POLLACK PERIODICA
An International Journal for Engineering and Information Sciences
DOI: 10.1556/606.2017.12.1.2
Vol. 12, No. 1, pp. 17–27 (2017)
www.akademiai.com

A STOCHASTIC APPROACH FOR REGIONAL-SCALE SURFACE WATER QUALITY MODELING

[1] **Máté KARDOS**, [2]**László KONCSOS**

Department of Sanitary and Environmental Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3, H-1111 Budapest, Hungary
e-mail: [1]kardos@vkkt.bme.hu, [2]koncsos@vkkt.bme.hu

Received 6 February 2017; accepted 2 March 2017

Abstract: A methodology is proposed to calculate statistical average and standard deviation of long time water quality parameter series along a river network. The method considers the water network as a graph consisting of straight sessions and junctions. With a Taylor-series approximation, statistical values of an arbitrary point of the network can be calculated from upstream ones without the need to calculate the single downstream values. According to preliminary results of the first calculations on a pilot area, mean value of the downstream biological oxygen demand and the so called ‘transfer coefficient’ can be approximated with a relative accuracy of 10%.

Keywords: Water quality, Water quality monitoring, Water monitoring statistics, Point-source pollution, River monitoring, Water quality modeling

1. Introduction

The quality of waters of Europe remains an important issue despite the introduction of the EU Water Framework Directive (WFD) several years ago [1], [2]. According to the WFD models provide important means of process understanding [3]. However, some existing models are characterized by high data demand, which is difficult to satisfy, while others having less data requirements only reproduce water quality parameters with significant errors [4], [5]. As sampling times do not coincide inside the stream network in most of the cases, the consistent spatial profile for single water bodies can rarely be reproduced. This fact makes mass balance calculations extremely difficult [6], [7]. Nevertheless, many models require some kind of on-site mass-balance data for calibration, so often additional field measurements are conducted to patch the holes of routine monitoring databases [8], which increases the costs of regional-scale modeling. At the same time, there is a lot of useful information in routine monitoring databases [1], [9], [10], [11].

In Hungary, regular, comprehensive surface water quality sampling was launched in 1968 [12]. The 113 most important rivers were monitored in 300 monitoring stations. Sampling was done at least monthly, but on some stations so often as twice a week, with determination of about 50 parameters all over the country [12]. All the data was collected in the Water Quality 2000 (VM2000) database. The basics of this system remained unchanged until implementation of the EU WFD in 2006, when systematic monitoring of smaller reaches and creeks started [13]. At the same time, sampling stations with decade's long series were closed. Since 2006, data are collected in Surface Water Information System (SWIS). In *Fig. 1* a portion of the Hungarian river and monitoring network is visualized in order to give the reader an impression on their density and distribution.

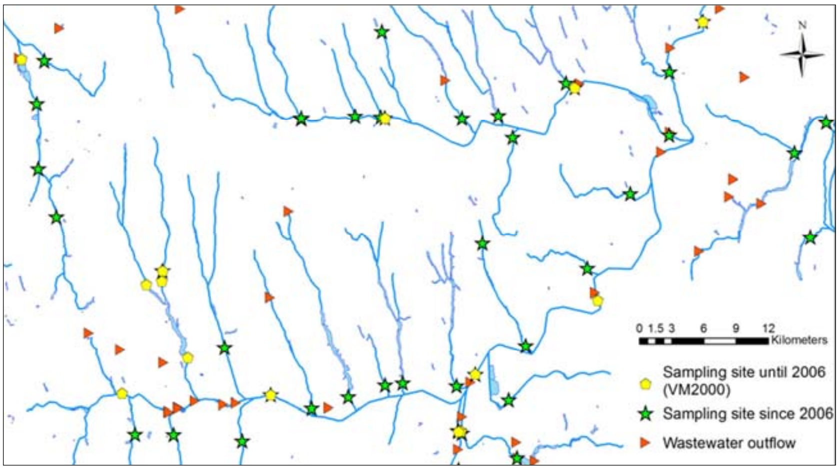

Fig. 1. Portion of the Hungarian river and monitoring network

Current work at the Department of Sanitary and Environmental Engineering, Budapest University of Technology and Economics aims to develop a method suitable for assessing performance of a water quality database. The means is a process-based statistical model that describes the connection between distributions of water quality parameters at junctions of the stream network in order to make predictions for junctions that are not covered by measurements. In a series of papers, it is intended to describe functioning and accuracy of the different elements (edge, node) of the method for several in-stream processes (settling, decay, etc.) and different geometry.

The present paper is the first one in this series, reporting first steps of the model development. A stand-alone river-reach without any sources/sinks in between the two endpoints is investigated. The water quality constituent is considered to follow first-order kinetics with no influence of other constituents. Also, results of testing the method against measured data are reported.

2. Methods

2.1. Water quality model

The model is based on a graph representation of the stream network. Confluences of rivers, wastewater inlets, and sampling stations are represented by graph vertices; stream reaches by edges (see *Fig. 2*). Edges are characterized by their physical properties: width, length, slope, etc. Internal changes (e.g. reactions, decay) are assigned to edges and increment/decrease of discharge and/or pollutant to nodes (supposing instant total mixing inside the node).

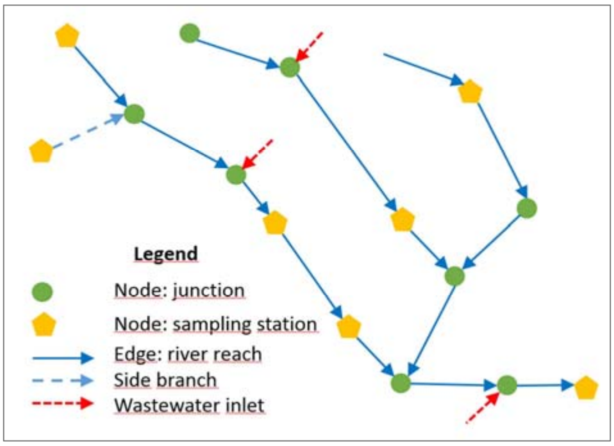

Fig. 2. Graph of an example river network

A single edge is a river reach without any side inflow/outflow but at the endpoints. Assuming first order decay kinetics along the edge, the downstream concentration of a water quality constituent can be calculated from the upstream concentration according to the well-known formula

$$C_1 = C_o \exp(-kt)\,, \tag{1}$$

where C_1 and C_0 are up- and downstream concentrations [g m^{-3}] (e. g. of Biological Oxygen Demand, (BOD)) respectively; k is the first-order decay coefficient; and t is the travel time [d], [14]. For the downstream end of an edge the nonlinear part of Eq. (1) can be compressed into a single transfer coefficient (α [–]):

$$C_1 = \alpha(Q,T)C_o\,, \tag{2}$$

where α can be derived from discharge and water temperature based on (2):

$$\alpha(Q,T) = \exp(\gamma \frac{\Theta^{T-20}}{Q^{0.4}}), \tag{3}$$

where T is the water temperature in [°C]; Q is discharge [m^3 s^{-1}]; Θ is the base number for temperature-dependence (usually Θ = 1.04); γ is a location-dependent calibration constant compressing the effects independent of $Q(t)$ or $T(t)$:

$$\gamma = \frac{-k_{20}\ L\ B^{0.4}}{k_{st}^{0.6}\ I^{0.3}}, \tag{4}$$

where k_{20} is decay coefficient in the 20 °C water; L is length of the reach [m]; B is characteristic width of the channel [m]; k_{st} is bed roughness [$m^{1/3}$ s^{-1}]; and I is mean slope of the reach [-]. Fixing the calibration constant also means supposing all it's components to be constant in time.

2.2. Statistical model

The objective was to develop a method for transforming statistics (mean and variance) of water quality parameters measured in a certain node to non-monitored ones.

As a first step, was intended to calculate the mean and variance of the transfer coefficient from the mean and variance of Q and T. Deriving long-term average of the 2nd-order Taylor-approximation for Eq. (3) around the mean values $(\bar{Q}, \bar{T})$ yields

$$\begin{aligned}\bar{\alpha} \approx\ & \alpha(\bar{Q},\bar{T}) + \frac{1}{2}\frac{\partial^2 \alpha(\bar{Q},\bar{T})}{\partial Q^2}\cdot\sigma^2(Q) + \frac{1}{2}\frac{\partial^2 \alpha(\bar{Q},\bar{T})}{\partial T^2}\cdot\sigma^2(T) \\ & + \frac{\partial^2 \alpha(\bar{Q},\bar{T})}{\partial Q \partial T}\cdot \operatorname{cov}(Q,T)\end{aligned} \tag{5}$$

and

$$\begin{aligned}\sigma^2(\alpha) \approx\ & \left(\alpha(\bar{Q},\bar{T}) - \bar{\alpha}\right)^2 + \frac{1}{2}\frac{\partial^2 f(\bar{Q},\bar{T})}{\partial T^2}\cdot\sigma^2(T) + \frac{1}{2}\frac{\partial^2 f(\bar{Q},\bar{T})}{\partial Q^2}\cdot\sigma^2(Q) \\ & + \frac{\partial^2 f(\bar{Q},\bar{T})}{\partial Q \partial T}\cdot \operatorname{cov}(Q,T),\end{aligned} \tag{6}$$

where

$$f(Q,T) = \left(\alpha(\bar{Q},\bar{T}) - \bar{\alpha}\right)^2, \tag{7}$$

σ is the standard deviation and the stroke above any letter means mean value throughout this paper.

According to Eqs. (2) and (3), there is a time-variable nonlinear transformation of concentration between the upstream and downstream nodes. This transformation distorts the distribution of C_0 into the distribution of C_1. Based on Eq. (2) the mean of C_1 is precisely:

$$\bar{C}_1 = \bar{C}_0 \cdot \bar{\alpha} + \mathrm{cov}(C_0, \alpha) . \tag{8}$$

If C_0 and α are independent, this simplifies to

$$\bar{C}_1 = \bar{C}_0 \cdot \bar{\alpha} , \tag{9}$$

which is the basic approximation.

Due to the nonlinear transform, there is no precise solution for the variance of C_1, but there is a possibility to use the 2nd-order approximation again:

$$\sigma^2(C_1) \approx (\bar{\alpha}\,\bar{C}_0 - \bar{C}_1)^2 + \bar{\alpha}^{\,2} \cdot \sigma^2(C_0) + \bar{C}_0^{\,2} \cdot \sigma^2(\alpha) + (4\,\bar{C}_0\,\bar{\alpha} - 2\,\bar{C}_1) \cdot \mathrm{cov}(C_0, \alpha) . \tag{10}$$

Again assuming independence of C_0 and α yield to:

$$\sigma^2(C_1) \approx (\bar{\alpha}\,\bar{C}_0 - \bar{C}_1)^2 + \bar{\alpha}^{\,2} \cdot \sigma^2(C_0) + \bar{C}_0^{\,2} \cdot \sigma^2(\alpha) . \tag{11}$$

In contrast to Eq. (9), Eq. (11) was built on 2 stages of approximations, first the 2nd-order Taylor-approximation and then the assumption of independence of C_0 and α. Accordingly, the error associated with Eq. (11) is expected to be larger than the error of Eq. (9). The magnitude of these errors was tested in the case study and was decided if the proposed approach is suitable for making predictions within an entire graph of the stream network.

3. Case study

3.1. Site description

To test the model, a river reaches monitored both at the inflow and the outflow point is needed. The outflow point sample series are needed to test the above equations' accuracy. According to the VM2000 database, only a few reaches, satisfying the above requirements, exist. One of them is a reach of the Kapos River between Dombóvár and Kurd (Western Hungary). The Dombóvár WasteWater Treatment Plant (WWTP) discharges 16 000 population equivalent, i. e. about 2000 m³ d^{-1} treated wastewater into the Kapos River 1.4 km upstream of the Dombóvár sampling site (see *Fig. 3*). Mean BOD concentration of the WWTP effluent is about 30 g m^{-3} (data from year 2007 - it used to be much higher in last decades of the 20th century). Yearly mean discharge of the Kapos increases by 20% (ranging from -6% to 75%, based on bi-weekly data) along this section due to base flow and the smaller side reaches (*Fig. 3*). Although this increment is not negligible, in the calculations it was disregarded. From the many water quality parameters only BOD was examined in this phase of the research.

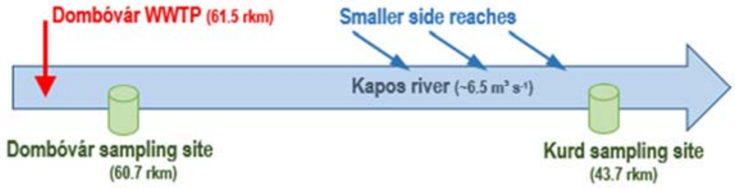

Fig. 3. River reaches and sampling stations at the pilot area

3.2. Input data

In the tests data from two sampling stations was used: 1. Dombóvár and 2. Kurd. Following parameters were involved in the calculations:

1. Discharge time series [m^3 s-1];
2. Water temperature time series [°C];
3. BOD time series [g m^{-3}].

Besides, transfer coefficient α has been calculated for each date after fixing the calibration constant γ of Eq. (3).

The original database contained 465 and 597 records in the period of 1968–1980 and 1994–2006 for Dombóvár and Kurd, respectively. Data from 1981–1993 were missing from the database. BOD values after 1994 were significantly lower than before 1980 (annual means decreased from 10.5-38.2 g m^{-3} to 4.6-6.3 g m^{-3} in Dombóvár and from 6.7-24.8 to 3.9-5.5 in Kurd). The discharge measurement was missing for some dates; however, only dates with all discharge, temperature and BOD concentration available were taken into account.

Following six (sub)series of data were generated: St is the test series; years 1968 and 1994:

S0: Full series;
S1: Series consisting of data from odd years;
S2: Series consisting of data from even years;
S3: Series consisting of data from years dividable by 3;
S4: Series consisting of data from years with remainder of 1 after division by 3;
S5: Series consisting of data from years with remainder of 2 after division by 3.

With this process independent pairs/triplets of subseries were produced while ruling out effects of systematic changes. Since S2 is independent from S1, and S4 and S5 are independent from S3, they are suitable for validation. All 6 series were divided into two periods: 1968 - 1980, and 1994 - 2005, and the whole time period (1968 - 2005) was also investigated. The number of records in the various data series and periods are summarized in *Table I*.

Table I

Number of records in the various data series and periods

	1968-1980		1994-2005		1968-2005	
	Dombóvár	Kurd	Dombóvár	Kurd	Dombóvár	Kurd
St	13	12	24	26	37	38
S0	157	296	293	286	450	582
S1	75	144	148	144	223	288
S2	82	152	145	142	227	294
S3	60	108	99	91	159	199
S4	51	93	94	92	145	185
S5	46	95	100	103	146	198

3.3. Calculations

The sequence of calculation is:

1. Determination of $\bar{\alpha}$ and $\sigma^2(\alpha)$ from Eqs. (5) and (6), respectively;
2. Determination of $\overline{C_1}$ from Eq. (8) or (9);
3. Determination of $\sigma^2(C_1)$ from Eq. (10) or (11);
4. Comparison of modeled and measured statistics of α and C_1.

The measurements for C_1 (downstream sample series) allowed to calibrate the parameter γ and control it's value calculated based on literature and real data.

4. Results and discussion

4.1. Calibration constant γ

Considering our series of equations as a model, the parameters of model calibration are constants γ and Θ of Eq. (3). According to literature data, Θ varies between 1.0,…,1.1; BOD decomposition is usually calculated with Θ=1.04 or Θ=1.047 [14], [15]. Θ=1.04. was set.

Substituting literature (k_{20} 0.35 d^{-1}, k_{st} 10 $m^{1/3}$ s^{-1}) and true geometrical data (L=17.0 km, B=20 m, I=0.04 %) into Eq. (4), it was found that γ=-0.6 $m^{1.2}$ $s^{-0.4}$. However, the above values were slightly modified to get an insight on the uncertainty. In *Table II*, the range of values for the calibration constant γ is shown.

Table II

Calculated and calibrated values for γ [$m^{1.2}$ $s^{-0.4}$]

γ	Minimal	Average	Maximal
Substituting literature and true geometrical values into Eq. (4).	-3.2	-0.6	0.0
Calibration	-1.5	-0.5	-0.2

4.2. Transfer coefficient α

The mean value of the transfer coefficient ($\bar{\alpha}$) can be calculated via Eq. (5). According to pilot calculations, the error caused by the approximation is always less than 11%; (2.4% in average). According to our investigations on the particular pilot area, there is a certain correlation between temperature and flow: low waters happen usually during summer. The correlation coefficient for the data series was usually around -0.2, ranging from -0.6 to 0. Still the weight of the covariance term in Eq. (5) is less than 1%, so there is no big advantage in calculating $\mathrm{cov}(Q,T)$.

Investigating Eq. (6) the following conclusions can be drawn. The covariance term here has a higher proportion (<12%), and the relative error for $\sigma(\alpha)$ ranged between 12–49% (32% in average), which is not negligible.

4.3. Outflow BOD concentration C_1

According to the assumptions behind Eq. (9) the dependence between C_0 and α might be negligible. For the data series used in the case study, the correlation between C_0 and α was between -0.08 and 0.36. The $\mathrm{cov}(C_0,\alpha)$ was between -0.26 and 0.17, so for the test dataset the cost of using Eq. (9) instead of Eq. (8) was maximally 2% relative error in $\overline{C_1}$.

Calibration and verification happened as follows:

- calibration with series S1, verification with series S2; or
- calibration with series S3, verification with series either S4 or S5.

The measured and calculated values can be found in *Table III*. The difference between calculation and measurement is below 10% except for one case. In that case the mean outflow BOD concentration was below 50% of the inflow one (the usual value was between 60-70%). This might be due to the time shift between upstream and downstream sampling in that particular series of years. As can also be seen, accuracy of the model decreases with decrease of the sample amount.

Table III

Measured (m) and calculated (c) values of mean outflow BOD concentration ($\overline{C_1}$) [g m^{-3}] and the difference between calculated and measured values in [%]

	S2			S4			S5		
	m	c		m	c		m	c	
1968-1980	11.85	11.37	(-4%)	10.38	16.68	(+38%)	15.28	17.03	(+10%)
1994-2005	4.53	4.83	(+6%)	4.83	4.92	(+2%)	4.57	4.69	(+3%)

For the calculation of the outflow BOD concentration's standard deviation according to Eq. (10), $\mathrm{cov}(C_0,\alpha)$ is needed, meaning that the transfer coefficient has to be calculated for each sample. This is in contradiction with the intention of using statistical parameters instead of unique values. Comparing the weight of unique terms of Eq. (10) to the exact value of $\sigma^2(C_1)$, it was found that the error of the approximation is below

10%, and that neglecting the covariance term would make an additional error of less than 6% for the test series. This means, that the error caused by using Eq. (11) instead of Eq. (10) was less than 16%. Both errors decreased with increasing series length.

The performance of the proposed method in estimating the standard deviation of the measured outflow BOD values was variable. In *Fig. 4* calculated σ values as function of measured values' deviation were plotted.

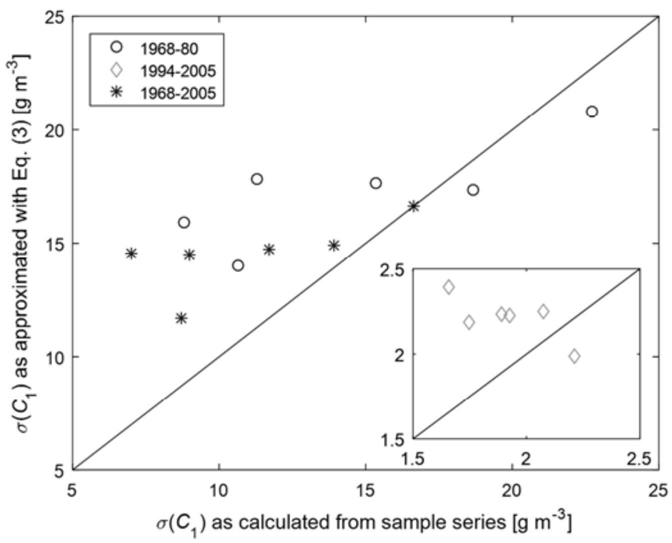

Fig. 4. Approximated standard deviation versus standard deviation of measured downstream concentration

According to calculations on the test series, Eq. (10) overestimates $\sigma(C_1)$ by about 30% compared to the measured std. deviation, the (over)estimation value ranging from -10 to 110%. Again, larger errors appear in the shorter series of S3-S5.

Summarizing, both mean values ($\bar{\alpha}$ and $\overline{C_1}$) are approximated with satisfying accuracy. However, approximation of both standard deviations ($\sigma(\alpha)$ and $\sigma(C_1)$) contains considerable errors. In general, errors can be caused by

- neglection of the covariance term in the 2nd-order Taylor-series;
- neglection of higher order terms of the Taylor-series;
- time-shift in upstream and downstream sampling dates.

In this paper, the first possibility was eliminated: it was shown, that neglecting of the covariance term does not lead to major errors. Thus, further investigation is needed to locate the errors learned.

5. Conclusions

A statistical method is proposed to calculate statistical average and variance of water quality parameter series along stream sections. With a Taylor-series approximation downstream values can be calculated from upstream ones. First tests of this novel analysis tool on the VM2000 gave promising results. Generally, the suggested simplifications did not cause unacceptable errors for the case study. Mean and variance of the downstream BOD and the transfer coefficient could be approximated with an accuracy of +/-10% in most cases. The accuracy can be potentially improved with the estimation of higher order terms of the Taylor-series.

Acknowledgements

The work presented in this paper was partially funded by the ENVISHEN project of the Hungarian National Development Agency.

References

[1] Chapman D. V., Bradley Ch., Gettel G. M., Hatvani I. G., Hein Th., Kovács J., Liska I., Oliver D. M., Tanos P., Trásy B., Várbíró G. Developments in water quality monitoring and management in large river catchments using the Danube River as an example, *Environmental Science & Policy*, Vol. 64, 2016, pp. 141–154.

[2] Fetter É., Sándor D., Fleit E. Implementation of a monitoring system on the ecological effects of anthropogenic interventions on the Danube morphology for the Gemenc-Béda-Karapancsa floodplain, *Pollack Periodica*, Vol. 5, No. 1, 2010, pp. 151–162.

[3] Tsakaris G., Alexakis D. Water quality models, An overview, *European Water*, Vol. 37, 2012, pp. 33−46.

[4] Tango P. J., Batiuk R. A. Chesapeake Bay recovery and factors affecting trends,: Long-term monitoring, indicators, and insights, *Regional Studies in Marine Science*, Vol. 4, 2016, pp. 12–20.

[5] Cox B. A. A review of currently available in-stream water-quality models and their applicability for simulating dissolved oxygen in lowland rivers, *The Science of the Total Environment*, Vol. 314-316, 2003, pp. 335−377.

[6] Jolánkai Zs., Koncsos L. Application of an integrated hydrological model at the southern Trans-Tisza region of Hungary as part of a conceptual phosphorus transport model framework, *Pollack Periodica*, Vol. 10, No. 2, 2015, pp. 133–144.

[7] Myers D. N., Ludtke A. S. Progress and lessons learned from water-quality monitoring networks, In Ahuja S. (ed.) *Chemistry and water, The science behind sustaining the world's most crucial resource,* Ch. 2, Elsevier, NY, 2017.

[8] Behmel S., Damour M., Ludwig R., Rodriguez M. J. Water quality monitoring strategies, A review and future perspectives, *Science of the Total Environment* ,Vol. 571, 2016, pp. 1312–1329.

[9] Trowbridge P. R., Davis J. A., Mumley T., Taberski K., Feger N., Valiela L., Ervin J., Arsem N., Olivieri A., Carroll P., Coleman J., Salop P., Sutton R., Yee D., McKeea L. J., Sedlak M., Grosso C., Kelly J. The regional monitoring program for water quality in San Francisco Bay, California, USA, Science in support of managing water quality, *Regional Studies in Marine Science*, Vol. 4, 2016, pp. 21–33.

[10] Somlyódy L., Pintér J., Koncsos L., Hanacsek I., Juhász I. Estimating averages and detecting trends in water quality data (lake Balaton, Hungary), *Proceedings of the Budapest Symposium on Monitoring to Detect Changes in Water Quality Series*, Budaopest, Hungary, 2-4 July 1986, publ. No. 157, pp. 61−69.

[11] Hatvani I. G., Kovács J., Márkus L., Clement A., Hoffmann R., Korponai J. Assessing the relationship of background factors governing the water quality of an agricultural watershed with changes in catchment property (W-Hungary), *Journal of Hydrology*, Vol. 521, 2015, pp. 460–469.

[12] Hock B., Schneider J. Storage and processing of the results of water quality investigations of streamflows, (in Hungarian) *Vízügyi Közlemények*, Vol. LXVIII, No. 2, 1986, pp. 260−268.

[13] Gaál E. Chemical qualification of surface waters, (in Hungarian) *Proceedings of the XXXIV. Country-wide Symposium of the Hungarian Hydrological Society*, Debrecen, Hungary, 6-8 July 2016, pp. 1-10.

[14] Chapra S. C. *Surface water-quality modeling*, McGraw-Hill, Singapore, 1997.

[15] Thomann R. V., Mueller J. A. *Principles of surface water quality modeling and control*, Harper &Row, NY, 1997.