

Crucial elements and technical implementation of intelligent monitoring networks

A. Winkelbauer, R. Fuiko, J. Krampe and S. Winkler

ABSTRACT

Growing complexity of water monitoring instrumentation leads to specialized solutions in respect to sensor integration across several measurement device suppliers. Despite efforts of standardization for data interfaces and protocols, problems regarding the combination of several devices to gain the complete picture in terms of water quality remain. This assessment, especially accomplished from the perspective of a catchment area, requires a transition from sole use of data collectors toward an implementation of intelligent measurement networks. Several challenges and bottlenecks concerning distributed data collection are discussed starting with data acquisition up to the user-scope of utilizing data processing software. Finally, experiences using automated data inspection and export tools are discussed and a brief overview of expectable long-term data availability is given.

Key words | automated online monitoring networks, data abstraction, data to information, plausibility assessment

A. Winkelbauer (corresponding author)

R. Fuiko

J. Krampe

Institute for Water Quality, Resources and Waste Management, Vienna University of Technology, Karlsplatz 13/E226-1, 1040 Vienna, Austria
E-mail: awinkelbauer@iwag.tuwien.ac.at

S. Winkler

HACH LANGE GMBH, Hütteldorfer Straße 299/Top 6, A-1140 Wien, Austria

INTRODUCTION

Operating water quality monitoring networks implies the requirement of several measurement probes for the assessment of the water quality state, which usually originate from different manufacturers. Even single manufacturers are applying different sensor interfaces in the analog and digital domain, partially based on manufacturer-customized control software. In addition, requirements for long-term measurement station operation referring to data stability have to be met. Besides that, automatic event-based cycle control, monitoring of proper operation and alerting in case of failure need to be included. Finally, the demands for proper data combination, presentation and export particularly with regard to the end-user perspective have to be satisfied.

One approach regarding the increase of data quality through so called 'functional analysis' for three parameters is shown in [Piñeiro Di Blasi *et al.* \(2013\)](#). The paper focuses on a mathematical discourse regarding outlier detection and less on the architecture and instrumentation of online stations. In [Ellis *et al.* \(2013\)](#) a strategy using online water quality monitoring for the prediction of bacteriological failures is shown. However, the implementation of data collection tasks and, just to mention one example, the implications regarding time-offsets

between probe sampling and data interpretation itself, fall short. [Strobl & Robillard \(2007\)](#) describe a distinguished approach on network design for water quality monitoring, giving efficient guidelines and recommendations, especially regarding the sampling location and frequency. Nevertheless, the remarks concern primarily the design phase of the monitoring network, and this paper, in contrast, is mainly based on a practical implementation.

To achieve the mentioned requirements the water monitoring network platform 'i^{TUW}mon' (intelligent information water monitoring networks) was developed and is described in detail by [Winkler *et al.* \(2012\)](#). The network platform consists of distributed measurement stations, a central database station and several data examination workstations in the form of workplace computers. This paper focuses on the pitfalls and bottlenecks during implementation of this technology. The domain is split into three logic layers: 'data acquisition', 'data to information' and 'data management and presentation'. The layers are described in detail after a brief overview of the therefore required measurement data channel abstraction. Following the description an overview of the implementation of the tool is given.

INTELLIGENT MONITORING NETWORKS BASED ON THE CONCEPT OF MEASUREMENT DATA CHANNEL ABSTRACTION

The simplified monitoring platform structure is shown in Figure 1. The data acquisition task is commonly performed directly at the monitoring station. In contrast, data from external sources, for example from meteorological stations and monitoring data for the verification of compliance of the corresponding water framework directive released by BMFLUW (2006), are collected or calculated based on already collected data, as database-administrated or database-calculated respectively, and therefore called ‘virtual channels’ inside the main memory. These channels are subsequently displayed directly on site, together with the sensor data from local probes (screenshot in Figure 2, left side: database icons in the channel list). ‘Data to information’ takes place directly at the monitoring station (aggregation, plausibility checks and alerting) and in the central database (metadata integration of calibration,

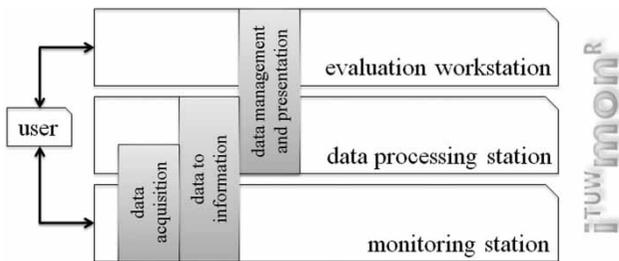


Figure 1 | iTUWmon monitoring platform architecture.

maintenance and laboratory data). Finally, data management and presentation is realized with iTUWmon.Examine (screenshot in Figure 2, right side). This tool is basically intended for data viewing and inspection; export routines to convert data into a standardized, interchangeable file format are also provided. The user interacts with each level of the structure by defining general setup and site-specific plausibility assessment parameters at the monitoring station and through performing further data inspection and processing data which are obtained on the user’s workplace computer, utilizing the connection of a virtual private measurement network.

The software architecture is implemented using the graphical system design language LabVIEW as an established standard for implementing measurement and automation technology systems. A brief overview of LabVIEW is given by Elliott et al. (2007). Rigorous considerations regarding software modularity and code reuse were obeyed so that for example the use of iTUWmon.Examine as a data inspection tool is possible for data sets which are imported semi-automatically as ‘virtual channels’ – even without the comprehensive use of a complete monitoring station installation.

The smallest common denominator to fulfill the monitoring tasks and meet the mentioned aspects is the abstraction of measurement values into abstract data channel objects (see Figure 3), which form the basic data structure in our solution. Every layer of the monitoring platform architecture works with a conglomerate of single data channels as basic units for measurement result processing and for metadata management during station operation.



Figure 2 | User interfaces: iTUWmon.Measurement (left) and iTUWmon.Examine (right).

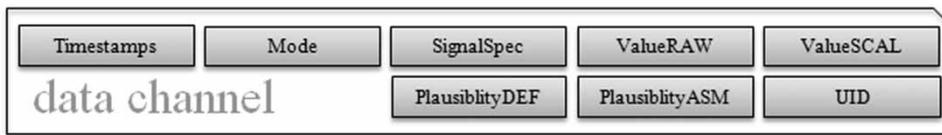


Figure 3 | Abstract data channel structure.

The data channel should be seen as an abstract, atomic data object, which includes the measurement timing behavior, actual conditions of operation and plausibility assessments across the named domains. The structure is instantiated multiple times in the form of a software object in all involved systems, starting at the basic data collection system until the end-user toolset for data viewing and export.

The data channel consists of the following parameters. 'Timestamps' contain the measurement timings in three time formats (time-UTC, -local and -MATLAB) and as a primary key element it acts as the first and most important data descriptor for the data structure. There are three measurement 'modes', which are selected statically in the case of modes 'General' and 'Maintenance' or dynamically in the case of exceptional measurement events. A typical event leading to more dense data recording in the time domain is a storm water event; during such events the mode 'Trigger' is selected automatically. Besides that, triggering of an automatic probe sampler is implemented and used to gain probes for further laboratory analysis. The 'Signal-Spec' (signal specification) contains information regarding measurement area and location, the physical data source, the parameter and unit definition as well as the data source description. The data source is segmented into several types of interfaces, e.g. analog input, diverse serial interfaces, ModbusTCP, Profibus, spectrometer interfaces or external data ('virtual channels'). The measurement values at a time instance are stored as 'raw' and 'scaled values' extended by information about configurable data aggregation timeframes, for example automatic aggregation of plausible single measurements to 1 hour averages. The structure is implemented as an array of values, enabling efficient storage including absorption spectra where one array element corresponds to an absorption value at a specific wavelength. Definitions and results of onsite plausibility assessments (for details see [Winkler et al. \(2013\)](#)) are stored in the plausibility structures facilitating automatic data filtering for near real-time reaction on events ('PlausibilityDEF' for definitions and 'PlausibilityASM' for assessment results). The type of assessments is user-configurable and the extent of calculations is only limited by the processing power on site. UID stands for unique measurement identifier and acts as a data

fingerprint for single measurements. Hardware serial numbers, network environment, elapsed time ticks since a reference date and further aspects are used for calculation of the UID. In case of data faults the complete measurement data set might be rebuilt using this UID and redundant repositories or databases which contain raw measurement data sets, as a sort of hot backup.

Data acquisition

The formal tasks of this layer are the sensor interface integration and the operational timing control. In the implemented monitoring network, sensor data are processed in terms of abstract monitoring data channel objects as shown above. The aim of device integration is to get as much information as possible in the shortest way from the sensor as possible. It is given that, the more digital the device interface is, the more information beside the raw measurement value is transmitted into the measurement system.

Additional information could be, for example, timestamps regarding the exact sampling instant, error messages generated by probe controller systems and calibration/maintenance demand. A wide range of sensor connection possibilities, starting from analog signal sampling up to complete remote control of sensing software using all-digital data transfer, are implemented through integration of data protocols offered by the manufacturer. Analog probe sensing together with simultaneous digital data transmission of measurement channels and operational parameters leads to a higher number of logical channels than physically installed water quality probes, but shows higher data strength against influences like noise and value range exceedance.

In the analog domain, data acquisition is a straightforward task using common analog voltage- and current-I/O interfaces (International Electronic Commission IEC 60381-1 (1982) and IEC 60381-2 (1987)) for sensor connection and signal processing. The drawback of easy integration is the necessity of providing further information to interpret the signal. The scaling of the analog input signal is usually fixed, and manual configuration on site should meet the range of the analog/digital-converters as effectively as possible to reduce noise impact. Owing to

the missing option of adjusting the physical signal scaling online or automatically and corresponding to the actual signal range on probe-side, the use of analog interfaces leads to unfavorable results for data quality. For example, in the case of rain events, thereby the value ranges are extended multiple times compared to the normal operation state, and the analog signal in this case is usually stuck at full-scale levels. Another problematic aspect is the need for additional digital/analog- and analog/digital-conversions; therefore, additional signal noise is generated. This noise is undesirable especially in the case of utilizing complex, up-to-date probe controllers, which already use digital data processing internally and therefore shall be incorporated into the network using existing digital interfaces.

In the digital domain, several pitfalls mentioned above are addressed through consistent digital signal propagation. Industrial standards for interfaces are applied and, in the majority of cases, the realization is not properly documented. The disadvantage becomes apparent in the lack of conformance in the practical realization of these standards; various dialects and modifications in the implementation of the standards can be found, although compliance is assured at vendor side. In practice an increased effort for signal integration is the consequence.

In our monitoring solution different digital interfaces like RS232/485, Profibus, ModbusTCP and TCP/UDP-port based access of measurement signals were used, and modular software structures for various sensor controller classes were implemented. Besides the implementation of data exchange protocols using industrial standards, data transfer from the probe controllers is occasionally possible only by using the vendor-supplied software. In the best case communication with this software via programmable software interfaces is found; in the worst case the implementation of text file parsers for the collection of results in order to establish data communication between the measurement aggregator and the probe controller is the only way to achieve proper data communication. In some cases watchdog-mechanisms, which detect and completely re-initialise the vendor-provided software tool in case of software failures, were constructed. This can be described as 'remote controlling of literally non-remote-controllable software' and led to time-consuming implementation tasks.

Time synchronization and operational timing control of measurements and cleaning cycles are a fundamental prerequisite for comparability of data within a distributed measurement network. The lack in treatment of the temporal dimension of measurement data is not least shown in missing system time synchronization of probe controllers, which is in

fact still not provided in the majority of cases. This leads to the situation whereby aspects regarding the time zone, daylight saving time and simply the temporal deviation of clocks are not observed. Besides that, information regarding the sampling time and the associated validity of the signal is lost due to quasi-continuous availability of measurement signals on the interface, often combined with the lack of an external measurement trigger. To cope with these pitfalls, time synchronization based on the Network Time Protocol (NTP) following the Internet Engineering Taskforce [IETF RFC 5905 \(2010\)](#) was implemented on suitable devices within the network. During measurement operation every signal channel is sampled depending on the specific delay-time (see international standard [ISO 15839 \(2003\)](#)) and automatically aligned to general measurement time instants, ignoring the non-synchronized controller timers. The controller time base therefore is interchanged with the network-wide, validated data acquisition timescale and measurement data collected at various points in a catchment area become comparable.

The implemented control of measurement, cleaning and maintenance cycles has to ensure correct measurement timings and therefore interference of active measurements is impossible through postponing of lower prioritized tasks a certain amount of time. For event-based cycle control the measurement cycles as well are adjusted dynamically. To reach this requirement an algorithm based on non-preemptive, dynamic scheduling (see [Kopetz 2011](#)) was implemented. The adaptive character is given by the possibility of rearrangement of already scheduled tasks due to necessary interventions based on fault removal; e.g. a measurement cycle is postponed because of a sample supply pump failure. In this case, an automatic pump sieve clean cycle is scheduled, after which the postponed measurement cycle is caught up.

Measurement abstraction utilizing the proposed data channel structure, modular software architecture for data sampling in the analog and digital domain, and strict consideration of timing aspects enable the transition of data flow into the direction of the data processing layer called 'data to information'.

Data to information

The main challenge in managing extensive data sets is the generation of information out of raw data, possibly utilizing automatic distinguishing of not relevant data sets, preferably in real-time and based on criteria related to the local station environment. The logical layer 'data to information' is responsible for data aggregation and evaluation through plausibility checks, combined with trigger condition generation and/or

automatic alerting in case of value range excess. Plausibility checks consist of statistical assessments like allowed minimum/maximum signal range, based on actual measurements; maximum step and minimum signal change, based on signal history; and coefficient of variance, and take place directly at the monitoring station. In addition, quality assessment procedures following dilution behaviors were examined. For detailed information regarding these checks see [Winkler *et al.* \(2012\)](#) and [Winkler *et al.* \(2013\)](#).

To get the full picture in regard to data completeness, beside water quality probes, several operating parameters like cabinet temperature, air pressure and pump flow have to be monitored. Together with the raw water quality measurement data, these parameters serve as input signals for event detection and triggering of fault-correction mechanisms in case of operational errors. Even more, these signals act as the basis of reliable long-term operation. Fault detection, near real-time reaction times and reduced maintenance demand can be achieved through implementation of statistical onsite data plausibility checks called 'CoarseChecks', combined with automated user notification. Compared to post-processing, these structures enable early detection of faults during data collection and facilitate the differentiation between real system dynamics and device faults. All checks are done online, in near real-time and directly at the station. The plausibility assessment is stored together with the raw data in the same data channel objects as used in the 'data acquisition' layer. Again, utilizing this approach, different data structures, typically vendor supplied, no longer occur, which leads to streamlined data processing throughout the system.

Data management and presentation

In this layer the combination of several data sources within a measurement network and the appropriate presentation of data to the end user, including export functionality, take place. Several requirements for proper operation of data synchronization and long-term provisioning of data archives have to be met. The key requirements, as seen from the user perspective, are sufficient data availability and short software reaction times combined with intuitive graphical user interfaces.

For data transfer and remote access, the measurement network is built using a virtual private network based on high-availability, industrial 3G-routers installed at every station. Data are stored in local databases and a central data archive in a redundant manner. Asynchronous data replication, independent of measurement cycle time instants, is responsible for balancing of data amounts

between the databases. The replication is on one hand used for post-synchronisation in case of connection errors and on the other hand used for passing measurement data to the project partner's databases (for mechanisms see [Wolfson & Jajodia \(1992\)](#)).

Short reaction times for data viewing and export utilizing easy to use software toolkits are expected from the end user nowadays. In the particular case of long-term trend examination, which means processing of a huge amount of data in a short period of time, sufficient dimensioning of database storage machines and software tuning are required.

To overcome these requirements, several database management systems were evaluated and PostgreSQL, issued by [The PostgreSQL Global Development Group \(2014\)](#), was chosen. This open-source enterprise database system provides, among other features, good basic performance, several opportunities of particular tuning and receives good support through the database online community.

For data inspection and export the tool *i^{TUW}mon.Examine* was developed. This software allows inspection of long-term behavior, based on the abstract channel concept described above. The work routine from end-user scope can be described as 'select, examine and export data'.

For data selection the graphical user interface offers all common signal representation and manipulation capabilities using free-transform, multi-axes graphs and optional overlying measurement cursors for highlighting extreme values.

Powerful export mechanisms allow the automatic combination of arbitrary channels using a two-dimensional table including automatic data alignment of all channel data into a common time-column for further processing. As the data basis for table-based data export, the user can select between the full data set containing all channels between start and end date and the currently displayed data set displayed as a graph, revealing a possibly interesting part of signal behavior. In addition, data export as a picture using common data formats is achievable. Again, timing of measurements play a leading role, and data sets across several monitoring stations become comparable in the aspect of signal sampling timestamps.

RESULTS: INTELLIGENT INFORMATION WATER MONITORING NETWORKS IN PRACTICE

In this section data availability and user experience of data inspection using *i^{TUW}mon.Examine* are discussed. The long-term data availability of three stations, one *in situ* and two more complex bypass stations, using this data

processing tool is compared. A typical monitoring station is equipped with probes for measurement of ammonium-nitrogen, chloride, conductivity, dissolved organic carbon, nitrate-nitrogen, pH-value, potassium, spectral absorption coefficient (SAC), total organic carbon derived from SAC, turbidity, diverse medium and device temperatures and further sensors for proper operation control like compressed air pressure, cabinet temperatures and air humidity as well as power supply measurement variables.

At an *in situ* station the sensors are placed on a submersible platform directly into the measurement medium. The bypass stations contain a pump supplying the medium into a measurement tub, where the probes are mounted.

The data availability in 2012 versus station complexity is shown in Figure 4 on the left side. The availability calculation is based on the summation of single measurement instants with applied versus not applied CoarseChecks. These checks offer an automatic quality assessment for collected data subsequent to every single measurement directly at the station. They are based on site-specific, statistical tests and provide a first indication of data quality achieved. The results of the checks are available for inspection and post processing in *i*^{TUW}mon.Examine. In this specific case, data availability using bypass stations is notably higher through the lower amount of rejected data sets. Applying the automated CoarseChecks around 15% of single measurements at the bypass stations compared to 25% of data at the *in situ* station are discarded. The difference in data rejection rate can be explained by:

- (1) ensured sample stream through flow rate monitoring and pipe trace heating at bypass stations, especially in winter time (frozen surface water);

- (2) environmental-sealed housing and enclosure of measurement apparatuses in measurement containers;
- (3) better intervention options in case of failure (e.g. remote pump clean cycle triggering).

Consequently, it can be said that, despite the higher effort and mechanical complexity of a bypass system, implementation is worthwhile in respect of data completeness. In 2012 on average 60,000 (passed data) out of 71,500 (raw data) single measurements per channel were collected. In the year 2012, utilizing about 35 data channels per station, a total of 6,200,000 (passed data) out of 7,600,000 (raw data) data sets in the described abstract channel structure are stored in the databases in a redundant way. These data sets represent the aggregation of all channels of all three measurement stations in the catchment area; typically around 2,000,000 checked data sets are gathered per single station. Raw measurement data are residing in the database and so postmortem analysis and retrospective refining of statements through the dynamic adaption of plausibility assessments are possible.

Results show that the availability of stations with increased station complexity through medium bypass situation and automated pump control is at least equivalent, and perhaps could be higher, than using an *in situ* station with reduced complexity. This is applicable especially for station 2 operating *i*^{TUW}mon since 2010. Extensive experiences in long-term water quality monitoring have been acquired.

The graph on the right hand side of Figure 4 shows the data processing benchmark using *i*^{TUW}mon.Examine as analysis tool. The objective of this benchmark is to give an impression of what the end user could expect from data

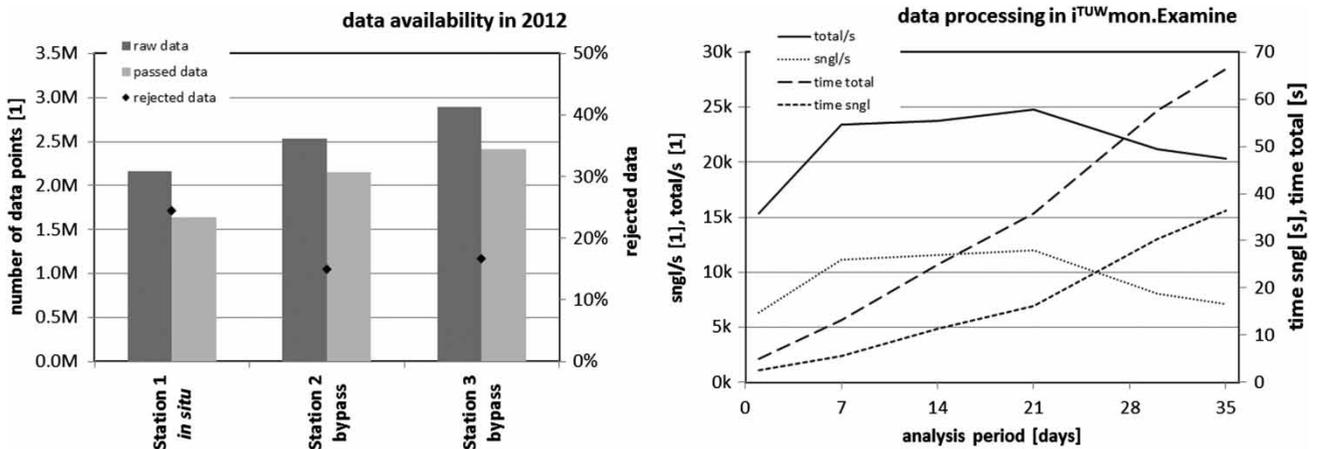


Figure 4 | Data availability (left) and data processing utilizing *i*^{TUW}mon.Examine (right).

processing within a reasonable amount of time. Depending on the user-chosen analysis period, an increasing number of data points have to be processed for graphical display and export. The processing task includes loading the complete data set of the selected time range, filtering regarding the plausibility assessments and, in the case of exporting tasks, preparation of export file structure and initialization of data selection, if only a subset of channels should be exported. The automatic alignment of data channel objects into a common timestamp column is designed as an online task, which takes place inside the database utilizing aggregation functions triggered by data insertion during measurement operation. Therefore, combining of time series is done directly after signal sampling at the station, and necessary processing time is multi-partitioned over the complete operation time domain, resulting in immediate data availability for the end user during an export task.

Conventional data are referred to as 'sngl' and conventional combined with spectral data, stored as vector objects, are referred to as 'total'. As can be seen in the figure the number of data points increases linearly in the first seven analyzed days and then remains constant. For analysis periods of more than 21 days the processing capacity decreases. As additional information, on the right axis the cumulative processing time is shown, which more or less increases linearly up to an analysis period of 21 days. For longer analysis periods the processing time increases more strongly due to database performance degrading.

The average data loading procedure on a machine equipped with an Intel Core 2 Duo E8400 processor for a 3 weeks data record including spectra lasts about 38 seconds and results in over 870,000 single measurements, which are processed and formatted for automation-assisted viewing. That means around 23,000 single measurements are processed per second, whereby spectral values of UV-visible spectrometers are processed around three times faster compared to conventional data. Spectral data are processed column-wise and database performance is notably higher for vectorial data sets. Results show that 21 days as a basis for trend analysis offers a good tradeoff between processing time and data extent. Similar results are obtained during export of merged data. In this case the memory-intensive data display task is bypassed and the processed data are written directly into the export data file, resulting in even higher throughput. Apart from manual adaption of database configuration to achieve higher database throughput and query performance, the implementation of solid-state drives as the storage basis on the database server was a great leap ahead. These efforts resulted in a performance

factor of approximately 40 for the data processing task compared to the use of standard hard-disks and therefore extensive data records become reasonably manageable.

Using the proposed monitoring platform, actual measurements are available on every data processing workstation within the institute approximately 1 minute after the onsite sampling takes place, and the water quality picture based on plausibility checked data sets is available more or less 'live'. Utilizing our proposed structure allows the operation of centrally managed but physically distributed data acquisition stations in a comfortable and resource efficient way, over years. Automatic alarming of faulty system states (e.g. pump failures) and automatic data plausibility assessments (e.g. sensor drifting estimation) offer technicians the ability to plan maintenance cycles and demands.

The specified software architecture allows the use of several software modules as a standalone system; an actual project is based on a small amount of probe controllers onsite and a separate measurement and data processing workstation located in-house. In this case, the software interface for the end user is available using the virtual private network, in the same transparent manner as before, although the measurement workstation is locally separated from the measurement area.

The novel character of the described approach can be seen in the sum of separate technical aspects addressed during the implementation since 2010. The developed tool includes comprehensive interfaces to a nearly complete range of market-available water quality probes and controllers. The so-collected dense, statistically approved and plausible data are available over a long period of time and ready for immediate 'online' use.

REFERENCES

- BMFLUW (Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management Austrian regulation on the Monitoring and Quality of water bodies) 2006 Gewässerzustandsüberwachungsverordnung (Austrian regulation on the monitoring and quality of water bodies). BGBl. Nr. 479/2006 i.d.g.F. Sektion VII/Abteilung 1, Nationale Wasserwirtschaft, Vienna, Austria.
- Elliott, C., Vijayakumar, V., Zink, W. & Hansen, R. 2007 [National instruments LabVIEW: a programming environment for laboratory automation and measurement](#). *Journal of Laboratory Automation* **12** (1), 17–24.
- Ellis, K., Mounce, S. R., Ryan, B., Templeton, M. R. & Biggs, C. A. 2013 Use of on-line water quality monitoring data to predict bacteriological failures. *Procedia Engineering* **70**, 612–621.

- IEC 60381-1 1982 Analogue signals for process control systems. Part 1: Direct current signals. Beuth Verlag, Berlin.
- IEC 60381-2 1987 Analogue signals for process control systems. Part 2: Direct voltage signals. Beuth Verlag, Berlin.
- IETF RFC 5905 2010 Network Time Protocol Version 4: Protocol and Algorithms Specification. <http://datatracker.ietf.org/doc/rfc5905/> (accessed 6 June 2014).
- ISO 15839 2003 *Water Quality – On-Line Sensors/Analyzing Equipment for Water – Specifications and Performance Tests*. International Organization for Standardization, Geneva, Switzerland.
- Kopetz, H. 2011 *Real Time Systems, Design Principles for Distributed Embedded Applications*, 2nd edn. Springer, New York.
- Piñeiro Di Blasi, J. I., Martínez Torresb, J., García Nietoc, P. J., Alonso Fernándezd, J. R., Díaz Muñizd, C. & Taboadaa, J. 2013 *Analysis and detection of outliers in water quality parameters from different automated monitoring stations in the Miño river basin (NW Spain)*. *Ecological Engineering* **60**, 60–66.
- Strobl, R. O. & Robillard, P. D. 2007 Network design for water quality monitoring of surface freshwaters: a review. *Journal of Environmental Management* **87** (4), 639–648.
- The PostgreSQL Global Development Group 2014 PostgreSQL – the world’s most Advanced Open Source Database. <http://www.postgresql.org> (accessed 6 June 2014).
- Winkler, S., Fuiko, R. & Winkelbauer, A. 2012 iTUWmon – a monitoring network platform for automated data plausibility assessment and data integration. In: *IWA New Developments in IT and Water Conference, Amsterdam*.
- Winkler, S., Kornfeind, L., Fuiko, R. & Winkelbauer, A. 2013 Separating real system dynamics from measurement failures and artifacts by means of automated data quality assessment methods. In: *ICA2013, 11th IWA Conference on Instrumentation Control and Automation, Narbonne*.
- Wolfson, O. & Jajodia, S. 1992 Distributed algorithms for dynamic replication of data. In: *ACM Proceeding: PODS '92 Proceedings of the Eleventh ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, New York*, pp. 149–163.

First received 17 February 2014; accepted in revised form 6 October 2014. Available online 30 October 2014