ORIGINAL ARTICLE

Machine learning techniques for monitoring the sludge profile in a secondary settler tank

Jesús Zambrano[1] · Oscar Samuelsson[2,3] · Bengt Carlsson[3]

Abstract
The aim of this paper is to evaluate and compare the performance of two machine learning methods, Gaussian process regression (GPR) and Gaussian mixture models (GMMs), as two possible methods for monitoring the sludge profile in a secondary settler tank (SST). In GPR, the prediction of the response variable is given as a Gaussian probability density function, whereas in the GMM the probability density function is built as a weighted sum of Gaussian distributions. In both approaches, a residual is calculated and a fault detection criterion is implemented via a recursive decision rule. As case study, GMM and GPR were tested using real data from a sensor measuring the suspended solids concentration as a function of the SST level in a wastewater treatment plant in Bromma, Sweden. Results suggest that GMM gives a faster response but is also more sensitive than GPR to changes during normal conditions.

Keywords Covariance function · Fault detection · Gaussian mixture models · Gaussian process regression · Monitoring

List of symbols

h_ε	Threshold of decision rule
H_0	Normal condition
H_1	Abnormal condition
I	Identity matrix
k	Covariance function
k_{**}	Autocovariance of test data set
$\mathbf{k}_*$	Covariance vector between test and training data set
$\mathbf{K}_\theta$	Covariance matrix of training data set
K	Number of Gaussian distributions
M	Number of profiles used as training data set
m	mth profile
N	Number of observations
p	Probability density function
r	Residual
S	Silhouette value
t	Time
t_{FA}	Duration when detector incorrectly indicates a fault
t_{FD}	Duration when detector correctly indicates a fault
Δt_{FD}	Time delay to detect a fault
$\mathbf{w}$	Data set $(\mathbf{x}, \mathbf{y})$
x_i	Single input
$\mathbf{x}$	Vector of inputs
y_i	Single output
$\mathbf{y}$	Vector of outputs

Greek letters

α_ε	Threshold factor of decision rule
ε	Decision rule
μ	Mean value
π_k	Mixing weight in Gaussian mixture model
σ	Standard deviation
θ	Vector of hyperparameters

Subscripts

gm	Gaussian mixture
gp	Gaussian process
k	k-distribution
*	Predictive

Superscripts

T	Transpose

✉ Jesús Zambrano
jesus.zambrano@mdh.se

1 School of Business, Society and Engineering, Mälardalen University, P.O. Box 883, 72123 Västerås, Sweden

2 IVL Swedish Environmental Research Institute, P.O. Box 210 60, 10031 Stockholm, Sweden

3 Department of Information Technology, Uppsala University, P.O. Box 336, 75105 Uppsala, Sweden

Springer

Introduction

Increasing demands on effluent water quality and resource efficient operation are important driving forces for wastewater treatment plants (WWTPs). Process monitoring and detection of abnormal conditions are crucial tasks, since they help in keeping a robust and efficient performance of the plant (Olsson et al. 2014). Furthermore, increasing the amount of sensors adds process information but also increases the complexity for plant operators. Hence, the need for fault detection methods is a priority.

The secondary settler tank (SST) is a key part of a WWTP since it provides two functions: clarification and thickening. In the clarification, the aim is the removal of suspended solids (SS) so to get a clarified effluent that meets the effluent SS goal. In the thickening, the aim is to get the concentration of the settled solids to be returned to the bioreactor. The SST uses gravity to separate the sludge (solid) component from the treated water (liquid). Due to the clarification and thickening functions, typically a concentration profile (also called sludge profile) has a low value for the concentration close to the effluent, and this value increases towards the bottom of the SST. There is a level where the solid concentration abruptly changes, which is called sludge blanket level. This level together with the SS concentration in the bottom or in the return sludge is the typical values to monitor in a SST.

Some examples of methods applied to monitor a SST include image analysis (Grijspeerdt 1997) and model-based approaches (Traoré et al. 2006; Yoo et al. 2002). Even though, the prediction of the sludge concentration profile is still far from satisfactory (Li and Stenstrom 2014), which makes it problematic to perform a good monitoring of the SST, mathematical models have also been used for predicting the sludge profile; see, for example, 1D model proposed by Diehl et al. (2016) and 3D model proposed by Xanthos et al. (2010).

In the last two decades, the research field *Machine Learning* has gained especial attention since with machine learning it is possible to develop methods that can automatically detect patterns in data (i.e. *learning*) and then to use this information to predict future data (Murphy 2012). There are many different techniques in machine learning including decision trees, data clustering, neural networks, Gaussian process regression, Gaussian mixture models, to mention a few.

From these machine learning techniques, Gaussian process regression (GPR) and Gaussian mixture models (GMMs) are two techniques that have started to gain interest in different applications. GPR is a regression method where a prediction of the response variable is given as a Gaussian probability density function. Thus, the predicted value of the response variable comes with a variance estimate, which is interpreted as an uncertainty measure of the prediction (Rasmussen and Williams 2005). It is worth to note that GPR is not a new concept, and it was originally known as kriging, with an origin from geostatistics in the 1950s (Cressie 1990).

GPR has several properties making it useful for monitoring and fault detection, such as probabilistic prediction including an uncertainty estimate, flexible regression in a nonparametric fashion, and it is relatively simple to implement in common programming languages. GPR has been used for monitoring and fault detection in different applications (Roberts et al. 2012), such as maritime vessel track analysis (Smith et al. 2012), change point detection (Garnett et al. 2010) and process monitoring (Serradilla et al. 2011). GPR has also been used in environmental applications, such as monitoring and fault detection in water monitoring signals (Samuelsson et al. 2017), modelling of an anaerobic wastewater treatment system (Ni et al. 2012), modelling nitrification process and biomass growth (Ažman and Kocijan 2007) and control of a sequencing batch reactor (Kocijan and Hvala 2013).

GMM is an alternative machine learning method to GPR for data monitoring. GMM is a parametric probability model for density estimation using a mixture of Gaussian distributions (Bishop 2007). In this way, the GMM can describe a set of data using the combination of Gaussian distributions. Applications of GMM in data monitoring can also be found in the literature, for example in sensor monitoring (Zhu et al. 2014), fault detection and diagnosis (Jiang et al. 2016; Yu 2012). Some other applications include data classification (Bouveyron 2014), image segmentation (Greggio et al. 2011) and many others.

The objective of this study is to compare the performance of GPR and GMM as tools for monitoring and fault detection of sludge profiles in a secondary settler tank. The paper is organized as follows. First, a general introduction to GPR and GMM is presented. Then, the case study is detailed, and a fault detection criterion based on GPR and GMM is then formulated. Next, results and discussion are shown, and some conclusions are drawn.

Gaussian process regression

A Gaussian process (GP) is a collection of random variables which has a joint Gaussian distribution. Assume we observe some input x_i and some output y_i from a certain process and that $y_i = f(x_i)$. Then, the GP is completely specified by its mean $\mu(x_i)$ and covariance function $k(x_i, x_j)$, and a distribution over the function $f(x_i)$ can therefore be expressed as

$$f(x_i) \sim \text{GP}\big(\mu(x_i), k_\theta(x_i, x_j)\big). \tag{1}$$

Typically, the covariance functions involve a vector $\theta = [\theta_1, \dots \theta_z]$ of parameters (called hyperparameters). Given a data set $\mathbf{w} = (\mathbf{x}, \mathbf{y})$, the simplest approach to optimize these hyperparameters is to maximize the log-marginal likelihood $\log p(\mathbf{y} \mid \mathbf{x})$

$$\hat{\theta} = \arg\max_\theta \Big(-\frac{1}{2}\mathbf{y}^{\text{T}}(\mathbf{K}_\theta + \sigma^2 I)^{-1}\mathbf{y} - \frac{1}{2}\log(\mid \mathbf{K}_\theta + \sigma^2 I \mid) - \frac{N}{2}\log(2\pi) \Big), \tag{2}$$

where $\mathbf{x} = \left[x_1, \dots, x_N\right]^{\text{T}}$ and $\mathbf{y} = \left[y_1, \dots, y_N\right]^{\text{T}}$ are vectors with N observed data with Gaussian noise of variance σ^2, $\mathbf{K}_\theta = k_\theta(\mathbf{x}, \mathbf{x})$ is a $N \times N$ covariance matrix of the training data set and I is a $N \times N$ identity matrix [see more details in Rasmussen and Williams (2005)].

A regression in a GP means that, based on the data set $\mathbf{w}$, and a new input x_*, we want to find the predictive distribution of the associated output y_*. The predictive distribution of y_* over $\mathbf{w}$ is Gaussian described by

$$p(y_* \mid (\mathbf{x}, \mathbf{y}), x_*) = \mathcal{N}(\mu_*(x_*), \sigma_*^2(x_*)), \tag{3}$$

with mean $\mu_*(x_*)$ and covariance $\sigma_*^2(x_*)$ given by (Rasmussen and Williams 2005)

$$\mu_*(x_*) = \mathbf{k}_*^{\text{T}}(\mathbf{K}_\theta + \sigma^2 I)^{-1}\mathbf{y}, \tag{4}$$

$$\sigma_*^2(x_*) = k_{**} - \mathbf{k}_*^{\text{T}}(\mathbf{K}_\theta + \sigma^2 I)^{-1}\mathbf{k}_*, \tag{5}$$

where $\mathbf{k}_* = \left[k(x_1, x_*), \dots, k(x_N, x_*)\right]^{\text{T}}$ is a $N \times 1$ vector of covariance between the test and the training data set and $k_{**} = k(x_*, x_*)$ is the autocovariance of the test data set.

To obtain the predictive mean and covariance, an appropriate covariance function has to be selected by the user. The covariance function determines the model structure of the GPR; therefore, a prior knowledge of the process is useful to determine a good candidate for this function. For example, a typical choice is the squared-exponential covariance function, which has the form $k_\theta(x_i, x_j) = \theta_1 \exp\left[-\theta_2(x_i - x_j)^2\right]$.

Gaussian mixture models

Assume again the data set $\mathbf{w}$ from a certain process, formed by N independent observations. One way to model these data is by mixture of models, where the aim is to represent certain subpopulations from the whole data set by means of a certain conditional probability density (binomial, exponential, etc.) (Južnič-Zonta et al. 2012). In the case of Gaussian mixture models, the distribution of the observation $\mathbf{w}_n$ is modelled as a sum (or mixture) of several Gaussian distributions (Murphy 2012)

$$p(\mathbf{w}_n) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{w}_n \mid \mu_k, \sigma_k^2), \tag{6}$$

where μ_k represents the mean and σ_k^2 represents the covariance matrix of the k-distribution. Hence, expression (6) is a combination of K Gaussian distributions, where each of them has a mixing weight π_k. These mixing weights must satisfy $0 \le \pi_k \le 1$ and $\sum_{k=1}^{K} \pi_k = 1$. The resulting function $p(\mathbf{w}_n)$ is a probability density function from observing the data $\mathbf{w}_n$.

Once the value K is specified, the GMM parameters π_k, μ_k and σ_k can be inferred by using the iterative expectation–maximization (EM) algorithm applied to Gaussian mixtures (Murphy 2012). EM is a technique used to find maximum likelihood solutions for probabilistic models containing variables that are not directly observed but can be inferred (Bishop 2007). EM applied to GMM is summarized in Algorithm 1.

Algorithm 1 EM for Gaussian mixtures

1: Collect a data set $\mathbf{w}$ with N observations
2: Initialize $\mu_k^1, \pi_k^1, \sigma_k^1$, set $i = 1$.
3: **while** not convergence of μ_k, π_k, σ_k **do**
4:      Expectation step: Compute

$$\gamma(z_{nk}) = \frac{\pi_k^i \mathcal{N}(\mathbf{w}_n | \mu_k^i, \Sigma_k^i)}{\sum_{j=1}^{K} \pi_j^i \mathcal{N}(\mathbf{w}_n | \mu_j^i, \Sigma_j^i)}, \tag{7}$$

for $n = 1, ..., N, \ k = 1, ..., K$.

5:      Maximization step: Compute

$$\mu_k^{i+1} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{w}_n, \tag{8}$$

$$\pi_k^{i+1} = \frac{N_k}{N}, \tag{9}$$

$$\Sigma_k^{i+1} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \Omega \Omega^T. \tag{10}$$

6:      $i \leftarrow i + 1$
7: **end while**
where $\Sigma_k^i = (\sigma_k^2)^i, \ N_k = \sum_{n=1}^{N} \gamma(z_{nk}), \ \Omega = \mathbf{w}_n - \mu_k^{i+1}$.

One way to assign a value for the number for Gaussian distributions K is using the silhouette criterion; see details in Rousseeuw (1987). This criterion calculates a silhouette value S which indicates how similar samples are in one cluster to samples in another cluster. S ranges from -1 (data misclassified) to $+1$ (data well clustered), whereas S close to zero means that the clusters are indistinguishable.

Case study: monitoring a SST in a WWTP

The present approach was evaluated using real data from a sensor installed in a SST at Bromma WWTP in Stockholm, Sweden. The sensor goes from the top to the bottom of the settler, passing through the clarification and the thickening zone. In this way, the sensor measures the level (m) and the SS concentration (g/L), as shown in Fig. 1a. The profile obtained is called *sludge profile*. A typical sludge profile is shown in Fig. 1b.

Fig. 1 **a** Experiment setup; **b** typical sludge profile plotted as level versus SS concentration
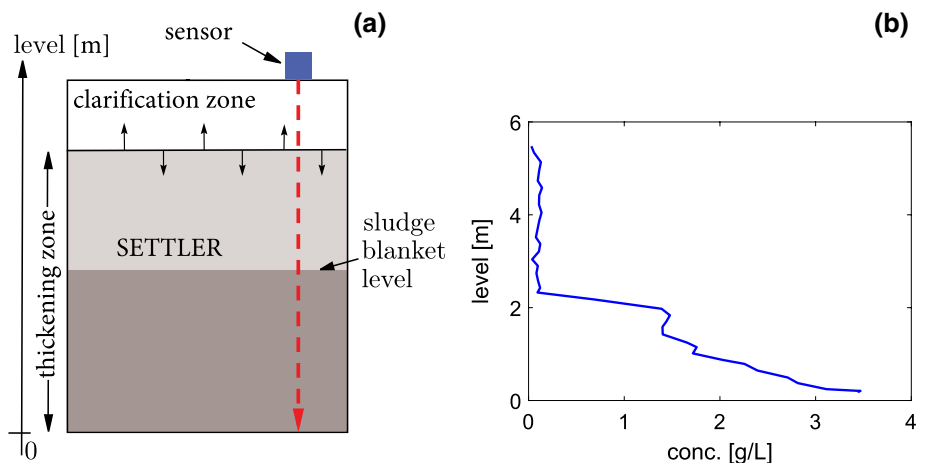
The sensor works periodically, and it means that a new sludge profile is automatically measured after a certain period of time. The sludge profile can be affected by different factors including: changes in the return and/or excess of sludge flow rates, sludge scape, large variations in the influent flow and composition, and sensor clogging or malfunctioning.

As part of the experiment, two additional measurements were recorded: the level at which the SS concentration is equal to 0.5 g/L (here referred as *fluff level*) and equal to 2.5 g/L (here referred as *sludge level*). We will refer to these levels during the results and discussion.

Fault detection criteria

Decision rule

The implementation of a fault detection (FD) method for a sensor signal y_m involves detecting any significant change in the dynamic of y_m when the sensor is subject to possible clogging or malfunctioning situation. A parameter v related to the dynamic of y_m is assumed to belong to one out of two conditions

$$
\begin{aligned}
H_0 &: v = v_0 \quad \text{(normal condition)}, \\
H_1 &: v = v_1 \quad \text{(abnormal condition)}.
\end{aligned}
\tag{11}
$$

To decide between H_0 and H_1, two FD methods are proposed. These methods are based on GPR and GMM (see the next subsections), which use the following recursive decision rule (Basseville and Nikiforov 1993)

$$
\varepsilon_m = \begin{cases} \varepsilon_{m-1} + s_m & \text{if } \varepsilon_{m-1} + s_m > 0 \\ 0 & \text{otherwise} \end{cases},
\tag{12}
$$

where $\varepsilon_0 = 0$ is set as the initial value and $s_m = \ln\left(p_{v_1}(y_m)/p_{v_0}(y_m)\right)$ is the log-likelihood ratio between the probability density function $p_{v_1}(y_m)$ and $p_{v_0}(y_m)$, in which m indicates the profile number at which ε_m is calculated.

A common way to detect possible changes in y_m is by analysing its *residual* r_m, which is related to the distance between data in normal and possible faulty conditions. Therefore, in normal condition r_m is equal or close to zero. Otherwise, it will increase and will belong to a possible faulty condition. Hence, the detection of a change in the residual can be done by detecting a change in the mean value of the sequence r_m, i.e. in this work, the parameter v referred in (11) corresponds to the mean value of the residual calculation of y_m.

Assume that r_m follows a Gaussian distribution $\mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$, where $\tilde{\mu}$ and $\tilde{\sigma}$ are the mean and standard deviation of r_m, respectively. The log-likelihood ratio test (Basseville and Nikiforov 1993) for a change in $\tilde{\mu}$ is expressed by

$$
s_m = \frac{(\tilde{\mu}_1 - \tilde{\mu}_0)}{\tilde{\sigma}^2}\left[r_m - \frac{(\tilde{\mu}_0 + \tilde{\mu}_1)}{2}\right],
\tag{13}
$$

where $\tilde{\mu}_0$ is the mean value in normal condition and $\tilde{\mu}_1$ is the possible change we want to detect. $\tilde{\mu}_1$ is calculated by collecting data in a moving window. Expression (13) can be used in (12) to calculate ε_m recursively.

A fault is decided if $\varepsilon_m > h_\varepsilon$, where $h_\varepsilon = \alpha_\varepsilon \max\left(\varepsilon_m\right)\big|_{1 \le m \le T_0}$ is a threshold value which is calculated taking the maximum value of ε_m obtained in normal condition for a certain predefined time T_0, and α_ε is a threshold factor.

The next subsections show how to calculate the residual r_m for GPR and GMM, denoted as r_{gp} and r_{gm}, respectively.

Residual calculation using GPR

Algorithm 2 is proposed to compute the residual value r_{gp} using GPR.


Algorithm 2 GPR-based residual calculation

1: Collect M profiles in normal conditions, each of them with N observations, set $m = 1$.
2: Select a covariance function and optimize the hyperparameters by maximizing Equation (2).
3: Obtain $p(y_* \mid (\mathbf{x}, \mathbf{y}), x_*)$ as described by Equations (3)-(5).
4: **while** monitoring a new profile $(\mathbf{x}_m, \mathbf{y}_m)$ **do**
5:     Compute

$$
r_{gp}(m) = \frac{1}{N}\sum_{i=1}^{N} |y_{i,m} - \mu_*(x_{i,m})|,
\tag{14}
$$

    where $\mathbf{x}_m = [x_{1,m}, ..., x_{i,m}, ..., x_{N,m}]$ and $\mathbf{y}_m = [y_{1,m}, ..., y_{i,m}, ..., y_{N,m}]$
6:     $m \leftarrow m + 1$
7: **end while**

Fig. 2 **a** Sludge profiles used as training data set; **b** sludge profiles in **a** plotted using dots; **c** predictive distribution over the training data set, showing μ_* (red line) and $\pm 2\sigma_*$ (grey zone); **d** contour of the GMM probability density function, with values indicated with colour scale
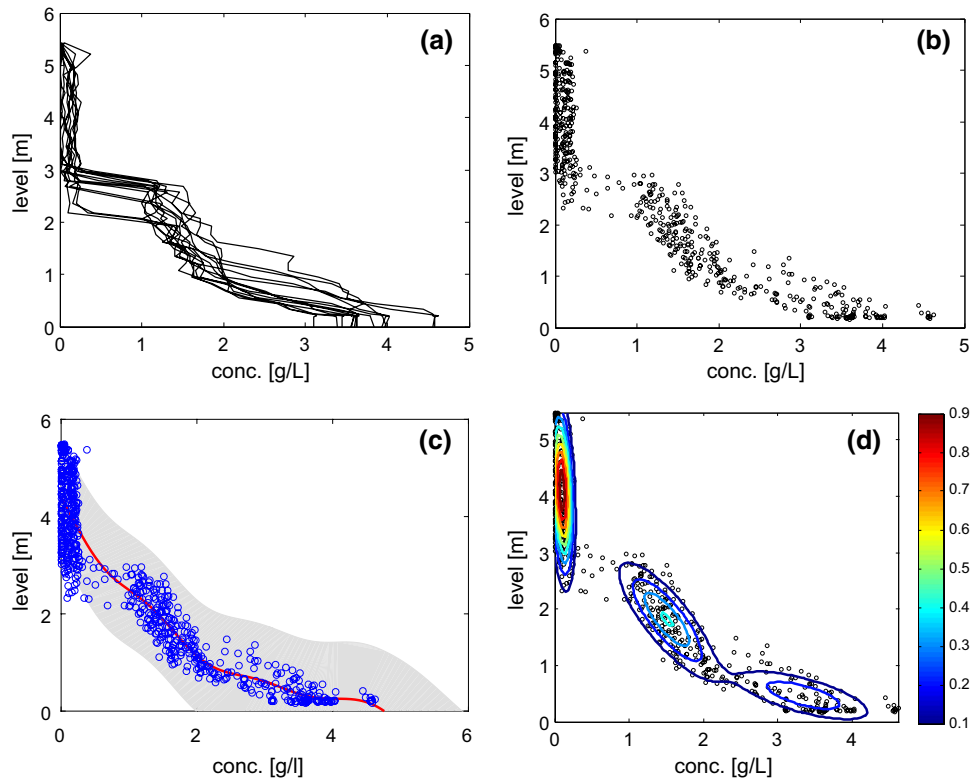
Note from (14) that the residual is calculated using the distance between the data and the predictive mean μ_*. Then, the more the data in the new profile that are outside the predictive distribution, the larger the residual r_{gp}.

Regarding the covariance function, the selection depends on the case study. A common choice is to use a squared-exponential function. However, if a particular dynamic of the data needs to be captured, a combination of different covariance function should be implemented, such as constant, linear and sinusoidal. See studies by Wilson and Adams (2013), Lloyd et al. (2014) and Samuelsson et al. (2017) for some examples showing the choice of covariance functions to different data sets. In this case study, the sum of a linear and a squared-exponential function was found feasible, that is,

$$k_\theta(x_i, x_j) = \theta_1 + \theta_2 x_i + \theta_3 \exp\left[-\theta_4\left(x_i - x_j\right)^2\right], \qquad (15)$$

where $(\theta_1, \theta_2, \theta_3, \theta_4)$ are hyperparameters.

Residual calculation using GMM

Algorithm 3 is proposed for the residual calculation r_{gm} using GMM.

Algorithm 3 GMM-based residual calculation

1: Collect M profiles in normal conditions, each of them with N observations, set $m = 1$.
2: Set K and compute Algorithm 1 to obtain (μ_k, π_k, σ_k), where $k = 1, ..., K$.
3: **while** monitoring a new profile **do**
4:    Compute

$$r_{\mathrm{gm}}(m) = \left(\sum_{n=1}^{N}\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{w}_n | \mu_k, \sigma_k)\right)^{-1}, \qquad (16)$$

5:    $m \leftarrow m + 1$
6: **end while**

Fig. 3 Sludge profiles during the SST monitoring. **a** after 5 days; **b** after 10 days; **c** after 15 days; **d** after 20 days. First days coloured in dark blue, last days in dark red
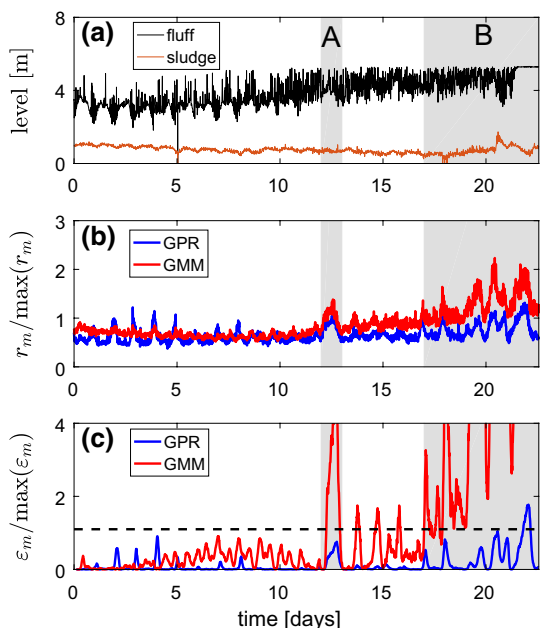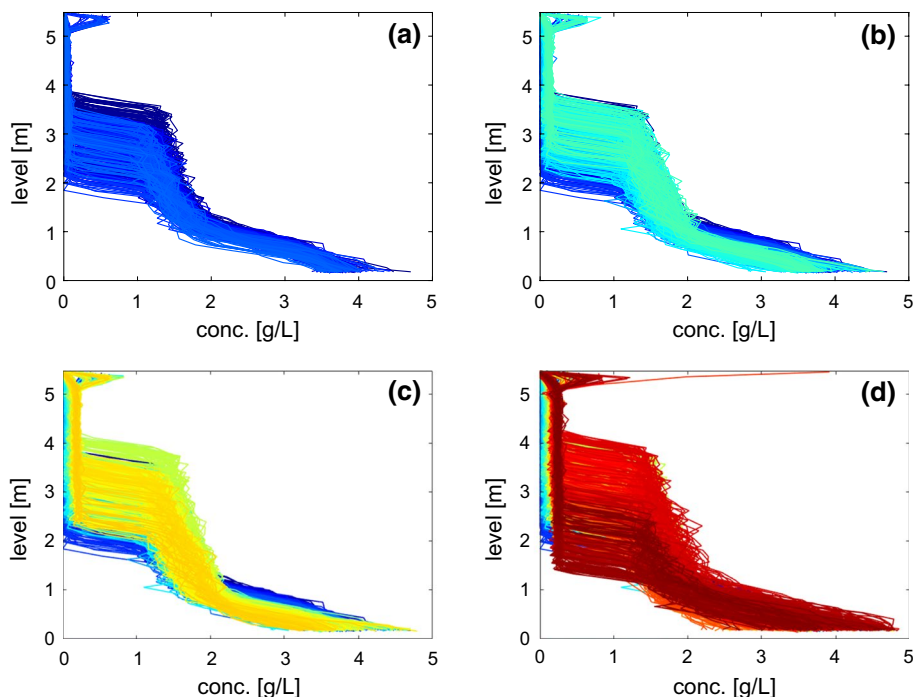

Fig. 4 **a** Fluff level (black line) and sludge level (brown line); **b** normalized residual response for GPR (blue line) and GMM (red line); **c** normalized decision rule response for GPR (blue line) and GMM (red line), threshold for $\alpha_\epsilon = 1.1$ (black dashed line). Periods A and B are shown as grey zones

the data in normal condition, the decrease in the probability density function and the increase in r_{gm}.

Software

The software MATLAB was used for all the calculations and simulations. For the GPR implementation, the GPML toolbox (Rasmussen and Nickisch 2010) was used, whereas the GMM implementation was done with the function `gmdistribution`.

Results and discussion

This section shows the comparison between the GPR and GMM performances applied to the case study. We decided to choose x representing the level of the sensor and $y = f(x)$ representing the SS concentration; hence, $\mathbf{w} = (x, y)$.

The training data set

A total of $M = 15$ sludge profiles in normal conditions were used as training data set for Algorithm 2 and 3; see Fig. 2a. Figure 2b shows the sludge profiles in normal conditions using dots.

The recursive decision rule [cf. (12)–(13)] was implemented with a moving window of 10 profiles, with $T_0 = 4$ d to compute the threshold value h_ϵ, and $\alpha_\epsilon = 1.1$ as threshold factor.

Note from expressions (16) that r_{gm} is the inverse of the summation of the probability density function of the entire data set $\mathbf{w}$. See that the farther the new profile data are from

Fig. 5 Group of sludge profiles for periods indicated in Fig. 4a. For GPR: **a** Period A, **b** Period B. The plots include the predictive mean of the training data set (black line). For GMM: **c** Period A, **d** Period B. The plots include the contours of the probability density function of the training data set
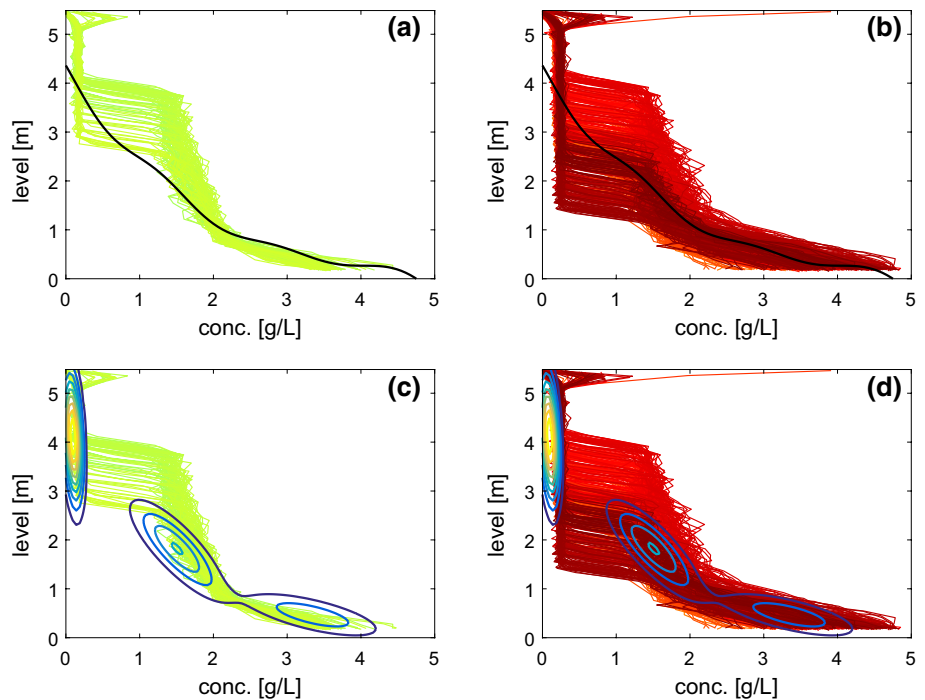

Table 1 Summary of GMM and GPR performance for different α_ε

α_ε	Method	$\Delta t_{FD}[d]/t_{FD}[d]$		FA[#]/$t_{FA}[d]$
		Period A	Period B	Rest of time
1.1	GMM	0.25/0.65	0.01/5.22	3/0.51
	GPR	–/–	4.81/0.35	0/–
1.25	GMM	0.26/0.62	0.02/4.74	3/0.21
	GPR	–/–	4.91/0.22	0/–
1.5	GMM	0.29/0.58	0.04/4.04	0/–
	GPR	–/–	–/–	0/–
1.75	GMM	0.41/0.40	0.06/3.66	0/–
	GPR	–/–	–/–	0/–
2.0	GMM	0.5/0.25	0.96/3.34	0/–
	GPR	–/–	–/–	0/–

The optimized GPR hyperparameters (cf. (15)) obtained were: $\hat\theta_1 = 3.88, \hat\theta_2 = -0.95, \hat\theta_3 = 0.35, \hat\theta_4 = 1.13$. Figure 2c shows the predictive mean value (red line) along with $\pm 2\sigma_*$ (the predictive distribution of the standard deviation).

For GMM, the highest silhouette value was $S = 0.77$ with $K = 3$, indicating that three is the optimal number of clusters, as shown by the three Gaussian distributions in Fig. 2d. Following the notation presented in expression (6), the values for π_k, μ_k and σ_k obtained for the GMM are expressed in the following probability density function

$$p(\mathbf{w}) = 0.43\,\mathcal{N}\left(\mathbf{w}\middle|\begin{pmatrix}4.11\\0.10\end{pmatrix},\begin{pmatrix}0.71 & -0.02\\-0.02 & 0.01\end{pmatrix}\right)$$
$$+\ 0.34\,\mathcal{N}\left(\mathbf{w}\middle|\begin{pmatrix}1.82\\1.51\end{pmatrix},\begin{pmatrix}0.36 & -0.18\\-0.18 & 0.15\end{pmatrix}\right)$$
$$+\ 0.23\,\mathcal{N}\left(\mathbf{w}\middle|\begin{pmatrix}0.47\\3.34\end{pmatrix},\begin{pmatrix}0.09 & -0.12\\-0.12 & 0.36\end{pmatrix}\right).$$

See in Fig. 2a–b that a typical sludge profile shows an abrupt change in the SS concentration from values close to zero to values larger than 1g/L, and then, this concentration keeps increasing as the sensor approaches the bottom of the SST. This change in the concentration was captured by the GPR and GMM. For GPR, the mean predictive decreases from top to bottom of the SST, passing through the data set, with the predictive standard deviation covering almost all the points. In the case of GMM, it classifies data before and after the jump as two separate Gaussian distributions. Note also that data for high concentration and low SST level were classified with another Gaussian distribution.

Monitoring of the SST

Several trials were done to monitor the settler. As illustration, we show one trial which lasted for 23 days, where a new sludge profile was collected every 15 min, giving a total of 2208 profiles. The evolution of the profiles during time is shown in Fig. 3a–d after 5, 10, 15 and 20 days of the experiment, respectively.

Figure 4 shows the profile of the fluff and sludge level, and the profile of the residuals (r_m) and the decision rule (ε_m) calculated via GPR and GMM approach.

There are two periods to highlight in Fig. 4, marked as Periods A and B. The sludge profiles of these periods are shown in Fig. 5, which also includes the GPR predictive mean and the GMM probability density function. Period A corresponds to variations in the residual profiles observed in days 12–13. The profiles of this period show that the concentrations between 1 and 2 g/L are in a higher level with respect to the level in the training data set (see Fig. 5a and c). This gave a certain variation in the fluff level as seen in Fig. 4a. This behaviour was mainly captured by the GMM approach (see Fig. 4c).

Period B refers to an event related to sensor clogging occurring after day 17. This event was confirmed by an in situ ocular inspection of the sensor and the presence of floating sludge at the surface level of the settler, causing sludge scape in the effluent. The profiles of this period are shown in Fig. 5b and d, where concentrations in the range of 0–2 g/L are far from the GMM probability density region and far from the GPR predictive mean. As seen in Fig. 4c, this event was first captured by the GMM which shown a persistent increasing in the decision rule. This event was also captured by the GPR after day 20.

The sludge level profile (see Fig. 4a) did not show any significant change during Period A; however, it showed a change late in Period B after day 20. Compared to the fluff level (level at 0.5 g/L), the sludge level (level at 2.5 g/L) did not fluctuate from its initial position, as it can be observed from the sludge profiles in Fig. 3, where most of the level fluctuations in the sludge profiles were in the range of 0–2 g/L.

When comparing the performances of GPR and GMM, the GMM gives a more fluctuate dynamics during the experiment. This can be seen in the behaviour of the residuals in Fig. 4b. Since a Gaussian function decreases faster than a linear function, then data far from maximum values of the GMM probability density function will give a higher residual than data far from the GPR predictive mean (cf. expressions (14) and (16)).

See that when using $\alpha_\varepsilon = 1.1$, GMM produces higher values than GPR during normal conditions (i.e. false alarms); see the responses between days 13 to 17. However, this behaviour will depend on the α_ε used for the evaluation of the decision rule. A higher α_ε might reduce the events of false alarms. To see this more in detail, the performance of GMM and GPR was also evaluated for different values of α_ε using the following indicators:

- Delay of fault detection (Δt_{FD}): time spent by ε_m to reach the threshold.

- Time in fault detection (t_{FD}): duration when the detector correctly indicates a fault, i.e. time spent by ε_m above threshold in fault detection condition.
- Number of false alarms (FA).
- Time in false alarm (t_{FA}): duration when the detector incorrectly indicates a fault, i.e. time spent by ε_m above threshold in false alarm condition.

The evaluation was done for Periods A, B and the rest of the experiment time. The results are summarized in Table 1.

See that the table includes the case shown in Fig. 4 when $\alpha_\varepsilon = 1.1$. As expected, when α_ε increases, the time delay to detect a fault also increases, whereas the time the decision rule expends in faulty condition decreases. See also that when α_ε increases, the time in false alarm decreases. When $\alpha_\varepsilon = 1.5$, the GMM performance is superior to GPR, i.e. it gave fault detections in both periods with relatively short time delay and with no false alarms. A threshold factor above 1.5 will also avoid false alarms but will also increase the fault detection time of GMM.

Concluding remarks

An important aspect in the GMM method is the definition of the amount of Gaussian distributions that describe the data set. In a given data set, the parameters involved in the GMM should be determined together with a value that indicates how well clustered is the data set, in our case by using the silhouette criterion. This criterion should be evaluated for different numbers of clusters, in order to find the optimum data clustering.

Regarding the GPR method, a key aspect for determining the predictive mean is the selection of the covariance function. In our case study, the covariance function (cf. expression (15)) was formed by two functions: a linear and a squared-exponential function. A linear function was required to capture the shape of typical sludge profiles. The squared-exponential function can be seen as a smoothing function. For any other process profiles, a new definition of the covariance function should be made.

Note that the data in the sludge profiles include outliers, defined as sharp changes between two successive data points. These outliers mean that measure concentrations are far from the values shown by profiles in normal condition. If there are few outliers in a profile, a possible task in the fault detection is to perform data correction, i.e. to replace outliers using data from the training data set. In this work, the correction of outliers was not part of the study.

Another possible situation in data monitoring is missing data. This situation did not happen in our case study, but it is common in other process monitoring applications. As

discussed for the case of dealing with outliers, a missing data can be reconstructed by using data from the training data set. See Samuelsson et al. (2017) where the case of missing data is evaluated for some GPR-based approaches.

It is important to recall that two sensors measuring the same variable in the same reactor will give two non-identical data sets. It means that each sensor will have a unique predictive mean and standard deviation for the case of GPR, as well as a unique probability density function for the case of GMM. It follows that the present methodology has the advantage that it can be applied to sensors in diverse areas.

One of the several applications of the current approach could be to use the decision rule of the FD algorithms as a tool for control actions. Therefore, new control strategies could include this variable as useful information to perform further tasks, for example, changing the recycle flow rate of the WWTP in order to keep the sludge profiles in a predefined level, or to give an early alarm that the SS sensor of the settler tank might need a supervision or a cleaning action.

Conclusions

This study tested two machine learning techniques, GPR and GMM, for monitoring the sludge profiles (level vs. suspended solid concentration) of a secondary settler tank in a wastewater treatment plant. The main idea was to train these two methods by using a set of sludge profiles in normal conditions and then perform the test by monitoring new sludge profiles.

The results show that GMM gave a fast fault detection than GPR, but GMM also shows to be more sensitive to false alarms. Nevertheless, it was possible to avoid the false alarm condition with a proper setting of the threshold factor.

Both methods have shown to be potential tools for monitoring sludge profiles. They could be applied for getting useful information about the performance of processes with repetitive profile data and to detect possible abnormal conditions.

Compliance with ethical standards

References

Ažman K, Kocijan J (2007) Application of Gaussian processes for black-box modelling of biosystems. ISA Trans 46(4):443–457

Basseville M, Nikiforov IV (1993) Detection of abrupt changes: theory and application. Prentice-Hall Inc, Upper Saddle River

Bishop C (2007) Pattern recognition and machine learning (information science and statistics). Springer, Berlin

Bouveyron C (2014) Adaptive mixture discriminant analysis for supervised learning with unobserved classes. J Classif 31(1):49–84

Cressie N (1990) The origins of kriging. Math Geol 22(3):239–252

Diehl S, Zambrano J, Carlsson B (2016) Steady-state analysis of activated sludge processes with a settler model including sludge compression. Water Res 88:104–116

Garnett R, Osborne MA, Reece S, Rogers A, Roberts SJ (2010) Sequential Bayesian prediction in the presence of changepoints and faults. Comput J 53(9):1430–1446

Greggio N, Bernardino A, Laschi C, Dario P, Santos-Victor J (2011) Fast estimation of Gaussian mixture models for image segmentation. Mach Vis Appl 23(4):773–789

Grijspeerdt K (1997) Image analysis to estimate the settleability and concentration of activated sludge. Water Res 31(5):1126–1134

Jiang Q, Huang B, Yan X (2016) GMM and optimal principal components-based Bayesian method for multimode fault diagnosis. Comput Chem Eng 84:338–349

Južnič-Zonta Ž, Kocijan J, Flotats X, Vrečko D (2012) Multi-criteria analyses of wastewater treatment bio-processes under an uncertainty and a multiplicity of steady states. Water Res 46(18):6121–6131

Kocijan J, Hvala N (2013) Sequencing batch-reactor control using Gaussian-process models. Bioresour Technol 137:340–348

Li B, Stenstrom M (2014) Research advances and challenges in one-dimensional modeling of secondary settling tanks—a critical review. Water Res 65:40–63

Lloyd JR, Duvenaud D, Grosse R, Tenenbaum JB, Ghahramani Z (2014) Automatic construction and natural-language description of nonparametric regression models. In: Association for the advancement of artificial intelligence (AAAI)

Murphy KP (2012) Machine learning: a probabilistic perspective (adaptive computation and machine learning series). The MIT Press, Cambridge

Ni W, Wang K, Chen T, Ng WJ, Tan SK (2012) GPR model with signal preprocessing and bias update for dynamic processes modeling. Control Eng Pract 20(12):1281–1292

Olsson G, Carlsson B, Comas J, Copp J, Gernaey KV, Ingildsen P, Jeppsson U, Kim C, Rieger L, Rodríguez-Roda I, Steyer J-P, Takács I, Vanrolleghem PA, Vargas A, Yuan Z, Åmand L (2014) Instrumentation, control and automation in wastewater—from London 1973 to Narbonne 2013. Water Sci Technol 69(7):1373

Rasmussen CE, Nickisch H (2010) Gaussian processes for machine learning (gpml) toolbox. J Mach Learn Res 11:3011–3015

Rasmussen CE, Williams CKI (2005) Gaussian processes for machine learning (adaptive computation and machine learning). The MIT Press, Cambridge

Roberts S, Osborne M, Ebden M, Reece S, Gibson N, Aigrain S (2012) Gaussian processes for time-series modelling. Philos Trans R Soc A Math Phys Eng Sci 371(1984):20110550

Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65

Samuelsson O, Björk A, Zambrano J, Carlsson B (2017) Gaussian process regression for monitoring and fault detection of wastewater treatment processes. Water Sci Technol 75(12):2952–2963

Serradilla J, Shi J, Morris A (2011) Fault detection based on Gaussian process latent variable models. Chemom Intell Lab Syst 109(1):9–21

Smith M, Reece S, Roberts S, Rezek I (2012) Online maritime abnormality detection using Gaussian processes and extreme value theory. In: 2012 IEEE 12th international conference on data mining. Institute of Electrical and Electronics Engineers (IEEE)

Traoré A, Grieu S, Thiery F, Polit M, Colprim J (2006) Control of sludge height in a secondary settler using fuzzy algorithms. Comput Chem Eng 30(8):1235–1242

Wilson AG, Adams RP (2013) Gaussian process kernels for pattern discovery and extrapolation. In: International conference on machine learning, pp 1067–1075

Xanthos S, Gong M, Ramalingam K, Fillos J, Deur A, Beckmann K, McCorquodale JA (2010) Performance assessment of secondary settling tanks using CFD modeling. Water Resour Manag 25(4):1169–1182

Yoo CK, Choi SW, Lee I-B (2002) Adaptive modeling and classification of the secondary settling tank. Kor J Chem Eng 19(3):377–382

Yu J (2012) A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes. Chem Eng Sci 68(1):506–519

Zambrano J, Samuelsson O, Carlsson B (2016) Monitoring a secondary settler using Gaussian mixture models. In: 9th EUROSIM congress on modelling and simulation. Oulu, Finland

Zambrano J, Samuelsson O, Chistiakova T, Liu H, Carlsson B (2015) Gaussian process regression for monitoring a secondary settler. In: 2nd new developments in IT & Water. Rotterdam, The Netherlands

Zhu H, Chen S, Han C (2014) Fusion of Gaussian mixture models for possible mismatches of sensor model. Inf Fusion 20:203–212