Proceedings of H04, IAHS-IAPSO-IASPEI Assembly, Gothenburg, Sweden, July 2013 (IAHS Publ. 361, 2013).

Freshwater quality monitoring systems: ways towards improvements

MARINA G. ERECHTCHOUKOVA, PETER A. KHAITER & DENNIS P. KHAITER
York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada
marina@yorku.ca

Abstract Investigation and management of the quality of freshwater resources require data on the status of aquatic environments. These data are traditionally collected by routine monitoring systems which are supported by federal and/or local governments operating under limited budgets. This makes the optimization of water quality monitoring systems, in general, and monitoring designs, in particular, an urgent issue. The paper presents a model-driven approach to the development of efficient multi-parameter temporal designs for stream monitoring systems. The approach is based on an operation research model developed using cost-effectiveness analysis and non-linear regression models describing multi-parameter interactions. The developed model does not require site specific parameterization other than a series of water constituent concentrations measured at a given observation site. This approach was tested on a small river with a highly urbanized watershed and the results were compared with existing monitoring programmes.

Key words freshwater quality; monitoring designs; model driven approach; constraint optimization; rivers

INTRODUCTION

The scarcity of usable freshwater on our planet and its uneven distribution makes the sustainable management of this life-preserving natural resource a necessity. Freshwater is a key resource in societal development, playing important roles in the global hydrological cycle, natural ecosystem functioning, the economy, human health and various social activities. Sustainable management of this precious natural resource requires reliable information on its availability and the implications of decisions made with respect to the resource usage. Water availability is determined by the quantity and the quality of the resource. Only water of a certain quality is suitable for a designated use. Water quality is described by a set of parameters reflecting the physical, chemical and biological characteristics of a water body. The values of these parameters along with estimated volumes of water must be taken into account in water-related decision making or in measuring environmental performance and determining the attainment of regulating standards (Yakhou & Dorweiler, 2004). Data on the quality of freshwater resources become an important and essential component of management activities and policy making.

One of the main sources of data on the status of the aquatic environment is water quality monitoring systems collecting the data through direct observations and measurements in a systematic and standardized way. Automatic samplers registering values of concentrations of water constituents with high frequencies are installed at some observation sites. However, they cannot replace routine water sample collection followed by laboratory analysis due to the following reasons. Firstly, these automatic tools can determine values for only a limited set of water constituents. Secondly, the limited budgets of monitoring systems cannot afford a large number of such devices, thus leaving many monitoring sites to operate under routine sampling programmes. In monitoring systems, data collection is conducted in accordance with various programmes which are expected to provide sufficient data for a specific set of scientific, environmental, and/or managerial objectives.

Monitoring systems have a complex infrastructure supporting the entire process from a water sample collection through its processing and reporting. They can be designed based on different approaches. An overview of the main types of water quality monitoring systems can be found in USEPA (2003). According to the Canada-wide framework for water quality monitoring, identified monitoring objectives pre-determine important aspects of the system such as monitoring programme design, field sampling programmes, laboratory analysis and procedures, data analysis and interpretation, reporting and follow-up (WQTG, 2006). Specific monitoring objectives differ for various monitoring programmes. However, a generic set of monitoring objectives comprises of: (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 (Whitfield, 1988).

Monitoring objectives dictate sampling programmes. 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). Whitfield (1988) suggested investigating an attainment of water quality standards by sequential sample collection whereby the number of observations is determined based on the outcome of observations as they are made. Estimations of mass discharge require sampling programmes which take into account not only concentrations of water constituents but also the duration and magnitude of the violation of existing standards (Shabman & Smith, 2003) as well as properties of selected estimators (Robertson & Roerish, 1999; Erechtchoukova & Khaiter, 2009). Since monitoring data are expected to support different types of assessment and projection, some of which were not taken into account during the data collection, the data sets should exhibit simple structure. In order to satisfy multiple data needs, Overton & Stehman (1995) suggested data collection based on simple random designs. However, these designs require high frequencies of water sample collection at an observation site, and the numbers of samples required deviate significantly for different water quality parameters, even at the same observation site. Given that the systems operate under budgetary constraints, the recommendation becomes unattainable.

This paper discusses an approach to optimization of temporal water quality monitoring designs based on formal methods and cost-effectiveness analysis, which is applied to the development of monitoring designs. Although there are other components of a monitoring system affecting its efficiency, they are not considered in the present study. The approach is extended to the designs for multiple water quality parameters determined simultaneously from the same water sample. It was tested on a small river from a highly urbanized area. The results were compared with existing recommendations for water sample collection.

FRAMEWORK FOR OPTIMIZATION OF MONITORING DESIGNS

Existing monitoring designs

Many monitoring systems use networks of fixed stations to collect water samples and process them according to a tiered approach. The approach requires splitting all water quality parameters observed at a given site into two sets: core and supplemental. Water constituents from the core set are determined based on designated uses. Concentrations of these constituents are monitored routinely to check for the attainment of water quality standards. The supplemental set of water quality parameters reflects site- or project-specific needs. Usually, the parameters from this set are observed with lower frequency. Table 1 illustrates data collection programmes in one of the largest water quality monitoring systems of the Russian Federation (Tsirkunov, 1995).

The water quality networks of the Canadian Regional Conservation Authorities are also composed of fixed observation sites with tiered sampling programmes. For these systems, Statistics Canada (2008) recommends collection of as few as 4–6 observations at some monitoring sites or 6–12 samples over a three-year period in order to evaluate the water sustainability index. The Water Framework Directive (EU, 2000) prescribes sampling intervals of three months for all physico-chemical water quality parameters with the exception of priority constituents, which must be observed on a monthly basis. The existing recommendations produce 4–12 values of concentrations of water constituents per year. Such low frequencies of observation are justified mainly by limited budgets allocated for monitoring activities. At the same time, the systems are expected to provide sufficient data for a wide range of scientifically-valid conclusions with levels of error not exceeding 10%. Under these conditions, investigation of monitoring designs, their efficiency and possible ways for improvement becomes an important problem with clear practical significance.

Table 1 An example of monitoring programs for a tiered monitoring system with fixed station.

Frequency of observations	Programme *vs* Different categories of sites: Category I	Category II	Category III	Category IV
Every day (programme 1)	Hydrological parameters, visual observations, temperature, dissolved oxygen	Visual observations	N/A	N/A
Every 10 days (programme 2)	pH, suspended substances, biochemical oxygen demand, concentration of 2-3 typical pollutants		N/A	N/A
Every month (programme 3)	Concentrations of all pollutants for the particular sites			N/A
Major hydrological events (mandatory programme)	Programme 2 plus reduction–oxidation potential, dissolved gases, major ions, nutrients, widely distributed pollutants			

Mathematical articulation of the problem

Exhaustive sample collections should be made to satisfy various information needs. However, sample collection and subsequent processing are conducted under financial and logistical constraints. The optimal number of samples and their distribution over an investigation period can be obtained based on cost-effectiveness analysis (Groot & Schilperoot, 1983). Assuming that for a monitoring design, its cost and effectiveness can be evaluated in the same measuring units (e.g. monetary terms), the designs with minimum cost and maximum effectiveness should be easily obtained. Unfortunately, both required estimates are not readily available. There are different approaches to the assessment of the cost of environment-related activities. Even for the assessment of the direct cost of monitoring activities, many cost components are unknown (e.g. Loftis & Ward, 1980).

Deriving estimates of the effectiveness of a monitoring design in monetary terms is even harder. Given that both the cost and effectiveness of a design should depend on the number of observations suggested by the design, optimal designs can be derived from a constraint optimization problem, minimizing the cost of the design within an acceptable level of its effectiveness (Erechtchoukova & Khaiter, 2012). It is worth mentioning that all currently available approaches to the cost estimation of monitoring systems support the assumption that the cost is a non-negative, monotonically increasing function of the number of observations. This assumption justifies replacement of the actual cost function by a linear function of the number of observations in the optimization problem. The effectiveness of a design measures the extent to which generated data satisfy information needs and, thus, is described by the quality of information generated from monitoring data. According to the Quality Assurance Plan (USEPA, 2003), it is important to understand and quantify the uncertainty and incorporate its estimates into environmental assessment.

In general, the uncertainty of an environmental indicator depends on the applied estimator and its mathematical properties, the variability of an investigated environmental parameter and an available data set. The larger the set, the lesser the uncertainty of values calculated based on the set. Following the theory of design 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. Therefore, the optimization problem can be re-formulated in the following way:

min n subject to

$$\left|\frac{D(I_k(n))}{I_k(n)}\right| \cdot 100\% \leq V_k,\ k = 1,\ldots,K \tag{1}$$

where $I_k(n)$ is the estimator of the k-th water constituent on a set of n observations, $D^2(I_k)$ is the variance of the estimator I_k, V_k is the acceptable level of uncertainty in this estimate, and K is the total number of water constituents of interest. In such articulation, the solution always satisfies the constraint for the most variable water quality parameter resulting in significant oversampling of other parameters determined from the same water sample. In many cases, the series of concentrations of water constituents detected at the same observation site exhibit some relationships. This can be explained by the fact that these concentrations were formed under common environmental conditions. If such relationships are registered and their approximations are defined, they can be used to reduce the number of water samples required to achieve the established level of uncertainty in the estimates through the substitution of the following formulae into the constraint function of model (1):

$$C = f(C_{CMV}) \tag{2}$$

$$D(I_C) = D(I(f(C_{CMV}))) \tag{3}$$

where C_{CMV} is the concentration of the base water constituent, f is a regression function identified based on the least squares fitting. Application of linear regression functions has been investigated by Erechtchoukova & Khaiter (2012, 2011). Although linear functions (2) have an obvious advantage in calculating the variance of an estimator (3), their utilization is not always justified by the basic fitting procedure. The present study considered polynomial regression functions whose variance was estimated using the Delta method (e.g. Hosmer, 2008).

CASE STUDY

The model (1) with polynomial regression functions (2) has been investigated using data collected at the Old Mill Road station of the Humber River (Ontario, Canada). This site belongs to the Toronto and Region Conservation Authority monitoring network. It is located at the lower part of the main section of the Humber River. The main branch of the Humber River flows more than 120 km through a 908 km^2 watershed covering the Niagara Escarpment, the rolling hills and kettle lakes of the Oak Ridges Moraine, the high-quality agricultural lands of the South Slope and Peel Plain, and the ancient Lake Iroquois shoreline. The Humber River is classified as a small river with the annual water discharge of 0.20–0.32 km^3/year (Fig. 1).

The river flows in Southern Ontario from Georgian Bay to Lake Ontario through the Greater Toronto Area, the most urbanized centre in Canada. Its waters experience significant anthropogenic impact. The upper reaches of the river are situated in the areas with permeable soils facilitating infiltration and notably reducing overland runoff. One third of the baseflow of the Humber River comes from clean groundwater discharge. The lower reaches of the river are surrounded by urbanized areas with an impervious surface estimated as 27% of the entire river catchment. Precipitation runoffs wash out accumulated pollutants from the surface including bacteria, nutrients, pesticides, toxic chemicals, heavy metals, oil, grease, road salt and hydro

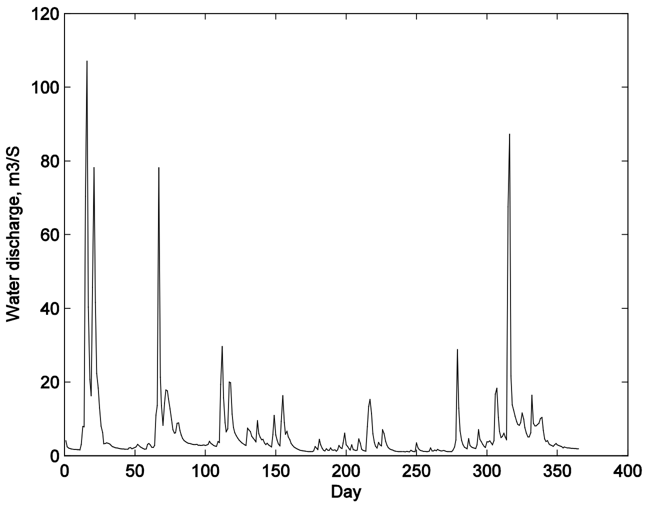

Fig. 1 The hydrograph of the Humber River, Ontario, Canada.

Table 2 Investigated water quality parameters at Humber River, Ontario.

Water constituent	Mean (mg/L)	Variance (mg^2/L^2)	Coefficient of variance
Ca	80.04	520.42	0.28
C	4.37	1.62	0.29
Mg	15.59	19.38	0.28
Cl	220.80	85042.7	1.32

carbons. Water of the upper sections of the river is relatively clean, while the Lower Humber River exhibits the poorest water quality. As a result of de-icing salt applications, chloride ions often exceed the existing standards for wildlife protection. Accidental spills from the chemical, manufacturing and transportation sectors have become the major source of organic and inorganic contaminants present in the Humber River waters (TRCA, 2008).

Data collection at the Old Mill Road observation site is implemented on a regular basis, providing from 14 to 20 concentrations per year, with a rare exception in 1995 when 80 values of concentrations of some water quality parameters were collected. These series of concentrations of total calcium (Ca), organic carbon (C), total magnesium (Mg), and total chloride (Cl) ions were used in the study due to the availability of the relatively long series of concentrations. Table 2 presents the basic statistics of the selected water constituents.

The constraint function of model (1) is to be evaluated based on a selected indicator. In general, the choice is conditioned by the task at hand. However, since the attainment of water quality standards is based on the concentrations of water constituents in a water column, the average concentrations of each water quality parameter over the period of interest were chosen as indicators in this study. For these indicators, simple random designs were developed.

Simple random designs for individual water quality parameters

The designs were developed using model (1) with K = 1 for each investigated water quality parameter for different levels of acceptable uncertainty in the estimates of the average concentrations. These designs vary significantly between parameters (Fig. 2).

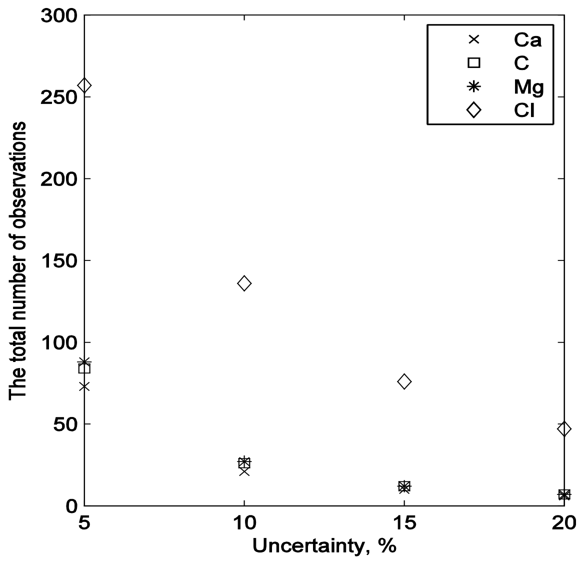

Fig. 2 The annual total number of observations *vs* uncertainty.

Simple random designs for multiple water quality parameters

The designs presented in Fig. 2 lead to conflicting recommendations for water sample collection. When designs common to all four investigated water constituents are determined from model (1), the solution corresponds to the number of observations required for the most variable water constituent (Erechtchoukova & Khaiter, 2012). For the Old Mill Road observation site, this results in at least three-fold oversampling of all other parameters. To generate a design common for all water quality parameters derived from the same sample, without oversampling, the regression functions were applied. The relationships between the series of concentrations of water constituents were analysed using least square fitting in MATLAB 7.1. The concentration of Ca ions was chosen as the base water quality parameter. This parameter served as the independent variable of the regression functions for the three other parameters. The best fitting was achieved on the fourth-order polynomial functions. The solution of model (1) with regression functions (2)–(3), applied to all four investigated water quality parameters, yielded some improvements (Table 3).

Table 3 Designs for multiple water quality parameters.

Uncertainty level, %	Total number of observations	The improvement in common designs, %
5%	206	20%
10%	89	35%
15%	46	39%
20%	27	43%

DISCUSSION AND CONCLUSIONS

The proposed approach is based on the assumption that the series of water quality parameter concentrations collected at the same site are related. This assumption is supported by the fact that these concentrations are formed under the same environmental and anthropogenic conditions. Admitting that the assumption is not always valid for all water quality parameters or all observation sites, it is necessary to mention that such relationships are, however, registered at many sites and are justified by the regression analysis.

It is worth noting that the regression functions (2) used in model (1) do not replace the values of concentrations of non-base water quality parameters derived from a water sample. These functions are used to complement the values of observed concentrations. To obtain statistically valid polynomial coefficients, sufficiently large data sets are required. The rule of thumb for practical application suggests maintaining no less than 45–50 values in the set for this purpose (Cochran, 1963). The simple random designs with fewer numbers of observations are insufficient for the proposed approach.

The comparison of individual designs developed for each investigated water quality parameter with the designs common for all of them, justifies the application of regression functions (2) in the operation research model (1). The proposed approach with linear regression functions was presented in (Erechtchoukova & Khaiter, 2012). The designs common for Ca, C, and Mg ion concentrations are presented in Table 4.

Table 4 Designs developed with linear regression function.

Uncertainty level, %	Total number of observations	The improvement in common designs, %
5%	73	40.2%
10%	21	48.6%
15%	10	47.1%
20%	6	50.0%

The utilization of linear regression functions simplifies estimation of both the variance of selected indicators and constraint functions. At the same time, series of water constituent concentrations are not always highly correlated and regression functions of higher degrees are required. The relationship between Ca and the most variable water constituent in the study, Cl, is an example. Non-linear regression functions broaden the applicability of model (1) for developing efficient designs for multiple water quality parameters.

The designs presented in this paper were developed for evaluation of the same indicator with different but common levels of uncertainty using sets of multiple water quality parameters. The proposed model (1) allows us to derive designs which generate data sufficient to evaluate different indicators and at different levels of uncertainty. An obvious advantage of model (1) with equations (2) and (3), is that it does not require site-specific parameterization. The constraint functions are evaluated based on the time series of concentrations of the selected parameters collected at the investigated sites.

The proposed approach can be useful for tiered monitoring systems where water quality parameters are split into groups according to their priority, which can be reflected in the acceptable levels of uncertainty found in the results.

Comparison of the efficient designs developed based on model (1) with the monitoring designs of existing monitoring systems shows that the declared 10% of uncertainty in data in most of the cases is not attainable. Model (1) cannot be used to evaluate the uncertainty of the estimates for a particular design. Moreover, the designs are site-specific. However, the results of subsequent calculations of the design with different levels of uncertainty support some conclusions. Thus, the data sets used in this study support estimates with 10% of uncertainty only for less variable parameters. However, monthly observations are not sufficient even for that. To estimate average concentrations of chloride ions, sample collection must be increased by 10%. It is worth noting that high variability of concentrations of chloride ions is anthropogenic in nature and can be explained by salt application in the Greater Toronto region during snowfall periods. To improve the common designs for this region, stratified designs should be investigated.

The importance of data on the quality of freshwater resources can hardly be overestimated. In many cases accurate estimates are mandatory. While certain improvements in the quality of the monitoring data can be made by selecting specific monitoring designs, it is necessary to increase the frequency of observations in existing monitoring systems.

Acknowledgements The authors would like to thank the Toronto and Region Conservation Authority (Ontario, Canada) for providing data for the current study. The authors are grateful to Angela Wallace for her work on data files and valuable comments on data. The authors are also thankful to anonymous reviewers for their helpful comments and thoughtful suggestions on the improvement of the manuscript.

REFERENCES

Cochran, W. (1963) *Sampling Techniques*, 2nd edn. Wiley, New York.

Erechtchoukova, M. & Khaiter, P. (2009) Investigation of monitoring designs for water quality assessment. In: *18th IMACS World Congress - MODSIM09 International Congress on Modelling and Simulation* (ed. by B. Anderssen, *et al.*), (13–17 July 2009, Cairns, Australia), 3612–3618.

Erechtchoukova M. & Khaiter P. (2011) A Model-Driven Approach to Uncertainty Reduction in Environmental Data. In: *Information Technology in Environmental Engineering. New Trends and Challenges* (ed. by P. Golinska, M. Fertsch & J. Marx-Gomez). Springer-Verlag, Berlin. 107–122.

Erechtchoukova, M. & Khaiter, P. (2012) Model-driven approach to optimization of monitoring designs for multiple water quality parameters. In: *Managing Resources of a Limited Planet* (ed. by Seppelt, R. *et al.*) (Proc. 2012 International Congress on Environmental Modelling and Software, Sixth Biennial Meeting, Leipzig, Germany, 1–5 July 2012).

EU (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000, Establishing a framework for Community action in the field of water policy.

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

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

Hosmer, D.W., Lemeshow, S. & May, S. (2008) *Applied Survival Analysis: Regression Modelling of Time-to-Event Data*, 2nd edn. John Wiley & Sons.

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

Loftis, J.C. & Ward, R.C. (1980) Sampling frequency selection for regulatory water quality monitoring. *Water Resourses Bulletin* 16(3), 501–507.

Overton, W.S. & Stehman, S.V. (1995) Design implications of anticipated data uses for comprehensive environmental monitoring programmes. *Environmental and Ecological Statistics* 2, 287–303.

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

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

Statistics Canada (2008) Canadian environmental sustainability indicators 2007: Freshwater quality indicator. Data sources and methods. (Cat. N 16-256-X) URL http://www.statcan.gc.ca/pub/16-256-x/16-256-x2008000-eng.pdf. Accessed 10.10.12

Toronto and Region Conservation Authority (TRCA) (2008) Humber River. State of the Watershed Report – Surface Water Quality. Online URL http://www.trca.on.ca/dotAsset/50153.pdf. Accessed 15.01.13.

Tsirkunov, V. (1995) Water quality monitoring in Russia. In: *Proc. Fourth International Symposium – Fish Physiology, Toxicology and Water Quality* (ed. by V. Thurston) (19–21 September 1995, Boseman, Montana, USA).

US Environmental Protection Agency (USEPA) (2003) Elements of a state water monitoring and assessment program (EPA 841-B-03-003), online URL http://www.epa.gov/owow/monitoring/elements/index.html. Accessed 10.10.12.

Water Quality Task Group (WQTG) (2006) *A Canada-wide framework for water quality monitoring*. PN 1369, online URL http://www.ccme.ca/assets/pdf/wqm_framework_1.0_e_web.pdf. Accessed 10.10.12

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

Yakhou, M. & Dorweiler, V.P. (2004) Environmental accounting: an essential component of business strategy. *Business Strategy and Environment* 13, 65–77.