mon*EAU*: a platform for water quality monitoring networks

L. Rieger and P. A. Vanrolleghem

ABSTRACT

Continuous monitoring of water quality creates huge amounts of data and therefore requires new concepts to guarantee high data quality and to prevent data graveyards. Monitoring stations commonly used in practice today suffer from insufficient flexibility and a lack of standardization. That is, although a lot of monitoring tasks are comparable and should lead to robust and powerful platforms, most monitoring stations are case specific developments.

In this paper the underlying ideas of a new generation of monitoring networks is described. First a problem analysis of monitoring stations typically seen in current river monitoring practice is outlined, then the mon*EAU* vision on monitoring networks will be discussed together with an overview of a planned system set-up with innovative data evaluation concept.

Key words | diagnosis, fault detection, on-line sensors, remote control proactive maintenance

L. Rieger
P. A. Vanrolleghem
model*EAU*, Département de génie civil,
Université Laval, Pavillon Pouliot,
Québec (QC), G1K 7P4,
Canada
E-mail: *leiv.rieger@gci.ulaval.ca*;
peter.vanrolleghem@gci.ulaval.ca

INTRODUCTION

With the increasing use of on-line sensors for water quality measurements, a shift can be seen from not enough data (due to the time consuming sampling and lab analysis) to plenty or even too much data. Whereas the accuracy of the lab measurements is normally sufficient, urgently needed data quality evaluation concepts for the continuous measuring devices are not available or inefficient in day-to-day operation. Several statistical methods have been developed in other fields but only recently has research (e.g. Lee & Vanrolleghem 2003; Rosen *et al.* 2003) shown their applicability in the water/wastewater sector with its special demands on durability and accuracy of sensors as a result of the often harsh conditions. However, only few of these methods have been implemented in software platforms for practical use and many have not proven their potential to detect measurement errors or other equipment failures. The goal of the investigations discussed here is therefore to bring these methods from an academic level into practice. This must be done in a way that is as user-friendly as possible yet as rigorous as possible to detect potential failures, quantify uncertainties and finally solve the problem or deal with the wrong or uncertain measurements.

Initiated by immission-based legislation, (e.g. WFD in the EU or the TMDL approach in the USA) monitoring networks will be essential tools to monitor pollutants, to (better) understand the ongoing processes and finally improve the water quality of our water courses. Whereas the state-of-the-art is still stand-alone monitoring stations, ongoing research is focussing on the development of monitoring networks that integrate the information from different locations into knowledge about whole river basins (Strobl *et al.* 2006). The development of monitoring networks instead of individual stations leads to new demands on bidirectional data exchange, i.e. various telemetry options, safety issues and accessibility.

In this paper we describe our vision of the next generation of water quality monitoring networks. This mon*EAU* (monitoring of water, “eau” in French) vision is being realized in an ongoing project involving research groups, public organizations and private companies from

doi: 10.2166/wst.2008.135

North America and Europe. Besides the focus on new data evaluation methods, this monitoring network concept will combine state-of-the-art technology with the highest possible flexibility in terms of connectable sensors, measuring locations and monitoring goals.

PROBLEM ANALYSIS

Although monitoring stations and more recently monitoring networks increasingly have been built in parallel to the development of continuously measuring devices (in-situ: directly in the liquid: on-line: in a parallel sample line) these stations nearly always suffer from the same problems. Starting with our own experience over the past 10 years, Figure 1 shows a monitoring station set-up that is a typical example for river monitoring stations used today. A submerged pump feeds the whole set-up including filtration unit, in-situ sensors and an auto-sampler through a fast hydraulic loop. Passing a transmitter with some visualization capabilities, the data is collected on a data logger and then sent out to a server via SMS text messages.

van Griensven *et al.* (2000) report that the first reliable data set was only collected after one year of operation due to problems with clogging (filter, tubes and membranes), pump failures, incorrect temperature sensor locations, insufficient auto-cleaning capabilities among other things. Furthermore, not all the problems could be solved and adhesion of silt and clay prohibited the planned operation even though maintenance was done on a weekly basis requiring two team members. In summary, van Griensven *et al.* state that "Buying an AMS [Automated Measuring Station] is like buying a new house: one should first have bought one to have the necessary experience to buy one".

A comparison with other monitoring projects reveals that most stations face similar problems (e.g. Beck *et al.* 1998; Vandenberghe *et al.* 2005). Unfortunately, much of the failure information is only available from personal communication or it has to be deduced from "between the lines" of any publications. Often, three major reasons limit the use of monitoring stations: i) the lack of standardization, ii) data quality problems, which lead to data graveyards that do not provide the required information, iii) insufficient flexibility of the stations being evaluated leading to problems when new or better sensors should be connected or when the focus of the project changes.

Whereas the lack of standardization might be due to the fact that water quality monitoring is still often seen as an individual challenge with no need for a generic platform, the data quality problem can be related to the lack of adequate tools for fault detection. The lack of flexibility is often caused by the typical design procedure starting with one or a few very special objectives followed by individual design specifications without considering possible changes in the objectives, sensor choice etc. that might be needed. Harmancioglu & Alpaslan (1994) give an overview of a structured design procedure leading to individual solutions.

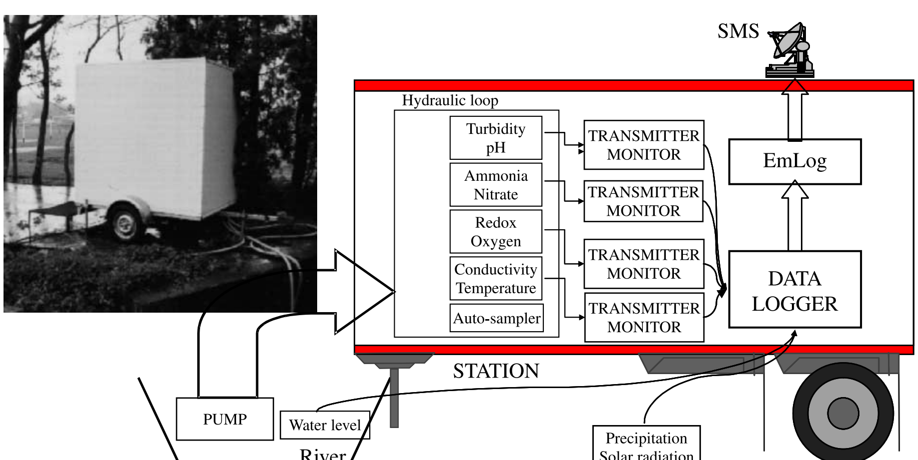

Figure 1 | Typical monitoring station set-up (van Griensven *et al.* 2000).

These individual solutions were probably necessary in the past, but the current task should be to develop standardized platforms based on the experience from hundreds of individual monitoring stations. From our review, we see the following typical reasons for inadequate designs:

- limited budget (and focus on sensors rather than the station itself)
- lack of knowledge (data communication, databases, automation…)
- harsh conditions (WWTP: fat, clogging, electrical interferences; Rivers: changing water levels, flood, heat/cold, vandalism…)
- unsuitable hardware (e.g. pumps, filtration units, data communication, damp protection…)
- inflexible design with respect to the selection of sensors (no change for better sensors or other monitoring objectives)
- one-sided focus on special monitoring objectives
- closed/locked source code and therefore limited freedom to adapt the system to changing project needs (not modular)

In-situ sensors have improved significantly during the past 5–10 years and now provide a real alternative for water quality monitoring by combining low maintenance with sufficient accuracy (Jeppsson *et al.* 2001; Vanrolleghem & Lee 2003). However, standardized data transmission protocols allowing full access to sensor configuration and meta-data are still missing. Access to configuration data would give additional information to the end-user as well as providing remote control options. Meta-data (data about data) enables the extraction of usable information out of a (measuring) signal. For instance, without the unit, the measuring location and a description of the compound measured, the signal only provides a data value without any context for the end-user. In addition, some sensors already provide self-diagnosis or other status information, which if available could be used for data quality evaluations.

Similar difficulties are being experienced by SCADA system providers who are interested in data quality evaluation but are facing the same lack of information from the sensors. Unfortunately, most sensor companies do not allow full bidirectional communication with the sensor or they see their control unit as the last point in the measuring chain.

THE mon*EAU* VISION

For the development of new products a clear vision is necessary. The key design elements are:

- A Flexible System: The goal here is to create a system that can be used:
  - for different monitoring and research goals.
  - at different locations (e.g. river, WWTP, sewer or use for collection of meteo data).
  - with all types of sensors and sampling methods (in-situ/on-/off-line)
  - with all standard communication protocols for sensor connections (if possible, use of bidirectional bus protocols).

More specifically, the idea is to create a station that can be used for multiple goals irrespective of the location (i.e. it should be easily transportable and suitable for set up in a trailer/housing for use in cities (vandalism) or at WWTPs). Similarly, the station should include automatic configuration or at least user-guided configuration procedures through graphical user interfaces.

- An Open and Modular System: To create a modular design, a robust software framework with basic functionality is being created. The addition of functionality is accomplished with ‘plug-in’ modules. In this way it is possible to adapt the system to special demands and to add special features (e.g. data evaluation tools, data visualization methods) and at the same time guarantee the integrity of the framework itself. Modules can be written in standard software languages and it should be feasible for end-users to easily extend the station’s functionality.
- A High Quality/Performance Database: Critical to the station is the database system used for storing the large data series. The database structure needs to provide sufficiently fast access but be flexible enough for any monitoring task and further developments of the station’s functionality.

- Remote Use: Monitoring stations can not always be placed in convenient location as such, the design must consider:
  - reduced maintenance requirements (attendance on site max. twice per month)
  - minimization of energy demand in combination with different power options (e.g. solar power)
  - various telemetry options (e.g. telephone or satellite)
  - remote access to sensors (e.g. for diagnosis, configuration, calibration)
  - remote access to monitoring station operation (e.g. backwash of filtration units or auto-cleaning equipment).
- Automatic Data Quality Assessment: Data quality assessment is essential but will be based on different information sources providing redundant data over time, space and determinants:
  - reference samples (to be compared with the sensor measurements)
  - sensor status/diagnosis data (e.g. auto-calibration factor or self-diagnosis)
  - time series information (to be used for univariate and multivariate statistical tools).
- User-Friendly and User-Oriented Software Concept: A user-based concept should guarantee that the required information is provided and visualized depending on the user level and the location. For instance, the operator at the base station needs information criteria for maintenance and recalibration, whereas the expert working at the central data repository should full access to all information of all connected base stations allowing him to decide, for instance, about remote calibration, changes in operational settings or relocation of a base station.
- Proactive and Flexible Maintenance Concept: Based on the available information the required maintenance is determined and a schedule for the operators is provided. This concept is based on three information sources including:
  - sensor self-diagnosis
  - company or user experience
  - a proactive set of station-triggered experiments (e.g. evaluation of different auto-cleaning cycles).

THE mon*EAU* SYSTEM

Software

The heart of the mon*EAU* system is a robust software framework serving as the backbone of the stations and server network allowing the simple connection of various modules through a specified API (Application Programming Interface). Some modules will provide basic functionality like data input or output but the main reason for this framework structure is the ability to integrate new developments or to connect third party modules. In this way, robust operation (the framework is not open to the end-users) is combined with the required flexibility. Figure 2 shows the mon*EAU* concept.

Data transmission protocols

Sensor ↔ Base Station: Although several communication protocols are of interest in water quality sensor systems (Profibus, HART, 4–20 mA, USB…), to make use of all available information coming from the sensor itself (e.g. sensor self-diagnosis or other meta-data) a bus protocol is used. However, as a fall-back strategy 4–20 mA connections are also provided.

Base Station ↔ Central Server: Different telemetry modules will be available to connect the base stations with the central server, for instance:

- Telephone line, xDSL, cable TV
- GSM/UMTS (e.g. using SMS text messages as a safe and cheap way for data transmission)
- Dedicated radio link
- Satellite

Database Structure and Safety: To guarantee the readability of the data by different software programs and also to allow data use in the future, the IEEE 754 floating point standard format has been chosen. Different export functions will allow easy exchange of data with other software platforms.

To guarantee data safety, different safety measures are integrated in the base stations as well as on the central server. These measures include RAID (redundant array of independent disks), and sufficient hard disk space at the base stations to bridge communication breakdowns or store

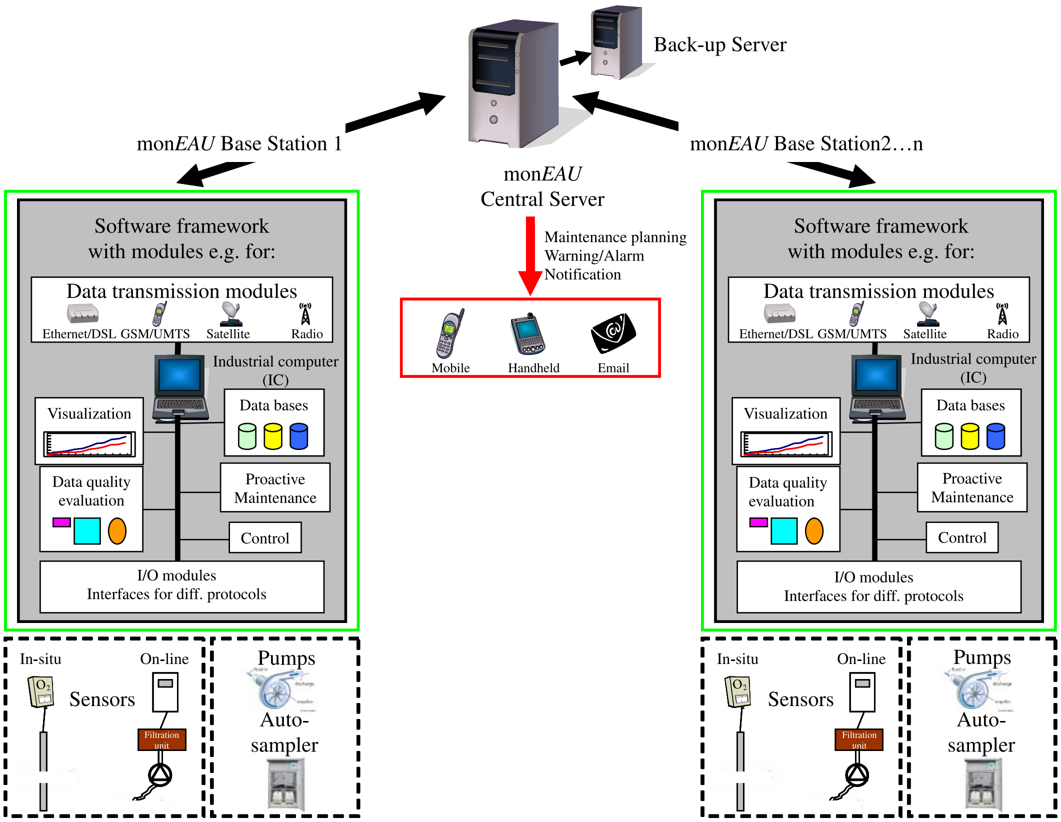

Figure 2 | Set-up of the mon*EAU* water quality monitoring network.

measurement data. At the central data repository, a RAID level 5 is envisioned.

Besides the raw data, processed (data quality evaluation), condensed (averages) and/or derived data (e.g. load, sludge retention time, statistical indexes) are stored. The data evaluation will add quality aspects to the database. All subsequent data manipulations are marked with the raw data remaining untouched, for other processing should the need arise.

Data Quality Evaluation: A major problem when developing a data quality evaluation concept is that it should use as much information as possible while still being flexible for use at different locations. Problems arise if process or expert knowledge, specific for only one measuring location, is integrated. The mon*EAU* system includes a generic concept but encompasses the possibility of further integration of knowledge-based approaches. Three evaluation levels will be implemented and user-selectable:

- Level 1: Basic data evaluation using only univariate methods (only one measurement signal is used for the evaluation)
- Level 2: Multivariate time series analysis methods but without expert or process knowledge.
- Level 3: Advanced data evaluation including expert and process knowledge (data mining and control actions are foreseen, in this evaluation level).

Level 1: The basic concept uses three independent sources of information (Figure 3): i) comparative measurements for off-line analysis coming from reference samples taken at the same location as the sensor samples while considering sensor response times to allow a time-corrected comparison, ii) time series information, iii) sensor status or self-diagnosis data together with process knowledge and information from the electrical signal itself (e.g. frequency, out of bounds, etc.) to create additional data quality information.

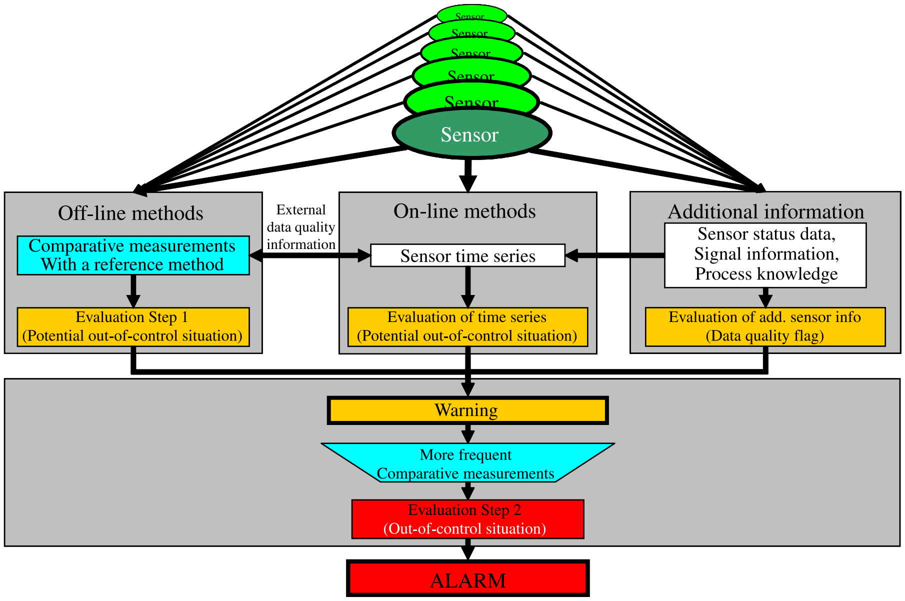

Figure 3 | Univariate/multivariate data quality evaluation concept.

Under normal conditions (no warning) a reference sample should be taken approx. once every two weeks (depending upon the variability of the monitored system). This frequency is sufficient only in combination with the on-line evaluation methods of the mon*EAU* system. For off-line concepts commonly applied today (e.g. ATV 2000; Thomann *et al.* 2002), it is suggested that a reference sample be taken once a week. Every single block of the mon*EAU* data quality evaluation concept will create a warning, which demands more frequent comparative measurements. An alarm is triggered only if the warning is validated with additional measurements. This approach will require man-power but guarantees the high data quality we expect from the system.

Level 2: Envisioned here are pre-configured graphical user-interfaces that allow the end-user to train the more advanced, but also more powerful multivariate models for time series analysis. These statistical methods (e.g. PCA, ICA, see Rosen *et al.* 2003; Yoo *et al.* 2004) will principally use different signals from one station but extension to data coming from other base stations is also foreseen. The training will be fully automated after preliminary quality checks using reference samples. In this way we guarantee that the training data set is of sufficient quality. Also in Level 2, the outcome of the data evaluation is a warning, which will trigger additional comparative measurements according to Figure 3.

Level 3: The last step is an advanced data evaluation and control concept, which includes additional knowledge for the specific measuring location and monitoring network. This might include process knowledge (e.g. biological processes at WWTPs, rates of change), known correlations between different measurements and also information from other measuring locations (e.g. by calculating mass balances or simpler: flow upstream < flow downstream) or redundant information. If sufficient validated experience is available, the end-user will have the option to use the available information for *data mining* and/or later for *control*. This would allow not only the detection of potential failures but also the identification of the source of the error. Based on this information, measures could be triggered to solve the problem or to react by changing to a safer operation of the monitored system (fault-tolerant control, see e.g. Devisscher *et al.* 2000; Lardon *et al.* 2004).

Maintenance on demand: To reduce the maintenance and service effort good maintenance planning is required. In addition to the static maintenance intervals (e.g. by exchange intervals of chemicals or spare parts), a dynamic monitoring concept (Thomann *et al.* 2002; Rieger *et al.* 2004) will be integrated (the left part of Figure 3). A new proactive maintenance concept is being designed to trigger experiments to evaluate, for instance, clogging of the sensor (by analysing different cleaning cycles) or calibration experiments where automated standard additions are carried out. The planning schedule is located on the main server but will be accessible from each point of the monitoring network. Maintenance actions will be announced in advance and have to be confirmed by the operator.

Messaging of warnings, alarms and status is an important feature for monitoring stations and particularly for stations in remote areas. A good warning system enables better service and maintenance intervals with more safety against unrecognized station failures. Different transmission types (e.g. SMS, pager, email) and user levels will be selectable.

Hardware

In the first step of the project, specifications were developed using the highest standards in terms of durability, robustness and data safety. As the mon*EAU* platform should be the same for all measuring locations, the set-up must consider all demands concerning space, energy demand and environmental conditions. That is the basic unit (a box with computer and I/O units) will be the same, but the power supply, data transmission and climate control options will vary so the station itself could be housed in a trailer or delivered as a stand-alone box, to be used directly with in-situ probes.

Sensors and actuators: Sensors and actuators are not seen as part of the mon*EAU* system. Our concept is to build the station flexible enough that all types of sensors and protocols can be connected. The base stations and also the central server are designed in a way that the connection of a new sensor triggers a procedure to provide storage capacity and standard visualization. Where possible, meta-data from the sensor (sensor configuration, dimension, measuring range…) is used to limit the required effort for the installation of new measuring devices. As plug-and-play is not feasible (due to standardization problems), a list with pre-configured sensors will be made available to facilitate the connection of new devices.

CONCLUSIONS

The mon*EAU* system will provide a high-level platform for all kinds of monitoring tasks, and eliminate the same design errors numerous other groups have done before. The flexibility of this new monitoring network concept enables different monitoring tasks as well as different measurement locations. As the most commonly used data transmission protocols (between sensor and base station) are provided, the user can select the best suited sensor for his application independent of specific monitoring station capabilities.

The software structure with a fixed framework as the backbone and connected modules (via API) combines robust operation with the flexibility to include new developments or to add individual applications. Free access to the interface settings allows easy development and implementation of user modules, but by locking the framework users are guaranteed robust basic functionality.

The hardware will fulfil the highest requirements regarding data safety, environmental conditions and robustness. Various telemetry options, low energy demand and proactive maintenance concepts enable remote use of the monitoring network.

However, the most important step forward is the advanced data quality evaluation concept helping to relate the measurements to the processes under evaluation and not to guesswork about data meaning. Most importantly, this evaluation concept will eliminate the danger of building more data graveyards.

ACKNOWLEDGEMENTS

The ideas behind mon*EAU* evolved from discussions with numerous people. We would especially like to thank Stefan Winkler (TU Vienna), Kris Villez (BIOMATH, Ghent University), Jens Alex / Mario Thron (ifak Magdeburg) and Lina Belia / John Copp (Primodal Inc.) for their most valuable input. Peter A. Vanrolleghem holds the Canada Research Chair in Water Quality Modelling.

REFERENCES

ATV-DVWK 2000 Merkblatt ATV-DVWK (now DWA) M 269: Prozessanalysengeräte für N, P und C in Abwasseranlagen. Hennef: GFA-Verlag [in German].

Beck, M. B., Watts, J. B. & Winkler, S. 1998 An environmental process control laboratory: at the interface between instrumentation and model development. *Water Sci. Technol.* **37**(12), 353–362.

Devisscher, M., Harmand, J., Steyer, J.-Ph. & Vanrolleghem, P. A. 2000 Control design of an industrial equalization system - Handling system constraints, actuator faults and varying operating conditions. In: *Proc. IFAC 4th Symposium on Fault Detection, Supervision and Safety for Technical Processes. Budapest, Hungary, June 14–16*.

Harmancioglu, N. B. & Alpaslan, N. 1994 Basic approaches in design of water quality monitoring networks. *Water Sci. Technol.* **30**(10), 49–56.

Jeppsson, U., Alex, J., Pons, M. N., Spanjers, H. & Vanrolleghem, P. A. 2001 Status and future trends of ICA in wastewater treatment–a European perspective. *Water Sci. Technol.* **45**(4–5), 485–494.

Lardon, L., Punal, A. & Steyer, J. Ph 2004 On-line diagnosis and uncertainty management using evidence theory–experimental illustration to anaerobic digestion processes. *J. Process Control* **14**, 747–763.

Lee, D. S. & Vanrolleghem, P. A. 2003 Monitoring of a sequencing batch reactor using adaptive multiblock principal component analysis. *Biotechnol. Bioeng.* **82**, 489–497.

Rieger, L., Thomann, M., Joss, A., Gujer, W. & Siegrist, H. 2004 Computer aided monitoring and operation of continuous measuring devices. *Water Sci. Technol.* **50**(11), 31–39.

Rosen, C., Röttorp, J. & Jeppsson, U. 2003 Multivariate on-line monitoring: challenges and solutions for modern wastewater treatment operation. *Water Sci. Technol.* **47**(2), 171–179.

Strobl, R. O., Robillard, P. D., Shannon, R. D., Day, R. L. & McDonnell, A. J. 2006 A water quality monitoring network design methodology for the selection of critical sampling points: part I. *Environ. Monit. Ass.* **112**, 137–158.

Thomann, M., Rieger, L., Frommhold, S., Siegrist, H. & Gujer, W. 2002 An efficient monitoring concept with control charts for on-line sensors. *Water Sci. Technol.* **46**(4–5), 107–116.

Vandenberghe, V., Goethals, P. L. M., van Griensven, A., Meirlaen, J., De Pauw, N., Vanrolleghem, P. A. & Bauwens, W. 2005 Application of automated measurement stations for continuous water quality monitoring of the Dender river in Flanders, BE. *Environ. Monit. Ass.* **108**(1–3), 85–98.

van Griensven, A., Vandenberghe, V., Bols, J., De Pauw, N., Goethals, P., Meirlaen, J., Vanrolleghem, P. A., Van Vooren, L. & Bauwens, W. 2000 Experience and organisation of automated measuring stations for river water quality monitoring. Proceedings 1st IWA World Water Congress, Paris, France, July 3–7 2000 (on CD-ROM).

Vanrolleghem, P. A. & Lee, D. S. 2003 On-line monitoring equipment for wastewater treatment processes: state of the art. *Water Sci. Technol.* **47**(2), 1–34.

Yoo, C. K., Lee, J.-M., Lee, I.-B. & Vanrolleghem, P. A. 2004 Dynamic monitoring system for full-scale wastewater treatment plants. *Water Sci. Technol.* **50**(11), 163–171.