece, provided at least some elements on assessment methods (simulation models and trend analysis) for the evaluation of trends in the water quality evolution.

Progress in the quality of action programs has been significant in the recent years and should contribute to the improvement of water quality in future reporting periods.

The TSA for water quality is a part of the retrospective modeling of water ecosystem functioning carried out with deterministic methods [1, 2]. The modified TSA is applied for modeling and prognosis of the Strouma river waters quality [6, 8].

In revolution investigations of the collective statistical treatment of the experimental data is developed and an analysis of the Strouma river waters quality in the Bulgarian part [7].

The aim of this work is developing and generalization of the algorithms and programs of the modified TSA.

2. A MODIFIED TIME SERIES ANALYSIS METHOD

For modeling 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, 2]:

$$\mathbf{x} = \mathbf{x}_T + \mathbf{x}_p + \varepsilon_t \tag{1}$$

where $\mathbf{x}$ – 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 as is a polynomial to high degrees:

$$\mathbf{x}_T = \sum_{j=0}^{r} a_j t^j \tag{2}$$

where a_j – coefficients of polynomial, j = 0, 1, 2, …, r; r – degree of the polynomial.

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

$$\mathbf{x}_T^M = A_0 + A_1 t \tag{3}$$

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:

$$\mathbf{x}_P = \sum_{k=0}^{p} b_k \sin\left(\frac{2\pi}{c_k} t + d_k\right) \tag{4}$$

where: p – 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 and the polynomial degree (3) is determined on basis the statistical criteria experimental Fisher function – F_E and experimental correlation coefficient – R^2_E.

Then the model (1) for analysis and prognosis has the following form ($\varepsilon_t = 0$):

$$\mathbf{x} = \sum_{j=0}^{r} a_j t^j + \sum_{k=0}^{p} b_k \sin\left(\frac{2\pi}{c_k} t + d_k\right) \tag{5}$$

An algorithm and program is developed of the modified TSA.

2.1. The main algorithm of the Modified TSA

The generalized algorithm scheme is:

1. BEGINNING
2. Statistical treatment and analysis of main indexes for the water quality. Computing of statistical the experimental data. For this aim an algorithm and program. Output results from subroutine program are: the real and centered experimental data – $x(i, j)$ and $\tilde{x}(i, j)$. The centered or the normalized data are computing is:

   $\tilde{x}_{i,j} = x_{i,j} - \bar{x}_j$ or $\tilde{x}_{i,j} = \left(x_{i,j} - \bar{x}_j\right) / \sqrt{s_j^2}$,

   where: $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{i,j}$; $s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{i,j} - \bar{x}_j\right)^2$; n – number of time partitions, (i = 1, …, n); m – number of investigation parameters of the pollutions, (j = 1, …, m). Subprogram also makes the following: *check-ups for presence of trend, test for stationary, computing of autocorrelation, correlation matrix, and significance of correlation coefficients.*
3. The computing time in year [4, 5]: $t = Year + \frac{1}{12}\left(Month - 0.5\right)$.

4. The coefficients A_0, A_1 are determined in the main trend (3), and the coefficients a_j, $j = 0, \ldots, r$ in trend (2). The procedure is carried out with the help of the subroutine POLRG from IMSL library of FORTRAN.
5. The experimental values of the periodical function: $\mathbf{x}_p = \mathbf{x} - \mathbf{x}_T$ are calculated.
6. The coefficients b_k, c_k and d_k in periodical functions (4) and their numbers are determined. Comparison according to Fisher function – F. The procedure is carried out with the help of the subroutine for nonlinear regression RNSQ from IMSL library of FORTRAN.
7. The models (5) are determined by the programs for SIMPLEX method for optimization. The optimization program for direct search of the minimum of the criteria. The minimization criterion represents the sum of the squares of weighed residues ($SSWR$) [9]:

$$SSWR = \sum_{i=1}^{n} \frac{\left(x_i^m - x_i\right)^2}{w_i^2}, \text{ where: } w_i = \max(x_i, x_i^m).$$

8. Computing validation of the models: the correlation coefficient – R^2_E, the experimental value of the Fisher's function – F_E and the relative error S_L. The S_L is calculated by [10, 11]:

$$S_L = \sqrt{\frac{1}{(n-\nu)} \sum_{i=1}^{n} \left(\frac{x_i^e - x_i^m}{x_i^e} \right)^2},$$

   where x_i^e are the experimental values, x_i^m are the values calculated with the help of the model; ν – degrees of freedom; n – number of the experimental data points.
9. Comparison with the theoretical correlation coefficient R^2_T and Fisher's function F_T is made. *IF* $R^2_E > R^2_T$ and $F_E < F_T$, then the model is adequate. The lower the values of the relative error S_L are the higher the validity of the model for the investigated indexes.
10. Write: R^2_{E}, F_E, T' and S_L.
11. Write: Time, x_M^T from (2), x_T from (3), model for analysis and prognosis from (5).
12. END

The main algorithm is shown in Fig. 1, where: POLRG is a procedure of the polynomial regression, RNSQ is a procedure for nonlinear regression, SIMPLEX is a procedure for optimization. SIMPLEX serves for specifying of the model coefficients, if it is necessary.

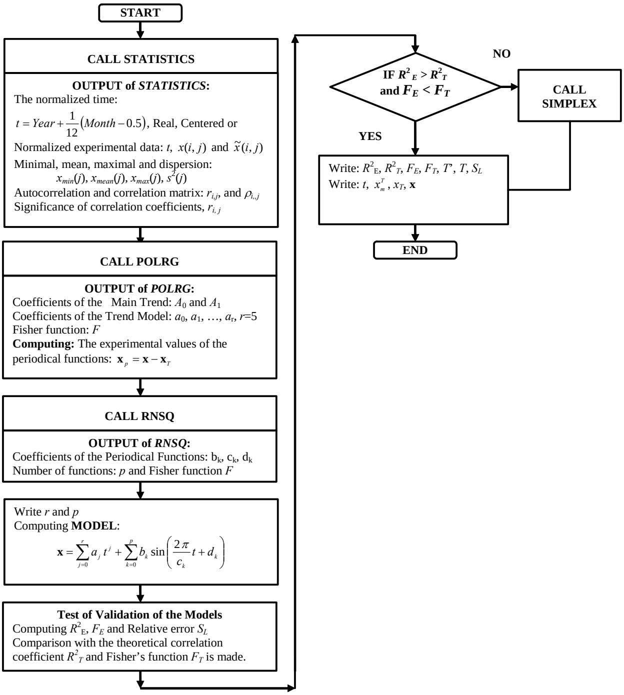

Fig. 1. Main algorithm of modified TSA

2.2. Algorithm for statistical treatment data

The block scheme of the algorithm is shown in Fig. 2, where: SPLINE is a procedure for extrapolation of the experimental data in the time; STATIONARY is a procedure for check-up of the *stationary* of the experimental observations; TREND is a procedure for test for presence of a *trend* (check-up for normally distribution of the observations); CORR is a procedure for calculating of the autoand correlation functions; STUDNT is a procedure for determining of the correlation coefficients.

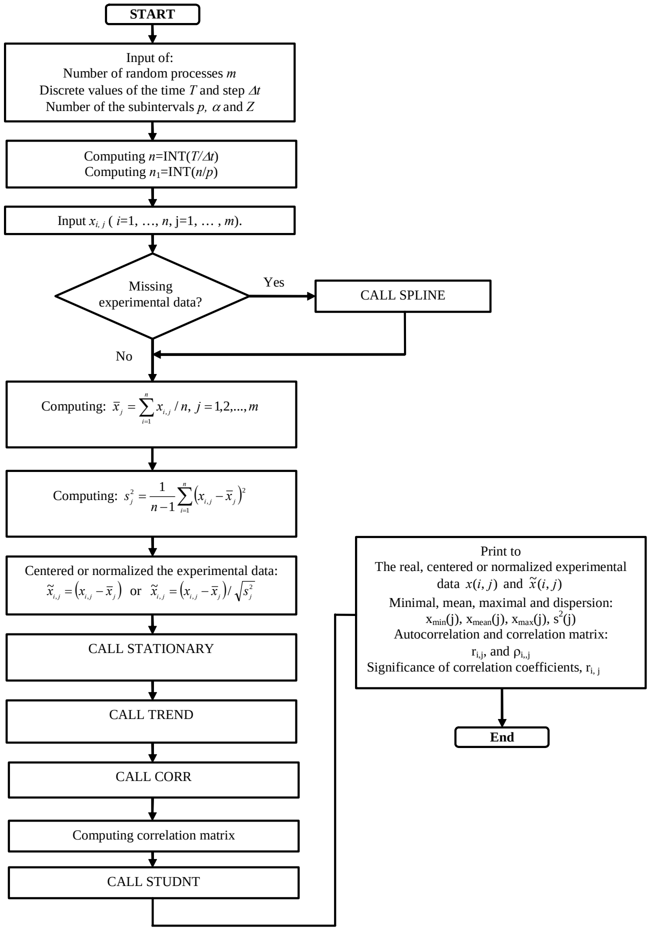

Fig. 2. Block scheme of the algorithm

All programs are developed on COMPAQ Visual FORTRAN Pro [3].

3. CONCLUSION

1. A model for analysis and prognosis of the river pollution, based on a modified time series analysis is developed. This approach can be used at modeling and prognosis of different water ecosystems. The proposed modified approach for time series analysis can be used on solving different problems connected with ecosystem management.
2. A method, algorithm and program for statistical treatment and analysis of experimental observation for the river water quality are developed. The statistical data treatment is used at modeling of the quality of the Struma river. The program can be used at the statistical data treatment also river ecosystems.
3. An algorithm and program for modified times series analysis is developed. Thus the offered algorithm and developed program are used for analysis, modeling and prognosis of the river water quality.

Acknowledgments

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 D., G. Djenkins, The Time Series Analysis Method, Mir, Moskva, 1974, (in Russian).
2. Box G., N. Draper, Empirical Model Building and Response Surface, J. Wiley and Sons, N.Y., 1987.
3. COMPAQ Visual FORTRAN Programmer's Guide v. 6.6, Compaq Computer Corporation Houston, Texas, 2001.
4. http://en.wikipedia.org/wiki/Time_series
5. http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm
6. Ikova T., M. Petrov, Modeling and Prognosis of the Strouma River Pollution, *Proceedings of the Int. Symposium BioPS'07*, Sofia, 6-7 November 2007, II.15-II.24.
7. Ikova T., M. Petrov, Statistical Treatment and Analysis of Main Indexes for the Water Quality of the Struma river, *Ecological Engineering and Environment Protection*, 2006, 3-4, 10-18, (in Bulgarian).

8. Ilkova T., M. Petrov, A Modified Times Series Analysis Method for Modeling and Prognosis of River Water Quality: *Application of the Strouma River*, *Proceedings of the Int. Conference Automatic and Informatics*, 3-6 October, Sofia, Bulgaria, 2007, I, I-17-I-20.
9. Sachs L., Angewandte Statistik. Anwendung Statistischer Sethoden, Springer Verlag, Berlin, 1992.
10. Stegmann H.-J. Ehrig, P. Liem, Anwendung Einer Zeitreihenanalyse in der Wassergutewirtschaft, *Wasser und Boden*, 1978, 3, 50-54.
11. Topnikov V., V. Vavilin, Sravnitelnaja Ocenka Matematicheskih Modelej Samoochishchenija Rek, *Vodnie ressursi*, 1992, 1, 59-76, (in Russian).