Detecting anomalous air flow-ammonia load ratios, using Gaussian process regression

* Division of System and Control, Department of Information Technology, Uppsala University, P.O. Box 337, 751 05, Uppsala, Sweden
**IVL Swedish Environmental Institute, Process modelling & IT, P.O. Box 210 60, 100 31 Stockholm, Sweden

Keywords: Fault detection, monitoring, Gaussian process regression

Summary and key findings
In this paper we propose a method to detect abnormal air flow-ammonia load ratios in active sludge basins. The purpose is to detect faulty sensors and process disturbances affecting the air flow-ammonia load ratio. The proposed method uses Gaussian process regression with sequential Monte Carlo optimization which gave good results on real plant data. I.e. drift in an ammonia on-line sensor, over flow during storm events and changed sludge properties were detected by the proposed method.

Background and relevance
Accurate sensor readings are essential to monitor and control a wastewater plant (WWTP) in a robust and resource-efficient way. However, plant data include faulty sensor data, partly due to harsh measurement conditions. On-line ammonia sensors, important for aeration control, is one example of sensors which are prone to have erroneous measurements (Åmand 2014).

Several monitoring and fault detection methods have been proposed, although rarely used in full-scale applications (Olsson et al. 2014). Air flow ratio between parallel active sludge lines has recently been used to detect faulty DO sensors (Carlsson & Zambrano 2013). Further, control strategies based on air flow and ammonia load have successfully been implemented (Svardal et al. 2003). In this paper, we make use of a non-parametric probabilistic data based method, Gaussian process regression (GPR), to monitor and detect anomalous air flow-ammonia load ratios. Anomalies include drifting sensor values, in particular from on-line ammonia sensors, and process disturbances. Further, we evaluate the applicability of GPR on real WWTP data using a standard GPR implementation (Rasmussen & Williams 2005) and a novel GPR with a sequential Monte Carlo implementation (Svensson et al. 2015). Both methods were used to map the potentially non-linear relationship between ammonia load and air consumption.

Methods
One straightforward approach would be to simply monitor the ratio of air flow and ammonia load, with upper and lower critical limits. However, the specific air flow, i.e. used air per treated amount of ammonia, is affected by e.g. ammonia load and diurnal variation. A general approach should include this information, only detecting anomalies relevant to actual conditions.

In this study we model the correlation between air flow and ammonia load as a Gaussian process.

Gaussian process regression A Gaussian process is a stochastic process
\[ f(x_i) \sim GP \left( m_{\theta_1}(x_i), cov_{\theta_2}(f(x_i), f(x_j)) \right), \]
which is fully described by its mean, \( m_{\theta_1}(x_i) \), and covariance function, \( cov_{\theta_2}(f(x_i), f(x_j)) \). In this study, \( f(x_i) \) is the modelled air flow at ammonia load \( x_i \). Since \( f(x) \) is modelled as a GP, any collection of air flow values are assumed to be jointly Gaussian
\[ [f(x_1), ..., f(x_N)] \sim \mathcal{N}(\mu, \mathcal{C}) \]
with mean values $\mu_i = m_{\theta_1}(x_i)$, and a $N \times N$ size covariance matrix $C$. In this study, mean values $\mu_i = m_{\theta_1}(x_i)$, correspond to a specific air flow at a given ammonia load. In Figure 2, the predicted mean is shown (solid black line) together with its confidence boundaries (dashed blue lines). For a detailed description of Gaussian processes, see (Rasmussen & Williams 2005).

The covariance function, also known as kernel, can be considered as a regularization matrix (Chen et al. 2012), in terms of regression. Here, we use the standard kernel which is an exponential covariance function. Although originally claimed to be non-parametric, most kernels contain a set of hyper parameters, $\theta_2$, which in the simplest approach are found by maximizing the likelihood function (ML). A non-parametric approach involves a prior probability distribution over the hyper parameters, which can be approximated by e.g. Monte Carlo methods. In this paper, we make use of both methods. For the ML approach, the GPML toolbox (Rasmussen & Nickisch 2010) was used. For the non-parametric approach, a novel algorithm based on a sequential Monte Carlo method (GPR-SMC) (Svensson et al. 2015) was used.

**Method workflow** In this study, the following steps were applied:

1) Define training (normal) data, $x_T$, for air flow, influent flow and ammonia concentration

   $$x_T := \{Q_{\text{air}1...t}, Q_{\text{flow}1...t}, C_{\text{NH}_41...t}\}$$

2) Compute training (normal) air flow-ammonia load ratio using GPR

   $$f(x_T) \sim \mathcal{GP}\left(m_{\theta_1}(x_T), \text{cov}_{\theta_2}\left(f(x_T), f(x_T)\right)\right)$$

3) Evaluate test data, $x_E$, and mark observations outside 95 percent confidence interval as potentially faulty.

   “Above normal airflow”:
   $$f(x_E) > m_{\theta_1}(x_T) + 1.96 \sqrt{\text{cov}_{\theta_2}\left(f(x_T), f(x_T)\right)}$$

   “Below normal airflow”:
   $$f(x_E) < m_{\theta_1}(x_T) - 1.96 \sqrt{\text{cov}_{\theta_2}\left(f(x_T), f(x_T)\right)}$$

All calculations were performed on batch data, i.e. not recursively.

**Data set** The data set included 1 hour measurements of airflow, ammonia concentration (on-line sensor) and influent water flow to the biological treatment step at Bromma WWTP, during 2014-05-05 to 2014-08-04. A higher sampling frequency will be evaluated in future studies, however in this initial study, 1 hour samples were considered to capture the main process features.

During the measurement period, three storm weather events, with overflow, occurred. During time $t = 350, ..., 1000$, a negative drift in the ammonia sensor was identified, see Figure 1, and will be referred to as the Faulty period. During the subsequent period, the Non-faulty period, $t = 1001, ..., 1800$, a similar negative trend was identified, due to a lower wastewater load during summer holidays. However, this trend could be explained by a decrease in ammonia concentration, confirmed by weekly lab analyses (black solid line).

**Results and discussion**

**Faulty ammonia sensor** In Figure 1, several detections during the faulty period were indicated by the proposed method, indicated as black circles. It is logical that the detections are marked as “Above normal airflow”, since the ammonia load is underestimated by a faulty ammonia sensor.

**Overflow detection during storm weather** During the non-faulty period, some sharp valleys at $t = 1080, 1810, 1900$ were detected as “Below normal air flow”, indicated by red circles. The detections coincide with peak flow during storm weather, and a by-pass of 5-30% of the influent flow. Thus, only a reduced part of the measured ammonia influent was treated, and subsequently the load was overestimated.

**Changed sludge properties** The detections “Above normal airflow” at $t = 1820, ..., 1900$ were originally considered as false positive detections. However, a closer examination of the original data
showed an increased airflow during the period, and an increase in suspended solids concentration (not shown here) caused by release of sludge from an anaerobic digester.

**Valve conditioning** Every third day, all air valves which are not used during normal operation, are operated during one hour to prevent valve stiction. This temporarily increases the total airflow by approximately 2 percent. It was not possible to detect the valve opening events, most probably because the additional airflow was within the range of what could be considered normal operation. I.e. the sensitivity of the method was too low to detect such small changes in airflow.

![Ammonia concentration Bromma WWTP](image_url)

**Figure 1.** Ammonia concentration measured by on-line sensor (blue solid line) at Bromma WWTP. Weekly lab analyzes (black solid line) were used to identify faulty sensor measurements. Detections by the proposed method are marked with black circles (higher airflow than normal), and as red circles (lower airflow than normal).

**Gaussian process regression – sequential Monte Carlo** Initially, the standard approach involving ML to estimate the hyper parameter values was used. However, optimizations quickly converged to local optimal solutions, resulting in mostly meaningless predictions (not shown here). Further, it was clear that the initial hyper parameter values were strongly linked to the values after optimization, suggesting several local minima. Predictions made by the non-parametric method, GPR-SMC, can be seen in Figure 2 (solid grey lines). The predictions are similar, indicating the robustness of the method. One randomly selected mean prediction from GPR-SMC was used to map the air flow-ammonia load ratio (solid black line).

In Figure 2, observations from both training data, \( x_T \), and test data, \( x_E \) are plotted as air flow with respect to ammonia load. The estimated mean value of the non-linear mapping by GP, \( m_{\beta}(x_j) \), together with 2 standard deviations confidence interval is indicated by dashed blue lines. Observations below (above) confidence limit indicate a lower (higher) air flow-ammonia load ratio than predicted by the training data, and are marked with red (black) circles.
Figure 2. Detections of data which deviate from training data. Black circles (red circles) indicate higher (lower) air flow per ammonia load.

It is clear that the cluster of faulty observations in Figure 2 lies outside the training data considering the mean prediction of the GPR and the confidence limits. It can be seen that the confidence interval increases for extreme loads, i.e. below 15 g/s and above 80 g/s. This is a property of GPR where the uncertainty estimate increases in regions with sparse amount of data, i.e. unknown regions. One could argue that this would make this method restrictive in the sense, not to make false positive detections.

References


Åmand L. (2014). *Ammonium Feedback Control in Wastewater Treatment Plants*. Uppsala Dissertations from the Faculty of Science and Technology.