

Drinking water quality monitoring using trend analysis

Jani Tomperi, Esko Juuso, Mira Eteläniemi and Kauko Leiviskä

ABSTRACT

One of the common quality parameters for drinking water is residual aluminium. High doses of residual aluminium in drinking water or water used in the food industry have been proved to be at least a minor health risk or even to increase the risk of more serious health effects, and cause economic losses to the water treatment plant. In this study, the trend index is developed from scaled measurement data to detect a warning of changes in residual aluminium level in drinking water. The scaling is based on monotonously increasing, non-linear functions, which are generated with generalized norms and moments. Triangular episodes are classified with the trend index and its derivative. The severity of the situations is evaluated by deviation indices. The trend episodes and the deviation indices provide good tools for detecting changes in water quality and for process control.

Key words | non-linear scaling, residual aluminium, water treatment

Jani Tomperi (corresponding author)

Esko Juuso

Kauko Leiviskä

University of Oulu,
Control Engineering Laboratory,
P.O. Box 4300,
FIN-90014 University of Oulu,
Finland
E-mail: jani.tomperi@oulu.fi

Mira Eteläniemi

Finnsugar Ltd,
Sokeritehtaantie 20,
FI-02460 Kantvik,
Finland

INTRODUCTION

In water treatment processes, surface waters are most commonly treated by chemical coagulation. Aluminium salts are widely used as coagulants to reduce the organic matter, colour and turbidity of raw water because aluminium has good ability to coagulate and flocculate both organic and inorganic compounds. Using aluminium salts in the water treatment process may lead to an increased concentration of aluminium in drinking water, if the aluminium is overdosed or the water treatment process is dysfunctional. The total intake level of aluminium from drinking water varies also according to the aluminium concentration in raw water, which depends on various physicochemical and mineralogical factors (Driscoll & Letterman 1995; WHO 2008).

The residual aluminium concentration in drinking water is affected by raw water temperature, raw water KMnO_4 (potassium permanganate), Al/KMnO_4 -ratio, silicate concentration, pH of coagulation and raw water pH, and it is also proportional to the turbidity of water (Driscoll & Letterman 1995; Juntunen *et al.* 2010; Tomperi *et al.* 2011, 2013). The aluminium concentration can be minimized by optimizing pH during the coagulation process, avoiding excessive dosing of aluminium, good mixing of coagulants, optimum paddle speed in the flocculation process and efficient flocculation (WHO 2008).

Many process measurements are available in water treatment plants (WTPs) but the residual aluminium is not measured until the drinking water is leaving the WTP to be distributed to consumers. In the case of a dangerous level of aluminium, distribution to consumers is cancelled, consumers are warned not to use water as it is and the water network is rinsed. This causes harm to the consumers and economic losses to the WTP. A high level of residual aluminium increases water turbidity and has some health effects on consumers. Reported minor symptoms of a high level of residual aluminium in drinking water are nausea, vomiting, diarrhoea, mouth and skin ulcers, rashes and arthritic pain (WHO 2003). Symptoms are generally mild and short-lived. More serious health effects of aluminium in drinking water have been studied widely and the results are conflicting. The environmental factors, the target groups, etc. in different studies vary greatly and therefore it cannot be said with certainty that the residual aluminium in drinking water is the main affecting factor for serious health effects like Alzheimer's disease. It has been hypothesized that aluminium exposure is an increased risk factor for the development or acceleration of onset of Alzheimer's disease in humans (McLachlan *et al.* 1996; WHO 2008; George *et al.* 2010). On the other hand, the Canadian

Study of Health and Aging claims that residual aluminium in drinking water does not increase the risk of Alzheimer's disease (Leakey 2004). However, it is vital to be able to predict the rising aluminium level in the drinking water, give a warning to the process personnel and control the process differently.

The time series of measurements contain effects of many distinct contributions: slow trends, different kinds of equipment faults, periodic disturbances and changing sensor noise (Stephanopoulos & Han 1996). Also outliers and structural breaks need to be analysed to find true trends (Metz 2010). Typical reasoning systems have three components: a language to represent the trends, a technique to identify the trends and a mapping from the trends to operational conditions (Dash et al. 2003). The fundamental elements are modelled geometrically as triangles to describe local temporal patterns in data (Figure 1). The elements are defined by the signs of the first and second derivative, respectively. These elements, which are also known as triangular episodic representations (Cheung & Stephanopoulos 1990), have their origin in qualitative reasoning and simulation (Forbus 1984; Kuipers 1985). An interval-halving scheme to facilitate the automatic extraction of temporal features from sensor data

in terms of the trend language of primitives was presented by Dash et al. (2001).

Temporal reasoning is a valuable tool for diagnosing and controlling slow processes. Manual process supervision relies heavily on visual monitoring of the characteristic shapes of changes in process variables, especially their trends. Although humans are very good at detecting such patterns visually, it is a difficult problem for control system software (Kivikunnas et al. 1996; Kivikunnas 1999). In these cases, the appropriate time window is not always recognized.

Detecting changes is important for data compression, process control and fault diagnosis. Linear regression combined with fuzzy reasoning can be used to detect significant changes up or down in process variables (Poirier & Meech 1993). The episodes shown in Figure 1 provide features for more detailed analysis. Trend extraction methods are based on polynomials (Konstantinov & Yoshida 1992), linear segments, wavelets and B-splines, and neural networks are also used. Similarities between trends are analysed, for example, with sequence matching, decision trees, pattern recognition and stochastic hidden Markov models; see Maurya et al. (2007). Resolution, fine or coarse, has a strong effect on results (Stephanopoulos & Han 1996).

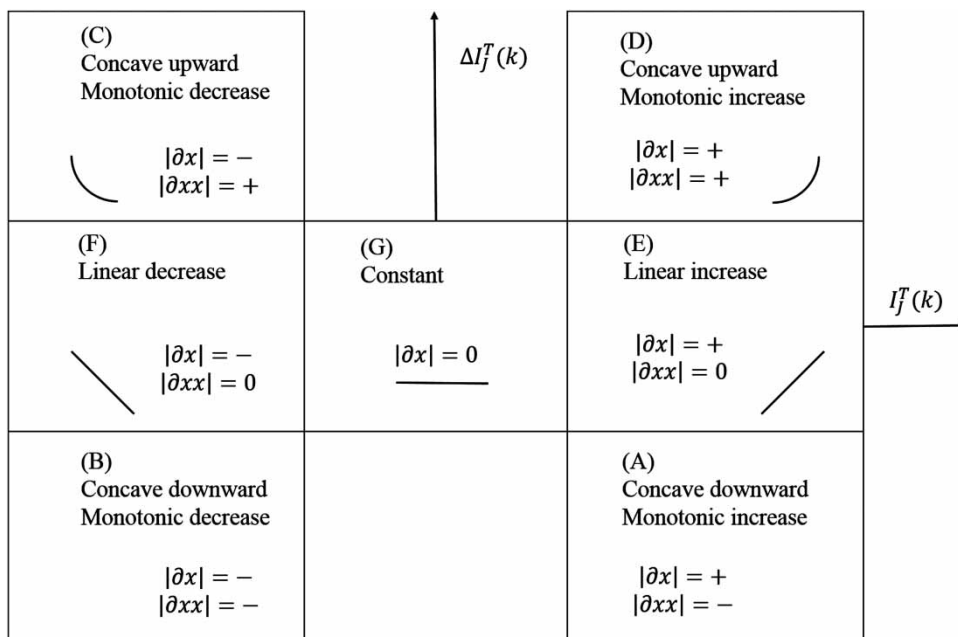


Figure 1 | Triangular episodic representation defined by the index $I_j^T(k)$ and the derivative $\Delta I_j^T(k)$: seven basic types of episodes used for interval description, each episode type is denoted by a letter from the set {A, B, C, D, E, F, G}. Redesigned from Kivikunnas (1999) and Juuso (2011).

The qualitative scaling defined by episodes shown in Figure 1 generates a hierarchical multiscale structure of trend descriptions over different time scales from process data (Stephanopoulos *et al.* 1997). Generalized norms provide tools for analysing different effects (Lahdelma & Juuso 2011a, b). These norms and the non-linear scaling approach (Juuso & Lahdelma 2010) are combined with the method introduced in Juuso *et al.* (2009) to form the trend analysis method presented in Juuso (2011).

In this study, the above mentioned trend analysis is applied for residual aluminium in drinking water and variables that have been found to affect the residual aluminium level. However, the method can be used for analysing any other quality parameter of drinking water as well. The trend episodes and the deviation indices provide a warning about the changes in water quality. First, all measurements are scaled with the non-linear scaling approach, which is based on generalized norms and skewness. Then the trend index is developed from scaled measurements. Finally, the trend analysis provides useful indirect measurements for water quality monitoring and high level control.

MATERIAL AND METHODS

The water treatment plant

The data were collected from the WTP of Finnsugar Ltd in Kirkkonummi, Finland. This chemical treatment plant uses

surface water from a nearby lake (Humaljärvi), an artificial lake (Pikkala) or a mixture of these two sources as raw water. Before addition of the coagulation chemical, the pH of raw water is adjusted to the optimal value with calcium hydroxide. Water is treated by chemical flocculation and filtration. An aluminium-based coagulation chemical is used in the coagulation process. The coagulation chemical dose is controlled as a function of raw water KMnO_4 value. After the filtration, water pH is again adjusted to the optimal level for distribution. UV-radiation and sodium hypochlorite are used for disinfection. Figure 2 shows the process stages of the Finnsugar Ltd WTP where high amounts of residual aluminium in drinking water are not a serious concern. During the period of data collection, the long-term mean value of residual aluminium in produced drinking water was less than half of the maximum target value of the quality recommendation (0.2 mg/l) defined in the Health Protection Act of Finland's Ministry of Social Affairs and Health (FINLEX 2000). Even so, Finnsugar Ltd has a great interest in being able to act in advance of changes in water quality and the residual aluminium level. Early warnings enable control of the water treatment process to occur differently and prevent the distribution of poor quality water.

The quality of the data analysis or modelling of the water treatment process depends highly on the quality of the source dataset. The dataset must fully represent the full spectrum of all possible conditions. For instance, the temperature of surface water varies depending on the season of the year and

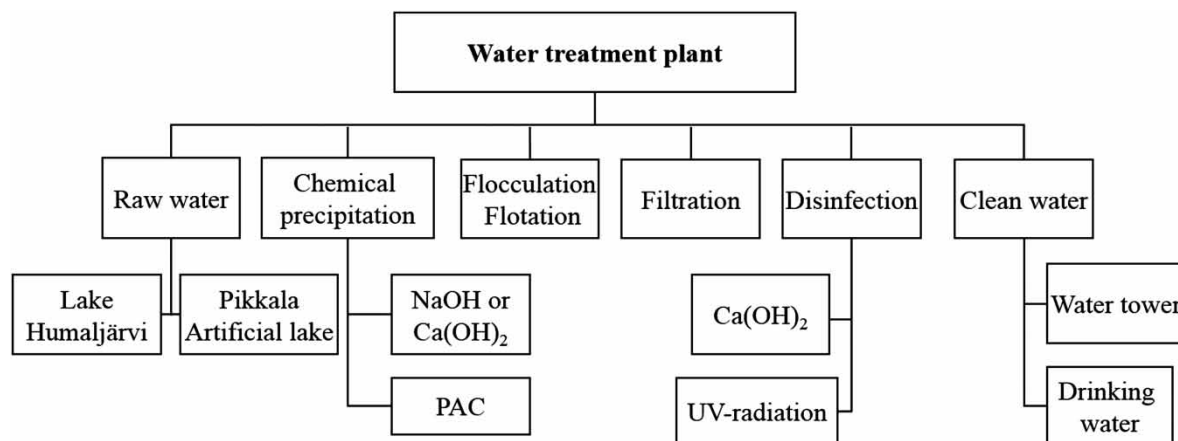


Figure 2 | The processing stages and used chemicals in Finnsugar Ltd water treatment plant. PAC: polyaluminum chloride.

therefore, the source dataset must encompass at least 1 full year of measured data (Baxter et al. 2001).

The dataset used in this work covered a period of 16 months. Tens of on-line measurements and several laboratory measurements of raw and drinking water were included originally. The final dataset was reduced leaving out variables unnecessary for this work. The laboratory measurements (pH, KMnO₄, turbidity, colour, conductivity, chlorine, bacteria and aluminium) of raw and drinking water were done at least once in every working day. If some measurement showed anomalous values, new samples were taken and analysed. The on-line measurements were originally stored at 5 minute intervals. For this study, all on-line measurements were averaged to 1 hour data before combining with the laboratory measurement data. Laboratory measurement values were combined to the corresponding hour of on-line measurement data. Evident outliers (not realistic values) were manually filtered out and missing data values were added by linear interpolation.

Statistical background of non-linear scaling

The expected value, E , of a random variable, X , is a central concept in probability and statistics. The moments, M , can be defined around some central value, e.g. the mean, median and mode. Variance $\text{Var}(X) = \sigma_X^2$ can be represented by

$$\sigma_X^2 = M^2 = E[(X - E(X))^2] = E(X^2) - [E(X)]^2 \quad (1)$$

The positive square root of the variance σ_X , the standard deviation, is used more often since it has the same dimension as the variable in question. Dimensionless features can be obtained by normalizing the moments, for example, by standard deviation, σ_X

$$\gamma_k = \frac{E[(X - E(X))^k]}{\sigma_X^k} \quad (2)$$

The feature γ_3 is called the coefficient of skewness, or briefly skewness, and the feature γ_4 as the coefficient of kurtosis. Skewness is a measure of asymmetry: $\gamma_3 = 0$ for a symmetric distribution. If $\gamma_3 > 0$, skewness is called positive

skewness and the distribution has a long tail to the right, and vice versa if $\gamma_3 < 0$. Kurtosis is a measure of the concentration of the distribution near its mean. For a Gaussian signal $\gamma_4 = 3$. Flatter, also described as long-tailed or heavy-tailed, distributions have $\gamma_4 < 3$ and for distributions with high peakedness, $\gamma_4 > 3$. For a sinusoidal signal, $\gamma_4 = 1.5$. An alternative definition of kurtosis reduces the ratio by 3 to give the value zero for the normal distribution (Lahdelma & Juuso 2011a).

A norm defined by

$$\|{}^\tau M_\alpha^p\|_p = ({}^\tau M_\alpha^p)^{1/p} = \left(\frac{1}{N} \sum_{i=1}^N \|x_i^{(\alpha)}\|^p \right)^{1/p} \quad (3)$$

where $\alpha \in R$ is the order of the derivative, the order of the moment $p \in R$ is non-zero. The signal is measured continuously and the analysis is based on consecutive equally sized samples. Duration of each sample is called sample time, denoted by τ . The number of signal values in a sample is $N = \tau N_s$, where N_s is the number of signal values in a second. This norm, which has the same dimensions as the signal $x^{(\alpha)}$, is defined in such a way that $-\infty < p < \infty$, $p \neq 0$, i.e. the definition includes the l_p norms defined for $1 < p < \infty$. This norm presented in Lahdelma & Juuso (2008) is also known as the power mean. The norm values increase with increasing order, i.e. for the generalized norms it holds that

$$({}^\tau M_\alpha^p)^{1/p} \leq ({}^\tau M_\alpha^q)^{1/q} \quad (4)$$

if $p < q$. The increase is monotonous if all the signals are not equal.

The absolute mean, the root mean square value and the absolute harmonic mean are special cases where the order is 1, 2 and -1 , respectively. The norm in (3) represents the norms from the minimum to the maximum, which correspond the orders $p = -\infty$ and $p = \infty$, respectively. When $p < 0$, all the signal values should be non-zero, i.e. $x \neq 0$. Therefore, the norms with $p < 0$ are reasonable only if all the signal values near zero are removed (Lahdelma & Juuso 2008, 2010, 2011a).

The normalized moments in (2) are here generalized by replacing the expectation value with the norm (3) as the

central value

$$\gamma_k = \frac{E \left[\left(X^{(\alpha)} - \| \tau M_{\alpha}^p \|_p \right)^k \right]}{\sigma_X^k} \tag{5}$$

where σ_X is calculated about the origin, and k is a positive integer (Juuso & Lahdelma 2010).

Non-linear scaling in modelling

The main concepts used in this section are presented in Figure 3: the feasible range and the membership definition. It also shows the connection between the membership definition and the corresponding fuzzy membership functions.

The concept of the feasible range is defined as a trapezoidal membership function, which is based on the support and core areas as defined in the fuzzy set theory; see Zimmermann (1992). The support area is between the

minimum and maximum values of the variable. The central value divides the whole value range of X into two parts. The core area, $[(c_l)_j, (c_h)_j]$, is limited by the central points of the lower and upper parts.

Scaling functions

Non-linear scaling in this context uses the concept of membership definitions, which provide the non-linear mappings of each variable from the operation area of the (sub) system inside a real-valued interval $[-2, 2]$. In this way, each variable is scaled between $[-2, 2]$, and the variables can be compared with each other on a similar basis.

Mathematically, the non-linear scaling consists of two second order polynomials: one for negative values, $X_j \in [-2, 0]$, and one for positive values, $X_j \in [0, 2]$

$$\begin{cases} f_j^- = a_j^- X_j^2 + b_j^- X_j + c_j, X_j \in [-2, 0] \\ f_j^+ = a_j^+ X_j^2 + b_j^+ X_j + c_j, X_j \in [0, 2] \end{cases} \tag{6}$$

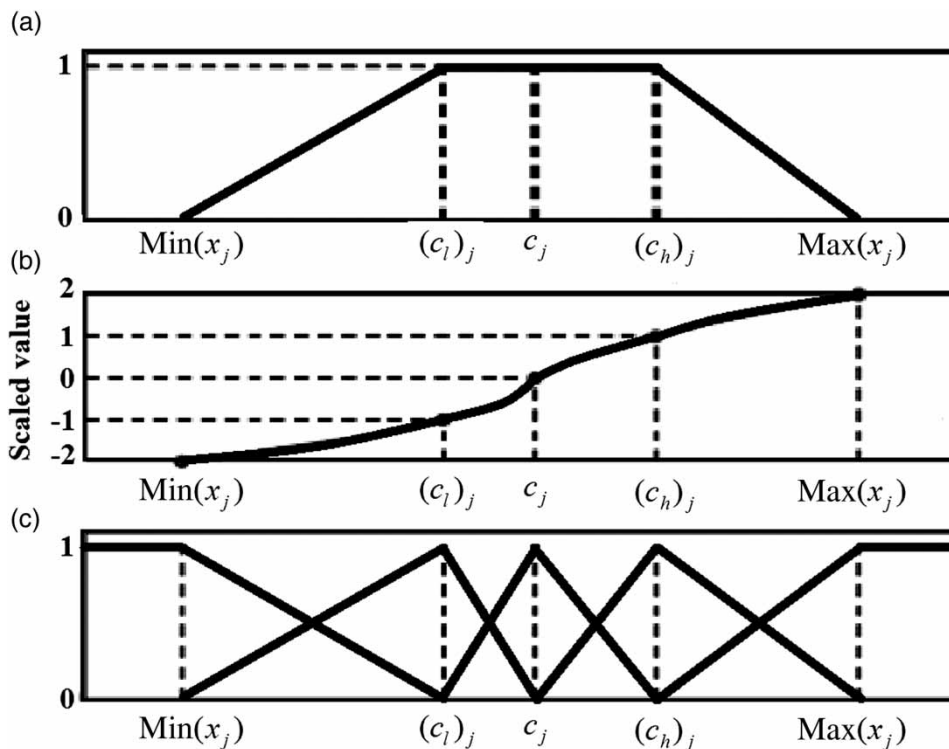


Figure 3 | (a) Feasible range, (b) scaled value, and (c) membership functions. Redrawn from Juuso (2004).

The coefficients of the polynomials are defined using the corner points of the feasible range

$$\left\{ (\min(x_j), -2), ((c_l)_j, -1), (c_j, 0), ((c_h)_j, 1), (\max(x_j), 2) \right\} \quad (7)$$

Tuning

As the membership definitions are used in a continuous form, the functions $f_j^-(X_j)$ and $f_j^+(X_j)$ should be monotonous, increasing functions in order to produce realizable systems. In order to keep the functions monotonous and increasing, the derivatives of the functions f_j^- and f_j^+ (Figure 4) should always be positive. Derivatives Dj can be presented in three groups: (1) decreasing and increasing, (2) asymmetric linear and (3) increasing and decreasing. The functions are monotonous and increasing if the ratios

$$\alpha_j^- = \frac{c_j - (c_l)_j}{(c_l)_j - \min(x_j)} \quad \text{and} \quad (8)$$

$$\alpha_j^+ = \frac{(c_h)_j - c_j}{\max(x_j) - (c_h)_j}$$

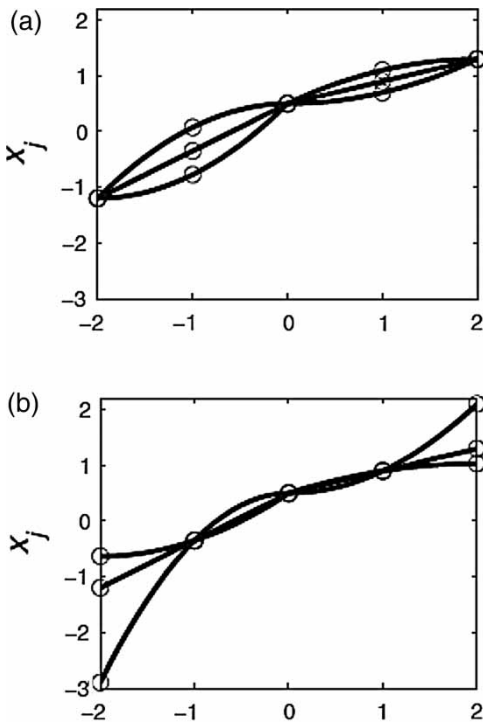


Figure 4 | Feasible shapes of membership definitions f_j : (a) coefficients adjusted with core and (b) support (Juuso 2009).

are both in the range $[1/3, 3]$. If needed, corrections are done by changing the core area, the support area or the centre point. The coefficients of the polynomials can now be calculated as

$$a_j^- = \frac{1}{2}(1 - \alpha_j^-) \Delta c_j^-,$$

$$b_j^- = \frac{1}{2}(3 - \alpha_j^-) \Delta c_j^-,$$

$$a_j^+ = \frac{1}{2}(\alpha_j^+ - 1) \Delta c_j^+ \quad \text{and}$$

$$b_j^+ = \frac{1}{2}(3 - \alpha_j^+) \Delta c_j^+,$$

where $\Delta c_j^- = c_j - (c_l)_j$ and $\Delta c_j^+ = (c_h)_j - c_j$. Membership definitions may contain linear parts if α_j^- or α_j^+ equal to 1 (Juuso 2009).

The typical way to tune the membership definition is to first define the central value c_j and the core $[(c_l)_j, (c_h)_j]$, then the ratios α_j^- and α_j^+ from the range $[1/3, 3]$, and finally calculate the support $[\min(x_j), \max(x_j)]$. The membership definitions of each variable are configured with four parameters $\{a_j^-, b_j^-, a_j^+, b_j^+\}$ and the central value c_j . The upper and the lower parts of the scaling functions can be convex or concave, independent of each other (Figure 4). Simplified functions can also be used, e.g. a linear membership definition requires two and an asymmetrical linear definition three parameters. Additional constraints can be taken into account for derivatives, e.g. a locally linear function results if the continuous derivative is chosen in the central value (Juuso 2009).

Corner points with generalized moments and norms

The value range of x_j is divided into two parts by the central tendency value c_j and the core area (as presented in Figure 3). One problem in tuning the scaling functions is in calculating the corner points of the feasible range (Juuso 2009). An approach based on (5) was introduced for estimating the central point and the core area (Juuso & Lahdelma 2010). The measurements are not derivate, i.e. $\alpha = 0$ in (5). The central point is chosen as the point where the skewness changes from positive to negative. Then the dataset is divided into two parts: a lower part and an upper part. The same analysis is done for both

subsets. The estimates of the corner points, $(c_l)_j$ and $(c_h)_j$, are the points where skewness goes to zero. The iteration is performed with generalized norms. Then the ratios α_j^- and α_j^+ are forced inside the range $[1/3, 3]$ by changing the corner points $(c_l)_j$ and $(c_h)_j$ or the upper and lower limits $\min(x_j)$ and/or $\max(x_j)$. The linearity requirement at the central point c_j is taken into account, if possible.

Trend indices applied

Trend analysis provides an indirect measurement for high level control. For any variable x_j trend index $I_j^T(k)$ is calculated from the scaled values X_j

$$I_j^T(k) = \frac{1}{n_S + 1} \sum_{i=k-n_S}^k X_j(i) - \frac{1}{n_L + 1} \sum_{i=k-n_L}^k X_j(i) \quad (10)$$

which is based on the means obtained for a short and a long time period, defined by window lengths n_S and n_L , respectively. The index value in range $[-2, 2]$ represents the strength of both decrease and increase of the variable x_j (Juuso et al. 2009).

The derivative of the index $I_j^T(k)$, denoted as $\Delta I_j^T(k)$, is used for analysing triangular episodic representations (Figure 1). The derivative is calculated simply as a difference of subsequent values of the trend index. An increase is detected if the trend index exceeds a threshold $I_j^T(k) > \varepsilon_1^+$. Correspondingly, $I_j^T(k) < \varepsilon_1^-$ for a decrease. The trends are linear if the derivative is close to zero: $-\varepsilon_2^- < \Delta I_j^T(k) < \varepsilon_2^+$. Concave upward monotonic increase (d) and concave downward monotonic decrease (b) are dangerous situations. Concave downward monotonic increase (a) and concave upward monotonic decrease (c) mean that an unfavourable trend is stopping.

In process monitoring, the concave upward monotonic increase requires special attention. New phenomena are activated when the system moves from the linear increase to this area. The scaling functions should be defined for the whole operating area. However, a wide range of measurements is only available after failure. Initial estimates are needed at the beginning for the corner points. Parameters for the lower part $[-2, 0]$ are updated when a change to episode (d) is detected with the trend analysis

based on the initial estimates of the scaling functions. Parameters for the upper part $[0, 2]$ can be updated recursively.

The severity of the situation can be evaluated by a deviation index

$$I_j^D(k) = \frac{1}{3} (X_j(k) + I_j^T(k) + \Delta I_j^T(k)) \quad (11)$$

This index has its highest absolute values, when the difference from zero is very large and is getting still larger with a rapidly increasing speed (Juuso et al. 2009). This can be understood as an additional dimension in Figure 1.

RESULTS AND DISCUSSION

In Tomperi et al. (2011, 2013) the concentration of residual aluminium in drinking water was modelled using artificial neural networks and multiple linear regression methods. The baseline of the residual aluminium level could be predicted reliably using only three important measurements: raw water temperature, raw water KMnO_4 value and coagulant chemical dosage. Prediction of peak values was possible only if some measurements from distributed water were included in the variable subset. The scaled values of residual aluminium and mentioned three variables are shown in Figure 5. Compared with the original data trends the non-linear scaling method improves the ability to identify interactions between measurement variables and changes in specific trend line. Seasonal changes in raw water temperature are clearly seen in Figure 5(b). In winter, the raw water temperature is at its lowest level and after almost a linear increase it reaches the highest value in the middle of summer. The amount of residual aluminium is high when raw water is cold even if the raw water KMnO_4 value and coagulant chemical dosage are at a low level. When the raw water temperature is higher and the raw water KMnO_4 value rises, the amount of residual aluminium is relatively low. This shows that the effectiveness of the water treatment process on the level of residual aluminium is better when raw water is warmer. Generally, winter and cold raw water cause challenges for WTPs. In addition, it seems that autumn rains and snow melt in the spring affects the quality of raw water and the efficiency

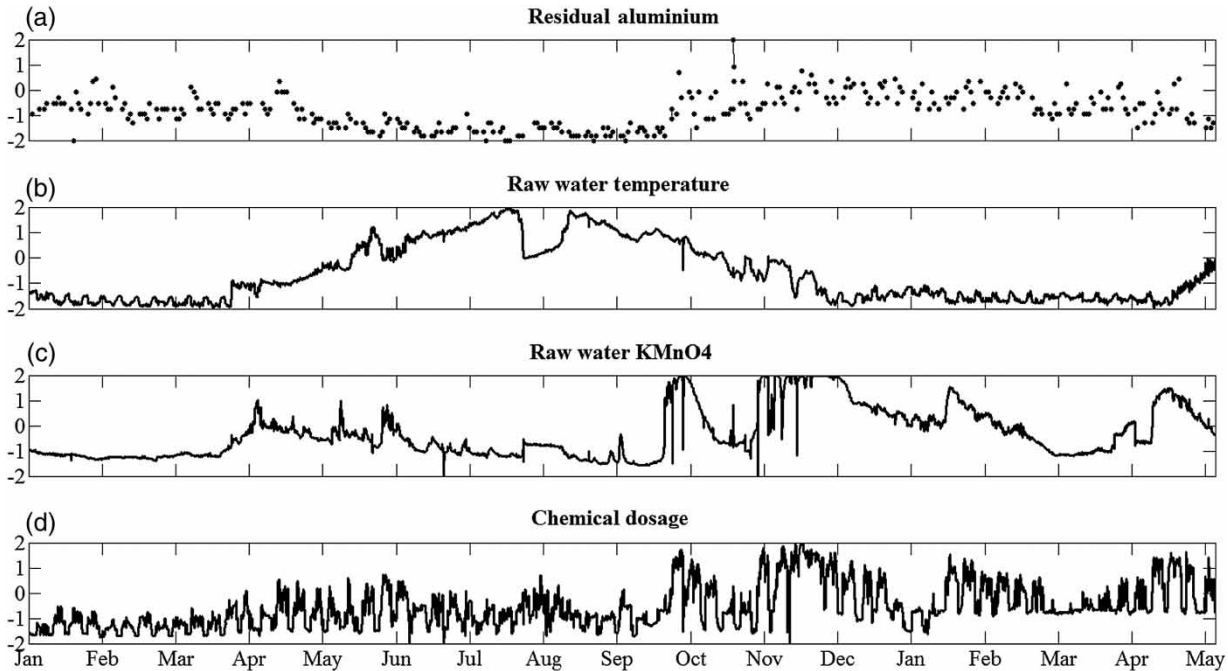


Figure 5 | Scaled trends of residual aluminium and measurement variables: (a) residual aluminium, (b) raw water temperature, (c) raw water KMnO_4 , and (d) chemical dosage.

of the water treatment process. The raw water KMnO_4 value is at a higher level in the autumn, late winter and spring due to heavy rain and snow melt.

The scaled measurement values of residual aluminium, raw water temperature and KMnO_4 are used in the trend analysis to calculate trend indices to find trend episodes and deviation indices for a short period. In simple terms, trend episodes and deviation indices show the direction, speed and the severity of the change. They provide a warning of the changes in water quality and environmental conditions that affect the residual aluminium level. A human eye is not very efficient in detecting changes in process variables, especially if the changes are slow and the monitoring window is short. The length of the time window in trend analysis depends on the process and the purpose of the analysis. Long time windows are usually applied in monitoring the process and short time windows are more suitable in control, where faster responses are demanded. In this study, the short time period n_S was 24 hours and the long time period n_L was 168 hours (1 week).

The warnings of the changes of residual aluminium level in drinking water are provided by the trend episodes and the deviation indices calculated from scaled measurements

(Figure 6). The lower the amount of aluminium in water the better. During the 16-month research period, the residual aluminium rose over the alarm limit twice: the first occasion was at the end of September and the second in the middle of October, when the level was dangerous for a short while. Here, the warning limit is set to +1 and the alarm limit to +1.5 (see the deviation index figure). The trend episodes are visualized step by step during the operation: the sequence of the points is not seen when all the points are presented in the same plot, but the position and direction of the change in the residual aluminium trend episode before it crosses the warning limit the first time is marked with an grey arrow. Residual aluminium is in a constant state (centre area of the left figure) but the state rapidly changes to the critical concave upward monotonic increase (upper right area) and returns through the linear and concave downward monotonic increase areas to the constant area. After the peak value is reached, the trend episode changes to decreasing area. The second peak that crosses the alarm limit in the deviation index figure is also observed in the increasing areas of the trend episodes, from where it rapidly moves to the decreasing area.

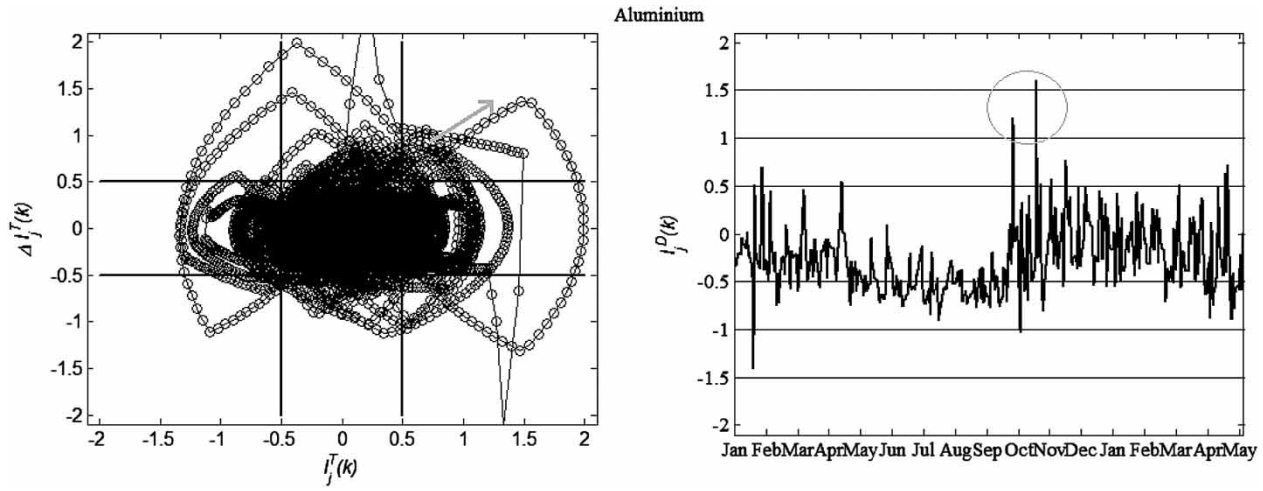


Figure 6 | The trend episode (left) and the deviation index (right) of the measured residual aluminium in drinking water.

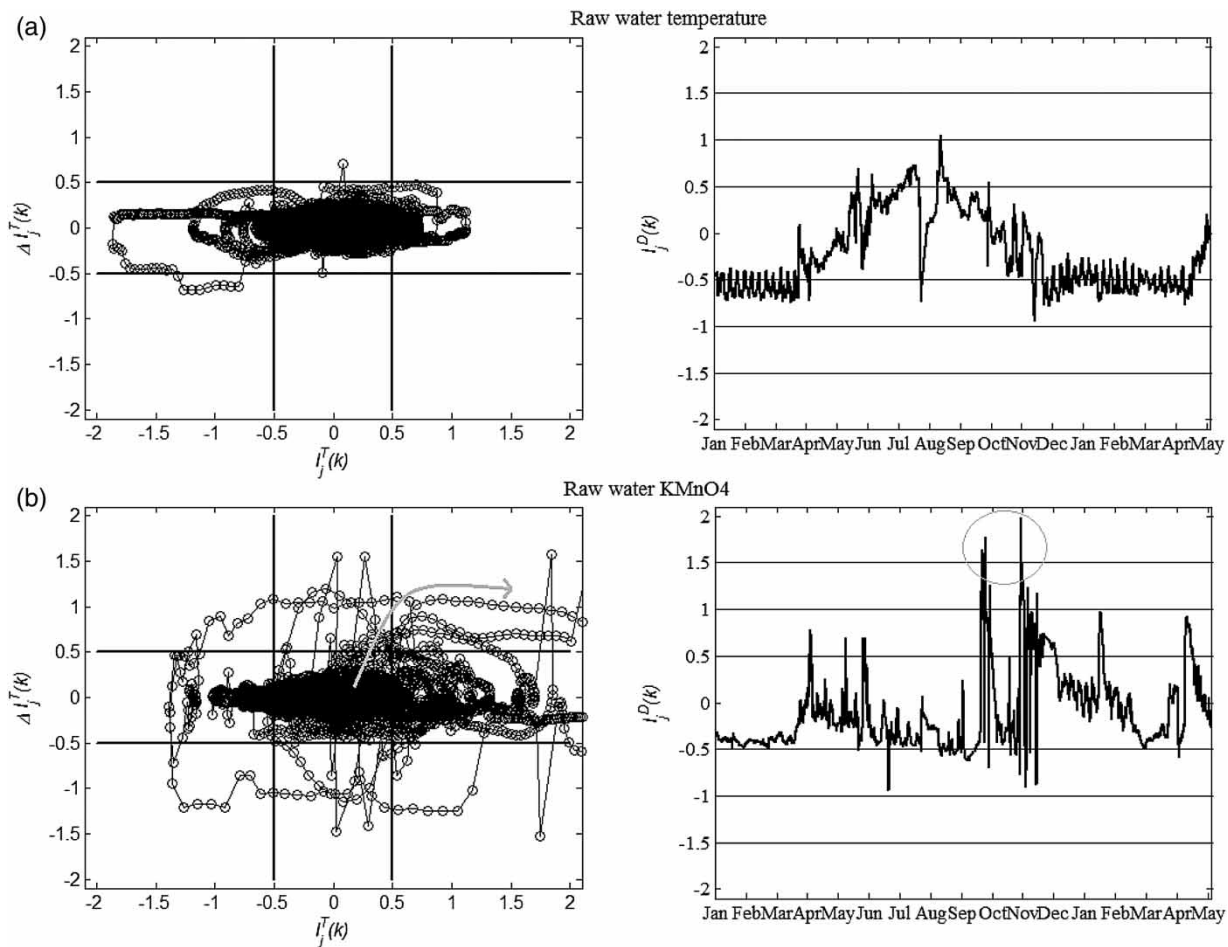


Figure 7 | Trend episodes (left) and deviation indices (right) of (a) the raw water temperature and (b) the raw water KMnO_4 value.

Raw water temperature and KMnO_4 value are not controllable variables, but their values depend on the season of the year, heavy rains, snow melt and other influences. The coagulant chemical is controlled as a function of KMnO_4 value. These variables can be used to predict the level of residual aluminium in drinking water and therefore their trend episodes and deviation indices are also studied. Changes of temperature are provided by the trend episode and the deviation index shown in Figure 7(a). The higher the temperature is, the better the water treatment coagulation chemical works. The changes in temperature were slow and almost linear. The KMnO_4 value should be as low as possible but seasonal changes and heavy rains affect it. The KMnO_4 value is stable in winter but snow melting in spring and heavy rains in summer and autumn causes the KMnO_4 value to rise rapidly and oscillate. The trend episode and the deviation index clearly show these changes (Figure 7(b)). The direction of the trend episode before the first alarming peak is marked with an arrow. The alarm limit is crossed twice during the study period, which causes the chemical dosage to rise and oscillate, and eventually the residual aluminium level to increase. The chemical dosage oscillated rapidly during the whole period as the water quality changed and crossed the warning limit couple of times, at the same time as the KMnO_4 value. The rapid rise of KMnO_4 value is observed 140 hours before the residual aluminium level crosses the warning limit. This gives a remarkable advance to the process control to keep the drinking water quality at required level.

Based on the information received from the trend analysis the water treatment process could have been controlled differently and the increased residual aluminium level would have been prevented. Because the KMnO_4 value and the raw water temperature are not controllable variables, the process optimization must be done by controlling the variables mentioned in the Introduction of this paper (raw water pH, optimal mixing and aluminium dose, etc.).

CONCLUSIONS

In this study, trend analysis was used to detect changes in the residual aluminium level, one of the most common

quality parameters of drinking water, and the operation conditions which affect it. The method of analysis can be used as well for any other quality parameter or measurement in a WTP. The increased residual aluminium level in distributed drinking water can cause health effects to consumers and financial losses to the WTP, so it is vital to be able to respond as early as possible to the changes of water quality and operate the process accordingly. It has been found that the residual aluminium level can be predicted using raw water temperature, KMnO_4 value and the coagulant chemical dosage. Seasonal changes, heavy rains and snow melt affect the water treatment process and the residual aluminium level. The non-linear scaling method improves the ability to identify interactions between the measurement variables and changes in specific trend line compared with original data trends. It was shown that the trend episodes and the deviation indices provide a good tool for detecting changes in water quality and operation conditions. The trend analysis of the measured variables has great advantages over the common approximate examination done by the process operators thus the visual monitoring is not efficient to detect the slow changes of process variables, especially if the viewed monitoring period is short.

ACKNOWLEDGEMENTS

The data used in this study were collected during the project 'Development of a comprehensive water quality management system (POLARIS)'. The writing of this paper was financially supported by KAUTE Foundation which is thereby greatly acknowledged. Co-workers in the Control Engineering Laboratory are also acknowledged for valuable comments during the writing process.

REFERENCES

- Baxter, C. W., Zhang, Q., Stanley, S. J., Shariff, R., Tupas, R.-R. T. & Stark, H. L. 2001 *Drinking water quality and treatment: the use of artificial neural networks*. *Can. J. Civil Eng.* 28 (Suppl. 1), 26–35.

- Cheung, J. T.-Y. & Stephanopoulos, G. 1990 Representation of process trends – part I. A formal representation framework. *Comput. Chem. Eng.* **14**, 495–510.
- Dash, S., Rengaswamy, R. & Venkatasubramanian, V. 2001 A novel interval-halving algorithm for process trend identification'. In: *Preprints of 4th IFAC Workshop on On-line Fault Detection and Supervision in the Chemical Process Industries*, June 7–8, Jeju Island, Korea, 173–178.
- Dash, S., Rengaswamy, R. & Venkatasubramanian, V. 2003 Fuzzy-logic based trend classification for fault diagnosis of chemical processes. *Comput. Chem. Eng.* **27**, 347–362.
- Driscoll, C. & Letterman, R. 1995 Factors regulating residual aluminium concentrations in treated water. *Environmetrics* **6**, 287–309.
- FINLEX 2000 Decree of the Ministry of Social Affairs and Health. Relating to the quality and monitoring of water intended for human consumption. No. 461. Issued in Helsinki on 19 May 2000. <http://www.finlex.fi/en/laki/kaannokset/2000/en20000461.pdf> (accessed 6 December 2013).
- Forbus, K. 1984 Qualitative process theory. *Artif. Intell.* **24**, 85–168.
- George, S., Pandit, P. & Gupta, A. B. 2010 Residual aluminium in water defluoridated using activated alumina adsorption – Modeling and simulation studies. *Water Res.* **44**, 3055–3064.
- Juntunen, P., Liukkonen, M., Pelo, M., Lehtola, M. & Hiltunen, Y. 2010 Modelling of residual aluminium in water treatment process. In: *EUROSIM 2010–7th EUROSIM Congress on Modelling and Simulation*, September 6–10 2010, Prague, Czech Republic.
- Juuso, E. K. 2004 Integration of intelligent systems in development of smart adaptive systems. *Int. J. Approx. Reason.* **35**, 307–337.
- Juuso, E. K. 2009 Tuning of large-scale linguistic equation (LE) models with genetic algorithms. In: *Revised selected papers of the International Conference on Adaptive and Natural Computing Algorithms – ICANNGA 2009, Kuopio, Finland, Lecture Notes in Computer Science* (M. Kolehmainen, ed.) Heidelberg: Springer-Verlag, 2009, LNCS 5495, 161–170.
- Juuso, E. K. 2011 Intelligent trend indices in detecting changes of operating conditions. In: *2011 UKSim 13th International Conference on Modelling and Simulation*, Cambridge, UK, 30 March – 1 April 2011. IEEE Computer Society, 162–167.
- Juuso, E. & Lahdelma, S. 2010 Intelligent scaling of features in fault diagnosis. *Proceedings of 7th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, 22–24 June 2010, Stratford-upon-Avon, UK.
- Juuso, E., Latvala, T. & Laakso, I. 2009 Intelligent analysers and dynamic simulation in a biological water treatment process. In: *6th Vienna Conference on Mathematical Modelling – MATHMOD 2009*, February 11–13 2009, Argesim Report no. 35, Argesim, 999–1007.
- Kivikunnas, S. 1999 Overview of process trend analysis methods and applications. In: *Workshop on Applications in Chemical and Biochemical Industry*, Aachen, Germany, September 15 1999, 443–452. <http://www.erudit.de/erudit/events/tc-a/erudit-tca-2-Kivikunnas-13050.PDF>.
- Kivikunnas, S., Ibatci, K. & Juuso, E. K. 1996 Process trend analysis and fuzzy reasoning in fermentation control. In: *Proceedings of IWISP'96 – Third International Workshop on Image and Signal Processing on the Theme of Advances in Computational Intelligence* (B. G. Mertzios & P. Liatsis, eds). November 4–7 1996, Manchester, UK. 137–140.
- Konstantinov, K. B. & Yoshida, T. 1992 Real-time qualitative analysis of the temporal shapes of (bio)process variables. *Am. Inst. Chem. Eng. J.* **38**, 1703–1715.
- Kuipers, B. 1985 The limits of qualitative simulation. In: *Proceedings of Ninth Joint International Conference on Artificial Intelligence (IJCAI-85)*, 128–136.
- Lahdelma, S. & Juuso, E. 2008 Signal processing in vibration analysis. In: *Proceedings The 5th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, July 15–18 2008, BINDT, Edinburgh, UK, 867–878.
- Lahdelma, S. & Juuso, E. 2010 Generalised lp norms in vibration analysis of process equipments. In: *Proceedings The 7th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, June 22–24 2010, Stratford-upon-Avon, UK.
- Lahdelma, S. & Juuso, E. 2011a Signal processing and feature extraction by using real order derivatives and generalised norms. Part 1: methodology. *Int. J. Cond. Monitor.* **1**, 46–53.
- Lahdelma, S. & Juuso, E. 2011b Signal processing and feature extraction by using real order derivatives and generalised norms. Part 2: applications. *Int. J. Cond. Monitor.* **1**, 54–66.
- Leakey, H. 2004 Aluminium Residual Management in Drinking Water Using Selection Criteria of Aluminium Based Primary Coagulants. MSc Thesis, Royal Roads University, Victoria, BC, Canada.
- Maurya, M. R., Rengaswamy, R. & Venkatasubramanian, V. 2007 Fault diagnosis using dynamic trend analysis: a review and recent developments. *Eng. Appl. Artif. Intell.* **20**, 133–146.
- McLachlan, D. R. C., Bergeron, C., Smith, J. E., Boomer, D. & Rifat, S. L. 1996 Risk for neuropathologically confirmed Alzheimer's disease and residual aluminium in municipal drinking water employing weighted residential histories. *Neurology* **46**, 401–405.
- Metz, R. 2010 Filter-design and model-based analysis of trends and cycles in the presence of outliers and structural breaks. *Cliometrica* **4**, 51–73.
- Poirier, P. J. & Meech, J. A. 1993 Using fuzzy logic for online trend analysis. In: *In 2nd IEEE Conference on Control Applications*, Vancouver, BC, 83–86.
- Stephanopoulos, G. & Han, C. 1996 Intelligent systems in process engineering. *Comput. Chem. Eng.* **20**, 743–791.
- Stephanopoulos, G., Locher, G., Duff, M. J. & Kamimura, R. 1997 Fermentation database mining by pattern recognition. *Biotechnol. Bioeng.* **53**, 443–452.
- Tomperi, J., Pelo, M. & Juuso, E. 2011 Predictive model for residual aluminium in a water treatment process. In: *SIMS 2011 the 52nd International Conference of Scandinavian Simulation Society*. September 29–30 2011, Västerås, Sweden.

Tomperi, J., Pelo, M. & Leiviskä, K. 2013 [Predicting the residual aluminum level in water treatment process](#). *Drinking Water Eng. Sci.* **6**, 36–46.

WHO 2003 *Aluminium in Drinking-water: Background document for Development of WHO Guidelines for Drinking-water Quality*. World Health Organization, Geneva.

WHO 2008 *Guidelines for drinking-water quality*. [Electronic resource]: incorporating 1st and 2nd addenda, Recommendations, Vol. 1, 3rd edn, World Health Organization, Geneva.

Zimmermann, H.-J. 1992 *Fuzzy Set Theory and its Applications*. Kluwer Academic Publishers, Boston, MA.

First received 15 April 2013; accepted in revised form 12 October 2013. Available online 27 December 2013