International Conference on Hydroinformatics

2014

Soft Sensing The Potential Amount Of Calcium Carbonate Precipitate In Drinking Water Distribution Infrastructure And Warm Water Household Appliances

Dirk Vries

Joost Van Summeren

Benjamin Van Den Akker

Alex Van Der Helm

Ignaz Worm

See next page for additional authors

Authors

Dirk Vries, Joost Van Summeren, Benjamin Van Den Akker, Alex Van Der Helm, Ignaz Worm, and Peter van Thienen

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

SOFT SENSING THE POTENTIAL AMOUNT OF CALCIUM CARBONATE PRECIPITATE IN DRINKING WATER DISTRIBUTION INFRASTRUCTURE AND WARM WATER HOUSEHOLD APPLIANCES

D. VRIES (1), J. VAN SUMMEREN (1), B. VAN DEN AKKER(1), A.W.C. VAN DER HELM(2), G.I.M. WORM(3,4), P. VAN THIENEN(1)

(1): KWR Watercycle Research Institute, P.O. Box 1072, 3430 BB Nieuwegein, the Netherlands

(2): Waternet, P.O. Box 94370, 1090 GJ, Amsterdam, the Netherlands

(3): PWN, PO Box 2113, 1990 AC, Velserbroek, the Netherlands

(4): Delft University of Technology, Stevin Laboratorium, PO Box 5048, 2600 GA, Delft, the Netherlands

A soft sensor is developed to predict the potential amount of precipitation of calcium carbonate (CCPP) in warm water household devices and scaling or corrosive behavior in water distribution networks. With the aid of a water supply network model, it is shown that the soft sensor is able to predict CCPP levels at pre-specified downstream nodes using only measurements at a limited set of upstream nodes. Furthermore, the soft sensor consists of a data assimilation algorithm to provide for best estimates of the CCPP and confidence intervals.

INTRODUCTION

Full automization in drinking water production can help ensure the demand of reliable, high quality drinking water. On-line sensoring to aid automation, monitoring and control will thus become increasingly important [1]. Relevant techniques include numerous numerical and statistical methods that enable filtering of noisy sensor signals, the estimation of unmeasured state variables from measured signals by process models and the computation of confidence intervals. Together, these techniques enable the development of advanced, so-called soft sensors that can further extend the applicability of currently available and installed sensors.

In this work, a soft sensor for a customer-oriented parameter is developed. This soft sensor predicts the potential amount of precipitation of calcium carbonate (scaling) in warm water household devices and scaling or corrosive behavior in water distribution networks, in order to take appropriate measures when predicted values exceed acceptable levels. The accuracy of the soft sensor crucially depends on models for transport and chemical reaction, as well as on the quality (variance) of input to the soft sensor. Furthermore, it is good practice to use data assimilation to synchronize the estimation of soft sensor forecast variability and water quality parameters. However, the estimation of the water quality parameters in a water supply network using a trade-off with measurements is a challenge, because (i) the network typically exhibits more than a thousand nodes and (ii) scaling and speciation reactions are non-linear and (iii) the update of the filter gain to enable an optimal trade-off between measurement data and model predictions involves the inverse of the model state and data covariance matrix. Especially the combination of (i) and the evaluation in (iii) leads to a computational burden with conventional data assimilation techniques like the (extended) Kalman filter. Such computations involve $O(n^3)$ operations [5], with n equal to the number of model states: that is, a multiplication of n_x the number of water quality parameters by n_n the number of nodes ($n=n_x n_n$). Alternatively, the ensemble Kalman filtering (enKF) algorithm [3,4] can deal with a (very) large number of model states and sparse observations. Importantly, enKF is computationally far less intensive because computation of the inverse of the state covariance matrix can be circumvented by running q parallel copies of the model (the so-called ensemble set with q members) and calculating the ensemble mean. The latter operation is of the order $O(qpn)$ [5], where p is the dimension of the measurement vector (in this case the water quality parameters measured at a subset of nodes).

In this work, the enKF is implemented to provide best estimates of the Langelier Saturation Index (SI) and the Calcium Carbonate precipitation potential (CCPP), using on-line measurements of water quality parameters and a transport-chemical speciation model. Additionally, the soft sensor algorithm provides confidence regions around the estimates.

METHOD

The development of the soft sensor is preceded by an assessment of which water quality parameters affect the *SI* and *CCPP* most. The analysis is based on an evaluation of measurement data within the water supply network of Amsterdam, the Netherlands.

The layout of the soft sensor consists of two components: (1) a model core that calculates transport of water and the chemical speciation and reactions within the water supply network, and (2) a data assimilation algorithm that (a) recursively updates model estimates at *all nodes*, including the ones at which measurement data are collected and (b) calculates the associated covariance matrices. The model (1) is an EPANET-MSX [6] module that is source linked to PHREEQC [7] to enable accurate speciation and lime precipitation calculations following the guidelines as specified in [2]. The enKF data assimilation is performed by a Monte Carlo approach that exploits several (possibly parallel) model calculations in order to provide best estimates and the covariance on the state variables, here the values of the water quality parameters specified on the water supply network grid. The workflow of the soft sensor is designed as follows.

Soft sensor workflow and assumptions

The workflow of the sensor is as follows:

- For simulation purposes, the user specifies a set-point or estimated pressure pattern over the distribution network with time horizon $h = [t_k, t_{k+h}]$ for the input (or source) nodes. Horizon h is typically a couple of hours to one or few days. Here, it will be 24 hours. In practice, the soft sensor is fed with real time inputs at their corresponding nodes;
- Input (source) nodes i are specified, i.e. the locations where upstream measurements of water quality parameters (temperature, pH, conductivity) are collected on-line;
- At observation nodes j, measurements are collected to improve soft sensor estimates of the water quality parameters (typically temperature, calcium, sodium, magnesium, carbonate, chloride, phosphate and sulfate concentration levels);
- Based upon the (assumed) on-line measurements at the source input and observation nodes, the soft sensor then recursively

i. calculates the transport of water and chemical speciation in the time frame [t_k, t_{k+1}] over the *whole* distribution network using updated water quality *inputs* at nodes *i* and a previous estimate of the water quality parameters at *all the other* nodes $n_n = N_t - i$, where N_t is the total number of nodes;

ii. collects the model outputs and updates the (previous) estimates of water quality parameter values, *SI* and *CCPP* (defined as *state variables*) and an estimate of their covariance matrix at *all* nodes n_n and, simultaneously, updates estimates at the observation nodes *j* whenever a measurement becomes available. Typically, measurements at nodes *j* are water quality measurements sampled and analyzed off-line at time instants t_{obs}, hence t_{obs} is typically a much smaller subset within the time horizon *h*.

The set-up of the model is schematically outlined in Figure 1, using the following definition of the nodes.

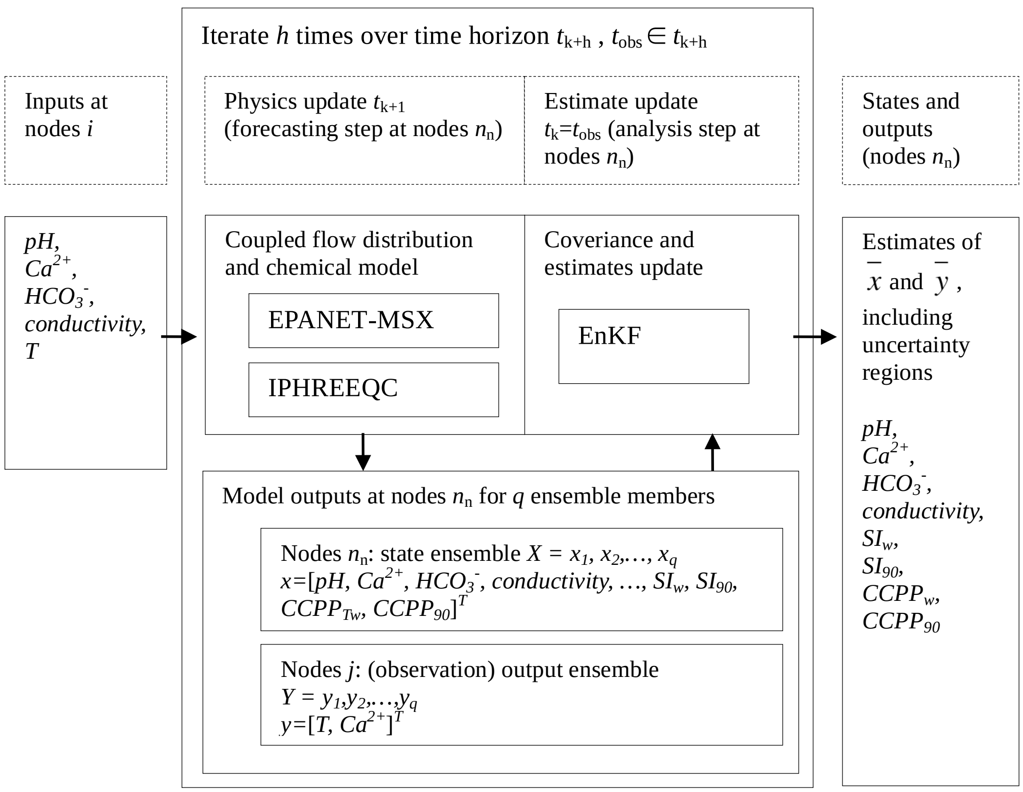

Figure 1. Overview of data assimilation that calculates the CCPP and SI for all nodes in the network *p*, fed by measured parameters at nodes *i* and *j*.

The total node set of network is defined as *L*, where soft sensor observation nodes (the measurement locations) are defined as a predefined subset $J = \{j_1, j_2 \ldots j_m\}$, sensor input nodes $i \in I$ (the sources) and all the water quality parameters of interest are calculated at nodes $n_n \in N_n$. Hence: $I, J, N_n \subseteq L$ and $N_n = L \cap I$.

Network and water quality at the source

The soft sensor is tested with a small example network with 5 nodes including a source node. The network model is adopted from Shang *et al.* [6] and is shown in figure 2.

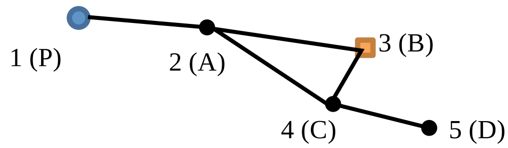

Figure 2. Example network model, with source node *i*=1 (blue circle P), nodes where state variables are calculated $n_n \in N_n = \{2,3,4,5\}$ and the observation node *j*=3 (orange square B).

Soft sensor simulation settings

The soft sensor simulation is executed by the following (fictuous) assumptions and settings: at the source, the calcium, carbonate and temperature vary during a day horizon (24 hours) with sinusoidal shape, whereas the pH is assumed constant at a value of 8.15. Furthermore, the ensembles for concentrations, temperature and pH are perturbed with normally distributed white noise around zero mean ($N(0,\sigma)$). All other signals (EGV, SI and CCPP) are not perturbed. The ensemble set is filled with16 members.

The set-up of the simulation run is summarized in Table 1.

Table 1. Simulation settings of the soft sensor: input signals, (generated) noise parameter values and sampling instants of measurements. N.a.: not applicable.

Water quality parameter	**Perturbation signal**	**Signal at P (on-line)**	**Measured at node B**
T	$N(0,\sigma=0.5)$	$20 + 5\sin(\pi t_k/12 - \pi/2)$	$t_k=[0,6,12,18]$
pH	$N(0,\sigma=0.001)$	8.15	n.a.
Ca^{+2}	$N(0,\sigma=0.2)$	$40 + 4\sin(\pi t_k/12 - \pi/2)$	$t_k=[0,6,12,18]$
HCO_3^-	$N(0,\sigma=0.1)$	$180 + 3\cos(\pi t_k/12 - \pi/2)$	n.a.
Specific conductivity	none	n.a.	n.a.
SI	none	n.a.	n.a.
CCPP	none	n.a.	n.a.

Numeric notation of the nodes is omitted in the following, instead Latin characters are used.

RESULTS

Data analysis of water quality parameters

Data analysis of a real water supply network of Amsterdam shows that lime precipitation (*SI* and *CCPP*) depends primarily on (and increases with) measured calcium and carbonate concentrations, pH-values, and temperature. Precipitation levels increase substantially when drinking water is heated from room temperature to near boiling temperature in household appliances. In Figure 2 it is shown, that the dependence of SI on calcium, carbonate, pH-values, and temperature outweigh the influence of measuring errors and measured concentrations of other drinking water species such as chloride (ions), carbon dioxide, magnesium, nitrate and sulfate.

The data analysis shows that trends in *SI* (Figure 3) and *CCPP* (not shown here) are comparable, but values are substantially higher at higher temperature.

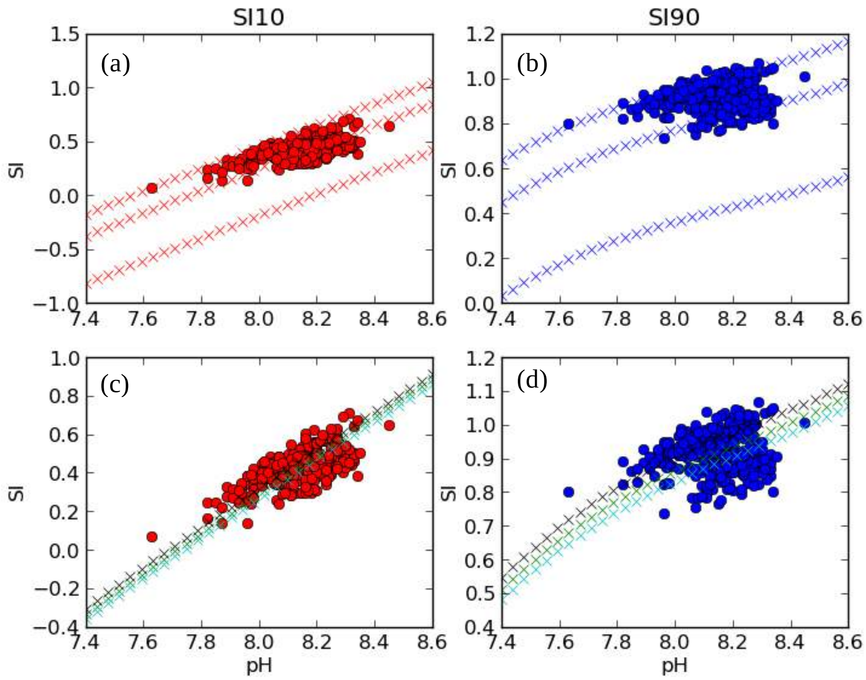

Figure 3. *SI* calculated for water quality parameters measured in an existing drinking water distribution system. Red dots in panel (a) show *SI* values corrected to a temperature of 283K (10°C) and pH values as indicated by the horizontal axis. Curves with cross-symbols represent *SI* values calculated for Ca x HCO_3 concentration products of 1, 3 and 5 $mmol^2/l^2$ (bottom to top), and this covers most of the distribution of SI values. Panel (c) relates the influence of Ca^{+2}, HCO_3^-, and pH (red dots, same data as in panel a) to the influence of measured Cl^-, CO_2, Mg^{+2}, NO_3^-, and SO_4^{-2} concentrations. Curves with cross-symbols represent SI values calculated with Ca^{+2} and HCO_3^- concentrations averaged over all measurements and with Cl^-, CO_2, Mg^{+2}, NO_3^-, and SO_4^{-2} concentrations of 0x (cyan), 1x (green), and 2x (black) the values averaged over all measurements. Panels (b,d) show the same as panels (a) and (c) but for a temperature of 90°C.

Soft sensor simulation

The soft sensor is run with the example network (ref. Figure 2) and set with water quality parameters at the source node (P) and measurements at node B (*j*=3) as shown in Table 1. We show the true (simulated) water quality parameter values together with the forecasts of the water quality parameters and their ensemble members in Figure 4. The results of the specific conductivity are omitted since it holds comparable information as the calcium and carbonate signals.

Notice that the ensemble forecasts diverges from the true signal when further away from the source nodes, which is expected since any model errors (i.e. the introduced perturbations) are transported through the network and are added at all the nodes every (new) time instant. Furthermore, Figure 4 reveals that the ensemble diverges within the time periods that no measurements are fed to the soft sensor (in the periods between time instant 0 and 6; 7 and 12, etc.). Figure 5 shows the influence of the water quality on the SI and CCPP at 10 and 90°C. The CCPP scales (non-linearly) with elevated temperature and is strongly related to the SI.

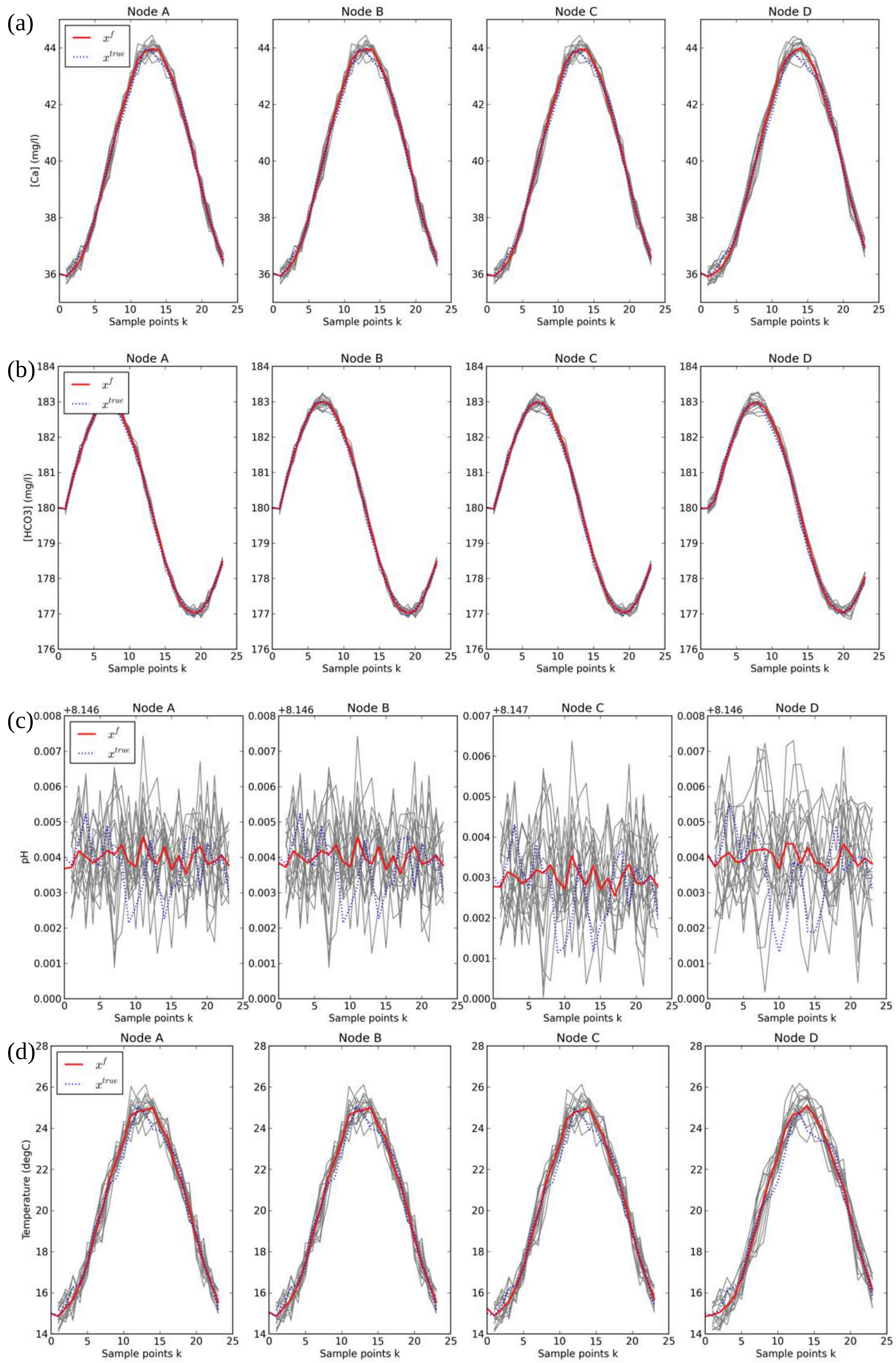

Figure 4. From upper (a) to lower panel (d): Ca^{+2} (mg/l), HCO_3^-(mg/l), pH (-) and Temperature (°C) with 'true' values at different nodes (blue dashed), forecasted values (red solid lines) and calculations by the 16 ensemble members (thin gray lines). From left to right, hourly forecasts are shown for Node A, B, C and D.

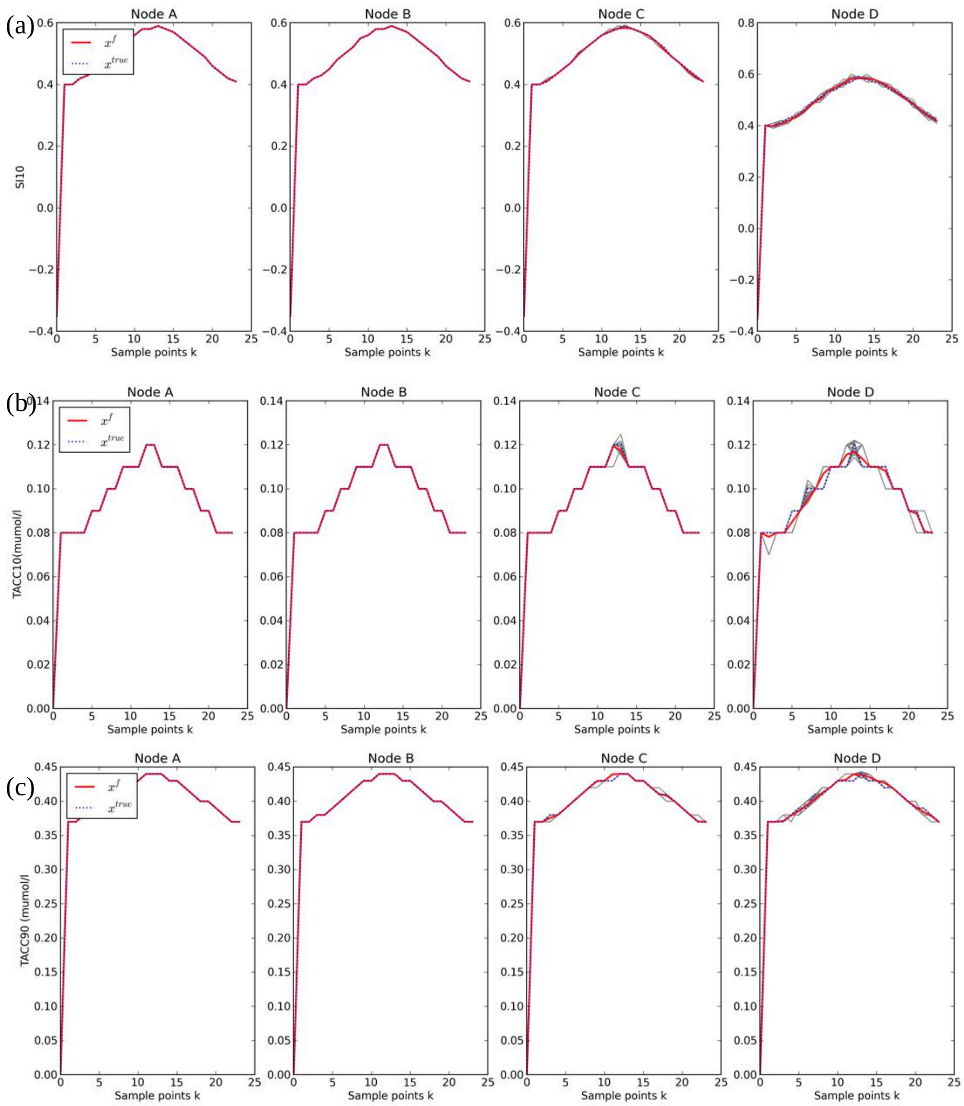

Figure 5. From upper (a) to lower (c) panel: SI10 (-), CCPP10 and CCPP90 (mmol/l) respectively, with ‘true’ values (blue dashed), forecasts (red lines) and calculations by the 16 ensemble members (thin gray). Left to right, hourly forecasts for Node A, B, C and D.

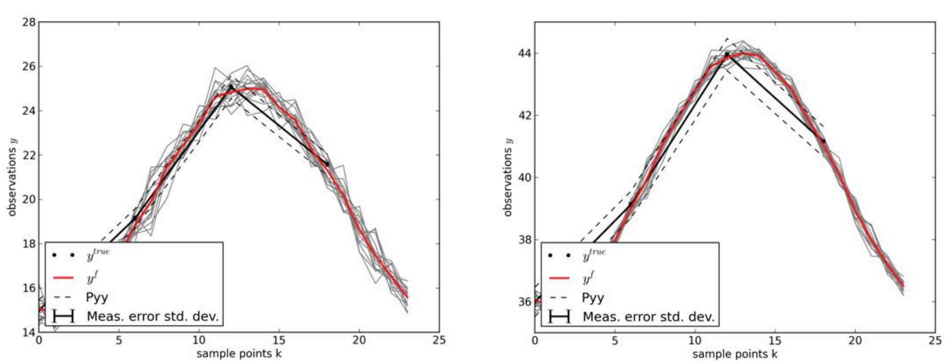

Figure 6. Left: temperature (°C), right: Ca^{+2} concentration (mg/l), both measured at node B.

Samples are taken at t_k= 0, 6, 12 and 18 (black dots with std. deviation bars), forecasts (red line), the 16 ensemble members (gray lines) and their variance (black dashed).

The observed temperature and Calcium concentration are shown in Figure 6, along with their standard deviation and ensemble variance (P_{yy}).

CONCLUSIONS AND OUTLOOK

Data analysis of a real network reveals that calcium and carbonate concentrations, pH-values, and temperature are the most relevant water quality parameters to influence lime precipitation. A soft sensor was developed by coupling a transport model with a chemical reaction model and wrapped in an ensemble Kalman filtering routine to provide best estimates on the most important set of water quality parameters, *SI* and calcium carbonate precipitation potential (*CCPP)*. A proof-of-principle was shown by using this data assimilation routine to forecast the water quality within a small water supply network.

The developed soft sensor is a promising tool for monitoring and decision support regarding appropriate measures when predicted values exceed acceptable levels. Testing of the soft sensor with a real water supply network, speed-up of the soft sensor calculations and an ensemble perturbation strategy to improve the estimated water quality parameters, are aspects that the authors would like to work on in future.

Acknowledgements

This work has been made possible by financial support via the Joint Research Program of Dutch water companies (BTO). The authors would like to thank prof. Luuk Rietveld, Rein Wuestman, Perry van der Marel and Luc Zandvliet for contributing in fruitful discussions.

REFERENCES

[1] Worm G.I.M., Oudmans J., van Schagen K. and Rietveld L.C., “A company specific simulator of a drinking water treatment plant”. *EU-ROSIM 2007* (2007).
[2] de Moel P. J., van der Helm A. W. C., van Rijn M., van Dijk J. C. and van der Meer W. G. J., “Assessment of calculation methods for calcium carbonate saturation in drinking water for DIN 38404-10 compliance”. *Drink. Water Eng. Sci. Discuss*, 6, (2013), pp. 167-198.
[3] G. Evensen, The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53, (2003), pp. 343-367.
[4] Houtekamer, Peter L., and Mitchell H.L. "Data assimilation using an ensemble Kalman filter technique." *Monthly Weather Review* 126, no. 3 (1998), pp 796-812.
[5] Gillijns S., Mendoza O. B., Chandresakar J., De Moor B.L.R., Bernstein D.S., and Ridley A., What is the ensemble Kalman Filter and how well does it work? *Proc. Of the 2006 American Control Conference*, Minneapolis, USA, (2006), pp 4448-4453.
[6] Shang F., Uber J. G., and Rossman L. “EPANET Multi-Species Extension Software and User’s Manual”. *EPA/600/C-10/002*, U.S. EPA, Washington, DC (2008).
[7] Parkhurst D.L., and Appelo C.A.J. “Description of input and examples for PHREEQC version 3—A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations”. *U.S. Geological Survey Techniques and Methods*, book 6, ch. A43 (2013).