International Congress on Environmental Modelling and Software

5th International Congress on Environmental Modelling and Software - Ottawa, Ontario, Canada - July 2010

Jul 1st, 12:00 AM

Efficiency Criteria for Water Quality Monitoring

Marina G. Erechtchoukova

Peter A. Khaiter

International Environmental Modelling and Software Society (iEMSs)
2010 International Congress on Environmental Modelling and Software
Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada
David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.)
http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

Efficiency Criteria for Water Quality Monitoring

Marina G. Erechtchoukova[1] and Peter A. Khaiter[1]

[1]*York University, Toronto, Canada (marina/pkhaiter@yorku.ca)*

Abstract: The issues of possible improvements, increased efficiency and/or optimization of a monitoring system in general, and a monitoring design in particular, are urgent. Since monitoring activities are always limited by financial and logistics constraints, algorithms of constrained optimization are deemed more suitable for this purpose. Monitoring designs are developed as solutions of an operation research model. In order to formulate such model the effectiveness function has been introduced. The effectiveness reflects the extent to which a monitoring design meets the objectives of the monitoring program and can be used for comparison of different monitoring designs. The effectiveness function depends on the investigated water quality parameters, selected indicators of water quality and their estimators. The function properties suggest the selection of an optimization algorithm. The proposed approach has been applied to a case study in order to develop temporal monitoring designs. It has been shown that the designs differ significantly only when the levels of the effectiveness are high. With the effectiveness of 80% or less the designs for different water quality parameters and the same indicator can be compromised. Since monitoring data are usually used for various purposes, the preference should be given to simple monitoring designs or to the designs which support efficient reconstruction of chemographs of investigated water quality parameters

***Keywords**:* Monitoring design; Water quality; Effectiveness; Constrained optimization.

1. INTRODUCTION

Monitoring data reveal important information about the current status of the aquatic environment and the anthropogenic impact water ecosystems are exposed to. Collected periodically at the same locations, these data can be used for projections and evaluation of consequences of management decisions. Monitoring systems have a complex infrastructure supporting all of their sampling and data processing activities. Monitoring systems comprise several components related to different aspects of their functioning. These components include the collection and analysis of physical, chemical, and biological data, as well as quality assurance and control programs which ensure that the data are scientifically valid. A key component of a monitoring system is its network of sampling sites where water quality observations and measurements are conducted. Monitoring activities are always limited by financial and logistics constraints. At the same time, the systems must provide data sufficient for a wide range of scientifically valid conclusions. Therefore, the issues of possible improvements, increased efficiency and/or optimization of a monitoring system in general, and a monitoring design, in particular, are urgent.

Multidisciplinary nature of a monitoring system explains different and sometimes contradicting aspects of the system which must be taken into account during its optimization. It justifies various approaches which are aimed to replace intuitive improvements of monitoring systems in order to increase their efficiency. Thus, the problem of monitoring optimization can be solved as a multi-objective mixed integer programming model with constraints [Ning and Chang 2002], or as a constrained optimization using generic algorithms [Icaga 2005; Cieniawski et al. 1995]. Fuzzy optimization approach is also used in optimization of a water quality monitoring network [Ning and Chang 2002].

The current study investigates the approaches to optimization of monitoring networks and, particularly, the development of temporal monitoring designs for water quality sampling. Application of formal optimization algorithms requires formulation of an objective function which reflects the goal of optimization and at the same time allows for quantitative comparison of different monitoring designs. Mathematical properties of the objective function restrict or even determine computational algorithms to be applied to derive monitoring designs which are at least satisficing. Hence, the articulation of an objective function affects the designs deemed to be efficient. Monitoring designs are considered improved when they either are more efficient than other known designs under the same, usually limiting, conditions or the designs have the minimal cost compared to all known alternatives with an acceptable level of efficiency. In both articulations, it is necessary to evaluate the efficiency of a monitoring design. The paper considers different criteria of efficiency of water quality temporal monitoring designs and their implications to suggested sampling programs. Although there are other important aspects in optimization of a monitoring program such as location of sampling site, analytical or sampling methods, they are not taken into account and the study is focused on sampling frequencies.

2. MONITORING FRAMEWORK

Groot and Schilperoort [1983] proposed a framework for developing an efficient monitoring network. The framework consists of five main steps: (1) to identify monitoring objectives; (2) to identify water quality indicators and relevant processes in order to choose approaches to data analysis; (3) to identify sampling programs and determine the effectiveness of a monitoring network; (4) to obtain cost estimates of the monitoring network; and (5) to implement cost-effectiveness analysis. This framework is very generic and can be applied to create an entire monitoring program or its part.

The development of a water quality monitoring system is significantly guided by monitoring objectives formulated for a waterbody. Whitfield [1988] specified the following five general objectives: (1) assessment of trends in variables (i.e. values of water quality parameters) of interest; (2) attainment of water quality standards; (3) estimation of mass discharge; (4) assessment of environmental impact; and (5) general surveillance. In addition to these objectives, monitoring data are also required for specific project and management needs. Each objective implies specific computational procedures employed to derive important information from a set of collected data.

The second step also relies on monitoring objectives since the latter determine important characteristics of selected water quality indicators required for decision making. For example, implementing the total maximum daily load process, it is important to know not only concentrations of water quality constituents and violations of water quality standards, but also the duration and magnitude of the violation [Shabman and Smith 2003]. Monitoring objectives dictate sampling programs. Thus, trend detection requires sampling of selected water quality indicators with a fixed frequency at the same location and at the reference site [Lettenmaier 1978]. Attainment of water quality standards can be investigated by sequential sample collection when the number of observations is determined based on the outcome of observations as they have been made [Whitfield 1988]. Estimations of mass discharge require sampling programs which take into account properties of selected estimators [Robertson & Roerish 1999; Erechtchoukova & Khaiter 2009].

The effectiveness of a monitoring program should quantitatively express the extent to which monitoring results meet the objectives. Although monitoring systems generate series of data, the main outcome of a monitoring program is information which supports decision making in accordance with the monitoring objectives. This information is usually obtained via processing monitoring data using simple or complex models. Hence, the evaluation of the effectiveness of a monitoring program should include mathematical models transforming observation data into information. It justifies the role of models in developing efficient monitoring designs [Erechtchoukova and Khaiter 2008].

Since monitoring systems operate under financial constraints, cost estimates of a monitoring system are important. They allow for straightforward comparison with available budget and between different monitoring programs. Although Groot and Schilperoort [1983] considered the process of deriving the financial estimates simply as a technical exercise, absolute values of the estimates or their components are not always available. At the same time, the cost of a monitoring program definitely depends on the number of collected samples.

If the effectiveness of a monitoring program could also be expressed in monetary form, the cost-effectiveness analysis can be done based on a direct comparison. However, such estimates are usually not available. In this case, the comparison can be replaced by constrained optimization techniques which do not require expressing a goal function and constraints in the same measuring units, although both the goal function and constraints must have common variables.

3. OPERATION RESEARCH MODEL

The constrained optimization approach can be applied to the development of an efficient monitoring network. For this purpose, it is necessary to define the goal of optimization. With respect to the framework mentioned above there are two possible articulations of the problem of the development of efficient monitoring design: (1) to maximize the effectiveness of the design under limited budget and (2) to minimize the cost of the design within an acceptable level of effectiveness. In both cases it is necessary to provide quantitative estimates of the cost and the effectiveness of a monitoring design.

Although the cost of a monitoring network comprises various components which may not be independent, it is reasonable to assume that the cost of a monitoring network increases monotonically with the number of samples collected at the monitoring sites and the number of sites. This assumption validates the replacements of a cost estimate by the total number of design samples in an operation research model for the development of a monitoring design. This assumption also leads to articulation (2) of the problem of an efficient water quality monitoring design, since articulation (1) requires explicit estimates of the cost of a monitoring design.

Monitoring objectives impose additional requirements on data sets used for data analysis. Thus, trend detection requires data collected with a fixed sampling frequency at the same observation sites for a long period of time. Attainment of water quality standards can be checked using different schemes including fixed frequency, sequential or Markov sampling. For mass transport estimation, simple random sampling or stratified random sampling are preferable. The designs supporting environmental assessment are very project specific. They must be compliant with the type of analysis employed for estimation of the effects of an investigated project. In all the cases, the designs are supposed to provide reliable estimates of selected environmental indicators.

The number of collected samples with additional requirements about sample distribution determines the extent, to which derived data reflect water quality conditions and hence affects the effectiveness of a network. Formalizing the effectiveness as a function of the number of observations ties together cost and effectiveness of the network. Then, the problem can be described by the following operation research model:

$$\min n \quad \text{subject to} \tag{1}$$

$$E(n) \geq V \tag{2}$$

where n is the number of required observations in a monitoring design, $E(n)$ is the effectiveness of the design, and V is the acceptable level of the design effectiveness.

Following the theory of designs of experiments [Fisher 1971], information derived from the observations is a reciprocal of the variance of an estimator which in its turn depends on the number of observations used in estimation. This implies that information increases infinitely when accurate estimates are obtained. Such representation may restrict the set of automated procedures which can be used to find a solution. Since observations of many water quality parameters cannot be done continuously, daily sample collection is considered as a design providing the best possible estimates with zero error. Under this assumption, the effectiveness can be expressed via variance of an estimator used in the assessment in the following way:

$$E(n) = (1 - D(I(n)) / \overline{I(n)}) \cdot 100\%, \qquad (3)$$

where I is the selected estimator, $\overline{I(n)}$ is its estimate on a set of n observations and $D(I)$ is its variance. The analytical expression of $D(I)$ depends on the selected estimator I which can be a regular statistics or a simulation model describing the selected water quality indicator.

.

It is worth noting that the effectiveness of the design may depend not only on the total number of observations, but also on the temporal and/or spatial distribution of sampling points. Only temporal monitoring designs for data collection at a given site will be considered further. An efficient distribution of samples over an investigated period depends on an indicator of water quality, seasonal variations in values of water quality parameters and the type of the function chosen to describe the effectiveness of the designs.

While formula (3) is based on the variance of the selected estimator, another approach appraises the effectiveness of a monitoring design by evaluating the extent to which the designs allow to reproduce water quality indicator's dynamics for a period of observations [Erechtchoukova et al. 2009]. Monitoring data are used to interpolate missing values of water quality parameters and after that an estimate of an investigated water quality indicator is calculated. The relative difference between the estimate and actual value of the indicator evaluates the uncertainty of the estimate. The latter can be used to measure the effectiveness of the design:

$$E(n) = \left(1 - \frac{|I_c - I_a|}{I_a}\right) \cdot 100\%, \qquad (4)$$

where I_c is the indicator estimate calculated using interpolation, and I_a is the actual value of the selected indicator.

According to formulae (3) and (4), the effectiveness can range from -∞ to 100%. The upper boundary is achieved when the design includes all possible observation points. In order to take into account particular formulae for the indicator I, operation research models (1),(2) and (3) or (1),(2) and (4) must be transformed further. Formulae determine to a great extent computational algorithms used to obtain a solution. Quantification of the efficiency based on formula (3) can result in an analytical expression for calculation of the required number of observations. Then the monitoring design contains sampling points randomly distributed over the segments of the investigated period. However, search techniques may give a set of certain points when the data must be collected in order to achieve the desired level of the effectiveness. Since data collection at specified dates is hardly possible, the formula (4) has been modified to allow for data collection during a certain interval or 'sliding window' around proposed dates [Erechtchoukova et al. 2009]:

$$E(n) = \left(1 - \max_{sw}\left(\frac{|I_c - I_a|}{I_a}\right)\right) \cdot 100\% \ , \qquad (5)$$

where *sw* is the size of the sliding window (i.e., the number of days) around a proposed date when a sample can be collected. Formulae (3) – (5) do not exhaust all possible representations of $E(n)$. They have been chosen from perhaps the infinity of alternatives due to their simplicity and mathematical properties allowing for application of automated procedures. The criterion of effectiveness (3) has been investigated with respect to the basic statistics of individual water quality parameters: chloride ions, hydrocarbonate ions and total dissolved solids observed at the same sampling site. The results of investigation have been compared with the criterion (5) studied in [Erechtchoukova et al. 2009].

4. CASE STUDY

Given general monitoring objectives, the average annual concentration and the total annual chemical load have been chosen as water quality indicators of interest. Both indicators were evaluated using the population mean and stratified mean estimators. The series of values of instantaneous chemical load were calculated as the multiplication of values of concentrations and corresponding water discharge. In order to verify developed monitoring designs, very detailed monitoring data are required. It was assumed that daily values of concentrations of the selected water quality parameters and water discharges at the selected site would be sufficient to obtain accurate estimates. The data were collected at the cross-section of the Vyatka River near the town of Vyatskiye Polyany. Two years have been selected for calculation: Year_1 with the unimodal type of hydrograph (only spring-summer high flow events) and another one Year_2 with the bimodal type of hydrograph (high flow events took place twice in spring-summer and late fall). Both hydrographs exhibit distinct hydrological seasons with sharp rising and falling limbs for spring-summer high flow events [Erechtchoukova and Khaiter 2007]. The average annual water discharge is estimated about 22.6 km^3. Chemographs of the selected water quality parameters in Year_2 are presented in Figure. 1.

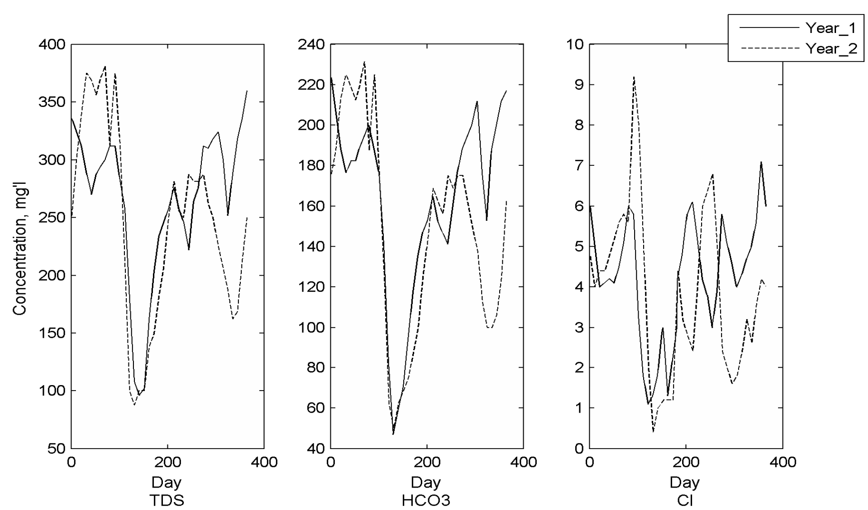

Figure 1. Investigated chemographs of the selected water quality parameters.

The observed concentrations in different samples varied from 0.4 to almost 9.0 mg/l for chloride ions, from 47.0 to 231.0 mg/l for hydrocarbonate ions and from 88 to 381 mg/l for total dissolved solids. The last two water quality parameters exhibited similar variability in both years (the coefficients of variation were about 0.25 in Year 1 and 0.34 in Year 2), while the coefficient of variation for chloride ions had higher values (0.34 in Year 1 and 0.51 in Year 2). Time series for all investigated parameters demonstrated correlation with water discharge values indicating that the dilution process affects concentrations of the investigated constituents to a great extent. The correlation coefficient was higher for total dissolved solids and hydrocarbonate ions (about -0.8) relatively to chloride ions (about -0.6). The correlation coefficients varied from year to year. This means that hydrological conditions change the contributions of individual processes underlying the formation of the constituent concentrations in the water column. The strong correlation between the series of concentrations of total dissolved solids and hydrocarbonate ions showed that the latter represent a significant portion of the total dissolved solids.

Monitoring designs have been developed for evaluation of the annual average concentration and the total annual chemical load of selected water quality parameters with different levels of effectiveness using basic and stratified estimators. The summary of non-stratified designs is presented in Table 1. In this case, samples are randomly distributed within investigated period of time.

Table 1. Simple random designs for estimation of the annual average concentration (*C*) and the total annual load (*L*).

Year	Acceptable level of the design effectiveness (*V*), %	The total number of required observation (*n*)					
		Cl		HCO_3		TDS	
		C	*L*	*C*	*L*	*C*	*L*
Year_1	95	120	274	85	244	71	253
	90	40	157	26	122	21	131
	85	19	92	12	67	10	73
	80	11	58	7	41	5	45
	75	7	39	4	27	4	30
Year_2	95	169	147	121	240	118	239
	90	65	52	40	117	39	118
	85	32	25	19	64	18	64
	80	19	15	11	39	11	39
	75	12	10	7	26	7	26

Seasonal variations in water discharges and concentrations of the investigated water quality parameters suggested the application of stratified estimates. As it has been shown previously [Erechtchoukova and Khaiter 2007; Erechtchoukova and Khaiter 2009], stratified estimators require a lesser number of observations to achieve the same effectiveness of a monitoring design. Table 2 presents the summary of stratified designs for the Year_2. The temporal stratification has been implemented according to different criteria in order to reduce the variance of the population means. Thus, for a single water quality parameter and its two indicators, the three factors can be considered: the water discharge, concentrations, and instantaneous loads.

Table 2. Stratified random designs for evaluation of the annual average concentration (*C*) and the total annual load (*L*) in Year_2

Stratification criterion	Acceptable level of the design effectiveness (*V*), %	The total number of required observation (*n*)					
		Cl		HCO_3		TDS	
		C	*L*	*C*	*L*	*C*	*L*
Water discharge	95	125	174	119	142	120	137
	90	60	82	47	59	50	61
	85	35	46	24	31	26	34
	80	21	29	15	20	15	21
Concentration	95	90	219	60	226	53	231
	90	37	107	21	125	18	124
	85	21	61	10	75	9	74
	80	12	38	5	48	4	48
Instantaneous load	95	245	72	140	99	146	99
	90	135	21	63	33	71	34
	85	84	10	34	16	40	15

	80	55	5	21	9	25	10

The temporal stratification was implemented manually and can be considered only as a "good enough" solution. The series of concentrations, instantaneous loads and water discharges have been divided into four strata according to the values of the selected criteria. The exact strata boundaries affect the designs to a great extent. The identified strata are both constituent and indicator dependent. However, they can be used to investigate criteria of effectiveness (3) and (5).

5. DISCUSSION

Monitoring designs presented in Tables 1 and 2 have been developed using model (1), (2), and (3). These designs are sufficient to achieve established levels of effectiveness by providing accurate estimates of the selected water quality indicators. They vary significantly with respect to the total number of observations and required observations per each stratum. From statistical perspective, the deviations can be explained by different variability of the investigated water quality indicators. Such variability depends on water quality parameters and mathematical properties of the selected estimators. As it was expected, the largest numbers of observations are required for estimation of chloride ions indicators, since it is the most varying water quality parameter in the study. Surprisingly, efficient estimation of the total annual load of chloride ion in Year_2 requires a lesser number of observations than those needed for the annual average concentration. It can be explained by the fact that the water flow diminished variations in chloride ion concentrations. In general, stratified estimators are more efficient for this type of hydrological and hydrochemical conditions. However, an efficient stratification must satisfy certain conditions, and strata boundaries are constituent, indicator and site specific [Erechtchoukova and Khaiter 2007]. It is worth noting that temporal stratified designs create additional obstacles in data collection since they require switching the frequencies of sample collection over an investigated period. The stratum boundaries must be determined *a priori*, whereas dates of typical hydrological and hydochemical events vary from year to year making the stratification very inaccurate and the corresponding recommendations hard to follow due to a human factor.

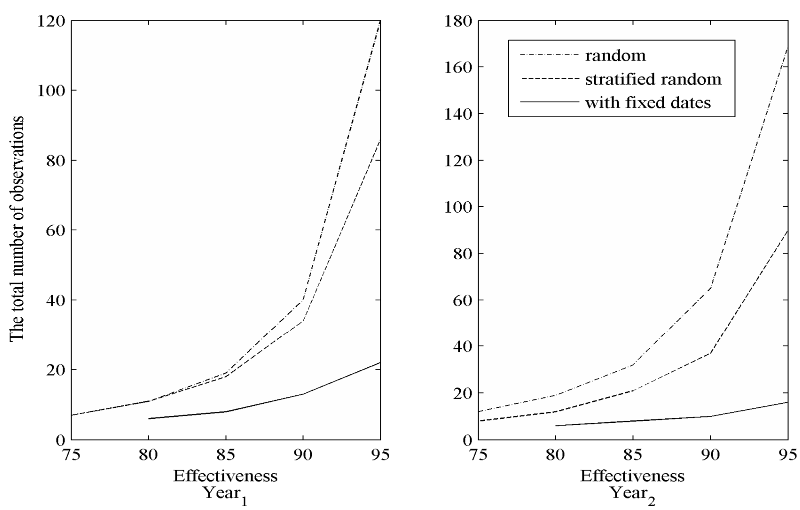

Figure 2. The total number of observation required for estimation of the annual average concentration of chloride ions versus the levels of effectiveness

For chloride ions, the designs developed from the operation research model (1), (2), and (3) have been compared with the designs derived from the model (1), (2), and (5) with the size of the sliding window of 31 days which means that each observation can be implemented within a month of a suggested date [Erechtchoukova et al. 2009]. The criterion (5) generated the designs with significantly lower numbers of required observations from 22 (for the effectiveness of 95%) to 6 (for the effectiveness of 80%) in Year_1 and from 16 (for the effectiveness of 95%) to 6 (for the effectiveness of 80%) in Year_2. The high efficiency of these designs has been achieved by selecting the points in time which support the interpolation of the main sections of the chemograph very well. Thus, the way the designs have been derived makes them site and constituent specific. The comparison of the designs developed using formulae (3) and (5) is sown in Figure. 2

Concentrations of different water quality parameters have been determined from the same set of water samples. That is why suggested monitoring designs must be common for all investigated water quality parameters. For non-stratified estimators, only the highest number of observations from the three identified ones supports the established level of effectiveness for all investigated water constituents. For stratified estimators, although this number can be lower, it is necessary to take into account the differences in temporal strata boundaries for the investigated parameters. Only stratification based on the water discharge values is common for all three water quality parameters. As it shown in Table 2, such stratification is not the best for all investigated water constituents.

The total number of required observations significantly decreases with the lower levels of effectiveness of the designs for all investigated water quality parameters and years. Thus, the numbers of observations suggested by the corresponding designs with the level of effectiveness of 80% are very close.

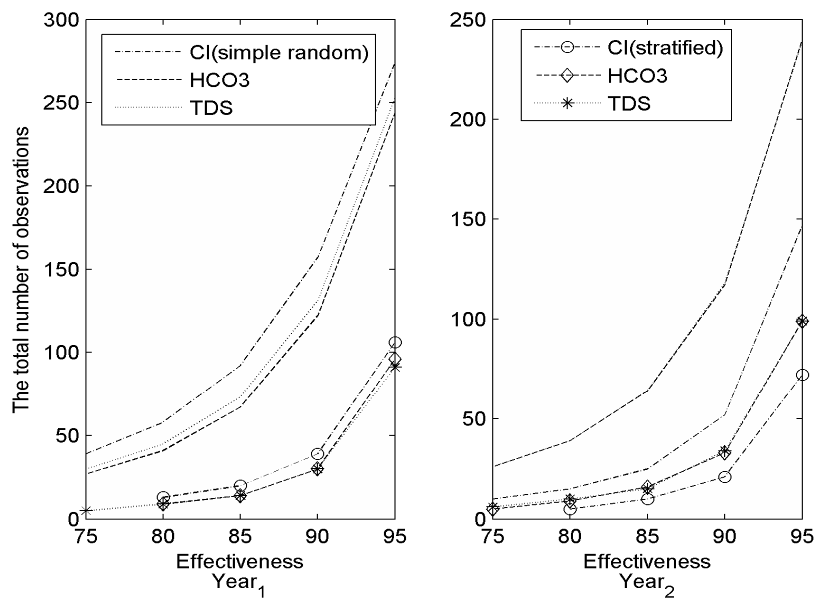

Figure 3. The total number of the required observations for estimation of chemical load of the investigated water quality parameters versus effectiveness

6. CONCLUSIONS

Application of formal procedures to the development of efficient monitoring designs requires quantitative assessment of the design effectiveness. The paper presented an attempt to express the effectiveness as a function of the number of required observations in order to provide necessary information which can be derived from the collected data. This information is obtained via application of statistical procedures or mathematical models which also affect the effectiveness of monitoring programs. The proposed approach implies an articulation of the problem as an operation research model. The efficient designs can be obtained as solutions of this model.

The presented case study demonstrated the dependency of the solutions of the operation research model on selected water quality parameters, indicators, models used for estimates and observation sites. However, the designs differ significantly only when the levels of the effectiveness of monitoring designs are high. With the effectiveness of 80% or less the designs for different water quality parameters and the same indicator can be compromised.

Since monitoring data are usually used for various purposes, some of which are even not known at the time of data collection, the preference should be given to simple monitoring designs or to the designs which support an efficient reconstruction of chemographs of investigated water quality parameters.

ACKNOWLEDGEMENTS

The authors are thankful to anonymous reviewers for their thoughtful suggestions and valuable comments on the manuscript. The study was conducted based on the data sets prepared in Hydrochemical Institute, the Russian Federation.

REFERENCES

Cieniawski, S., Ehart, J., and S. Ranjithan, Using genetic algorithms to solve a multiobjective groundwater monitoring problem, *Water Resources Research 31*(2), 399–409, 1995.

Erechtchoukova M.G., Chen S.Y., and P.A. Khaiter, Application of optimization algorithms for the improvement of water quality monitoring systems, Athanasiadis, I.N., Mitkas, P.A., Rizzoli, A.E., and Marx Gomez, J. (Eds.), Information Technologies in Environmental Engineering Proc. of the 4th International ICSC Symposium Thessaloniki, Greece, May 28-29, 2009. Berlin Heidelberg: Springer, 176-188, 2009.

Erechtchoukova, M.G., and P.A. Khaiter, Uncertainty reduction in modelling of chemical load in streams, Oxley, L. and Kulasiri, D. (Eds) MODSIM 2007 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2007, 2445 – 2451, 2007.

Erechtchoukova, M.G., and P.A. Khaiter, Role of Models for Building an Efficient Monitoring Design, Sànchez-Marrè M., Béjar J., Comas J., Rizzoli A. and G. Guariso (Eds.) iEMSs2008: International Congress on Environmental Modelling and Software. Integrating Sciences and Information Technology for Environmental Assessment and Decision Making, July 2008, 528-535, 2008.

Erechtchoukova, M.G., and P.A. Khaiter, Investigation of Monitoring Designs for Water Quality Assessment, Anderssen, B. et al. (Eds.),18th IMACS World Congress - MODSIM09 International Congress on Modelling and Simulation, 13-17 July 2009, Cairns, Australia, 3612–3618, 2009.

Fisher, R.A., *The Design of Experiments*, Hafner Press, 248 pp., New York, 1971.

Groot, S., and T. Schilperoort, Optimization of water quality monitoring networks, *Water Science and Technology 16*, 275–287, 1983.

Icaga, Y., Genetic algorithm usage in water quality monitoring networks optimization in Gediz (Turkey) river basin, *Environmental Monitoring and Assessment 108*, 261–277, 2005.

Lettenmaier, D.P., Design considerations for ambient stream quality monitoring, *Water Resources Bulletin, 14*:884–902, 1978.

Ning, S.K. and N.-B. Chang, Multi-objective, decision-based assessment of a water quality monitoring network in a river system, *Journal of Environmental Monitoring, 4,* 121–126, 2002.

Ning, S.K. and N.-B. Chang, Optimal expansion of water quality monitoring network by fuzzy optimization approach, *Environmental Monitoring and Assessment, 91*, 145–170, 2004.

Robertson, D.M., Roerish, E.D., Influence of various water quality sampling strategies on load estimates for small streams, *Water Resources Research, 35*(12): 3747–3759, 1999.

Shabman, L. and E. Smith, Implications of applying statistically based procedures for water quality assessment, *Journal of Water Resources Planning and Management, 129*(4), 330–336, 2003.

Whitfield, P.H., Goals and data collection designs for water quality monitoring, *Water Resources Bulletin, 24*(4), 775–780, 1988.