



Jul 1st, 12:00 AM

Role of Models for Building an Efficient Monitoring Design

Marina G. Erechtkoukova

Peter A. Khaiteer

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

Erechtkoukova, Marina G. and Khaiteer, Peter A., "Role of Models for Building an Efficient Monitoring Design" (2008). *International Congress on Environmental Modelling and Software*. 194.
<https://scholarsarchive.byu.edu/iemssconference/2008/all/194>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Role of Models for Building an Efficient Monitoring Design

M.G. Erechchoukova^a and P.A. Khaitea^a

^a*Atkinson Faculty of Liberal and Professional Studies, York University, Toronto, Ontario, Canada (marina/pkhaitea@yorku.ca)*

Abstract: Efficient assessment of water quality requires an integrated approach incorporating such tools as sampling programs supplying data for statistically valid assessment along with watershed and water quality models which are used to transform the collected data into information. It is important to quantify and possibly to reduce the uncertainty of the information supplied by a monitoring system. The paper describes the algorithm for an efficient temporal monitoring design for data collection at a given site of a fixed station tiered monitoring system. The algorithm takes into account features of models used to obtain estimates of investigated water quality indicators. Efficiency of a monitoring design is considered under the assumption that the cost of a sampling program is a monotonously increasing function of a number of observations. Application of the proposed algorithm to various water quality ingredients reveals a wide range of sampling frequencies needed for estimation of water quality ingredient load with a desired level of uncertainty. Monitoring design which is common for all observed water ingredients at a given site and supporting evaluation of their annual loads with the same level of uncertainty can hardly be attained due to financial and technical constraints. For tiered monitoring systems, different allowable levels of uncertainty can be recommended for water ingredients according to their importance for a particular site.

Keywords: Monitoring system; Water quality indicator; Chemical load; Monitoring design; Uncertainty

1. INTRODUCTION

Sustainable decision making requires adequate information about the present quality of the environment and its possible changes in the future. Environmental quality is usually determined via comparison of a set of values of selected indicators with existing standards and is aimed at verifying the suitability of an environmental resource for a designated use (e.g., recreation or drinking water supply). As a rule, environmental indicators are quantifiable variables reflecting physical, chemical or biological characteristics of natural ecosystems at a given moment in time and a certain point in space. With respect to the aquatic environment, its quality refers to physical conditions including temperature and presence of particulate matter and to chemical conditions described by concentrations of dissolved ingredients. Values of environmental indicators must be obtained by direct observation and measurement implemented under a certain program. Long-term standardized measurement, observation, evaluation and reporting of the aquatic environment in order to define status and trends of a water body are called monitoring.

Monitoring systems comprise of such components as collection and analysis of physical, chemical and biological data and quality assurance and control programs to ensure that the data are scientifically valid. Canada-wide framework for water quality monitoring [WQTG, 2006] identifies the following aspects which must be taken into account: monitoring program objectives, monitoring program design, field sampling program, laboratory

analysis and procedures, data analysis and interpretation, reporting and follow-up. While all these aspects are important for monitoring system functioning, the current study deals only with those which require mathematical tools. For efficient assessment of water quality, EPA recommends an integrated approach incorporating several techniques [USEPA, 2003]. These techniques include sampling programs supplying data for statistically valid assessment along with watershed and water quality models which are used to transform collected data into information.

One of the essential characteristics of information supplied by a monitoring system is uncertainty [Harmel et al., 2006]. According to Quality Assurance Plan [USEPA, 2003], it is important to understand and quantify the uncertainty and incorporate its estimates into environmental assessment. Uncertainty of an estimator utilized depends on its mathematical properties, the variability of an investigated environmental indicator and an available data set [Erechtkoukova, 2005]. The larger the set, the lesser the uncertainty of values calculated based on the set. However, for the majority of important water quality indicators and many sampling sites, extensive observations are not possible due to the logistic and financial constraints. Models can improve significantly the reliability of obtained information by reducing uncertainty of model outcomes if available data sets are sufficient for model application. Strictly speaking, the quality of generated information depends on the model used [Erechtkoukova and Khaite, 2007]. Investigation of model properties can give an insight about possible ways to reduce the uncertainty of the information and to suggest monitoring designs improving estimates of environmental indicators [Strobol et al, 2006; Urban, 2000].

Selection of a model depends to a great extent on investigated indicators. Shrestha et al. [2008] pointed out that for effective water quality management estimates of ingredient loads are more important than concentrations. Chemical loads can be used as an objective for effective design of a monitoring program [Hooper et al., 2001].

The paper investigates the role of models in water quality monitoring. It describes the algorithm for an efficient temporal monitoring design for data collection at a given site which takes into account features of models used for data analysis and uncertainty associated with the collected data. Chemical load is chosen as a water quality indicator of interest.

2. MONITORING SYSTEMS

Monitoring systems provide broad sets of data collected in accord with a program designed for a specific set of scientific, environmental or managerial objectives. Canada-wide framework for water quality monitoring includes three main phases: (1) planning which determines objectives and scope of the program; (2) collection/analysis incorporating field sampling, laboratory analysis, data management and processing, data analysis and interpretation, data reporting and quality assurance/quality control; and (3) information utilization for decision making, education and policy development and enforcement [WQTG, 2006]. Although monitoring objectives differ for various monitoring systems, in general, they include such common tasks as determination of water quality standards to be attained, attainment of the standards, identification of impaired waters, as well as causes and sources of water quality impairments and detection of long-term trends [USEPA, 2003]. Data collection must be conducted according to a proposed monitoring design. The monitoring design is to reflect these objectives as well as to provide reliable data for decision making. The extent, to which collected data represent real state of the aquatic environment, depends on a chosen spatial and temporal monitoring design. That is why its selection is important for many tasks of environmental assessment.

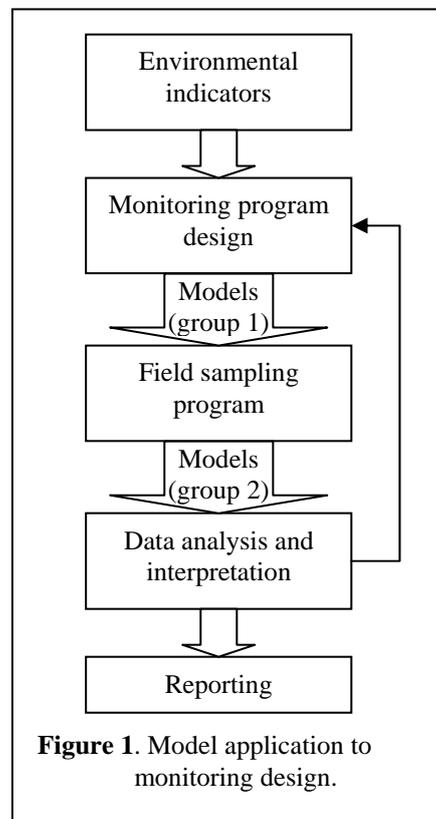
There are several approaches to a monitoring design. Fixed station approach assumes that the same sites are repeatedly sampled at regular time intervals over a long period of time. Short-term monitoring is a specific study which investigates particular water quality

problems and creates a ‘snapshot’ of the conditions in a given area. Rotating-basin approach is based on intensive short-term surveys which are conducted periodically. It may identify changes in water quality conditions over time. In a probability-based approach, sites are selected randomly from the total set of sites on water bodies in a selected area. An exhaustive approach requires sampling or surveying of all water bodies in the area. A tiered approach [USEPA, 2003] is adopted in many monitoring systems. The approach requires identification of a core set of water quality indicators which reflect designated uses and can be monitored routinely to assess attainment of applicable water quality standards. In addition to the core set of indicators of the aquatic environment, it is also necessary to identify supplemental indicators dictated by the site or project specific needs. In general, the selection of variables to be measured depends on such factors as monitoring objectives, site specific water quality issues and designated uses of interest which may result in significant variations of core sets for different sites across an investigated region. A consensus on core sets of indicators is strongly desirable since it creates compatible and sharable data sets for large scale analysis and data generalization [WQTG, 2006]. Water quality indicators from core and supplemental sets are observed with different frequencies at the same site. Moreover, core indicators can be observed with different frequencies at different cross-sections of the same section of the natural stream because of the importance of a particular location.

No single monitoring approach is sufficient to provide the data for all information needs. To meet the objectives, monitoring systems integrate several designs or programs of observations. Thus, fixed station approach along with tiered monitoring design coupled with sampling programs reflecting environmental heterogeneity is useful for long-term trend detection, for assessment of critical reaches of large streams, and at the same time provides site-specific water-quality data.

3. MODELS IN MONITORING DESIGN

Observation data supplied by monitoring systems are a mandatory component of any environmental decision making process, but this component can be useful only if the data are synthesized into information. Ideally, information has to be comprehensive and complete to meet multiple relevant needs. In many cases of environmental decision making, understanding of interactions of key environmental processes is vital. It cannot be achieved by observations along. This is the moment when models come into play. Environmental models form a diversified set of techniques based on different mathematical and computational methods [Jørgensen and Bendricchio, 2001; Straškraba and Gnauck, 1985]. Models transform observation data into information by extracting aggregate values from raw data, projecting values of selected environmental indicators and detecting trends to track changes in water quality. Models, in order to be applied, impose additional requirements on the way the data are acquired. A data set available for analysis must satisfy the assumptions that underlie mathematical techniques employed for data analysis (models from group 2 in Figure 1). These assumptions must accord with models from group 1. The latter can be used to determine frequencies of observations sufficient for deriving statistically meaningful results.



Strictly speaking, frequencies of observations conditioned by models from group 1 reflect variability of the model results rather than the natural environmental heterogeneity. They are derived from the model properties and seem to contain an error. Being a simplified representation of reality, no model can fully duplicate real system behaviour and, thus, introduces an error which is also referred to as model uncertainty. Model uncertainty, in its turn, transforms into the errors in recommended frequencies. These errors can be minimized by selecting a model which describes an investigated system better than others under given assumptions. Information extraction from observation data heavily depends on chosen models. At the same time, model selection is significantly restricted by available observation data and, hence, data collection must fit entire modelling process [Richardson and Berish, 2003]. This interdependency of data and models calls for the necessity to design sampling programs based on specific statistical or mathematical assumptions which must be consistent with the ways collected raw data are analysed and with the type of models to be used for this purpose.

4. UNCERTAINTY IN WATER QUALITY MONITORING

Conclusions drawn from monitoring data always contain uncertainty which is introduced by all components of a monitoring system. One of the sources of uncertainty, particularly model uncertainty, has been already mentioned. It is necessary to add observational artifact and the uncertainty introduced by selecting sampling sites. The very idea of the monitoring to describe continuous fields of environmental indicators by discrete samples collected from time to time implies the uncertainty since it is based on the assumption that values of observed indicators remain steady in a neighbourhood of a sampling site for some period of time which is not always valid. This type of uncertainty can be reduced by optimizing spatial and temporal monitoring design and most likely by introducing additional sampling sites with higher frequencies of observations, but can hardly be eliminated. Observational artifact is caused by measurement tools and analytical methods used in the laboratories in order to obtain values of environmental indicators of interest. Although some improvements of the results are possible, this type of errors is always present in monitoring data [Harmel et al., 2006]. Usually, monitoring guidelines recommend keeping both types of uncertainty under a 10% level.

Model uncertainty plays an important role in the analysis and interpretation of monitoring data. Models receive the uncertainty from previous monitoring phases and transform it into information required for decision making. The resulting uncertainty must be understood, quantified, and limited to a reasonable extent with respect to the cost of possible consequences of decision errors.

Commonly accepted definition of model uncertainty describes it as deviations of simulated system variables from their known or observed values [Campolongo et al., 2000]. There are various sources of model uncertainty. First of all, the uncertainty is caused by the model intrinsic feature as an abstraction of reality. Other sources of model uncertainty depend on model structure, mathematical formulae and approaches employed in model components. Model uncertainty undoubtedly influences the process of conversion of monitoring data into information.

It is necessary to evaluate the extent to which the model outputs are uncertain and to attempt to reduce this uncertainty in the results. Model uncertainty reduction is mainly achieved by selecting a particular model structure or parameter values. Although these two sources of model uncertainty are extensively investigated [e.g., Snowling and Kramer, 2001; van Nes and Scheffer, 2005], they create confounding effect which impedes identification of uncertainty caused solely by the model structure. That is why the following aspect must be taken into account during model development. The uncertainty of a chosen model under a given set of observation data must be less than the uncertainty delivered by any other model on the same data set. Therefore, the properties of the chosen model can be used for planning an efficient monitoring design. Strictly speaking, an

efficiency of a sampling program can only be determined against certain criteria. It is possible to relate these criteria with some kind of financial feasibility study or particular technical aspects that also suggest monetary estimates. Under the assumption that the cost of a sampling program is a monotonously increasing function of a number of observations, the criterion of efficiency can be rephrased as the problem of minimizing a required number of observations sufficient to keep the resulting uncertainty at a desired level.

5. TEMPORAL MONITORING DESIGN

Selection of a model for data analysis and interpretation is stipulated by monitoring objectives and an approach to a monitoring design. Chemical loads are important indicators of water quality. Using annual chemical load of water ingredients as selected water quality indicator is an effective way for organizing a large-scale monitoring system [Hooper et al., 2001]. For management purposes, Shrestha et al. [2008] also give preference to ingredient loads over their concentrations.

There are several approaches to load estimation. The overview of these approaches and their classification can be found, for example, in Aulenbach and Hooper [2006]. Different formulae for load calculation are presented in Preston et al. [1989]. Chemical load estimates usually use values of water discharge and concentrations of an ingredient determined from the samples collected at a particular cross-section of a water body over a period of time.

Uncertainty of a load estimate denotes possible deviations of values calculated based on available data sets and a selected formula from the actual value. Statistically, load uncertainty can be interpreted as the variance of the formula chosen for approximation. Thus, resulting uncertainty can be minimized by selecting an estimator (i.e. model) that outperforms others on the same set of observation data. At the same time, scarce observation data can magnify the resulting uncertainty of the selected estimator. Therefore, the chosen model must be used for planning an efficient sampling program.

Improvements in the performance of an estimator are usually achieved in two ways: additional sampling or modifications of the estimator bringing additional knowledge about natural phenomena. Majority of load estimates use water discharges which are described by more detailed series of values than concentrations of water ingredients. With this respect, regression and ratio estimates have to be considered. Regression models are probably more popular. They are used independently or as a part of the composite method [Aulenbach and Hooper, 2006]. A regression model describes the relationships between water discharge and concentrations of a water ingredient and is actually used to restore missing or predicted values of concentrations. It is assumed that the relationship is steady and the model is suitable for the forecasting. The main restriction in application of regression analysis is the number of available samples. If the relationship between concentrations and water discharge is linear and concentrations are normally distributed, the rule of thumb requires at least 50 samples to evaluate the reliability of regression coefficients. This number increases significantly for other types of regression equations and water quality indicators with high variability.

Ratio estimates are based on calculation of average water discharge using all available data and instantaneous load values only when concentrations are measured. Strictly speaking, the ratio estimator has a bias which must be corrected. Although ratio estimators have minimal uncertainty if the relationship between two investigated variables is linear, there is no need to make such a rigorous assumption for entire investigated period. Stratification of the investigated period can significantly improve the performance of load ratio estimators. The stratified estimator of an ingredient load dominates non-stratified one, if the variance of the instantaneous load within the strata is less than its variance between the strata. The efficiency of a stratified estimator certainly depends on a stratification scheme applied. It is worth to note that stratification reducing load uncertainty can be achieved for the majority

of natural streams with distinct hydrological seasons [Erechtkhoukova and Tsirkunov, 1989]. Then the following formula provides an efficient estimation of the load:

$$L = T \sum_{j=1}^k \frac{N_j}{N} Q_j \left(\frac{l_j}{q_j} + B_j \right), \tag{1}$$

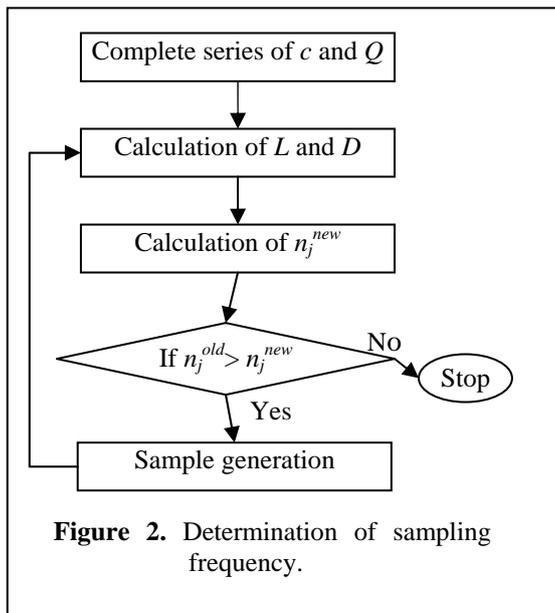
where Q_j is the average water discharge in j -th stratum calculated from the most frequent measurements, q_j is the average water discharge in j -th stratum calculated using values corresponding to the observed concentrations, l_j is the average instantaneous ingredient load calculated from the observed concentrations, N is the total number of water discharge observations over the period, N_j is the number of water discharge observations in j -th stratum, T is the duration of the investigated period, B_j is the bias of the ratio estimate. The formula for the bias estimation can be found in Cochran [1963].

The required number of concentrations which keeps the uncertainty of load estimator (1) under a given level can be obtained from the problem of mathematical programming [Bodo and Unny, 1983; Erechtkhoukova and Tsirkunov, 1989]:

$$\min \sum_{j=1}^k n_j, \text{ subject to } D(L) \leq V, \tag{2}$$

where $D(L) = T^2 \sum_{j=1}^k \frac{N_j^2}{N^2} S_j^2$.

Here n_j is the required number of concentrations per j -th stratum, V is the given level of uncertainty and $D(L)$ is the ingredient load variance, S_j^2 is the variance of average instantaneous load estimator in j -th stratum which depends on the number of observations, n_j , conducted during this stratum, k is the total number of strata. Such formulation is possible due to consistency of estimator (1). The given level of uncertainty V is determined using a desired accuracy and the calculated value of the ingredient load under the assumption of normal distribution of statistic (1).



Problem (2) can be solved numerically according to an algorithm utilizing the Lagrange multiplier method and presented in Figure 2. Initially, the algorithm requires the most complete series of concentrations of an investigated water quality ingredient. Such a series can be obtained via pilot sampling or interpolation in time between observed values. To improve robustness of the algorithm, at each iteration one hundred random samples were made and the average number of observations per each stratum was calculated.

The proposed algorithm was used in a case study to determine the number of observations per year in order to estimate annual load of major ions at the cross-section Vyatskiye Polyany of the Vyatka River. The Vyatka River is a large Eastern-European river with a length of 1,370 km and a watershed area of 129,000 km². The

selected cross-section is characterized by annual water discharge of about 22.6 km³. The year 1949 with unimodal type of hydrograph and sharp rising and falling limbs for spring-summer high flow events was selected. The annual loads of chloride ions, hydrocarbonate ions and total dissolved solids (TDS) were determined as 63,907 t, 2,465,922 t and 3,703,000 t, respectively.

Table 1. Frequency of observations guaranteeing a 5% uncertainty

| Ingredient | Number of samples per stratum | | | | Total per year |
|------------------|-------------------------------|----|-----|----|----------------|
| | I | II | III | IV | |
| TDS | 3 | 17 | 2 | 4 | 26 |
| Cl | 10 | 57 | 28 | 12 | 107 |
| HCO ₃ | 3 | 21 | 2 | 7 | 33 |

The stratification of annual series of concentrations and water discharges was implemented according to main hydrological seasons. Stratum I represented winter low water events. Stratum II combined low parts of rising and falling limbs of the hydrograph. Stratum III included pick discharges and upper parts of rising and falling limbs of

the hydrograph. Stratum IV corresponded to summer-fall low water events. Sampling frequencies for estimation of the annual loads of the selected water quality indicators with a 5% uncertainty are presented in Table 1.

6. DISCUSSION AND CONCLUSIONS

The investigated case study showed that the efficiency of formula (1) depends not only on the total number of concentrations available, but on their distribution among the temporal strata. Stratification is an important factor for the reduction of the required number of observations. In the presented case study, the stratification was implemented according to hydrological regime, not taking into account the variance of selected water quality indicators that presumably makes easier sampling recommendations. Such stratification gives lower values of the water discharge variance, but dates for each stratum vary from year to year making recommendations for sample collection still hard to follow. Temporal stratification common for several subsequent years is preferable, but it results in higher numbers of required observations.

Suggested monitoring designs are obviously model-dependent. If ingredient loads were estimated based on another formula, neither of the designs would guarantee the desired level of uncertainty. The determined number of observations not only conforms with the estimator (1), but was obtained by utilizing some additional information, particularly, about the shape of the hydrograph and boundaries of the main hydrological seasons. The even distribution of observations over a year would significantly increase the number of observations that are necessary to achieve the desired level of uncertainty. The proposed algorithm assumes that the accurate values of at least daily water discharges are available. In reality, the series of water discharges are not always accurate and, moreover, are not maintained for many cross-sections. In the latter case, the values can be obtained via hydrodynamic simulations using more sophisticated models.

As a rule, concentrations of the investigated indicators are obtained from the same sample, but recommended frequencies of their observations vary up to four times for different ingredients. A temporal monitoring design which is common for all observed water ingredients at a given site and supporting evaluation of their annual load with the same level of uncertainty is hardly attainable due to financial and technical constraints. For tiered monitoring systems, different allowable levels of uncertainty can be recommended for water quality indicators according to their importance for a particular site.

ACKNOWLEDGEMENTS

The authors are grateful to anonymous reviewers for their thoughtful suggestions and helpful comments on the manuscript. Part of the research was done based on the data sets prepared in Hydrochemical Institute, the Russian Federation.

REFERENCES

- Aulenbach, B.T., and R.P. Hooper, The composite method: An improved method for stream-water solute load estimation, *Hydrological Processes*, 20, 3029-3047, 2006.
- Bodo, B., and T.E. Unny, Sampling strategies for mass-discharge estimation, *Journal of Environmental Engineering*, 109 (4), 812 – 829, 1983.
- Campolongo, F., A. Saltelli, T. Sørensen and S. Tarantola, Hitchhiker's guide to sensitivity analysis. In A. Saltelli, K. Chan and E.M. Scott (eds.), *Sensitivity Analysis*, Wiley, Chichester, 15-47, 2000.
- Cochran, W.G., *Sampling Techniques*, 2nd ed., Wiley, New York, 1963.
- Erechtkoukova, M.G., and P.A. Khaite, Data-model issues in environmental impact assessment, *The International Journal of Environmental, Cultural, Economic and Social Sustainability*, 3 (6), 149-156, 2007.
- Erechtkoukova, M. G., Uncertainty transformation in ecological simulation models. In A. Zenger and R.M. Argent (eds.), *MODSIM 2005 International Congress on Modelling and Simulation*, Modelling and Simulation Society of Australia and New Zealand, 2477-2483, 2005.
- Erechtkoukova, M.G., and V.V. Tsirkunov, Determination of sampling strategy for calculation of mass-discharge in streams with a given accuracy. In *Proc. of the Conf. Problems of Surface Hydrology, February, 1987*, Gidrometeoizdat, Leningrad, 171-189, 1989.
- Harmel, R.D., R.J. Cooper, R.M. Slade, R.L. Haney and J.G. Arnold, Cumulative uncertainty in measured streamflow and water quality data for small watersheds, *Transactions of the ASABE*, 49 (3), 689-701, 2006.
- Hooper, R.P, B.T. Aulenbach and V.J. Kelly, The National Stream Quality Accounting Network: A flux-based approach to monitoring the water quality of large rivers, *Hydrological Processes*, 15, 1089-1106, 2001.
- Jørgensen, S.E., and G. Bendricchio, *Fundamentals of ecological modelling*, Elsevier, Amsterdam, 2001.
- Preston, S.D., V.J. Bierman, Jr. and S.E. Silliman, An evaluation of methods for the estimation of tributary mass loads, *Water Resources Research*, 25(6), 1379-1389, 1989.
- Richardson, J. R., and C. W. Berish, Data and information issues in modelling for resource management decision making: communication is the key. In V.H. Dale (ed.), *Ecological modelling for resource management*, Springer, New York, 167-179, 2003.
- Shrestha, S, F. Kazama and L.T.H. Newman, A framework for estimating pollutant export coefficients from long-term in-stream water quality monitoring data, *Environmental Modelling and Software*, 23, 182-194, 2008.
- Snowling, S.D., and J.R. Kramer, Evaluating modelling uncertainty for model selection, *Ecological Modelling*, 138, 17-30, 2001.
- Straškraba, M., and A.H. Gnauck, *Freshwater ecosystems: modelling and simulation*, Elsevier, Amsterdam, 1985.
- Strobol, R.O., P.D. Robillard, R.D. Shannon, R.L. Day and A.J. McDonnell, A water quality monitoring network design methodology for the selection of critical sampling points: Part I, *Environmental Monitoring and Assessment*, 112, 137-158, 2006.
- Urban, D., Using model analysis to design monitoring programs for landscape management and impact assessment, *Ecological Applications*, 10(6), 1820-1832, 2000.
- US Environmental Protection Agency (USEPA), 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>, 2003
- van Nes, E. H., and M. Scheffer, A strategy to improve contribution of complex simulation models to ecological theory, *Ecological Modelling*, 185, 153-164, 2005.
- Water Quality Task Group (WQTG), 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, 2006.