Analysis of Water Quality Data of St.Petersburg Small Rivers

Viktor N. Assaoul and Igor V. Bovykin
Institute of Limnology, Russian Academy of Sciences, Sevastyanov Str. 9, St.Petersburg, 196199, Russia

Abstract. The present data study is devoted to the analysis of different water quality series of the small rivers of St.Petersburg in connection with the main sources of pollutants. The first objective of the study is selection of statistical procedures with due regard for the features of water quality data. The second objective is analysis of the data and interpretation of the main results. The data analysis includes trend analysis and estimation of change points of the time series on different pollutants. The water quality parameters of the rivers are analysed and compared with appropriate values of maximum permissible concentrations. The analysis of the causes of high pollution levels is also carried out.

1. Introduction

One of the acute ecological problems of St.Petersburg is decline in water quality in the river Neva that depends on water quality in its tributaries and in Lake Ladoga. As a result the condition of the drinking water supply for 5 million inhabitants and water quality of the Neva Bay is estimated as critical.

The small rivers of St.Petersburg are an important element of the Lake Ladoga - Neva - Neva Bay water system. The rivers are the main ways of pollution transportation from numerous plants and the sewerage system into the river Neva and the Neva Bay. That is why they are the most vulnerable elements of the water system.

The analysis of changes of water quality of the rivers has been carried out for years. But only now the amount of data has become sufficient for the formal statistical analysis. Note that testing of water quality characteristics for trends is an important element of the monitoring. The main problems of the monitoring are not only estimation of the current water quality but also analysis of trends in indicators of ecological condition of the city. The trend analysis of water quality of the rivers is an indispensable stage for solving the problems of the management of the ecological state of the city.

The data analysis of the water quality series includes exposure of trends, estimation of change points and analysis of causes of the trends. It is important that visual analysis of time series plots usually precedes using statistical methods. The purpose of such kind of analysis is to formulate hypotheses of existence of trends.

2. Data Base

The data on the following rivers were studied: Fontanka, Okhta, Moyka, Chernaya Rechka, Karpovka, Zhdanovka, Izhora, Slavyanka and Tosna. The rivers Okhta, Chernaya Rechka, Izhora and Slavyanka are tributaries of the river Neva. The rivers Fontanka, Moyka, Karpovka and Zhdanovka are the branches of the river Neva. The selected rivers represent the diversity of all small rivers of the city.

The data analysis is concerned with the following series of regular observations:

- Dissolved Oxygen (DO);
- Biological Oxygen Demand (BOD-5);
- Suspended Matter;
- Nitrogen (Nitrates, Nitrites and Ammonium);
- Phosphorus;
- Phenols;
- Oil Products;
- Synthetic Surface-Active Matter;
- Metals (Copper, Lead, Marganese, Cadmium).

The complex indicator WPI (Water Pollution Index) is also analysed. It is determined as follows:

$$I = 1/6 \sum_{i=1}^{6} C_i/L_i, \qquad (1)$$

where C_i (i=1....6) are the concentrations of the pollutants with the biggest mean annual values; L_i (i=1....6) are the corresponding limit permissible concentrations.

3. Statistical Procedures

The following two statistical procedures are used for trend exposure in the series. The first trend test is based on linear regression of series Y versus time t

$$y_t = \alpha + \beta t + \varepsilon_t, \quad t = 1,2...n, \qquad (2$$

where y_t is the value of Y at the moment t; ε_t is a zero-mean random error and n is the number of observations. The coefficients α and β are calculated using methods of least squares.

The problem is to test the hypothesis H_0: $\beta = 0$. If the hypothesis is true then the trend is absent and vice versus if the hypothesis is violated then the trend exists. The statistic used in this procedure is

$$T = \sqrt{\frac{r(n-2)}{1-r}}, \qquad (3)$$

where r is the correlation coefficient between Y and t.

If the hypothesis H_0 is true then the statistic T follows the Student's distribution with n - 2 degrees of freedom. This method is appropriate for analysis of mean annual values.

Another currently acceptable statistical procedure is the Mann-Kendall test (Mann 1945, Kendall 1975) which has been used to detect trends in water quality data (Hirsch et. al. 1982, Hipel et. al. 1988) and is recommended for use in more general environmental applications (Gilbert 1987, Loffits et. al. 1989). The Mann-Kendall test statistic is

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \mathrm{sgn}(y_j - y_k), \qquad (4)$$

where

$$\mathrm{sgn}(y_j - y_k) = \begin{cases} 1, & y_j - y_k > 0 \\ 0, & y_j - y_k = 0 \\ -1, & y_j - y_k < 0 \end{cases}$$

The statistic S is asymptotically normally distributed with zero mean and variance $n(n-1)(2n+5)/18$.

This test is a robust procedure and it can be used not only for analysis of trend of single series but also for analysis of multivariate series (Gilbert 1987, Loffits et. al. 1989). The method can be applied to analysis of season water quality data as well (Hirsh et. al. 1982).

The problem of estimating change points in the series can be solved in the scope of the following linear model (Quandt, 1958, 1960):

$$y_t = X_t^T \beta + \varepsilon_t, \quad t=1...n, \qquad (4)$$

where $X_t = (x_{1t},..., x_{rt})^T$ is the vector of values of regressors at the moment t; $\beta = (\beta_1,... ,\beta_r)^T$ is a r-dimension vector of parameters; symbol T means transposition;

It is assumed that at a certain moment t_o the vector β changes. It means that

$$y_t = X_t^T \beta_1 + \varepsilon_t, \quad t=1...t_0-1,$$
$$y_t = X_t^T \beta_2 + \varepsilon_t, \quad t=t_0...n. \qquad (5)$$

To estimate t_o the statistic S_t is considered:

$$S_t = 1/2\,((t-1)\ln\sigma^2_1 + (n-t+1)\ln\sigma^2_2 - n\ln\sigma^2_0), \quad (6)$$

where σ^2_1, σ^2_2, σ^2_0 are estimates of residual variances using the first t-1 values of y_t, the last n-t+1 values and all values correspondingly.

The values of t_o are determined from the formula:

$$t_o = \arg\min (S_t) \qquad (7)$$

The following two partial cases of the model (4) are considered.

1. Let $r = 2$, $x_{1t} = 1$, $x_{2t} = t$. Then the model (4) comes to:

$$y_t = \alpha + \beta t + \varepsilon_t, \quad t=1....n$$

The problem is to determine the moment t_o, when the parameters α и β change. It also means that the kind of the trend changes at the moment t_0.

2. Let $r = 1$, $x_{1t} = 1$. Then (4) comes to:

$$y_t = \alpha + \varepsilon_t, t_o, \quad t=1...n$$

So there is a switch of y_t at the moment t_o.

After estimating the values of change points it is necessary to test whether the coefficients in two regressions (see formula (5)) are equal. For this purpose the test statistic can be used:

$$\left(\frac{(n-r)\sigma_0^2}{(t_0-r)\sigma_1^2 + (n-t_0-r)\sigma_2^2} - 1\right)\frac{n-2r}{r}$$

which follows the F-distribution with r and n-2r degrees of freedom when corresponding regression coefficients are equal.

4. Results of the Data Analysis

Below are some examples of application of statistical methods to water quality series.

The temporal variability of Water Pollution Index (WPI) is presented in Figures 1-3. Trend analysis of the series of Water pollution index WPI was fulfilled using the Student's t-test. The analysis of statistic t (Table 1) shows that the condition of the rivers has not become better during the last years. The WPI decreased only for the river Okhta which is the most polluted river of the city.

In the present example the statistical procedure is used only for trend detection. Application of statistical methods for estimation of the change point is illustrated by the copper concentration C series (Figures 4 - 6).

The analysis of the plots shows growth of the concentrations in the river Okhta till 1988 and decline of the concentrations after 1988. The same happened with the rivers Chernaya Rechka and Karpovka but in 1989.

Table 1. Trend analysis of Water Pollution Index.

River	Fontanka	Okhta	Moyka	Chernaya rechka	Karpovka	Zhdanovka	Izhora	Slavyanka	Tosna
t	-3.38	4.53	-2.52	-0.62	0.82	-3.15	-0.32	-1.04	-5.16
$t_{5\%}$	-1.94	1.94	-1.94	-1.94	1.94	-1.94	-1.94	-1.94	-1.94
Kind of trend	Increase	Decrease	Increase	Absent	Absent	Increase	Absent	Absent	Increase

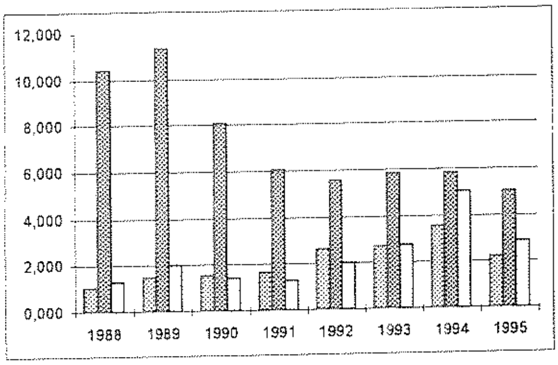

Figure 1. WPI (Fontanka, Okhta, Moyka).

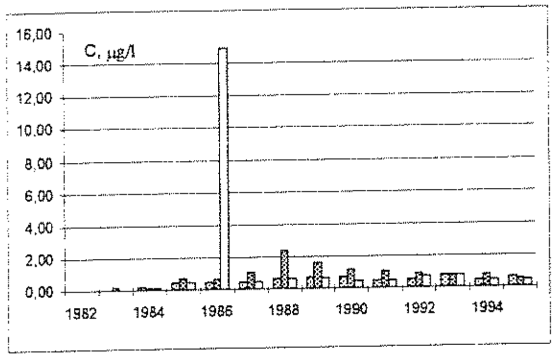

Figure 4. Copper (Fontanka, Okhta, Moyka).

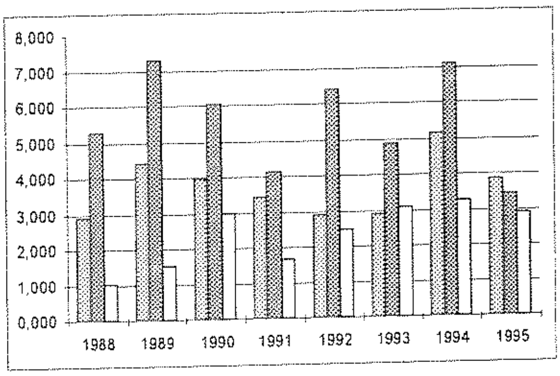

Figure 2. WPI (Chernaya Rechka, Karpovka, Zhdanovka).

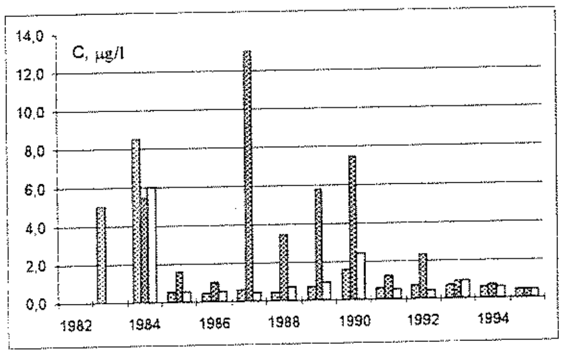

Figure 5. Copper (Chernaya Rechka, Karpovka, Zhdanovka).

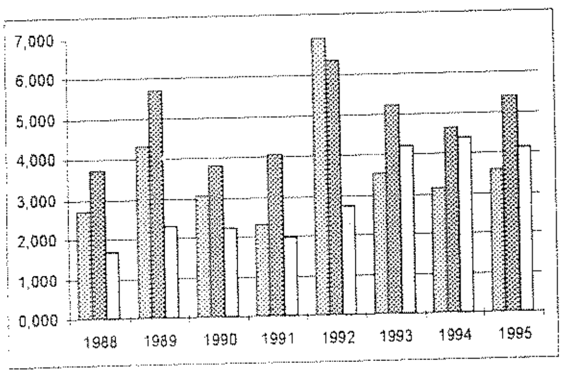

Figure 3. WPI (Izhora, Slavyanka, Tosna).

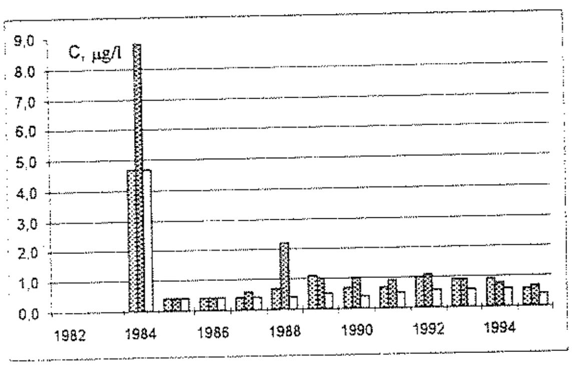

Figure 6. Copper (Izhora, Slavyanka, Tosna).

Analysis of the other rivers needs statistical methods. By the Student's t-test there is no trend in the series of Fontanka and Moyka. The trend is positive in the river Zhdanovka. The concentrations of copper are increasing in the river Tosna as well. In this series the change point can be determined using the formulas (5)-(7). It is seen from Figure 7 that the change of trend took place in 1989.

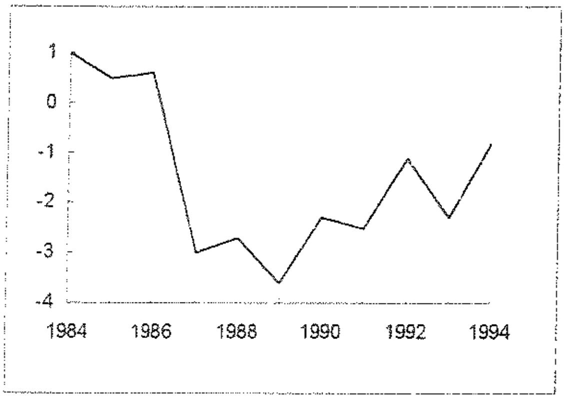

Figure 7. Function S_t for concentrations of copper in the river Tosna.

Below are the results of the data analysis presented in general form. Details of the calculations are omitted.

For the sake of convenience all the water quality parameters are divided into three groups. The first group includes the most toxic substances: phenols and oil products. The second group consists of metals (copper, lead, cadmium and marganese). BOD-5, phosphorus, DO and suspended matter were collected in the third group.

As for the substances from the first group the change of phenols and oil products concentrations are different. Since 1989 the concentrations of phenols have been increasing sharply and now sufficiently exceed the maximum permissible concentration (MPC). In turn the oil products concentrations in the rivers have been lowering during the last years. Moreover, the concentrations of oil products do not exceed the value of MPC.

The growth of the phenols concentrations is linked with the introduction of a new method of control in 1992. So the values corresponding to previous years are probably lowered. Decrease of the oil products concentration is associated with the work of new treatment stations. The main causes are the growth of oil prices and a sharp decrease of industry production due to the economical crisis.

Analysis of metals concentrations shows the following. Concentrations of copper sufficiently exceed the value of MPC (1 μg/l) for all rivers. In addition to previous analysis of the copper series it is necessary to emphasize that extremely high values of concentrations are observed in the rivers Okhta, Chernaya Rechka, Karpovka and Zhdanovka in the end of the 80's and the beginning of the 90's. The permanent growth of the copper concentrations in Fontanka and Moyka is determined by the analogous process in the river Neva because these rivers are branches of Neva.

The concentrations of marganese and cadmium in the rivers Okhta, Chernaya rechka, Karpovka and Zhdanovka in the end of the 80's and the beginning of the 90's were extremely high and exceeded the values of MPC. Now the situation has improved and the concentrations do not exceed the MPC values for all rivers except Okhta. Trends in the series of marganese and cadmium are not detected and the concentrations do not exceed the values of MPC in the rivers Moyka and Fontanka. The concentrations of marganese in Izhora, Slavyanka and Tosna exceed the value of MPC and do not decrease for the last years. The situation with cadmium in these rivers is normal.

The lead concentrations are lower than the value of MPC for all rivers except sharp peaks in 1983-1984 and in the end of the 80's and in the beginning of the 90's in Izhora and Slavyanka.

As for the substances of the third group the situation in general is normal. The phosphorus concentrations do not exceed the value of MPC and trend is absent. DO concentrations are also satisfactory except for the river Okhta where the concentrations are close to the critical values. State of the concentrations of suspended matter is stable and good. The maximum concentrations are in the rivers Okhta and Slavyanka.

5. Conclusions

The present study showed that the considered statistical procedures are effective for trend analysis of water quality series and estimation of change points.

The general characteristic of water quality WPI decreases only for the river Okhta. Analysis of other rivers did not expose the increasing of the WPI. Nevertheless some characteristics of water quality of the rivers decrease. It relates to oil products , the characteristics from the third group and partly to some metals.

Unfortunately, decreasing of some water quality parameters has not lead to lowering of the WPI yet because situation with some dangerous characteristics as phenols has not improved.

Lowering of concentrations of some substances is caused by the reduction of the industry production and only partly by improving of wastewater treatment system. The protection measures taken for reduction of the anthropogenic load on the rivers are not sufficient. Therefore the situation with water quality in the rivers of St.Petersburg is not satisfactory .

The study was supported by the Research Scheme of the Central European University.

6. References

Mann, H.B., Non-parametric test against trend.

Econometrica, 13, 245-259, 1945.
Kendall. M.G., Rank correlation methods. Charles Griffin, London, 1975.
Hirsch, R.M.,J.R. Slack, and R.A.Smith, Techniques of trend analysis for monthly water quality data. *Water Resources Research*, 18 (1),107-121,1982.
Hipel, K.W., A.I McLeod, and R.R. Weiler, Data analysis of water qality time series in lake Erie. *Water Resources Bulletin. American Water Resources Association*. 24 (3), 533-544, 1988.
Loffits, J.C., R.C. Ward,R.D. Phillips, and H.Taylor, An evaluation of trend detection technics for use in water quality monitoring programs. PA/600/3- 89/037, U.S. Environmental Protection Agency, Environmental Research Laboratory, Corvallis, OR., 1989.
Gilbert, R.O., Statistical methods for environmental pollution monitoring. Van Nostrand Reinhold Co., NY., 1987.
Quandt, R.E., The estimation of the parameters of a linear regression system obeying two eparate regimes. *J. Amer. Statist. Assoc.*, 53, 873-880, 1958.
Quandt, R.E. Tests of the hypothesis that a linear regression system obeys two separate regimes. J. Amer. Statist. Assoc., 55, 324-330, 1960.