

Characterizing long-term wear and tear of ion-selective pH sensors

Kito Ohmura, Christian M. Thürlimann, Marco Kipf, Juan Pablo Carbajal and Kris Villez 

ABSTRACT

Today, the development and testing of methods for fault detection and identification in wastewater treatment research relies on two important assumptions: (i) that sensor faults appear at distinct times in different sensors and (ii) that any given sensor will function near-perfectly for a significant amount of time following installation. In this work, we show that such assumptions are unrealistic, at least for sensors built around an ion-selective measurement principle. Indeed, long-term exposure of sensors to treated wastewater shows that sensors exhibit fault symptoms that appear simultaneously and with similar intensity. Consequently, this suggests that future research should be reoriented towards methods that do not rely on the assumptions mentioned above. This study also provides the first empirically validated sensor fault model for wastewater treatment simulation, which is useful for effective benchmarking of both fault detection and identification methods and advanced control strategies. Finally, we evaluate the value of redundancy for remote sensor validation in decentralized wastewater treatment systems.

Key words | data quality, drift, fault detection and identification, ion-selective electrodes, predictive maintenance, wastewater

Kito Ohmura
Christian M. Thürlimann
Marco Kipf
Juan Pablo Carbajal
Kris Villez  (corresponding author)
 Eawag,
 Überlandstrasse 133, CH-8600 Dübendorf,
 Switzerland
 E-mail: kris.villez@eawag.ch

Kito Ohmura
 Toshiba Infrastructure Systems & Solutions
 Corporation,
 Tokyo,
 Japan

Christian M. Thürlimann
Kris Villez
 ETH Zürich, Institute of Environmental Engineering,
 8093 Zürich,
 Switzerland

INTRODUCTION

By several accounts, the lack of online sensor data quality poses a long-standing challenge (Rieger *et al.* 2005, 2006, 2010; Rosén *et al.* 2008). It is therefore not surprising that considerable time and energy has been invested in methods for automated quality assessment and quality control of online instruments (e.g. Thomann *et al.* 2002; Thomann 2008; Corominas *et al.* 2011; Spindler & Vanrolleghem 2012; Alferes *et al.* 2013; Spindler 2014; Villez *et al.* 2016; Le *et al.* 2018). Moreover, the wealth of literature and the number of reviews on these and related topics (Venkatasubramanian *et al.* 2003a, 2003b, 2003c; Haimi *et al.* 2013; Corominas *et al.* 2018) suggest that the science and practice of fault detection and identification (FDI) is far from settled.

Data-analytical techniques can enable automated and remote detection of sensor faults. Without exception, such techniques rely on redundant relationships and can therefore be categorized by the type of redundancy that is used. A first category relies on reference laboratory measurements and computing a deviation between online sensor signal and a reference measurement. A second category relies on

hardware redundancy by placing multiple online sensors in the same location. A third category relies on temporal redundancy, the idea that meaningful changes in the sensor signal can only be smooth when measured with a sufficiently high frequency. Finally, the fourth category relies on spatial redundancy, relating signals produced at distinct locations or for different variables. Examples of this category include methods based on first principles; for example, balance equations, and methods rooted in statistical practice; for example, principal component analysis. Importantly, each of these advanced methods require tuning to find an optimal trade-off between correct alarms and false alarms. Invariably, such tuning is obtained by means of a historical, fault-free data set from which acceptable limits for computed residuals are derived. Consequently, optimal tuning relies on the availability of representative data of an acceptable quality. In addition, most techniques implicitly require that sensor fault symptoms appear independently from each other; that is, the probability that two sensor faults start at the same time is assumed to be zero.

In contrast to the tremendous amount of research on FDI methods, little is actually known about the cause-and-effect relationships between sensor ageing, the onset of sensor faults and failures, and the production of faulty data. This is explained by the fact that the availability of information describing the exact circumstances under which faults occur or faulty data are produced; that is, meta-data, is usually severely limited. For this reason, we report on a first-time study of sufficient scale to study the long-term effects of ageing on the performance of online water quality sensors. Additional insight is obtained by studying a variety of dynamic models to describe the empirical observations.

MATERIALS AND METHODS

We are particularly interested in the ageing process of typical pH sensors exposed to high-strength wastewaters. For this reason, a number of pH sensors are exposed to the contents of a reactor used to study process control strategies for nitrification of anthropogenic urine (Thürlimann *et al.* 2019). We study the effects of sensor ageing by regular evaluation of the sensors' offset and sensitivity through exposure of these sensors to calibration media. To do this, it is sufficient to record the raw sensor signal (electrode potential). This also avoids the need for regular calibration. The next paragraphs explain our approach in detail.

Theoretical and real-world behavior of ion-selective electrodes for pH measurement

The ion-selective measurement principle for pH measurement is understood rather well. In steady state, one measures an electric potential E (in mV) between a reference electrode and an electrode exposed to the medium. This potential is described by the Nernst equation (Westcott 2012) as a function of the proton activity ($[H^+]$) and temperature (T) of the medium to which the sensor is exposed:

$$E = E^0 + \frac{R \cdot T}{F} \ln([H^+]) \quad (1)$$

where E^0 is the reference potential, F is the Faraday constant (96485.33289 C/mol, Taylor *et al.* 2007), $[H^+]$ is the proton activity in the reference cell, R is the molar gas constant (8.3144598 J/(mol · K), Taylor *et al.* 2007), and T is the temperature measured in Kelvin. The pH is defined as $\log_{10}([H^+])$ (Buck *et al.* 2002) so that $S(T)$ is the

temperature-specific sensitivity, which can be computed as:

$$S(T) = \frac{R \cdot T}{F \cdot \log_{10}(e)} \quad (2)$$

Most typically, pH sensors are designed to deliver 0 mV at pH 7 so that E^0 is theoretically 0 mV. Similarly, the theoretical sensitivity at standard ambient temperature and pressure (SATP) is $S(298.15) = 59.1593$ mV per pH unit. The theoretical potential at pH 4.01 and SATP is 177.0 mV. Since the actual values of these parameters can deviate from their theoretical values, it is common to apply a 2-point calibration procedure to estimate them, often with buffered calibration media with pH 4.01 and 7.00.

Studied sensors

A total of 12 pH sensors belonging to five sensor types (T1-T5) are studied. All sensors are produced by Endress + Hauser (Reinach, Switzerland). The first eight sensors are pairs of four commercially available sensor types (T1-T4). The first (second) sensor in each pair is designated with an a (b), e.g. T1a, T1b. The remaining pH sensors (T5a, T5b, T5c, and T5d) are replicates of a recently developed prototype (T5). The T1, T2, and T3 sensors are built around a glass electrode while the T4 and T5 are built around an ion sensitive field-effect transistor (ISFET). The complete data series produced by these sensors is released as an open data set (Ohmura *et al.* 2019).

The first three sensor pairs (T1-T3) were in use throughout the exposure experiment, which lasted 731 days (4 October 2016 – 4 October 2018). An overview of this experiment is given in Figure 1. The fourth pair (T4) has been in use during the first half year and was replaced with the fifth pair (T5) on 3 April 2017 (day 182) (i) as the T4 sensors exhibit a long response time (not shown) and (ii) the T5 prototypes could then be tested. The T5a sensor stopped producing a meaningful signal on 30 June 2017 (day 270) while T5b became faulty (details below) on 31 August 2017 (day 332). These sensors were replaced with another sensor of the same prototype (T5) on 2 October 2017 (day 364). In this last pair, one sensor (T5d) failed within 1 day (day 365) while the other (T5c) was fully functional until the end of the experiment.

Long-term exposure experiment

The contents of a 12 L urine nitrification reactor are pumped through a recycle loop that includes a closed tube holding the

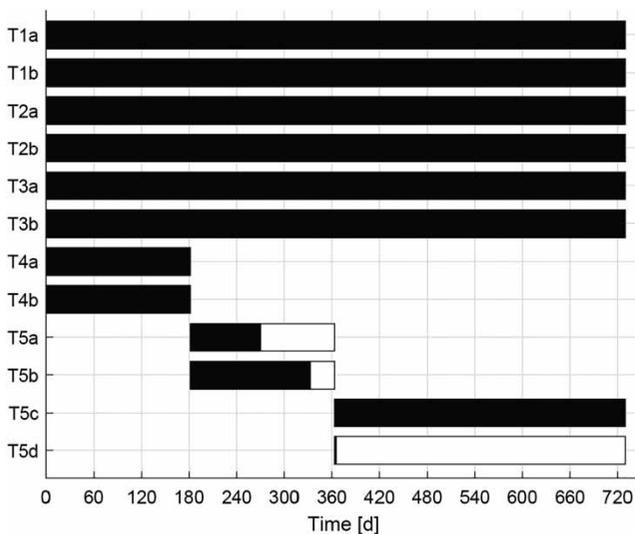


Figure 1 | Overview of the complete experimental campaign. The periods of sensor exposure are indicated by rectangles. The periods during which the sensors produced meaningful data are marked in black.

sensors in a fixed position. The design of this tube equipped with sensor-holding locks is shown in the Supplementary Information (Section B) (available with the online version of this paper). The recycle flow rate was 43 L/h.

The reactor was fed with urine collected from male lavatories in the Forum Chriesbach building at Eawag, with exception of the period from 30 April 2018 to 21 June 2018 (day 574–625), when it was collected from female lavatories in the same building. From 4 October 2017 to 24 November 2017 (day 366 to 417), the reactor was additionally fed with a nitrite stock solution. During the experimental period, the measured concentrations of nitrogen species in the nitrified urine ranged between 1,180 and 2,730 mgN/L (mg atomic nitrogen per liter) for total ammonia, 0 and 82 mgN/L for nitrite, and 1,290 and 2,720 mgN/L for nitrate (Thürlimann *et al.* 2019). These time series are visualized in the Supplementary Information (Section C) (available online). The pH value of the nitrified urine, as measured by two pH sensors installed directly in the reactor, ranged between 5.7 and 7.3.

Sensor characterization tests

At regular intervals, the sensors were removed from their normal position, cleaned with a wet sponge, and exposed to calibration media. This was executed 47 times in total to quantify the effect of long-term exposure on the sensors' offset and sensitivity. The exact times of these sensor characterization tests are listed in the Supplementary Information (Section G.1) (available online). Two pairs of tests were

executed on the same day to ensure acceptable experimental reproducibility (day 70: tests 11–12; day 351: tests 29–30). The selected media include (C4) pH 4.01 calibration solution (CPY20-C10A1, Endress + Hauser, Reinach, Switzerland); (C7) pH 7.00 calibration solution (CPY20-E10A1, Endress + Hauser, Reinach, Switzerland); (U4) nitrified urine at pH 4; (U7) nitrified urine at pH 7; and (W) tap water. For the present work, the exposure to W, C4, and C7 is relevant. This occurs in five distinct phases (P0–P4), each lasting at least 5 minutes and exposing the sensors to W, C4, C7, C4, and W (in this order). Exemplary results are shown in Figure 2 and discussed in detail below.

The offset (E^0) and two measurements of the sensitivity (S_D and S_R) are computed with raw potential measurements recorded during P1, P2, and P3. In line with Carr (1993), the following steps are applied for every sensor characterization test:

1. Compute the median value among the potential measurements collected in P1, P2, and P3 between 2 and 1 minutes before the start of the next phase (P2, P3, and P4). Refer to these values as E^{P1} , E^{P2} , and E^{P3} .

2. The sensor offset is defined as $E^0 = E^{P2}$.

3. The decay potential sensitivity is defined as

$$S_D = \frac{E^{P1} - E^{P2}}{7.00 - 4.01}$$

4. The rise potential sensitivity is defined as

$$S_R = \frac{E^{P3} - E^{P2}}{7.00 - 4.01}$$

These steps are demonstrated below with a practical example.

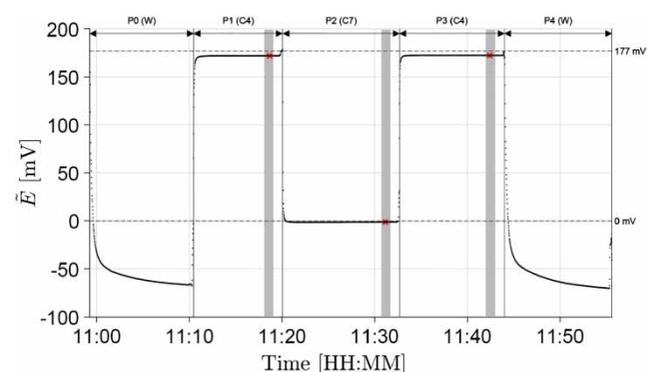


Figure 2 | Exemplary sensor characterization test. Raw data obtained in the first sensor characterization test with sensor T1a. The measured potential decays during P0, P2, and P4, while it increases during P1 and P3. Steady state is reached quickly in P1, P2, and P3. The theoretical potential values for P1, P2, and P3 are indicated with dashed horizontal lines. Grey shading indicates the data used to obtain the potential measurements (2 to 1 minute before phase change). The selected median potential values are shown with red crosses. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/wst.2019.301>.

Drift models

The results shown below indicate that the offset significantly varies over time while the sensitivity remains remarkably stable in all studied sensors. We describe the observed drift of the offset by means of three models.

Model 1 – constant trend followed by linear trend

For the first model, we apply a modified version of the excessive drift model proposed for the Benchmark Simulation Model family (Rosén et al. 2008). This model simulates $E^0(t)$, the sensor offset, as:

$$E^0(t) = d_o + r_d \cdot \max(t - t_f, 0) \quad (3)$$

with d_o the initial offset, r_d the drift rate parameter, t the time since sensor installation, and t_f the time of the drift onset. The applied modification consists of adding the parameter d_o . To fit this model, the offset measurements, $\tilde{E}^0(t_h)$, collected at discrete time instants t_h , are assumed to exhibit independently and identically distributed measurement errors, ϵ_h , drawn from a normal distribution with zero mean and standard deviation, σ_ϵ :

$$\tilde{E}^0(t_h) = E^0(t_h) + \epsilon_h, \quad \epsilon_h \sim N(0, \sigma_\epsilon) \quad (4)$$

Values for the four parameters d_o , t_f , r_d , and σ_ϵ are obtained independently for all sensors through maximum likelihood estimation (MLE). Once calibrated, the model is used to obtain the estimated mean and point-wise standard deviations for the sensor offset, $\mu_1(t) = E(E^0(t))$ and $\sigma_1(t)$, while using the estimates of t_f and σ_ϵ as fixed hyperparameter values.

Model 2 – integrated Brownian motion for a single sensor

In model 2, we assume instead that the recorded offset measurements are generated by an integrated Brownian motion. This is a continuous-time stochastic process, which captures the idea that the drift rate is subject to unmeasured disturbances:

$$\dot{r}_d(t) = \gamma(t) dt, \quad r_d(0) = r_{d,o}, \quad \gamma(t) \sim N(0, \sigma_\gamma) \quad (5)$$

$$\dot{E}^0(t) = r_d(t) dt, \quad E^0(0) = d_o \quad (6)$$

$$\tilde{E}^0(t_h) = E^0(t_h) + \epsilon_h, \quad \epsilon_h \sim N(0, \sigma_\epsilon) \quad (7)$$

This model also includes four parameters: the initial drift rate ($r_{d,o}$); the initial offset (d_o); an input noise standard deviation controlling the rate by which the drift rate changes (σ_γ); and an output noise standard deviation (σ_ϵ). As with model 1, parameter values are obtained through MLE. This is achieved by formulating the above process as a Gaussian process (Rasmussen & Williams 2006). This also enables computation of expected values and associated point-wise standard deviations, $\mu_2(t) = E(E^0(t))$ and $\sigma_2(t)$, with the estimates of σ_γ and σ_ϵ now used as fixed hyperparameter values.

Model 3 – integrated Brownian motion for multiple sensors

A third model is derived from Equations (5)–(7) by considering that two sensors of the same type may be characterized by distinct initial conditions ($r_{d,o}$, d_o) but the same noise parameters (σ_ϵ , σ_γ). This leads to one model with six parameters ($r_{d,o}^a$, d_o^a , $r_{d,o}^b$, d_o^b , σ_ϵ , σ_γ), describing the computed offsets for a pair of sensors, instead of two models with four parameters, each describing the computed offsets for one sensor only. Their values are again obtained via MLE and used to obtain calibrated predictions ($\mu_3(t) = E(E^0(t))$ and $\sigma_3(t)$), once again using the estimates of σ_γ and σ_ϵ as fixed hyperparameter values.

Model evaluation

The proposed models are evaluated through visual inspection of the measurements, predictions, and residuals between the measurements and predictions. In the present case, such a visual inspection is considered sufficient to select a suitable model.

Implementation

All data and code necessary to reproduce our results are added in the Supplementary Information (Section A) (available online). The complete data set can be found in Ohmura et al. (2019).

RESULTS

Sensor characterization tests: example

Figure 2 shows the results of the first sensor characterization test with sensor T1a (day 3). The raw potential measurement decreases during P0, increases to a steady value in P1,

decreases to a steady value in P2, increases to a steady value in P3, and decreases again in P4. The time intervals used for computation of E^{P1} , E^{P2} , and E^{P3} (in calibration medium, pH = 4, 7, and 4) are indicated by grey shading. One can see that the measured offset, E^{P2} , is slightly below 0 mV (-1.30 mV). The values for E^{P1} and E^{P3} are slightly lower than their ideal value (171.9 and 172.4 mV). The measured rise and decay sensitivities are therefore $S_D = 57.73$ and $S_R = 57.90$ mV per pH unit. The results of every sensor characterization test are available in the software package contained in the Supplementary Information (Section A).

Long-term trends in the offset measurements within the warranty period

Figure 3 displays the measured offsets in all sensors throughout the experimental period. The recorded values collected within the warranty period (1 year) range from approximately 0 mV (no offset) to roughly -70 mV. All commercially available sensors (T1-T4) produce a decaying trend in the offsets. To the authors' knowledge, this kind of drift is caused primarily by depletion of electrolyte in the cell containing the reference electrode. The firstly recorded offsets for the T1-T3 sensors are small in magnitude and concentrate around 0 mV. In contrast, the T4 sensors offset values indicate a shock effect producing a shift of -20 and -45 mV (T4a, T4b) within days from installation. According to the manufacturer, this is due to the high ammonium concentration in the medium and is typical for

this specific type of sensors. The accumulated drift in the T1 sensors is at most -25 mV after one year while the T2 and T3 sensors exhibit an offset of -75 mV after one year. Without calibration, this means the T1 sensors can produce a pH value as high as 7.4 when the true pH is 7. The T2 and T3 sensors will produce a pH value as high as 8.3 in the same circumstances. Due to failure of T5d, no offsets could be measured for this sensor. The remaining prototypes (T5a/b/c) do not produce a significant offset at any time, except for T5b which produces a dramatic shift in the offset during three sensor characterization tests executed prior to replacement. We speculate that the lack of drift is due to the use of a particularly robust reference electrolyte and reference junction in this sensor type. A detailed inspection of the T5b measurements revealed that the first symptoms of sensor degradation can be observed on day 332. This is however only obvious when comparing these measurements with the simultaneous T1b/T2b/T3b measurements (see the Supplementary Information, Section D, available with the online version of this paper). In all cases, except for the T4 and T5a/b pairs, the difference between offsets in sensors of the same type remains rather small with 1 year of installation, with a maximal difference of 16.7 mV recorded with the T2 sensors. Taking the 0.1 pH threshold discussed above as a guideline, one could propose to validate and calibrate the sensors when their potential measurements are 5.9 mV apart. This happens for the first time for the T1, T2, and T3 sensors on day 127, 79, and 309. By these times, the absolute offsets are already larger

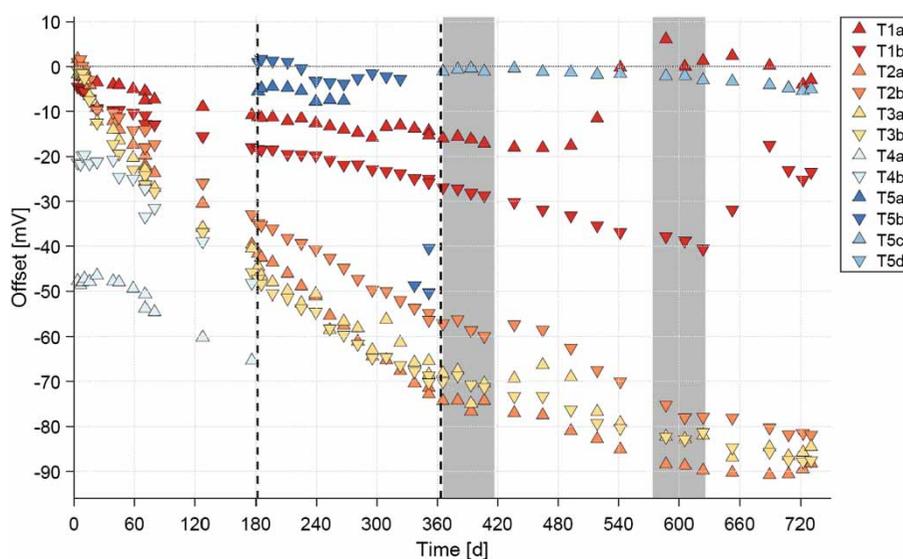


Figure 3 | Offset in all studied sensors as a function of time. Vertical lines indicate a change of installed sensors (see Figure 1). Grey bands indicate a change of reactor medium (see the section Long-term exposure experiment above). The commercially available sensors (T1-T4) exhibit drift from the start of installation while the prototypes (T5) exhibit close to no drift when otherwise functioning properly. A significant shock effect is observed for the T4 sensors at the start of the experiment but not for any other sensor.

than this accepted threshold. This means that the use of hardware redundancy when using sensors of the same type is unlikely to provide adequate drift detection performance.

Figure 4 shows offsets for the sensors T1a, T3a, and T3b collected in the first year of the experiment as a function of the difference in the offset between T1a and T3a (left panel) and between T3b and T3a (right panel). The left panel suggests that offset difference between sensors can be predictive of the offset in an individual sensor. The right panel shows that this is less likely to be successful for sensors of the same sensor type, as also described above. This is considered an important opportunity for further research, which we discuss further below.

Long-term trends in the offset measurements beyond the warranty period

The offset measurements obtained after the warranty period expired exhibit two phenomena that are surprising (Figure 3). The first phenomenon is the rise of the offset of the T1a sensor after 480 days of exposure and a similar rise of the offset of the T1b sensor after 630 days of exposure. Considering that this appears at distinct times in the lifetime of the T1 sensors, this cannot be explained as a direct effect of medium composition changes. Based on information provided by the sensor manufacturer, this type of drift rate sign reversal is unique for the T1 sensors and is unlikely for any other sensor type covered in this study. It is the opinion of the

authors that the time for this reversal is difficult to predict in advance. For this reason, this phenomenon is best handled as an unmeasured process disturbance.

The second phenomenon consists of the rather flat to increasing profile of the offset measurements in the T2 and T3 sensors between day 360 and day 480. Before and after this period, the drift rate in these sensors is visually similar. Given the synchronicity of this effect between 4 pH sensors, it is hypothesized that this change in the drift rate is influenced by the deliberate addition of nitrite in the form of NaNO_2 salt to the reactor contents from day 366 to 417. The nitrite addition affected the biomass concentration and the concentrations of all dominant nitrogen species (ammonia, nitrite, nitrate, see Supplementary Information, Section C) and may also have affected the ion strength and conductivity of the reactor medium. Due to this combination of effects, the available data only offer an incomplete understanding of the complete chain of causes and effects between the nitrite addition and the observed changes in the sensor drift rates. For this reason, the effects of changing media composition on the sensor drift rate is also considered an unmeasured process disturbance.

Long-term trends in the sensitivity measurements

Figure 5 displays the computed sensitivity measurements for the potential rise (S_R) during the complete experimental period. These measurements do not exhibit strong trends

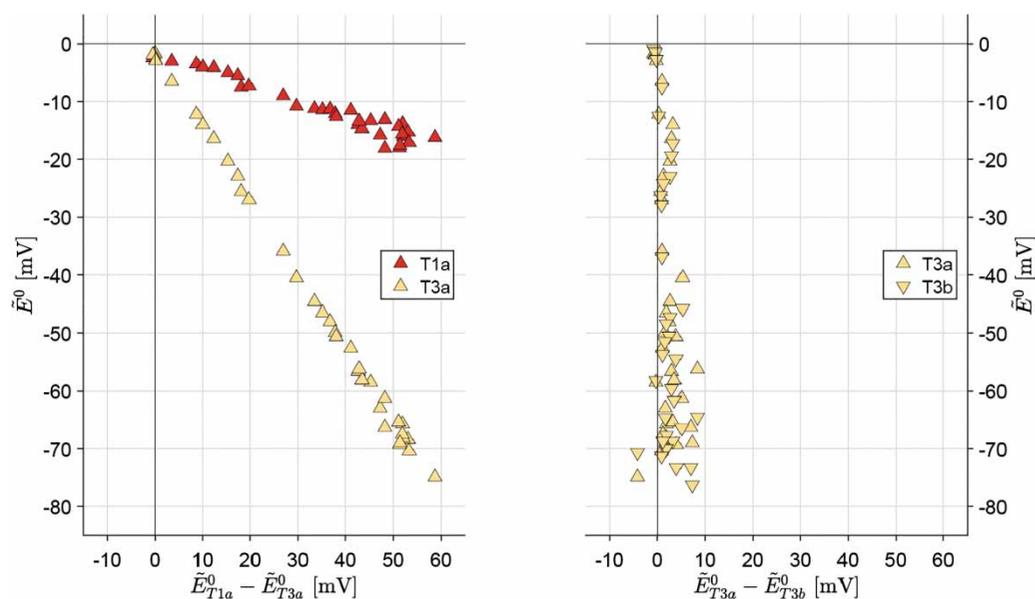


Figure 4 | Offset measurements as a function of relative deviations in the offset measurements. Left panel: Offsets of sensor T1a and T3a as a function of the difference of these offsets. These data are suggestive of a close to linear relationship between sensor offsets and the offset difference. Right panel: Offsets of sensors T3a and T3b relative to the difference of these offsets. The difference in offset remains small and there is no obvious relationship in this case.

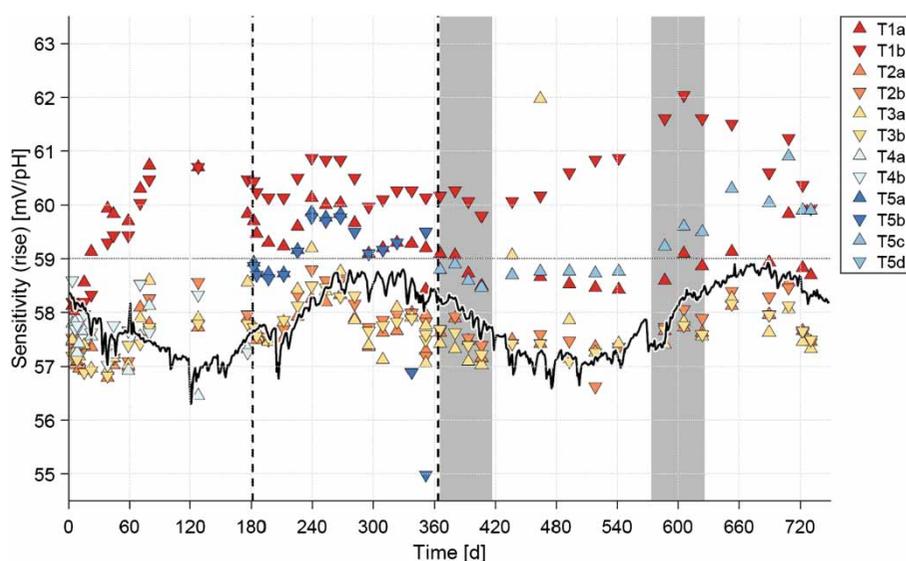


Figure 5 | Sensitivity measurements for the potential rise as a function of time. Vertical lines indicate a change of installed sensors (see Figure 1). Grey bands indicate a change of reactor medium (see the section Long-term exposure experiment above). A black line shows the theoretically expected sensitivity computed with Equation (2). Variations in the sensitivity are small and follow the theoretical sensitivity closely.

in any particular direction. The sensitivity measurements fall between 54.9 and 62.1 mV per pH unit. This means that one can expect to measure a pH value between 5.95 and 6.08 when (i) the true pH value is 6 and (ii) any offset is corrected for. This is likely acceptable for most wastewater applications. The same graph also shows the theoretical value of the sensitivity according to Equation (2) and the recorded temperature. This profile is very similar to the recorded sensitivity profiles and explains most of the variations in the sensitivity measurements, which are small anyway. The computed sensitivity measurements for the potential rise exhibit no meaningful differences with the computed sensitivity measurements for the potential decay (see Supplementary Information, Section E, available online).

Drift models

For practical intents and purposes, the sensitivity – when corrected for temperature variations – can be considered constant for the considered process and sensors. This is why only the offset measurements are used for modelling. While we discuss all modelling results next, we only display results obtained for the T2a and T2b sensors in view of space requirements. The remaining modelling results are shown in the Supplementary Information (Section F) (available online).

The left panel of Figure 6 shows the offset measurements for the T2a and T2b sensor together with the model predictions and their confidence bounds. The right panel of Figure 6 shows the prediction residuals. With Model 1, the

time of the drift onset (t_f) is always identified as a time before the first measurement was obtained (2.1 and 2.3 days), suggesting that drift occurs throughout the experiment. More importantly, however, is that Model 1 offers a rather poor description of the data. The confidence intervals are wide and the residuals are clearly auto-correlated. In contrast, Models 2 and 3 provide narrow confidence intervals and residuals that do not suggest the presence of autocorrelation. There are no clear differences in performance between these two models so that Model 3, which has fewer free parameters, is preferred. The same kind of model performances are obtained for sensor types T1 to T3. For the T4 sensors, all model types delivered the same, adequate performance. This may indicate that (a) the T4 sensors exhibit a drift which is influenced less by unmeasured disturbances and therefore occurs with a close to constant rate or (b) that the shortened exposure – 6 months in this case – was too short to capture the long-term effects of unmeasured disturbances.

DISCUSSION

This is the first publicly available study discussing long-term wear-and-tear on wastewater quality sensors in a systematic manner and at this scale (12 sensors). The results reveal that commonly held assumptions regarding sensor faults and fault symptoms are false. First, drift in pH sensors occurs simultaneously in all commercially available sensors

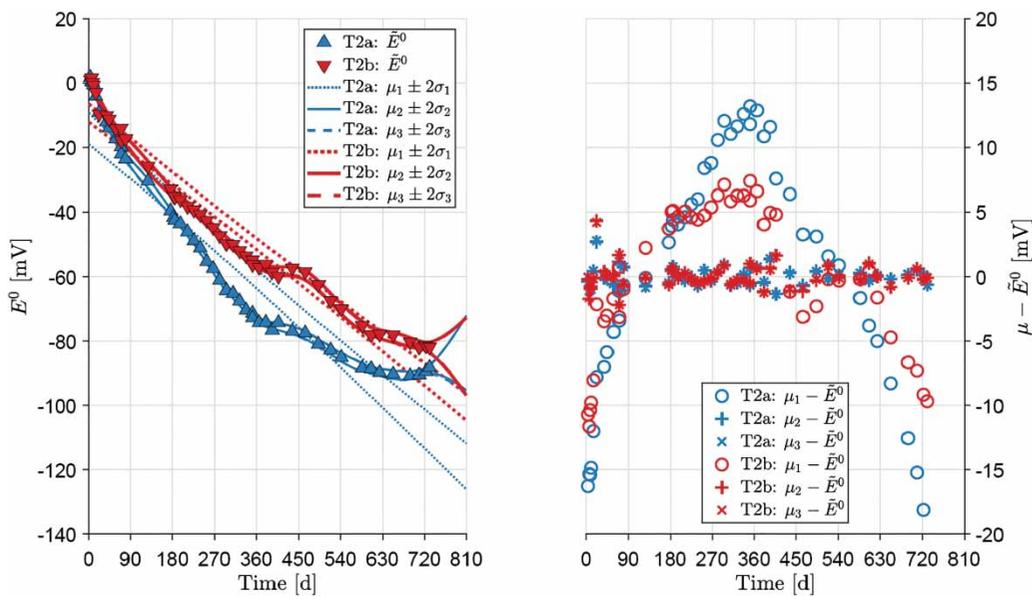


Figure 6 | Modeling results for the T2 sensors. Left panel: Offset confidence bounds ($\mu \pm 2\sigma$) obtained with models 1 (μ_1, σ_1), 2 (μ_2, σ_2), and 3 (μ_3, σ_3). Right panel: Residuals between expected values (μ) and measured potentials (E^0). Model 1 does not describe the data well, leading to larger confidence bounds and auto-correlated residuals. Models 2 and 3 fit the data well and their predictions are hard to distinguish from each other.

(T1-T4). Second, drift occurs as soon as a sensor is deployed in the measured medium. In some cases, drift is paired by a significant shift in the offset at the time of installation. Importantly, one can compute offsets and sensitivities as a function of time with data that is electronically available in modern measurement transformers.

These observations are consequential for the development of methods for FDI. For instance, it is our opinion that the development of FDI methods and model-based benchmarking (e.g. Downs & Vogel 1993; Barty et al. 2006) should be focused on methods that do not rely on wrongful assumptions. The Benchmark Simulation Model family could especially benefit from an expanded sensor model for water quality sensors, considering that the available drift model (Rosén et al. 2008; Gernaey et al. 2014) cannot adequately describe the naturally occurring drift in ion-selective electrodes. We believe the challenges associated with the FDI task can be simulated more accurately with the proposed stochastic sensor drift model, specifically an integrated Brownian process.

Fortunately, our results also reveal a number of opportunities for the use and maintenance of ion-selective measurements. First, the obtained stochastic models enable prediction of the expected offset measurement and associated confidence intervals beyond the last measurement. This means that such a model can be used for predictive sensor maintenance; for example, by planning sensor

validation and/or calibration before the predicted confidence interval for the reference potential (E^0) exceeds a predetermined tolerance. Exploring the utility of this idea is considered for future research. Second, the prototype sensors tested in this study exhibit a remarkably stable offset. While these sensors appear prone to failure, as one might expect from a prototype, this suggests that practically drift-free yet economical pH sensors will enter the market soon. Third, the recorded sensitivity measurements in all sensors hover around the ideal values and are remarkably stable throughout the experimental period. Such a stable sensitivity lends support for advanced monitoring and control strategies which are inherently robust to changes in the offset but still assume a rather stable sensitivity (Ruano et al. 2009, 2012; Villez & Habermacher 2016; Thürlimann et al. 2018, 2019; Schneider et al. 2019). Fourth, it was shown that the offset difference between two pH sensors in the same medium can be predictive of the offset of the individual pH sensors; however, only if two sufficiently distinct sensor types are selected. Combined with a stable sensitivity, this means that the deviation between two online pH sensor signals could be used as a proxy for the offsets in the individual sensors. Such a proxy measurement could be very useful for remote sensor quality assessment and predictive sensor maintenance, especially since one can compute such deviations between on-line sensor signals while the sensors remain in their normal measurement location.

CONCLUSIONS

Despite an abundance of literature on FDI methods, little information is actually known about the appearance of sensor faults and failures in wastewater quality sensors. This first long-term study of the ageing of 12 individual pH sensors gives valuable insights into this challenge:

- Commonly held assumptions in FDI method development and evaluation, such as the availability of fault-free historical data and fault independence, are invalid for commercially available pH sensors.
- Offset drift in redundant sensors is unlikely to be identified early if these sensors are of the exact same type and experience the same exposure.
- The observed sensor drift is different for every sensor, even when sensor types and exposure times are the same. This supports the idea that sensor drift is a stochastic process, rather than a deterministic one.

ACKNOWLEDGEMENTS

This research was made possible by the Swiss National Foundation (Project: 157097). Mr Ohmura's contributions are financially supported by Toshiba, Tokyo, Japan. Dr Carbajal's contributions are financially supported by Eawag Discretionary Funds (grant no. 5221.00492.011.05, project: DF2017/EMUmore). We thank Daniel Iten and Stefan Vogel from Endress + Hauser for their valuable advice during this study. The stochastic models were calibrated with the GPML toolbox v4.2 (Rasmussen & Nickisch 2010).

REFERENCES

- Alferes, J., Tik, S., Copp, J. & Vanrolleghem, P. A. 2013 *Advanced monitoring of water systems using in situ measurement stations: data validation and fault detection*. *Water Science and Technology* **68** (5), 1022–1030.
- Barty, M., Patton, R., Syfert, M., de las Heras, S. & Quevedo, J. 2006 *Introduction to the DAMADICS actuator FDI benchmark study*. *Control Engineering Practice* **14** (6), 577–596.
- Buck, R. P., Rondinini, S., Covington, A. K., Baucke, F. G. K., Brett, C. M., Camoes, M. F., Milton, M. J. T., Mussini, T., Naumann, R., Pratt, K. W. & Spitzer, P. 2002 *Measurement of pH. Definition, standards, and procedures (IUPAC recommendations 2002)*. *Pure and Applied Chemistry* **74** (11), 2169–2200.
- Carr, J. J. 1993 *Sensors and Circuits*. PTR Prentice Hall, Englewood Cliffs, NJ, USA.
- Corominas, L., Villez, K., Aguado, D., Rieger, L., Rosén, C. & Vanrolleghem, P. A. 2011 *Performance evaluation of fault detection methods for wastewater treatment processes*. *Biotechnology and Bioengineering* **108** (2), 333–344.
- Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortes, U. & Poch, M. 2018 *Transforming data into knowledge for improved wastewater treatment operation: a critical review of techniques*. *Environmental Modelling and Software* **106**, 89–103.
- Downs, J. J. & Vogel, E. F. 1993 *A plant-wide industrial process control problem*. *Computers and Chemical Engineering* **17** (3), 245–255.
- Gernaey, K. V., Jeppsson, U., Vanrolleghem, P. A. & Copp, J. B. 2014 *Benchmarking of Control Strategies for Wastewater Treatment Plants*. Scientific and Technical Report No. 23. IWA Publishing, London, UK.
- Haimi, H., Mulas, M., Corona, F. & Vahala, R. 2013 *Data-derived soft-sensors for biological wastewater treatment plants: an overview*. *Environmental Modelling and Software* **47**, 88–107.
- Le, Q. H., Verheijen, P. J., van Loosdrecht, M. C. & Volcke, E. I. 2018 *Experimental design for evaluating WWTP data by linear mass balances*. *Water Research* **142**, 415–425.
- Ohmura, K., Thuerlimann, C. M., Kipf, M., Carbajal, J. P. & Villez, K. 2019 *pH Sensor Ageing Experiment (SoDAN-dataset1)*, <https://zenodo.org/record/3229195>.
- Rasmussen, C. E. & Nickisch, H. 2010 *Gaussian Processes for Machine Learning (GPML) Toolbox*. *Journal of Machine Learning Research* **11**, 3011–3015.
- Rasmussen, C. E. & Williams, C. K. 2006 *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, USA.
- Rieger, L., Langergraber, G. & Siegrist, H. 2006 *Uncertainties of spectral in situ measurements in wastewater using different calibration approaches*. *Water Science and Technology* **53** (12), 187–197.
- Rieger, L., Takács, I., Villez, K., Siegrist, H., Lessard, P., Vanrolleghem, P. A. & Comeau, Y. 2010 *Data reconciliation for wastewater treatment plant simulation studies – planning for high-quality data and typical sources of errors*. *Water Environment Research* **82**, 426–433.
- Rieger, L., Thomann, M., Gujer, W. & Siegrist, H. 2005 *Quantifying the uncertainty of on-line sensors at WWTPs during field operation*. *Water Research* **39** (20), 5162–5174.
- Rosén, C., Rieger, L., Jeppsson, U. & Vanrolleghem, P. A. 2008 *Adding realism to simulated sensors and actuators*. *Water Science and Technology* **57** (3), 337–344.
- Ruano, M. V., Ribes, J., Seco, A. & Ferrer, J. 2009 *Low cost-sensors as a real alternative to on-line nitrogen analysers in continuous systems*. *Water Science and Technology* **60** (12), 3261–3268.
- Ruano, M. V., Ribes, J., Seco, A. & Ferrer, J. 2012 *An advanced control strategy for biological nutrient removal in continuous systems based on pH and ORP sensors*. *Chemical Engineering Journal* **183**, 212–221.
- Schneider, M. Y., Carbajal, J. P., Furrer, V., Maurer, M. & Villez, K. 2019 *Beyond signal quality: the value of unmaintained pH*,

- dissolved oxygen, and oxidation-reduction potential sensors for remote performance monitoring of on-site sequencing batch reactors. *Water Research* **161**, 639–651.
- Spindler, A. 2014 Structural redundancy of data from wastewater treatment systems. Determination of individual balance equations. *Water Research* **57**, 193–201.
- Spindler, A. & Vanrolleghem, P. A. 2012 Dynamic mass balancing for wastewater treatment data quality control using CUSUM charts. *Water Science and Technology* **65** (12), 2148–2153.
- Taylor, B. N., Mohr, P. J. & Douma, M. 2007 The NIST Reference on constants, units, and uncertainty.
- Thomann, M. 2008 Quality evaluation methods for wastewater treatment plant data. *Water Science and Technology* **10**, 1601–1609.
- Thomann, M., Rieger, L., Frommhold, S., Siegrist, H. & Gujer, W. 2002 An efficient monitoring concept with control charts for on-line sensors. *Water Science and Technology* **46** (4–5), 107–116.
- Thürlimann, C. M., Dürrenmatt, D. J. & Villez, K. 2018 Soft-sensing with qualitative trend analysis for control in full-scale wastewater treatment plants. *Control Engineering Practice* **70**, 121–133.
- Thürlimann, C. M., Udert, K. M., Morgenroth, E. & Villez, K. 2019 Stabilizing control of a urine nitrification process in the presence of sensor drift. *Water Research*. Accepted in press, <https://doi.org/10.1016/j.watres.2019.114958>.
- Venkatasubramanian, V., Rengaswamy, R. & Kavuri, S. N. 2003a A review of process fault detection and diagnosis – Part II: qualitative models and search strategies. *Computers and Chemical Engineering* **27** (3), 313–326.
- Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N. & Yin, K. 2003b A review of process fault detection and diagnosis – Part III: process history based methods. *Computers and Chemical Engineering* **27** (3), 327–346.
- Venkatasubramanian, V., Rengaswamy, R., Yin, K. & Kavuri, S. N. 2003c A review of process fault detection and diagnosis – Part I: quantitative model-based methods. *Computers and Chemical Engineering* **27** (3), 293–311.
- Villez, K. & Habermacher, J. 2016 Shape anomaly detection for process monitoring of a sequencing batch reactor. *Computers and Chemical Engineering* **91**, 365–379.
- Villez, K., Vanrolleghem, P. A. & Corominas, L. 2016 Optimal flow sensor placement on wastewater treatment plants. *Water Research* **101**, 75–83.
- Westcott, C. 2012 *pH Measurements*. Academic Press, New York, NY, USA.

First received 3 June 2019; accepted in revised form 23 August 2019. Available online 3 September 2019