Advanced Monitoring of Wastewater Quality: Data Collection and Data Quality Assurance

Janelcy ALFERES[1*], Coralie LAMAIRE-CHAD[1], Ravi Kumar CHHETRI[2], Carsten THIRSING[3], Anitha Kumari SHARMA[2], Peter Steen MIKKELSEN[2], Peter A. VANROLLEGHEM[1]

[1]modelEAU, Université Laval, 1065, Avenue de la Médecine, Québec, Canada QC G1V 0A6
[2]DTU Environment, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
[3]Lynettefælleskabet, Refshalevej 250Refshalevej 250, DK-1432 København K, Denmark
[*]Corresponding author
Email: Janelcy.Alferes@gci.ulaval.ca

ABSTRACT

Reliable water quality monitoring is increasingly recognised as an essential element of a strategy to reach current environmental quality objectives. In the last few years, continuous water quality monitoring has demonstrated to be useful in capturing the dynamics of sewer and wastewater systems and in evaluating the impact of discharges on the receiving water bodies. As measurements are carried out under arduous conditions, practical implementation of such monitoring systems entails several challenges, and automation of data collection and data quality assessment has been recognised as a critical issue. In this paper, a data quality assessment strategy is presented to achieve efficient water quality monitoring in real-world scenarios. Next to practical aspects concerning installation and maintenance of sensors, the paper also presents a software tool aimed at assessing the quality of the data being collected. In this paper, results showed the successful implementation of the proposed strategy for collection of water quality data at the inlet of a wastewater treatment plant.

KEYWORDS

Data quality assessment, filtering, online wastewater monitoring, wastewater systems

INTRODUCTION

Integrated management of water, stormwater and wastewater is nowadays required to fulfill current and future environmental quality objectives. Within the different water legislations existing worldwide, a consistent monitoring strategy is becoming a key component for integrated water qualityevaluation. The variability and complexity of the different processes and their interactions is promotingtheprogressive transition from measurements based on discrete sampling and lab analysis to acquisition of high frequency data with sensors. Possible benefits of larger databases are a better understanding of the ongoing processes and the impact of fluctuations in water quality, and the identification and description of pollution dynamics allowing remedial actions to be taken (Metadier and Bertrand-Krajewski, 2012).

Recent important technological developments regarding on-line water quality sensors and data acquisition systems have encouraged the implementation of situ monitoring stations to collect information-rich data sets (Winkler et al., 2002; Rieger and Vanrolleghem, 2008). However, such stations result in the generation of a huge amount of real-time data that are difficult to interpret without automated tools, eventually leading to data graveyards. Moreover, the collected data will only contribute to water qualitymanagement and evaluation if it is representative and correctly interpreted. Furthermore, since the sensors are usually subject to many disturbances, data reliability can suffer, leading to faulty conclusions and incorrect use of the data. This becomes particularly critical under wet weather conditionsand in sewers and in the primary stages of wastewater treatment systems where clogging, fouling and flooding are common situations. Regular maintenance of the sensors and application of a comprehensive data assessment methodology is therefore necessary to guarantee that high quality data is being collected.

Although some methods have been developed for data quality assessment purposes in different fields (Venkatasubramanian et al., 2003), practical application of those methods in the water sector faces important obstacles and in practice inefficient and time consuming manual inspection of the data is still the most applied procedure (Branisavljevic et al., 2010).This paper will illustrate these different issues for a full-scale situation, more specifically in the catchment area of the Lynette Wastewater Treatment Plant (WWTP), Denmark.Aspects like installation, comprehensive data collection and data quality assessment for further data exploitation will be addressed in the paper.

MATERIALS AND METHODS

Data collection

Extensive water quality data sets have been collected by two in-situ automated monitoring stations (RSM30, Primodal Systems, Canada)installed in raw sewage and at the outlet of a primary clarifier line of the Lynette Municipal Treatment Plant in Copenhagen, Denmark (Figure 1). Both monitoring stations include sensors for conventional physical-chemical parameters (temperature, pH, turbidity, conductivity), a UV/VIS spectrometer measuring total suspended solids (TSS), nitrate, total and filtered chemical demand (CODt, CODf) and ion selective electrodes for ammonia, potassium and chloride. Data were recorded at intervals of 5 to 60 seconds generating large information-rich data sets.The implementation of such a monitoring system has faced different challenges. The inlet of WWTPs probably being the hardest measurement location, especially under wet weather conditions, special attention must be taken concerning equipment installation, sensors maintenance and quality control of measurements to ensure an effective monitoring.

Installation of monitoring stations. Both monitoring stations have been placed close to the monitoring sites (raw sewage and outlet of a primary clarifier line). Two small wooden houses (Figure 1a) have been used to provide a robust, waterproof, corrosion and weather resistant protection for electronic and acquisition devices within the monitoring stations (Figure 1b). To facilitate the daily review of data and performance diagnostic for both the sensor and the system, both sites were provided with telemetry capabilities.

*Installation of on-line sensors.*A key aspect of the installation of sensors in the difficult environments they are required to work in, is the selection of a proper measuring location, which must be well representative of the water body under study under all flow conditions.Another aspect is the safeaccess to the sensors for maintenance and control sampling. The distance from locations with power supply and communication means must also be taken into account and sensors must be protected against damage caused by debris carried by the water. A fixed structure in aluminum with rails, mounting brackets and a lifting mechanism was chosen for sensor deployment providing a stable fit and easy access (Figure 2).

Figure 1.Installation of monitoring station: Housing (left); RSM30 monitoring station (right)

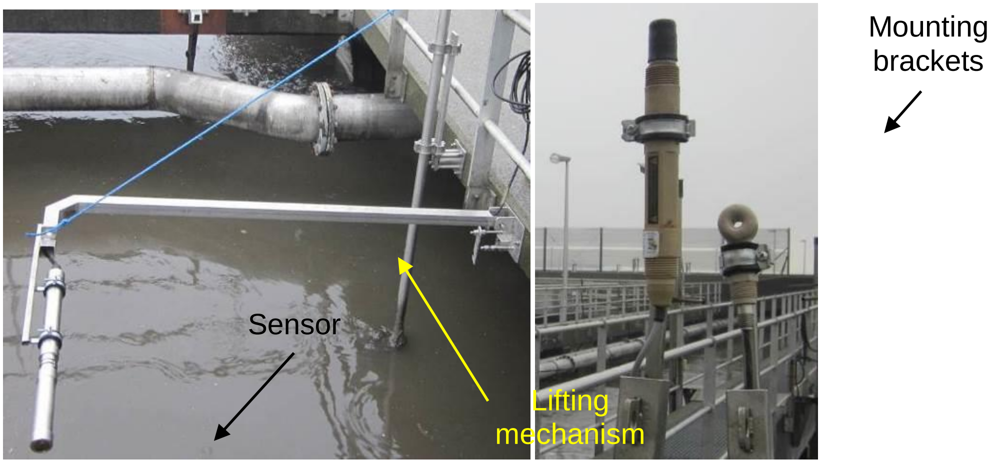

Figure 2.Details on sensors installation: Left: UV/Vis; Right: pH, conductivity.

Quality control of the measurements: Under the challenging measuring conditions that prevail in water systems (clogging, fouling, flooding, electrical interferences, etc.), sensors are subject to problems that compromise the precision and the reliability of the collected data (Figure 3a). To ensure good data quality of the on-line measurements a quality control strategy based on a proper maintenance routine is critical.

Figure 3.Maintenance routine:Clogging (left) and fouling (middle) of on-line water quality sensors, validation test of pH sensor (right).

The developed maintenance routine (summarized in Figure 4) includes site and instrument inspection, manual cleaning, validation, calibration and discrete sample collection. The validation comprehends the verification of sensors with a different method, e.g. by using handheld probes, standard solutions or by comparing results obtained from lab analysis of grab samplestaken at the sensor's location (example in Figure 3b). Control charts (Thomann et al., 2002)are then built with appropriate out-of-control criteria to detect systematic errors and to determine calibration needs and suggestions for corrective actions. In the experience of the authors the frequency and level of effort for maintenance varies from once/week to once/two weeks depending on the sensor type and location. Insufficient cleaning results in more extensive drift in the measured data, making them unusable for different purposes.

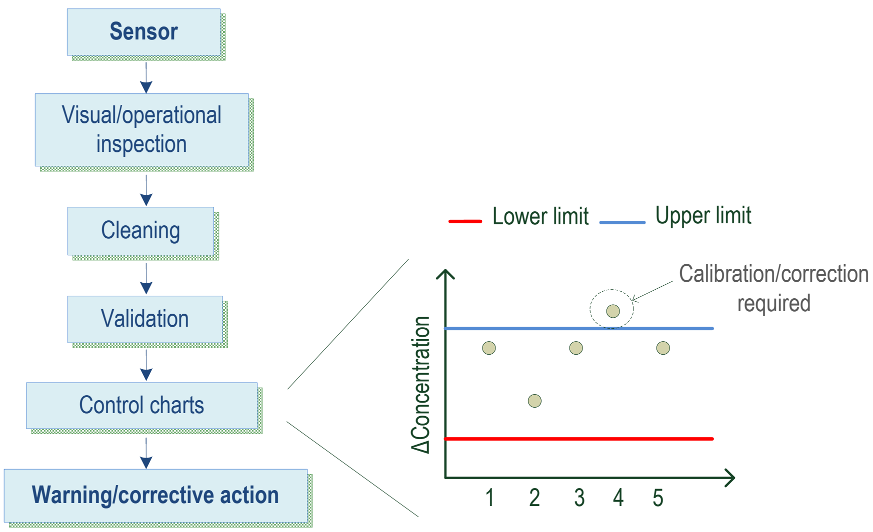

Figure 4.Maintenance routine developed and applied in this work.

Data quality assessment

To guarantee the desired high quality of data being collected, a complete quality assurance program will includea thorough quality assessment of the on-line time series so as to detect deviations from the normal state of the measuring system. For this purpose, advanced software tools for automatic data quality evaluation, with a practical orientation, were developed and applied. The developed methodology, based on statistical and forecasting methods, comprises two main steps (Figure 5): (1) pre-treatment of the data including detection of unreliable data, and (2) fault detection. The final objective is to come up with "quality validated" time series. In this paper the methods are limited to univariate series.

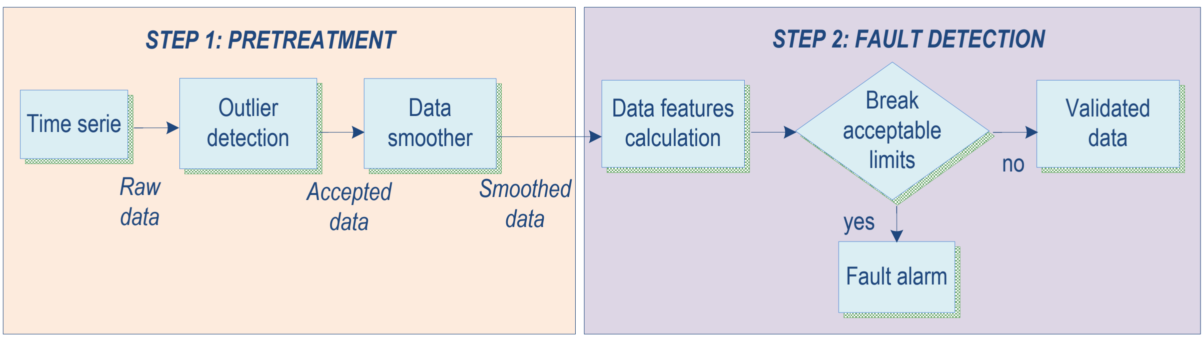

Figure 5. Proposed univariate time series analysis for data quality assessment.

The first step aims at detecting and removing doubtful data in the form of outliers, defined as data that appear to be inconsistent or that deviate importantly from the rest of the data (Maimon and Rockach, 2005).A univariate approach based on forecasting methods that use historical datato determine “normality” is implemented to create new pre-treated and smoothed time series that can subsequently be used for fault detection purposes. Specifically, with a minimum of *a priori* assumptions, an autoregressive model is identified for predicting the expected value of the data x and the standard deviation$\Delta$of the prediction error at the next time step, $T+1$. Supposing a normal distribution for the forecast error (Dochain and Vanrolleghem, 2001), the local variance σ_e^2 can be estimated as $\hat{\sigma}_e = 1.25\hat{\Delta}_T$, and a prediction interval is then definedas $xlim_T = \hat{x}_T \pm K\hat{\sigma}_{e,T}$, with K a user-defined proportionality constant. At time T, if a new data point is outside the prediction interval it is considered an outlier and it is replaced by its forecast value. After the outlier detection and replacement phase, the resulting time series is passed through a filter by using a kernel smoother with proper bandwidth (Sheather, 2004) to remove noise. This phase allows reducing the corruption of subsequent statistical calculations. In the second step, some statistical data features are calculated over the new filtered time series. These data features are aimed at evaluatingthe goodness of both the forecasting model and the resulting smoothed data. First, when initiating the fault detection step, acceptability limits have to be defined for each of the proposed data features. This is based on the evaluation of representative historical “good” data. Faults can now be detected in the new filtered time series by evaluating whetherone, several or all data features exceed the acceptability limits. According to that approach, each data point is classified as valid, not valid or doubtful (meaning that posterior analysis must be carried out). Validated data can subsequentlybe used to analyze the involved processes, e.g. to study the effect of dry and wet weather conditions in the sewer system, shock loads, and correlation between individual variables(e.g. Sharma et al.,2013).

RESULTS AND DISCUSSION

The quality assurance program described above has been followed during a five month (December 20th 2012 - May 20th 2013) measurementcampaignconducted at the Lynette WWTP, including dry, snow and wet weather periods. Thanks to high frequency sampling, data has allowed to identify, understand and describe the dynamic behaviour of the WWTP influent data for the period under study. For instance, Figure 6 shows the effect of road salt addition and snow melting on the conductivity measurements (shaded section a), and the effect of wet weather conditions on ammonia, conductivity and pH values (shaded section b).

The application of a rigorous maintenance routine for quality control has permitted to reduce the occurrence of bad quality data sequences and take corrective actions when needed. Detection of out-of-control situations by using control charts has allowed maintaining a good agreement with laboratory and portable sensor measurements. For example, Figure 7 nicely shows the effect of calibration of ammonia measurements around April 11th, when the control chart (in this case built with results from laboratory grab results) announced an out-of-control situation. Red points in Figure 7 represent results from both grab samples (for ammonia and chloride) and a handheld probe (conductivity) collected twice a week for research purposes.

Despite the extensive maintenance efforts, the harsh measurement conditions at the measuring sites (raw sewage) made that data was still affected by different kinds of faults, such asa large abundance of periods with extensive measurement noise. Figure 8 shows the result of the application of the data quality assessment method for a turbidity time series collected at the inlet of the primary clarifier, location characterised by its challenging measurement conditions. The first subplot represents the results for the first step of the data quality assessment procedure – pre-treatment of the data. The consecutive subplots illustrate the statistical data features calculated in the fault detection step.

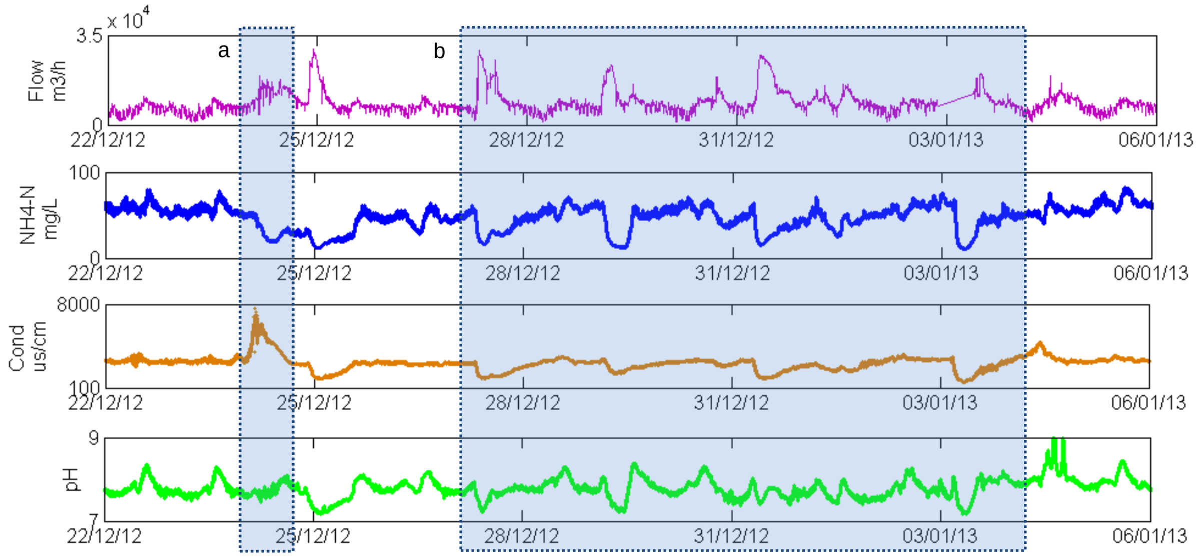

Figure 6.On-line raw sewage water quality measurements. Effect of road salt addition and snow melting (shaded section a), and wet weather conditions (shaded section b).

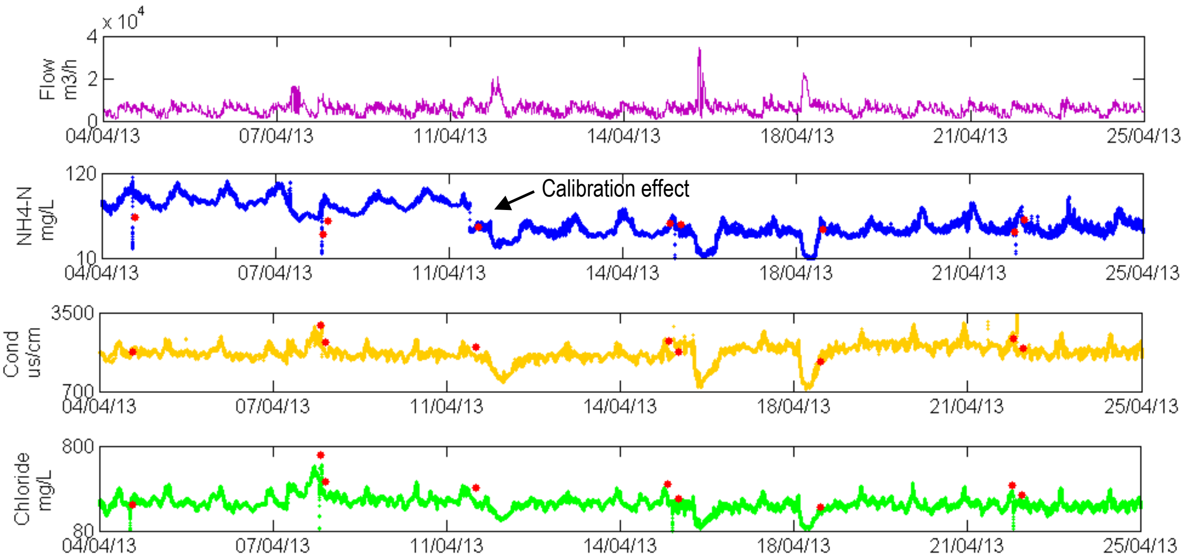

Figure 7. Flow, ammonia, conductivity and chloride measurements of raw sewage: Agreement between on-line (lines) and lab measurements (symbols) conducted twice a week.

The dynamic calculation of the prediction interval (blue and red lines) allows for adaptation of the signal processing to the time-varying measurement conditions and enables detecting some outliers, identified as black points in the first subplot. Once these have been removed and replaced by their forecast value, noise is removed, resulting in smoothed data (green line in first subplot) that allow interpreting the dynamics of the turbidity time series. Subplots 2 to 5 show the data features calculated over the smoothed time series which allow evaluating the degree of “confidence” of the data. The data features encompass the percentage of outliersreplaced by the forecast values, the autocorrelation of the residuals, the slope, the residuals' standard deviation (RSD) and finally the range of values deemed acceptable. Red horizontal lines represent the acceptability limits obtained from a "normal" operational period.

Some abnormal behaviour is detected for example in period I with RSD values exceeding those expected under normal conditions. This period also presents a high percentage of outliers. In period II an abnormal variation in turbidity measurements is detected by both RSD and slope values indicating unusual dynamics in the variable (increased more than three times compared to normal). After comparing the whole set of data features with the defined acceptability limits a data quality label is given to each value (Figure 8, last subplot). Data that passes all tests is classified as valid (marked as 0 - green), data that has failed some tests is considered doubtful (1 - orange) and finally data may be considered not valid (2 - red). For the period under study around 92% of the data was found valid, an excellent result compared to typical rejection rates (e.g. Thomann, 2008; Schilperoort, 2011). Figure 9 shows the validated time series obtained after applying the proposed data assessment method to a TSS time series. Most of the filtered data agree with the laboratory data.

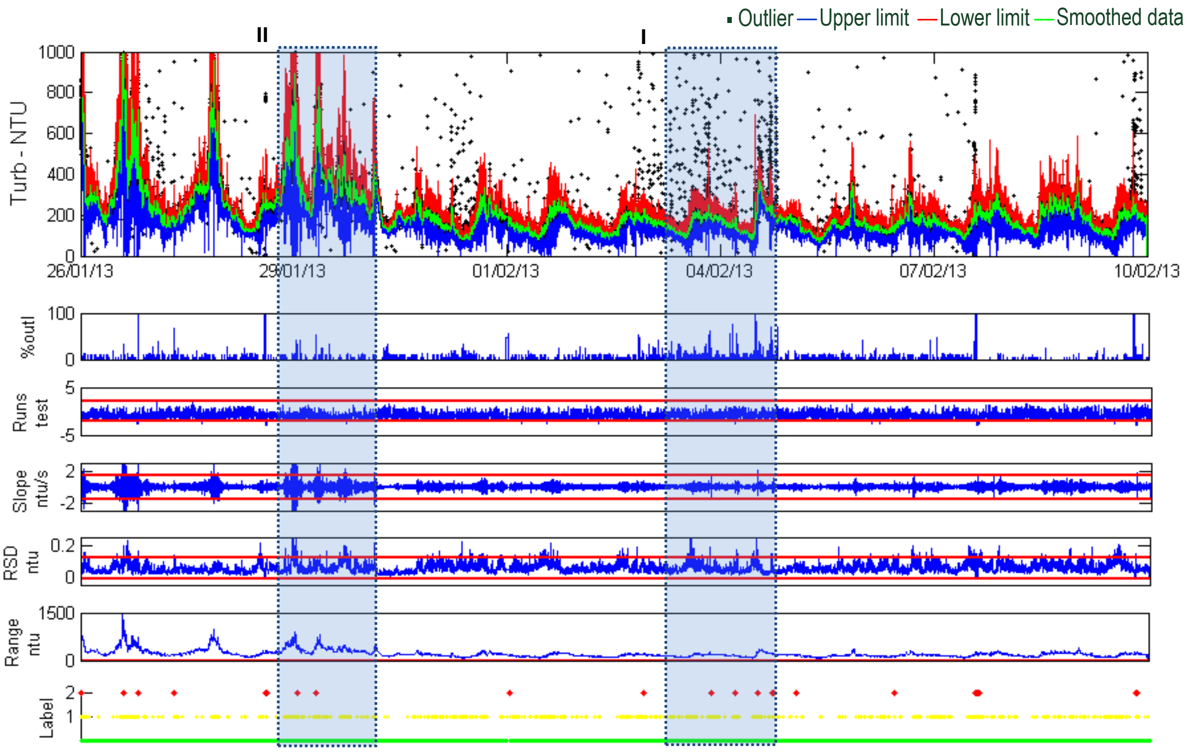

Figure 8. Application of the data quality assessment methodology for a turbidity time series.

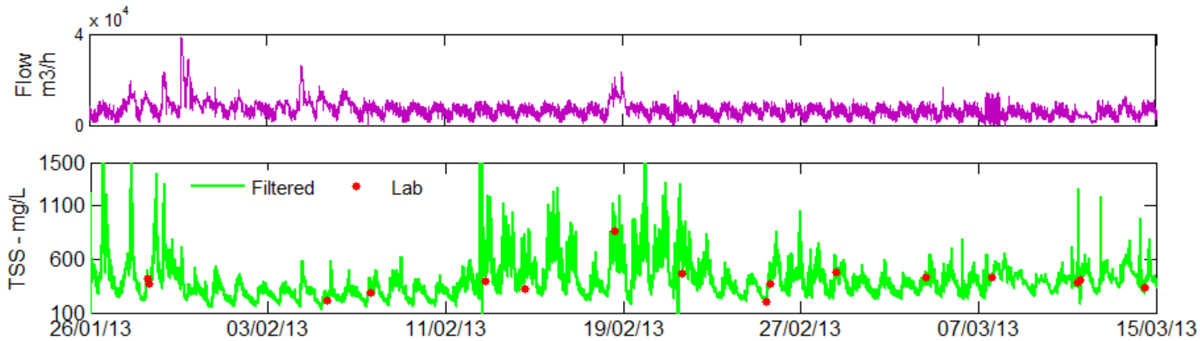

Figure 9. Validated time series for a long period of TSS data collected in raw sewage.

CONCLUSIONS

Within integrated wastewater managementa consistent water quality monitoring strategy and a complete assessment methodology needs to be applied. In practical applications the main issues to solve still are the automation of the collection of “good”quality data and the automatic assessment of the quality of the data being collected. In this paper an overall strategy was presented that aimsat promoting an efficient water quality monitoring program in real scenarios, taking into account aspects like installation and maintenance of on-line sensors, quality control of the data and, finally, validation of the collected data. Once the data has been validated it can then be used for different purposes including: identifying trends, understanding the cause-effect relationships between water quality and hydraulic variables, analyzing the impact of sewer overflows on the receiving water body, monitoring diffuse pollution, resource management, modelling and finally integrated control of the catchment.

ACKNOWLEDGEMENT

Peter Vanrolleghem holds the Canada Research Chair in Water Quality Modelling. The CFI Canada Research Chairs Infrastructure Fund project (202441) provided the monitoring stations. Peter Vanrolleghem was Otto Mønsted Guest Professor at the Technical University of Denmark in 2012-2013. Part of this research was co-financed by the Danish Council for Strategic Research (Storm and Wastewater Informatics project, SWI).

REFERENCES

Branisavljevic N., Prodanovic D., Pavlovic D. (2010) Automatic, semi-automatic and manual validation of urban drainage data.Wat. Sci. Tech., 62, 1013-1021

Dochain D. and Vanrolleghem P.A (2001) Dynamical Modeling and Estimation in Wastewater Treatment Processes.IWA Publishing, London, UK.

Maimon O. and Rockach L. (2010) Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers.Springer, London, UK.

Metadier M. and Bertrand-Krajewski J.L. (2012) The use of long-term on-line turbidity measurements for the calculation of urban stormwater pollutant concentrations, loads, pollutographs and intra-event fluxes. Wat. Res., 46, 6836-6856.

Rieger L. and Vanrolleghem P.A. (2008) monEAU: A platform for water quality monitoring networks. Wat. Sci. Tech., 57, 1079-1086.

Schilperoort R. (2011) Monitoring as a tool for the assessment of wastewater quality dynamics.PhD thesis, Technical University of Delft, The Netherlands. pp. 320.

Sharma A.K., Alferes J., Vezzaro L., Lamaire-Chad C., Vanrolleghem P.A. and Mikkelsen P.S. (2013) Effect of on/off pumping strategy on sewer sediment behavior elucidated by high frequency monitoring at the treatment plant inlet. In: Proceedings 7th IWA International Conference on Sewer Processes and Networks (SPN7). Sheffield, UK, August 28-30 2013.

Sheather S. (2004) Density estimation. Statistical Science, 9, 588-597.

Thomann, M., Rieger L., Frommhold S., Siegrist H. and Gujer W. (2002)An efficient monitoring concept with control charts for on-line sensors. *Water Sci. Technol*, 46(4-5), 107-11.

Thomann M. (2008)Quality evaluation methods for wastewater treatment plant data.Wat. Sci. Tech., 57(10), 1601-1609.

Venkatasubramanian V., Rengaswamy R., Kavuri S. and Yin K. (2003) A review of process fault detection and diagnosis. Part III. Process history based methods. Comp. Chem. Eng., 27, 327-346.

Winkler S., Kreuzinger N., Pressl A., Fleischmann N., Gruber N. and Ecker M. (2002). Innovative technology for integrated water quality measurement. In: Proceedings International Conference on Automation in Water Quality Monitoring (AutMoNet2002). Vienna, Austria, May 21-22, 202.