In: Proceedings WEFTEC2013. Chicago, IL, October 3-9 2013

Validating data quality during wet weather monitoring of wastewater treatment plant influents

Janelcy Alferes[1*], Anders Lynggaard-Jensen[2], Thomas Munk-Nielsen[3] , Sovanna Tik[1], Luca Vezzaro[3,4], Anitha Kumari Sharma[4], Peter Steen Mikkelsen[4] and Peter A. Vanrolleghem[1]

[1]modelEAU, Département de génie civil et de génie des eaux, Université Laval, 1065, Avenue de la Médecine, Québec, Canada QC G1V 0A6.
[2]DHI, Gustav Wieds Vej 10, DK-8000 Århus C, Denmark.
[3]Krüger A/S, Gladsaxevej 363, DK-2860 Søborg, Denmark.
[4]Department of Environmental Engineering (DTU Environment), Technical University of Denmark, Miljøvej, Building 113, DK-2800 Kgs. Lyngby, Denmark.
*Email: janelcy.alferes@gci.ulaval.ca.

ABSTRACT

Efficient monitoring of water systems and proper use of the collected data in further applications such as modelling, forecasting influent water quality and real-time control depends on careful data quality control. Given the size of the data sets produced nowadays in online water quality monitoring schemes, automated data validation is the only feasible option. In this paper, software tools for automatic data quality assessment with a practical orientation are presented. The developments from three organizations ranging from simple to more complex methods for automated data validation are described and evaluated for water quality measurements collected at the inlet of wastewater treatment plants, where probably the hardest measurement conditions are found. The objective of this collaborative effort is to come up with better tools and improved approaches for implementing a successful automatic data quality control procedure.

KEYWORKS: Fault detection, filtering, on-line monitoring, outlier detection, sensors.

INTRODUCTION

With flexibility and standardisation as main drivers of recent development, important advances have been made regarding several monitoring tasks and measurement applications in wastewater systems. However, besides the huge amount of real-time data collected by such measurement set-ups, the most important steps forward have been made in the field of data quality evaluation. As measurements are carried out under challenging conditions such as wet weather flow situations (clogging, fouling, flooding, etc.) raw data is frequently affected by faults like drift, bias, precision degradation or even complete failure, all of which cause the accuracy and reliability of the data to decrease. Those conditions may lead to erroneous conclusions and to the improper use of the data. For data analysis and further applications the collected data will thus be valuable only if the data is properly validated. Given the size of the data sets, only automated data validation is an option.

A wide range of methods have been developed for fault detection and isolation (FDI) in different fields (Venkatasubramanian et al., 2003). These cover methods as simple as evaluations of the range (min-max values), rate-of-change, cross-correlation of redundant measurements and

measurement noise to somewhat more complex often time series analysis based methods that allow eliminating outliers, trending, etc. There are also model-based methods that make use of the generation of residuals (the difference between a measured value and its prediction by a model) and their evaluation for data quality assessment. However, it is often difficult to identify and validate an accurate model that describes all physical and chemical phenomena occurring in the process that is monitored. As an alternative, data-driven methods consider the relationships between the process variables without the explicit expression of a process model.

Many of the aforementioned methods are considered "standard" and can be found in textbooks, but it turns out that their actual practical implementation in the water sector is not as straightforward as anticipated. Three teams that have developed actual implementations in measurement systems have combined their expertise and will present in this paper what the main principles and little details are that make for a successful on-line data quality validation scheme. All will be illustrated with extensive data sets collected at the inlet of wastewater treatment plants (WWTPs), probably the most challenging measurement locations, especially under wet weather conditions.

METHODOLOGY

Water quality sensors are typically disturbed by bias, drift, precision degradation or total failure effects that cause the reliability of measurements to decrease. The most common errors in the raw data include missing values, NaN values, measurement values out of range, peaks (outliers), noise and constant measurement values (indicating that the sensor is out of order). In the framework of practical water quality monitoring applications three different approaches for automatic data quality assessment are presented. While the three methods are aimed to detect doubtful and not reliable data based on different principles, their main strength is their possible integration into the monitoring and control schemes. The two first approaches, based on simple calculations, provide a short-term assessment of the data by using the immediate and recent data readings. The third approach, based on the analysis of the time series, uses a longer period of time (minutes, hour or days back from the actual time) to assess the quality of the data in the short-middle term.

Single data validation methods

Single data validation methods allow checking the data for some of the typical errors by using simple calculations. However, even if these methods are simple it is not common that they are implemented directly in programmable logical controllers (PLC) – the range check might be an exception. Those methods are usually applied as part of the interaction between the real time control (RTC) system and the PLC.

One-step control

From a control point of view, to guarantee a proper operation of the real time automatic controller for WWTPs, a continuous data quality control (DQC) of the available measurements being part of the control loops is essential. Basic one-step DQC techniques that are usually applied and implemented in practical control platforms are listed in Table 1. Despite the quite simple nature of these methods, they allow a quite efficient operation of the plant when more measurements are used simultaneously. These DQC techniques are applied at the level of the single value (e.g. test of minimum and maximum value), although they can also use some

information from previous measurements (e.g. test of the minimum and maximum change rate and test of standard deviation).

Table 1. One-step basic controls implemented in control platforms.

Method	Description
Error message from online meter	For example, calibration, cleaning, bad quality
Manual evaluation of measurement quality	Comparison of laboratory analysis of samples with values from online measurements
Constant value (TCV)	A combination of constant value and period. If the value is constant for a certain allowed period, DQC classifies the value as erroneous.
Range	Minimum and maximum allowable values. When the measured value exceeds the allowable value the DQC classifies the value as erroneous.
Rate of change	The minimum and maximum allowed rate of change in 2 minutes. When the measured rate of change exceeds the allowable value the DQC classifies the value as erroneous.
Running Variance	The minimum and maximum allowed standard deviation value. When the observed standard deviation exceeds the allowable value the DQC classifies the value as erroneous.

Confidence value

In this DQC method one or more tests are applied to the data that is read from the PLC resulting in a confidence value (number between 0 and 100) for each data point (Lynggaard-Jensen et al. 1998). If a measurement is within the limits of what can be expected the confidence will be 100, and if it is showing values that are highly unexpected or even impossible the confidence will be 0. However, between these two situations it might be difficult to judge and confidence for each single test will gradually decrease when the measurement is moving from expected values towards unexpected values (see Figure 1). If the confidence value is lower than a pre-set threshold, different actions can be taken (avoid using data for control, suspend the control based on the RTC-algorithm and fall back to default control by local control loops, calibrate/repair sensor, etc.).

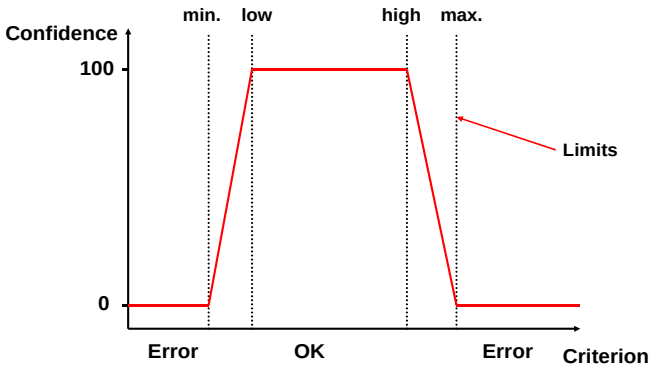

Figure 1. Principle of applying confidence to a measurement.

Table 2 gives an overview of the calculation of the confidence for the different (short term) validation methods, and as can be seen the setting of the parameters for the validation is quite important, and these should accommodate both sensor and process response.

Table 2. Calculation of the confidence for a measurement.

Method	**Confidence function**
Gap Filling (GF); N = no. of time steps, (proportion of data replaced)	$= 100 \cdot (1 - n_i / N); \quad n_i < N \, (n_i = 1, 2, ... N)$ $= 0, \quad n_i \geq N$
Range Check (RC); x = measurement value	$= 0; \quad x > L_{\max}$ $= 100 \cdot ((L_{\max} - x) / (L_{\max} - L_{high})); \quad L_{high} \leq x \leq L_{\max}$ $= 100; \quad L_{low} \leq x \leq L_{high}$ $= 100 \cdot ((x - L_{\min}) / (L_{low} - L_{\min})); \quad L_{\min} \leq x \leq L_{low}$ $= 0; \quad x < L_{\min}$
Rate of Change Check (RCC); x = change of measurement per time unit	$= 0; \quad x > L_{\max}$ $= 100 \cdot ((L_{\max} - x) / (L_{\max} - L_{high})); \quad L_{high} \leq x \leq L_{\max}$ $= 100; \quad L_{low} \leq x \leq L_{high}$ $= 100 \cdot ((x - L_{\min}) / (L_{low} - L_{\min})); \quad L_{\min} \leq x \leq L_{low}$ $= 0; \quad x < L_{\min}$
Running Variance Check (RVC); x = running variance of last n measurements	$= 100; \quad x > L_{low}$ $= 100 \cdot ((x - L_{\min}) / (L_{low} - L_{\min})); \quad L_{\lim} \leq x \leq L_{low}$ $= 0; \quad x < L_{\min}$
Overall Assessment	$= \min(C_{GF}, C_{RC}, C_{RCC}, C_{RVC})$

Univariate time series analysis

Using the online time series information the method developed by Alferes et al. (2013) integrates two main steps: outlier handling and fault detection. Since the presence of outliers can seriously affect the results of statistical tests on the data by altering, for example, the variance, the mean and the normality of the data set, they must be first detected and removed to avoid faulty conclusions. The core of the outlier treatment method lays in the use of univariate autoregressive models to forecast future expected time series data. Outliers are then identified by comparing the measured values with the forecast value with their dynamic prediction error interval.

Specifically, at time T two different exponential smoothing models are used to calculate in the next time unit T+1: (1) the forecast value of the data *x*, and (2) the forecast of the standard deviation Δ of the forecast error. The coefficients of both models are computed as function of the exponentially smoothed statistic S_T given by $S_T = \alpha x_T + (1 - \alpha) S_{T-1}$, with α a smoothing parameter between 0 and 1 that controls the speed at which the historical data is smoothed. The best value for the smoothing constant in each model is the one that results in the minimal residual between the model and a suitable set of calibration data. The calculation of the forecast value of the standard deviation provides a better estimation of the local variance σ_e^2. Assuming a normal

distribution for the forecast error (Dochain and Vanrolleghem, 2001), the estimate σ_e^2 is given by $\hat{\sigma}_e = 1.25\hat{\Delta}_T$, with $\hat{\Delta}_T$ calculated as $\hat{\Delta}_T = \alpha|e_T(1)| + (1-\alpha)\hat{\Delta}_{T-1}$, with $e_T(1)$ the one-step-ahead forecast error, defined by $e_T(1) = x_T - \hat{x}_T$. The prediction interval $x_{lim_T} = \hat{x}_T \pm K\hat{\sigma}_{e,T}$ is then defined by adding or subtracting a multiple *K* of the standard deviation of the forecast error to the forecast data value. At time T, the measurement data is evaluated to determine if it falls outside the prediction interval. In that case, it is considered an outlier and it is replaced by its forecast value. Once the outlier has been removed a new time series of *accepted data* is created. For fault detection purposes several statistical data features are calculated together with their acceptability limits. Since the presence of noise in the data can corrupt the calculations, the accepted data is first smoothed by using a kernel smoother (Schimek, 2000) with a proper bandwidth. The data features are then evaluated over the resulting smoothed or *treated data* time series. Figure 2 gives an overview of the complete method.

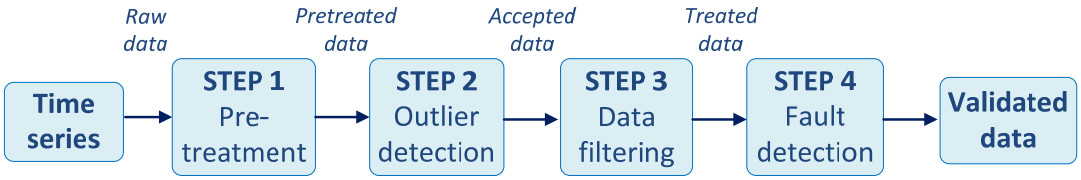

Figure 2. Univariate time series analysis according to Alferes et al. (2013).

Table 3 summarises the data features calculated over the smoothed data each time T. Acceptability limits are defined for each feature and have proven to be sensor, location and variable specific. Data is validated by given a proper mark: 0 - valid (all test are passed), 1 – doubtful (some tests have failed, posterior analysis is required), and 2 – not valid.

Table 3. Features used for fault detection purposes

Feature	Definition	Purpose
%replaced data	Fraction of forecast values used in the data set after outlier elimination	Evaluate the goodness of the smoothed data and the data features
Rate of change	Slope between two consecutive data points in the smoothed data	Evaluation of dynamics in the data (gradients and sudden changes)
Locally physical range	Range [min max] where values are normally observed in a specific location	Evaluate if the data lies in the expected range.
Residual standard deviation	Standard deviation of residuals (difference between the accepted data and the smoothed data)	Estimation of the variance of the data
Autocorrelation of residuals	Run test over the residuals (Dochain and Vanrolleghem, 2001)	Evaluate if residuals are randomly distributed

Case study

To illustrate the potential of the described data quality validation methods, the tools have been validated in different practical platforms. Data has been collected by online water quality sensors installed at the inlet of different WWTPs under the challenging measuring conditions that prevail in those environments. The first approach, one-step control, is integrated into the STAR® control platform developed by Krüger (Thomsen and Ønnerth, 2009). The second approach, confidence value, is part of the DIMS.CORE RTC system of DHI (Ingeduld , 2007). Finally, the third approach – time series analysis – has been implemented as part of the Primodal Systems RSM30 PrecisionNow software (Copp et al., 2010). Implementation of those tools in practical scenarios allows operators and process engineers following the status of current and past measurements and the detection of doubtful data and potential sensors faults. Posterior analysis must be carried out to identify the nature of the abnormal conditions and to apply the corrective actions in the field.

RESULTS

One-step

Figure 3 shows an example of Kruger's DQC applied to a single ammonia sensor over a time series of measurements collected at the Avedøre WWTP in Copenhagen (Denmark). In the specific case the Test of Constant Value (TCV) utilizes different time intervals according to the measurement values. For example, for low ammonia values (as in the case shown in Figure 3) TCV is performed over 120 minutes, while for higher values of NH_4, this interval is reduced to only 30 minutes. The NH_4-N concentration was constant for more than 120 minutes in the lower concentration interval; therefore the DQC test classifies the measurements as erroneous. There can be many reasons for the constant values. One of the reasons can be the calibration of the sensors, where the signal would be frozen. In this case the error message from the online meter can be used to classify the measurement as erroneous already just after 9 a.m. Another reason for constant value could be that the actual concentration is higher or lower than the online sensors measuring range. In this case the range value test shown in Figure 4 prevents the DQC to classify the measurements as erroneous. The other DQC test illustrated in Figure 3 is the detection of the excess change rate, where the rate of change in the concentration is higher than the allowed 1 mg NH_4-N/l/ 2 min.

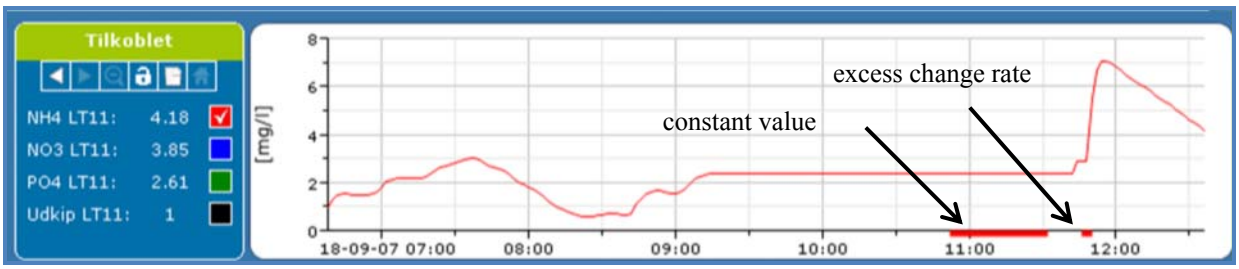

Figure 3. Example of erroneous NH_4 measurements. Constant value is detected around 11:00, while the maximum change rate allowed interval was exceeded just before 12:00. Erroneous measurements detected by DQC tests are illustrated by the coloured line on the x-axis.

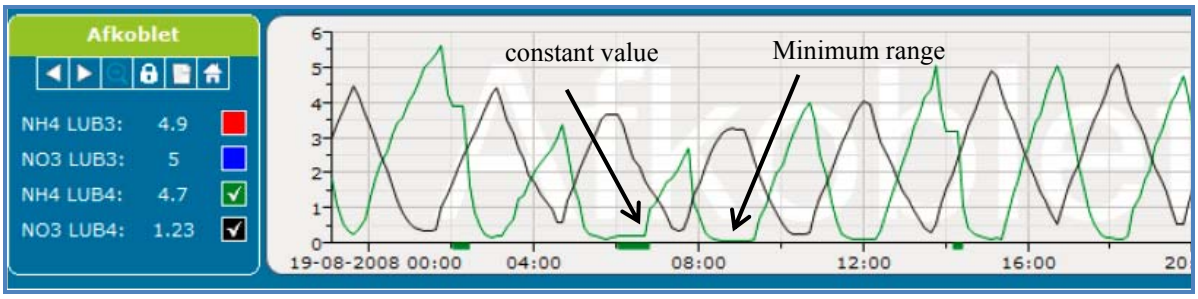

Figure 4. Example of erroneous NH_4 measurements using TCV and minimum range test. Erroneous measurements detected by DQC tests are illustrated by the coloured line on the x-axis.

These simple tests can also be applied to measurements that are related to another. Figure 5 shows an example from the Czajka – Warszawa WWTP in Poland, where the flow in the recycling flow is stopped (Qr), which results in the DQC classifying the SS in the return sludge as erroneous.

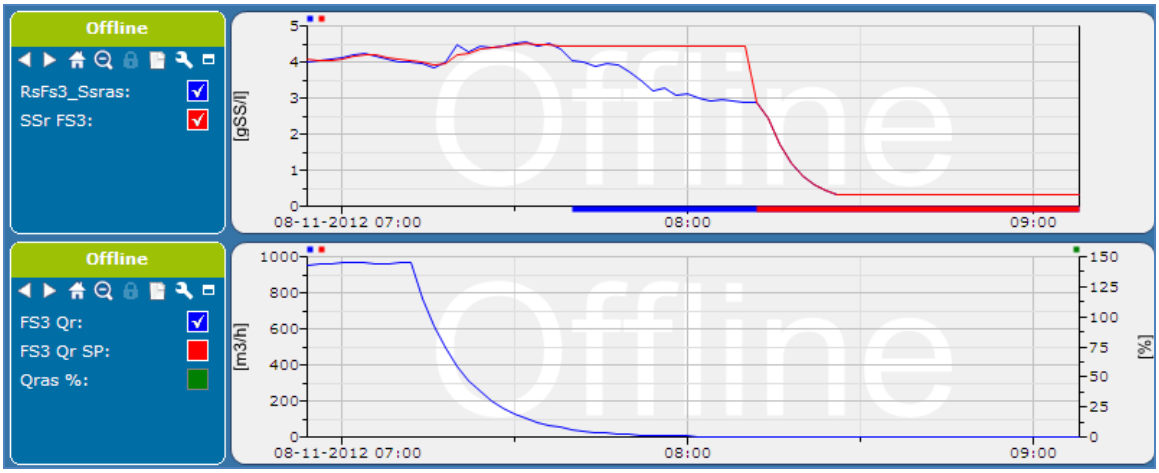

Figure 5. Example of DQC classifying SS in return sludge as erroneous based on the flow measurements in the return sludge flow.

An important element in making the control of WWTP reliable is the use of the integration of the results of different DQC tests to define the set-points sent to actuators. In the example shown in Figure 6, the results of simple DQC tests on different variables (NH_4, calculated OUR, communication signal) are utilized to define different fall-back strategies and to adopt different set-points.

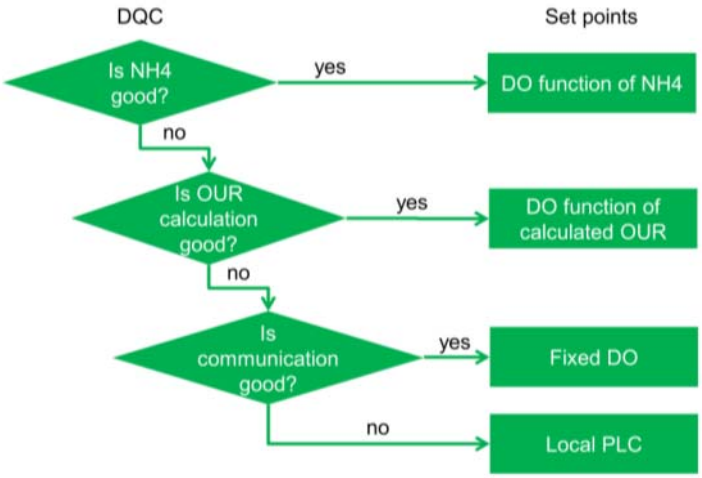

Figure 6. Example of integration between DQC controls and set points sent to actuators. Depending on the different number of failed DQC tests, different fall-back strategies are adopted.

Confidence value

Real time validation methods are also implemented in the DIMS.CORE RTC system from DHI and the following practical example is taken from an implementation at the Viby wastewater treatment plant in Denmark. Figure 7 shows the results from an ammonium measurement in the inlet to the Viby wastewater treatment plant together with the inlet flow for a period of 5 days including both dry and wet weather. Both the ammonium and the flow are logged every minute.

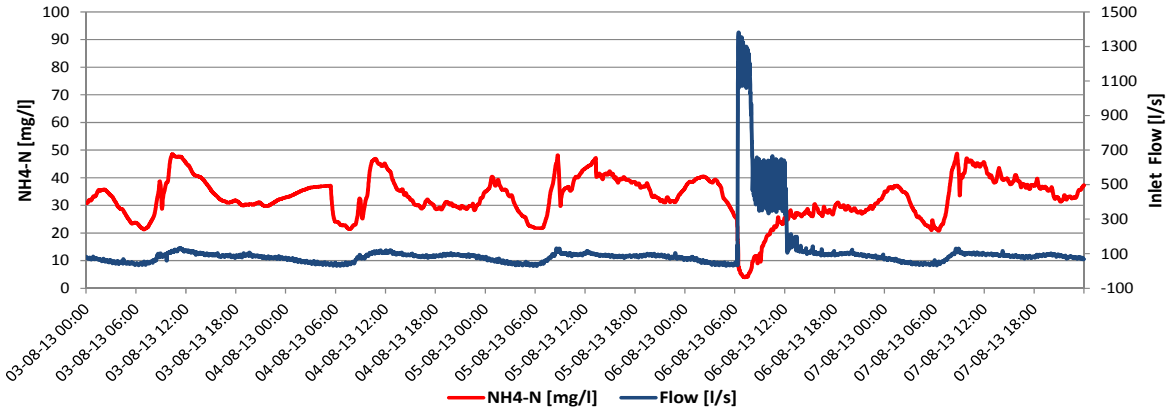

Figure 7. Measurements from the inlet to Viby wastewater treatment plant, Denmark.

The dry weather flow shows normal diurnal behaviour, and the wet weather flow is a result of a short but quite intense rain causing the flow to the treatment plant to increase rapidly to the maximum inlet pump capacity (1300 l/s). When the rain stops the inlet flow decreases to approx. 500 l/s, which is the capacity of the pumps emptying the retention basin holding the combined sewer overflow. During the rain the ammonium concentration decreases as expected and returns to normal dry weather values after the rain stops.

Data validation of the ammonium measurement requires configuration of the parameters for the selected methods, which here will be range check, rate of change check and running variance

check. In order to configure the parameters correctly, it is important to identify normal behaviour of the measurement and to look for deviations from normal behaviour over a longer time period. The aim is to make the data validation as sensitive as possible, and at the same time minimise false positives. The range check should reflect a normal working range of the measurement and Figure 7 suggests parameters to be configured as shown in Figure 8. Figure 8 is also an example on how configuration of data validation is done within the DIMS.CORE system.

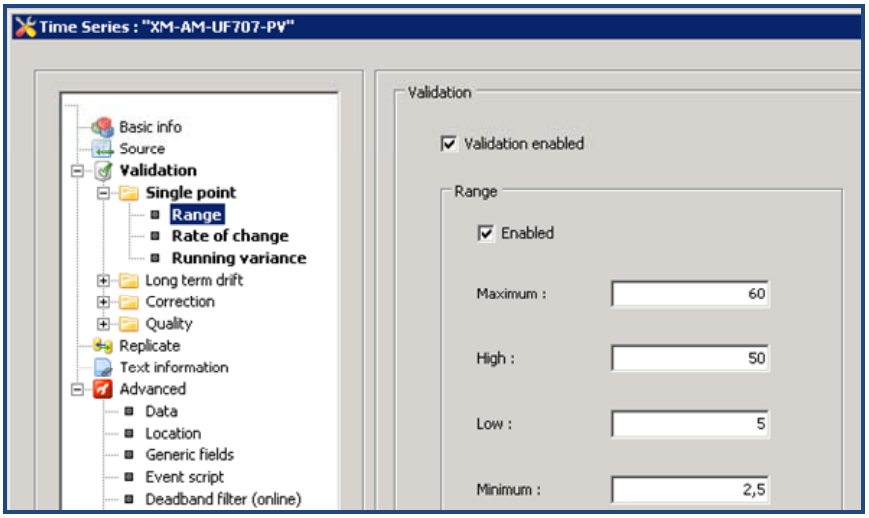

Figure 8. Configuration of data validation using range check.

Configuration of parameters for the rate of change check and the running variance check is not as straightforward as for the range check. However, as DIMS.CORE supports easy setup of software sensors it is possible to configure temporary software sensors calculating the rate of change and the running variance. These show that the rate of change varies between 1.3 and -4.6 and that the minimum for the running variance with 5 time steps is $2.1^{x}10^{-6}$. The resulting overall confidence for the ammonium measurement using these parameters is shown in Figure 9.

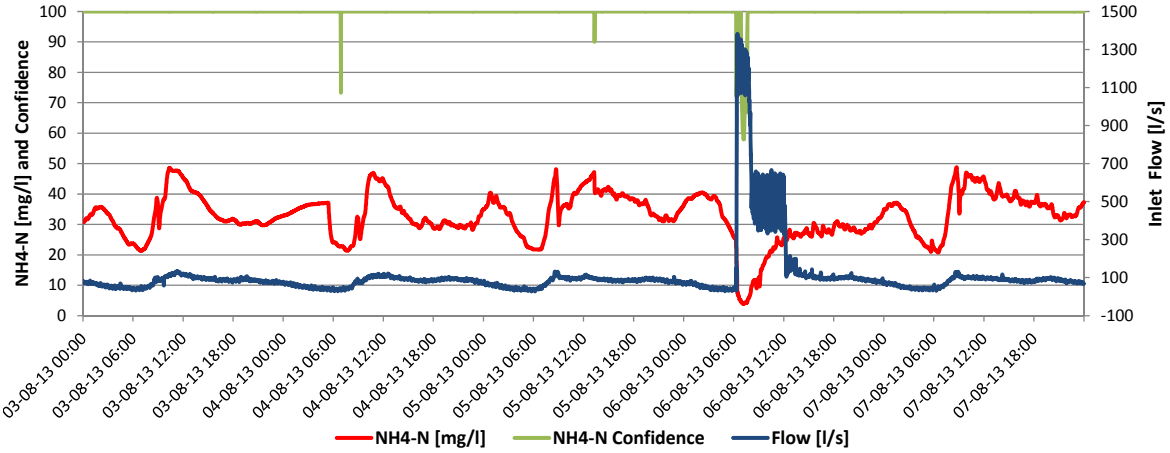

Figure 9. Overall assessment of the confidence in the ammonium measurement.

The first drop in the confidence in Figure 9 originates from the running variance check, the second from the rate of change check and the more complex decrease in the confidence

originates in the start from the rate of change check – the ammonium concentration decreasing rapidly due to the start of the rain event – followed by a lower confidence due to the range check.

In a real time control system it is not the confidence itself which is interesting – it is the automatic detection of a decreasing/(increasing) confidence crossing certain thresholds, and what automatic actions shall be taken, when these thresholds are crossed. This is handled by changing the quality of the measurement accordingly. This quality assessment gives an extra dimension to the data validation, because confidences for different measurements can be given different "weights". Figure 10 shows the quality assessment of the calculated confidence for the ammonium measurement. The quality is judged to be OK if the confidence is higher than 90. When the quality drops below 90 the warning state becomes active, below 75 the quality becomes critical and below 50 the quality enters the error state.

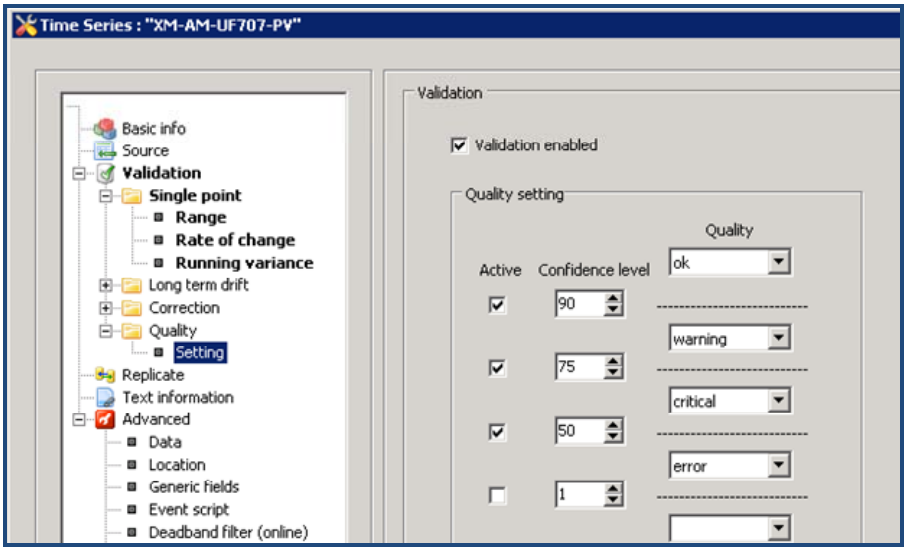

Figure 10. Quality assessment of calculated confidence for the ammonium measurement.

Each time a threshold is crossed the system generates an event, which in this case is used to send an email when a warning has to be issued and a SMS when the quality becomes critical – both to the staff being responsible for the maintenance of the ammonium sensor.

Time series analysis

Figure 11 shows the results of the application of Primodal Systems' time series analysis method to a TSS time series collected at the inlet of the primary clarifier from the Lynetten WWTP in Copenhagen (Denmark),. The method has been configured by using a representative training data set collected over a 7-day period. The band delimited by the red and blue lines represents the prediction interval within which normal data should fall according to the calculated time series models. Most of the TSS data fall into the prediction interval but a number of outliers and doubtful data is identified along the time series as indicated by the percentage of forecast values that have replaced the raw data in the calculation of the smoothed data (green line).

Although the TSS data respected the locally physical range for the whole period, the rest of the data features revealed some abnormal behaviour; for example around April 24th (period I in Figure 11) and May 3rd (period II in Figure 11). In both cases an atypical variation in the

dynamics of the TSS measurements is identified with excessive slope values. In these two periods the residuals standard deviation also exceeds the typical measurement errors indicating a larger variance in the data. Diagnosing the residuals (run test on a moving window) during these periods showed the presence of some non-randomly distributed residuals, and the percentage of replaced forecast values indicated also the existence of outliers. Once all the data features have been evaluated for each data point, data is validated according its degree of reliability. As illustrated in the label subplot, some of the measurements from the periods I and II are classified as not valid or doubtful data. For the whole period in Figure 11, about 8% of the data is considered as doubtful or not valid, which is quite acceptable in view of recent practical studies using partially automated data validation that report data losses of 5 to 60% (van Bijnen & Korving (2008): 40%; Thomann (2008): 5-15%; Metadier (2011): 40-60%; Schilperoort (2011): 25-50%).

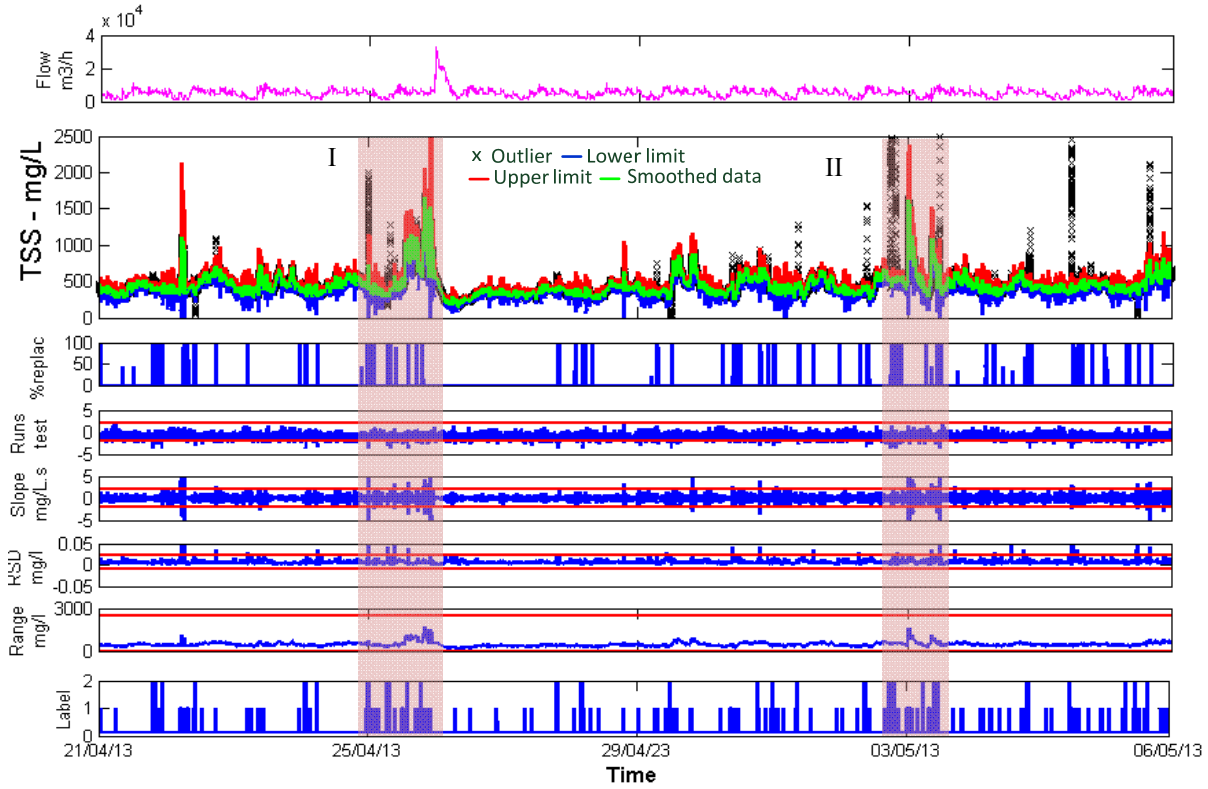

Figure 11. On-line outlier detection procedure applied to an actual TSS time series from Lynetten WWTP.

Figure 12 shows the results of the application of the same method to the ammonia time series at the same location. The impact of the rain event around April 26th is clearly observed in the NH_4 measurements. The time series analysis method was able to maintain its performance during both dry and wet weather conditions. However, higher variance and noise levels than those observed for the TSS measurements are revealed.

Although the forecasting model was able to describe the raw data in an efficient way for the whole period, the percentage of replaced forecast values also indicated the presence of outliers and abnormal data along the time series. The slope and residuals standard deviation values also

exceed their typical limits indicating atypical variations in the NH_4 data as shown for example in the periods around April 29th , May 4th and May 29th where some of the measurements have been classified as not valid or doubtful. Similarly to the TSS measurements example, for the whole period in Figure 12 about 11% of the data is considered as doubtful or not valid.

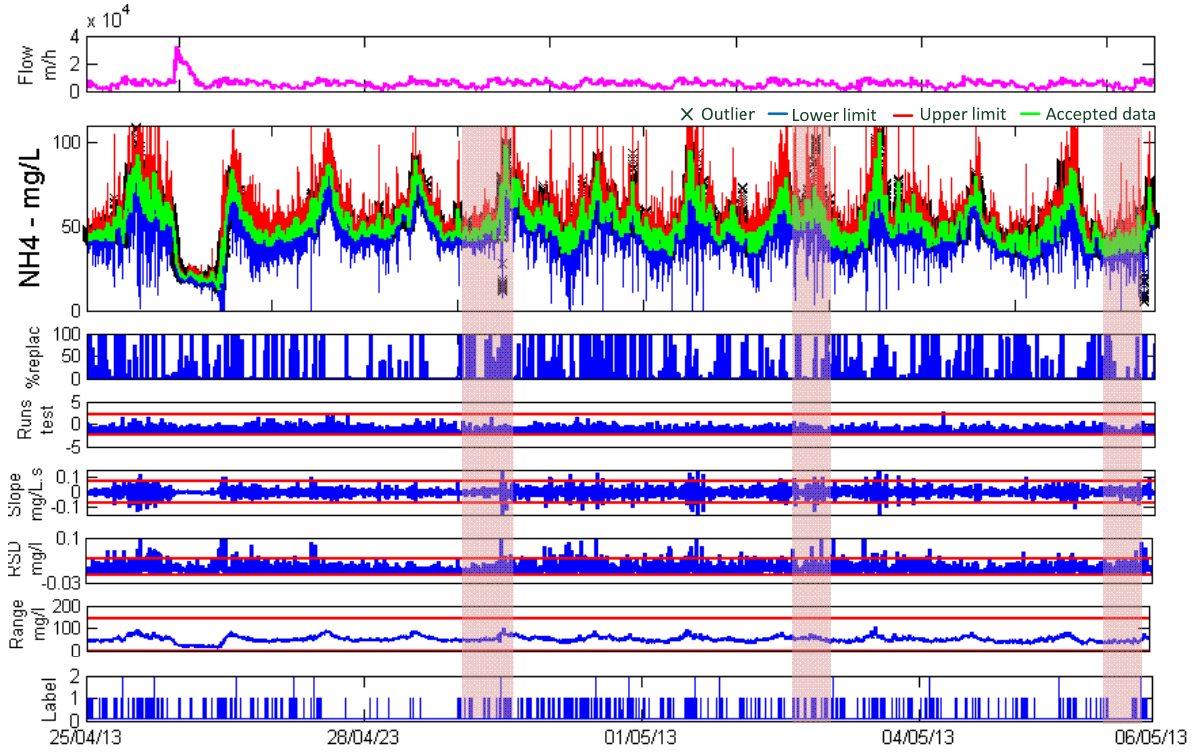

Figure 12. On-line outlier detection procedure applied to an actual ammonia time series from Lynetten WWTP.

DISCUSSION

The decision about when a value can be considered as "valid" or "not valid" is not simple. Properties of the water quality measurements (fast dynamics, autocorrelated time series, non-random noise, etc.) jeopardize the use of classical methods for data validation. A number of criteria based on the data characteristics, sensor, context of the process, final use of the data (including modeling, control, decision making, etc.) and expertise should be considered.

From the results previously presented, it can be observed how the single data validation methods, based on immediate readings, provide useful information about the current data value. Due to their computational simplicity they are usually integrated into the RTC systems. Since each typical single DQC method assesses a different property in the data, a robust evaluation would only be possible if a combination of those methods is applied. More complex methods need to be used to evaluate the quality of the data over a certain time period to identify for example gradual drift and presence of noise. The use of the time series analysis method allows dealing with those

situations and also with the critical process of detecting and removing outliers and noise from the raw data for posterior fault detection analysis.

The implemented methods are aimed to extract information from individual variables, and normally do not distinguish between a change in the sensors' properties or in the process variable itself. This requires more complex validation methods using more signals through the application of cross validation or multivariable methods that consider the correlation between different variables (Alferes et al., 2013; Villez et al., 2008). If more sensors are not available, it is also possible to configure software sensors as a real time calculation based on one or more sensors (Lynggaard-Jensen and Lading, 2006; Spindler and Vanrolleghem, 2012).

A sequential process that combines the implementation of simple and more complex univariate methods, together with multivariate methods would lead to a complete and successful data quality validation scheme. However, the key of a reliable data quality evaluation process will depend on the proper setting of the methods and on the right definition of the thresholds or acceptability limits for each test. Expert knowledge about expected data variability and sources of faulty situations should be combined to set the methods' parameters for each application.

CONCLUSIONS

Besides the huge amount of real-time water quality data that can be collected nowadays in water system monitoring applications, the challenge one is currently facing is the development of practical automated data quality evaluation tools that allow detecting and correcting doubtful data and that can help users in understanding, analysing and processing the data. Inefficient manual data evaluation procedures, typically used today in the water field, are not a feasible option to handle such data sets, given their specific characteristics. In fact, one of the difficulties for the joint use of monitoring and modeling, control and decision making applications has been the lack of good data affecting the proper use of the measurements. To answer that need and looking for an effective water quality monitoring scheme, three practical approaches for data quality validation that have been successfully implemented in different platforms and in diverse water systems have been presented in this paper and their main strengths have been highlighted. While the two first approaches provide information about the current single value based on simple calculations, the third approach allows evaluating the quality of the data by analysing the time series over a longer period of time. Crucial to the good performance of the methods is the proper setting of the algorithms for each specific application. Information that can be obtained from the two different orientations (short and middle term evaluation) is complementary. When combined with multivariable methods (considering different variables at the same time) an even more efficient data quality evaluation process can be anticipated.

ACKNOWLEDGMENTS

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). Luca Vezzaro is an industrial postdoc

financed by the Danish National Advanced Technology Foundation under the project "MOPSUS-Model predictive control of urban drainage systems".

REFERENCES

Alferes J., Tik S., Copp J. and Vanrolleghem P.A. (2013) Advanced monitoring of water systems using in situ measurement stations: Data validation and fault detection. Wat. Sci. Tech., (in press).

Copp J., Belia E., Hübner C., Thron M., Vanrolleghem P. A. and Rieger, L. (2010) Towards the automation of water quality monitoring networks. In: Proceedings 6th IEEE Conference on Automation Science and Engineering (CASE 2010). Toronto, Ontario, Canada, August 21-24, 2010.

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

Ingeduld P. (2007) Real-time forecasting with EPANET. In: Proceedings World Environmental and Water Resources Congress: Restoring Our Natural Habitat. Tampa, Florida May 15-17, 2007.

Lynggaard-Jensen A., Billington A., Mpe A., Wittig T., Rovira M.A. and Schmidt B. (1998) WaterNet - Distributed Water Quality Monitoring using Sensor Networks. In: Proceedings of Waste-Decision 98, International Workshop on Decision and Control on Wastes Bio Processing, Narbonne, France, February 25-27, 1998.

Lynggaard-Jensen A. and Lading L. (2006) On-line determination of sludge settling velocity for flux-based real-time control of secondary clarifiers. Wat. Sci Tech., 54(11-12), 249-56.

Métadier M. (2011) Traitement et analyse de séries chronologiques continues de turbidité pour la formulation et le test de modèles de rejets urbains par temps de pluie. PhD thesis, INSA-Lyon, France. pp. 419. (in French).

Schimek M.G. (2000) Smoothing and Regression: Approaches, Computation and Application, John Wiley&Sons, New York.

Spindler A. and Vanrolleghem P.A. (2012) Dynamic mass balancing for wastewater treatment data quality control using CUSUM charts. Wat. Sci. Tech., 65(12), 2148-2153.

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

Thomsen H.R. and Ønnerth T.B. (2009) Results and benefits from practical application of ICA on more than 50 wastewater systems over a period of 15 years. Keynote paper, 10th IWA Instrumentation, Control and Automation conference, Cairns, Queensland, Australia, 14-17 June 2009.

van Bijnen M. and Korving H. (2008) Application and results of automatic validation of sewer monitoring data. In: Proceedings 11th International Conference on Urban Drainage (ICUD2008). Edinburgh, Scotland, UK.

Venkatasubramanian V., Rengaswamy R., Kavuri S.N. and Yin K. (2003) A review of process fault detection and diagnosis. Part III. Process history based methods. Comp. Chem. Eng., 27(3), 327-346.

Villez K., Ruiz M., Sin G., Colomer J., Rosén C. and Vanrolleghem P.A. (2008). Combining multiway principal component analysis and clustering for efficient data mining of historical data sets of SBR processes. Wat. Sci. Tech., 57(10), 1659–1666.