Banks of Gaussian Process Sensor Models for Fault Detection in Wastewater Treatment Processes

H.L. Ivan[a] J.A. Ivan[b]

[a]*Mälardalen University, Sweden*
[b]*Örebro University, Sweden*
[a]heidi.ivan@mdu.se, [b]jean-paul.ivan@oru.se

Abstract

The harsh operating environment in a wastewater treatment process (WWTP) makes sensor faults commonplace. Detecting these faults can be challenging due to the complex process dynamics, unknown inputs, and general noise in the process and measurements. Comparing sensor readings against predictions from a physics-based or data-driven model of the WWTP is a common strategy for detecting such faults. In this work sensor measurements are directly modelled using Gaussian process (GP) regression, a data-driven multivariate approach. These GP sensor models are, with a generalised product of experts, combined into a dedicated fault isolation scheme resembling traditional observer bank methods. The residuals are monitored with a multivariate exponentially weighted moving average chart which is used for fault detection and isolation. The method is evaluated using simulated data generated with the Benchmark Simulation Model No. 1 WWTP. Fault detection performance is reported using several standard metrics such as false alarms, missed detections, time to detection, and successful fault isolations, with emphasis on reporting across a wide range of sensors and faults to provide a point of comparison for future studies. The proposed approach performs well across these metrics. Given sufficient data representative of normal operation, this approach can easily be adapted across a wide variety of plant configurations and can be used to create operator-friendly diagnostics resembling classical control charts.

1 Introduction

Wastewater treatment plants (WWTPs), like many critical components of public infrastructure, are gradually shifting to higher levels of automation in process operation. This is in part driven by global incentives to shift towards water resource recovery facilities, in conjunction with progress in regulation of the environmental impact of WWTPs, and typical cost incentives of reducing energy and material consumption. However, as in any process, automation depends on reliable process supervision; this is particularly challenging in WWTPs as sensors are often sparse and subject to harsh operating conditions.

A key task in a process supervision system is fault detection (FD). A common model-based FD strategy is to generate a residual signal from the difference between model predictions and actual sensor measurements (Chen & Patton, 1999). Predictions can stem from physics-based or data-driven models. Data-driven methods have risen in popularity as they generally require less extensive process-specific knowledge. However, forgoing process-specific knowledge means that more data is required for fitting data-driven models. Moreover, data-driven methods - neural networks as an archetypal example - can exhibit out-of-distribution overconfidence and in-distribution sensitivity to adversarial examples (Szegedy et al., 2014).

Gaussian processes (GPs) are a class of data-driven models which explicitly model uncertainty, and provide clear avenues - see for example (Jidling et al., 2017) - for reintroducing domain knowledge into learned models (Rasmussen et al., 2006). This principled treatment of uncertainty in the process model is useful as typically the residual generation process is complicated by requirements to reject model uncertainty and process disturbances while remaining sensitive to faults (Witczak, 2007).

Fault isolation (FI) - which requires FD - also requires further structure in the generated residuals. Two such structures are common: dedicated schemes - wherein each residual in a set is only *sensitive* to a single fault - and generalised schemes - wherein each residual is *insensitive* to only a single fault (Witczak, 2007; Chen & Patton, 1999). In WWTPs these schemes have previously been applied using physics-based state estimators configured in banks of observers (Nejjari et al., 2008; Nagy-Kiss & Schutz, 2013). However, the complexity of the process makes data-driven state estimation attractive - as in other fields (Palma et al., 2005; Sina Tayarani-Bathaie & Khorasani, 2015).

An issue which arises in using these schemes for FI is that the sets of residuals that need to be monitored for FD become large (one set per observer). The use of

multivariate statistical process monitoring techniques, such as the Hotelling T^2 chart (Hotelling, 1947) and the multivariate exponentially weighted moving average (MEWMA) chart (Lowry et al., 1992), can alleviate some of these difficulties. The latter is often more sensitive to smaller faults and slow drift faults (Montgomery, 2009). Creating a single, interpretable, and easily visualisable FD statistic for monitoring is of high priority in the WWTP industry; which is traditionally dominated by operator expertise.
In this work we illustrate the feasibility of the use of GP regression based sensor models combined using a generalised product of experts into an dedicated FI scheme for sensor fault detection in a standard biological WWTP. The ability to detect sensor faults on both controlled variable sensors, and ordinary monitored variables, is shown for two fault profiles and varying fault sizes and durations. The diagnostic performance is based on the number of faults detected, time to detection, number and duration of false alarms, as well as the number of faults correctly isolated.

2 Background

This section describes relevant ideas and theoretical perquisites to clarify the method section. In §2.1 FI schemes in general and the modification using direct sensor models proposed in this work are described. §2.2 concerns GP regression; used to create the aforementioned sensor models. Thereafter §2.3 covers products of experts, which combine several GP models into a dedicated FI scheme. Finally, §2.4 describes the MEWMA chart used for monitoring the residuals.

2.1 Dedicated FI Schemes without State Estimation

In a typical dedicated FI scheme residuals are generated from a bank of state estimators, where each estimator in the bank is ignorant of one of the sensors (Chen & Patton, 1999; Witczak, 2007). This is shown in (1): a sequence of measurements $\boldsymbol{y}_{-i,1:t}$ from time 1 to t (where $\boldsymbol{y}_{-i,t}$ denotes the vector $[y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_n]_t$) is used to estimate the state $\hat{\boldsymbol{x}}_t^{-i}$ from which a sensor model estimates the sensor measurements $\hat{\boldsymbol{y}}_t^{-i}$. The notation $-i$ in a superscript indicates a state/sensor estimate is ignorant of measurements from sensor i, whereas a subscript indicates a vector missing state/sensor i.

$$\boldsymbol{y}_{-i,1:t} \xrightarrow{\text{state estimation}} \hat{\boldsymbol{x}}_t^{-i} \xrightarrow{\text{sensor model}} \hat{\boldsymbol{y}}_t^{-i} \tag{1}$$

This estimate is used to compute a residual between the sensor estimates and measurements for FD/FI.
However, performing this state estimation in the WWTP is often difficult. In response to this difficulty the feasibility of directly estimating $\hat{\boldsymbol{y}}_t^{-i}$ from $\boldsymbol{y}_{-i,1:t}$ is considered. However, if the sequence $\boldsymbol{y}_{-i,1:t}$ is assumed to be Markovian, the problem can be further simplified by instead considering the estimation of $\hat{y}_{i,t}^{-i}$ directly from $\boldsymbol{y}_{-i,t}$. There is a problem with this: in the scheme shown in (1) if there is a fault in sensor i this appear as a) a residual in $\hat{y}_{i,t}^{-i}$, and b) in all $\hat{\boldsymbol{y}}_t^{-j}$ for $i \neq j$. This asymmetry is what allows FI from the residuals in (1). Directly estimating $\hat{y}_{i,t}^{-i}$ from $\boldsymbol{y}_{-i,t}$ eliminates this property - a fault on sensor i affects every estimate. This can be remedied by repeating the leave-out-one pattern in the original scheme. In this work we propose using a bank of $n(n-1)$ models $\mathcal{M}^{-ij}$ where $i \neq j$ where each model estimates $\hat{y}_{i,t}$ from all sensors except *i and j*. Given the Markov assumption the time subscripts are omitted:

$$\boldsymbol{y}_{-ij} \xrightarrow[\mathcal{M}^{-ij}]{\text{sensor model}} \hat{y}_i^{-ij}. \tag{2}$$

For each i this creates $n-1$ estimates $\hat{y}_i^{-ij}$, each insensitive to a fault in a different sensor $j \neq i$. This reestablishes the required asymmetry for FI.

2.2 Gaussian Process Regression

Gaussian process (GP) regression (Williams & Rasmussen, 1995; Rasmussen et al., 2006) is a non-parametric regression method which assumes the target function $f : \mathcal{Y}^{-ij} \to \mathbb{R}$ to be a stochastic (Gaussian) process and conditions this prior process on observations to obtain a posterior distribution over functions that explain the observations. A Gaussian distribution is fully specified by its mean vector and covariance matrix. Analogously, a GP is fully specified by a mean function $m^{-ij} : \mathcal{Y}^{-ij} \to \mathbb{R}$ and a covariance function $k_f^{-ij} : \mathcal{Y}^{-ij} \times \mathcal{Y}^{-ij} \to \mathbb{R}$. This is typically denoted $f \sim \mathcal{GP}(m^{-ij}, k_f^{-ij})$. Assuming the observations $y_i^{-ij}(\boldsymbol{y}_{-ij}) = f(\boldsymbol{y}_{-ij}) + \varepsilon$ of f are perturbed by Gaussian noise $\varepsilon \sim \mathcal{N}(0, \sigma_n^2)$, the measurement process is also Gaussian and is denoted,

$$y_i^{-ij} \sim \mathcal{GP}\left(m^{-ij}, k^{-ij}\right). \tag{3}$$

Here, the measurement covariance k^{-ij} is a sum of the 'base' covariance of f and the noise of the observation process: $k^{-ij}(y_{-ij}, y'_{-ij}) = k_f^{-ij}(y_{-ij}, y'_{-ij}) + \delta_{yy'}\sigma_n^2$ where $\delta_{yy'}$ is the Kronecker delta on $y_{-ij} = y'_{-ij}$.
Like many forms of Bayesian inference, GP regression has historically been associated with heavy computational costs. However, frameworks such as GPyTorch (Gardner et al., 2018), taking advantage of modern hardware and theoretical progress, allow practical use of GPs with standard covariance functions.

2.3 Generalised Products of Experts

Taken in combination, §2.1 and §2.2 suggest using GP regression to learn a bank of models $\mathcal{M}^{-ij}$. Structure in this bank can be exploited for FI, but in order to

allow for clear visualisation and interpretation by operators the signals from these $n \times (n-1)$ models must ideally be used to generate a single FD statistic.
One method of obtaining a combined predictive distribution $p_c(y|x)$ from several GP posteriors $p^i(y|x)$ is the (generalised) product of experts (GPoE/PoE) (Cao & Fleet, 2015),

$$p_c(y|x) = \frac{1}{Z}\prod_i p^{\alpha_i(x)}(y|x), \tag{4}$$

where Z is a normalisation constant. The annealing parameters α_i are used to amplify or diminish the importance of each model's contribution to the combined distribution. The simplest parameters, corresponding to a PoE, $\alpha_i(x) = 1$, are used in this study. If each model in the product is Gaussian the combined distribution is also Gaussian with mean and covariance

$$m_c^{-i}(x) = k_c^{-i}(x,x)\sum_j m^{-ij}(x)\alpha_j(x)\lambda^{-ij}(x), \tag{5}$$

$$k_c^{-i}(x,x) = \left(\sum_j \alpha_j(x)\lambda^{-ij}(x)\right)^{-1}. \tag{6}$$

Where $\lambda^{-ij}(x) := 1/k^{-ij}(x,x)$. Combining the predictions from the bank in this way produces a combined posterior for each sensor,

$$y_i^{-i} \sim \mathcal{GP}(m_c^{-i}, k_c^{-i}). \tag{7}$$

The vector $\hat{\boldsymbol{y}} = [y_1^{-1}, \ldots, y_n^{-n}]^T$ denotes the full estimate of the sensor state obtained from the GPoE.

2.4 Multivariate Process Monitoring

The multivariate exponentially weighted moving average (MEWMA) chart, first proposed by Lowry et al. (1992), utilises information from successive samples and is therefore relatively sensitive to small shifts in the mean of the monitored variable. In this application, that is the standardised residual vector $\boldsymbol{r}_t$, where the raw residuals are $\tilde{\boldsymbol{r}}_t = \boldsymbol{y}_t - \mathbb{E}[\hat{\boldsymbol{y}}_t]$. The relevant parameters are defined as (Montgomery, 2009)

$$\boldsymbol{Z}_t = \lambda \boldsymbol{r}_t + (1-\lambda)\boldsymbol{Z}_{t-1} \tag{8}$$

where $0 \leq \lambda \leq 1$ and $\boldsymbol{Z}_0 = \boldsymbol{0}$. The statistic monitored on the chart is

$$T_t^2 = \boldsymbol{Z}_t^T \boldsymbol{\Sigma}_{\boldsymbol{Z}_t}^{-1} \boldsymbol{Z}_t \tag{9}$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{Z}_t} = \frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2t}\right]\boldsymbol{\Sigma}. \tag{10}$$

$\boldsymbol{\Sigma}$ represents the covariance of $\boldsymbol{r}$ from a collection of samples when the process is known to be operating normally. The performance of the MEWMA chart is tuned by adjusting λ, the smoothing factor, as well as the limit H.
When $T_t^2 > H$ the limit is violated, indicating a fault. The source of the deviation can be determined by decomposition of the MEWMA statistic, described by VandenHul (2002). This requires recalculating T_t^2 for the value of t at which the limit is violated based on $\boldsymbol{r}^{-i} := [r_1, \ldots, r_{i-1}, r_{i+1}, \ldots, r_n]^T$, thus generating n values for $T_t^{2,-i}$. By observing which $T_t^{2,-i}$ decreases the most compared to T_t^2 the responsible residual can be isolated.

3 Methodology

The high-level strategy proposed for performing and evaluating FD/FI using banks of GPs is as follows. Data, with and without faults, is generated in simulation (§3.1) and used to train a bank of GP sensor models (§3.2). These models are combined in a GPoE, and the combined predictions used to generate a MEWMA chart on the residuals. FD/FI statistics are calculated over a set of 320 faults per sensor (12 total) in a typical sensor set - parameterised by fault type, size, duration, and start time.

3.1 Simulation

The Benchmark Simulation Model No.1 (BSM1) was used to simulate the operation of the WWTP. The simulation platform consists of two anoxic reactors of 2000 m^3 and three aerated reactors of 3999 m^3 followed by a secondary settler of 6000 m^3 (Gernaey et al., 2014). The BSM1 process contains two standard control loops: S_{NO} control in the second reactor with set-point of 1 $\mathrm{g\,N\,m^{-3}}$ and S_O control in the fifth reactor with set-point 2 $\mathrm{g\,O_2\,m^{-3}}$ (Gernaey et al., 2014).
Sensor measurements of dissolved oxygen (S_O), alkalinity (S_{ALK}), total suspended solids (TSS), nitrate/nitrite nitrogen (S_{NO}), and ammonium/ammonia nitrogen (S_{NH}), at several points in the process were recorded at 15-minute intervals. These are hereafter denoted: $S_{NH,1}$, $S_{NO,2}$, $S_{O,3}$, $S_{NO,3}$, $S_{NH,3}$, $S_{O,4}$, $S_{NH,4}$, TSS_4, $S_{O,5}$, $S_{NO,e}$, $S_{ALK,e}$, TSS_e where the subscripts denote that the measurements are taken in the indicated tank number (1 to 5) or in the effluent (e). The the two controller outputs, $u_{NO,2}$ and $u_{O,5}$ were also recorded. The sensors were selected based on the approach in Marais, Zaccaria, Ivan, & Nordlander (2022); Ivan (2023).
The simulations used the BSM1 dry weather influent file, simulating two weeks of operation. The first week of simulated data was used as training data for GP training (§3.2) and chart calibration (§2.4, §3.4). The second week was held-out for testing FD/FI, where data from the eighth day was used for chart tuning and faults were introduced starting in the ninth day. Two fault types, bias and drift, were used with

Table 1. Fault parameters used for testing performance. All 320 parameter combinations were tested.

Size	Bias [σ]	1.5	2	3	5	
	Drift [μ/Day]	0.1	0.25	0.5	1	
Direction		+	−			
Start Time [Day]		8.75	9.5	10.25	11	
Duration [Day]		0.5	1	1.5	2	2.5

varying parameters shown in Table 1. All combinations of size-direction-start-duration were simulated for each fault type for a total of 320 faults per sensor. The fault sizes are specified in proportion to the standard deviation (bias faults) and mean (drift fault) of the sensor signal as determined from the training data. All measurements are also standardised w.r.t. these statistics, $y_{i,t} = (\tilde{y}_{i,t} - \text{avg}_t\, \tilde{y}_{i,t})/\,\text{std}_t\, \tilde{y}_{i,t}$ where $\tilde{y}$ denotes the raw measurements.

3.2 Individual GP Sensor Models

The simulated sensor measurements $\mathcal{D} = (\boldsymbol{y}_t)_{t=1}^T$ are split into training and test data as described in §3.1. In order to create a supervised learning problem with FI asymmetry (see §2.1), $n \times (n-1)$ training sets are derived from this data. Each model $\mathcal{M}^{-ij}$ thus has associated datasets

$$\mathcal{D}^{-ij} \;=\; \{(\boldsymbol{x}_j, y_i) : \boldsymbol{x}_j = \boldsymbol{y}_{-ij}, \boldsymbol{y} \in \mathcal{D}\}. \quad (11)$$

As the measurements are standardised (§3.1), a GP prior with zero mean is placed over all models,

$$y_i^{-ij} \;\sim\; \mathcal{GP}(0, k^{-ij}). \quad (12)$$

The covariance k^{-ij} for each model is a sum of a linear kernel with noise parameter ψ^{-ij} and a squared exponential kernel with independent lengthscales ϕ^{-ij} for each input dimension and scale parameter σ^{-ij}:

$$\begin{aligned} k^{-ij}(\boldsymbol{x}_j, \boldsymbol{x}'_j) \;&=\; \sigma^{-ij} k_{\text{SE}}^{-ij}(\boldsymbol{x}_j, \boldsymbol{x}'_j) \\ &\quad + k_{\text{LIN}}^{-ij}(\boldsymbol{x}_j, \boldsymbol{x}'_j) + \sigma_n^2 \delta_{jj'}, \quad (13) \\ k_{\text{SE}}^{-ij}(\boldsymbol{x}_j, \boldsymbol{x}'_j) \;&=\; \exp\left(-\frac{1}{2}\boldsymbol{d}_j^T \Phi_{-ij}^{-2} \boldsymbol{d}_j\right), \quad (14) \\ k_{\text{LIN}}^{-ij}(\boldsymbol{x}_j, \boldsymbol{x}'_j) \;&=\; \psi^{-ij} \boldsymbol{x}_j^T \boldsymbol{x}'_j, \quad (15) \end{aligned}$$

where $\boldsymbol{d}_j = \boldsymbol{x}_j - \boldsymbol{x}'_j$ and $\Phi_{-ij} := \text{diag}\, \phi^{-ij}$. The initial covariance parameters are shown in Table 2. These GPs were defined using GPyTorch (Gardner et al., 2018) and each set of hyperparameters was independently optimised until convergence with respect to the marginal log-likelihood of each model. Optimisation was performed using the ADAM algorithm (Kingma & Ba, 2017) with learning rate 0.1 and learning rate decay 0.99.

Table 2. Initial covariance parameters and optimisation constraints.

Param.	Initial Value(s)	Optimisation Bounds
σ^{-ij}	1.0	$(0, 10)$
φ^{-ij}	Random $\in (0.5, 1.5)$	$(0, 10)$
ψ^{-ij}	1.0	$(0, 10)$
σ_n^2	1.0	$(0.01, \infty)$

3.3 Sensor Residuals via PoE

The GP posteriors are composed in a PoE, obtaining a combined estimate y_i^{-i} for each sensor:

$$\begin{aligned} y_i^{-i}(\boldsymbol{y}_{-i}) \;&\sim\; \mathcal{N}(m_c^{-i}, k_c^{-i}), \quad (16) \\ m_c^{-i}(\boldsymbol{y}_{-i}) \;&=\; k_c^{-i}(\boldsymbol{y}_{-i}, \boldsymbol{y}_{-i}) \sum_j m_\lambda^{-ij}(\boldsymbol{y}_{-ij}), \quad (17) \\ k_c^{-i}(\boldsymbol{y}_{-i}, \boldsymbol{y}_{-i}) \;&=\; \left(\sum_j \lambda^{-ij}(\boldsymbol{y}_{-ij})\right)^{-1}, \quad (18) \end{aligned}$$

where $m_\lambda^{-ij}(\boldsymbol{y}_{-ij}) := m^{-ij}(\boldsymbol{y}_{-ij}) \lambda^{-ij}(\boldsymbol{y}_{-ij})$. The model residuals are obtained from the PoE output at each timestep, $\hat{\boldsymbol{y}}_t$, as described in §2.4.

3.4 MEWMA Chart

As in §2.4, the residuals are standardised based on the training data:

$$r_{t,i} \;=\; (\tilde{r}_{t,i} - \text{avg}_t\, \tilde{r}_{t,i})/\,\text{std}_t\, \tilde{r}_{t,i}. \quad (19)$$

The covariance of the standardised residuals $\boldsymbol{\Sigma}$ was used to calibrate the chart according to (10). Small values for the smoothing factor, $\lambda = 0.05, 0.1, 0.2$, were evaluated; which in principle allow for the detection of smaller faults (Montgomery, 2009). For each value of λ the mean and standard deviation of the T^2 statistic was calculated during day 8 of the simulation data (§3.1) to determine an appropriate limit size H. Three values were tested for the limit size $H = \text{avg}_t\, T^2 + h\,\text{std}_t\, T^2$, for $h = 2, 3, 4$.
Charts with each combination of (λ, h) were used to monitor the performance of the process using (8) and (9). Every limit violation was treated as a fault alarm. In the case of correct fault detections, isolation was performed by constructing reduced charts of the persenor T^2-statistic, $T^{2,-i}$, as described in §2.4.

3.5 Diagnostics

As described in §2.4, a fault is detected when the chart limit H is crossed by the MEWMA T^2 statistic during the fault. For each pair of chart limit and smoothing factor the following statistics were recorded for all faults in Table 1: a) correct violation of H during a fault - fault detection (FD), b) number of incorrect

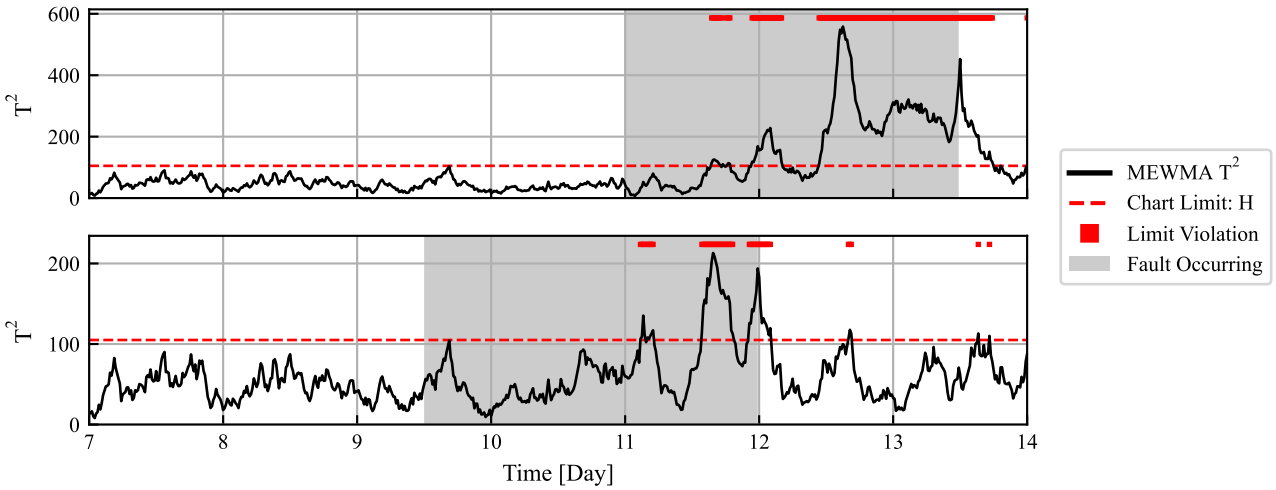

Figure 1. Illustration of the MEWMA chart showing the chart limit, the statistic, and limit violations. The time period during which the fault is occurring is highlighted. Top: Drift fault on $S_{O,5}$ with a rate of -0.25, Bottom: Drift fault on $S_{O,5}$ with a rate of -0.1.

chart limit violations - false alarms (#FAs), c) duration of false alarms (FA), d) time taken to detect the fault (TTD), and e) successful fault isolation (FI).

The MEWMA chart requires some time to return to normal after a fault stops. As such time spent above the chart limit immediately following a successful fault detection is not reported in the FA statistic. Note, defining a detection by *crossing* of the limit means that a false alarm preceding a fault which continues into the start of the fault does *not* constitute a detection.

Fault isolation was performed based on the mean of the T^2-decomposition during the first hour after violation. Only the $T^{2,-i}$ deviating most from T^2 was used for isolation. Isolation of faults of the controlled variable sensors was performed by monitoring the controller outputs, not the sensor measurements. For a discussion of 'fault hiding' on controlled variables see Marais, Zaccaria, & Odlare (2022).

4 Results and Discussion

Two MEWMA charts are shown in Figure 1 for two different drift faults on the $S_{O,5}$ sensor. It is clear that the smaller fault is harder to detect, shown by the longer detection time and the smaller values of T^2 relative to those of normal operation. Natural variation in the residuals, and therefore the T^2, can worsen the situation. For example, the T^2 statistic is low around day 10 - faults occurring near this point will be harder to detect due to the statistic being below its mean. This may be improved by reducing the nominal variance of the chart, requiring improved sensor estimates.

Overall the chart is clear and provides a good starting point for operator-friendly FD. With regard to FI, Figure 2 shows an example of an isolation plot, which could be shown to operators continuously using a rolling window on the decomposed T^2 statistic. The isolation chart shows clearly which residuals are contributing to variations in the MEWMA chart.

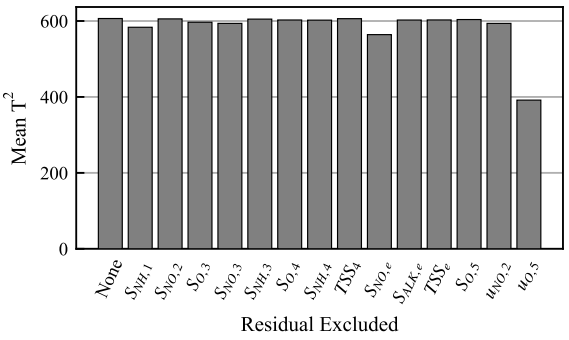

Figure 2. Example of an isolation plot for the drift fault in Figure 1-Top showing the mean of $T^{2,-i}$ (residual i excluded from T^2) during the isolation period. "None" denotes T^2: no residual excluded.

Figure 2 shows the mean value of $T^{2,-u_{O,5}}$ (i.e. the T^2 value calculated excluding $u_{O,5}$, the controller output for the $S_{O,5}$ controller) during the isolation period has the lowest value. This indicates the violation can be attributed to $u_{O,5}$, and therefore $S_{O,5}$.

The following sections present a more detailed analysis across chart parameters; fault types, sizes, and durations; and across different sensors.

4.1 Impact of MEWMA Chart Parameters

Broadly, the different values of (λ, h) affect FD/FI in accordance with theoretical expectations. As the limit size, h, increased detection becomes slower and less consistent, false alarms decrease, but FI becomes easier. As the smoothing factor, λ, increases the opposite occurs; smaller faults become detectable, but FI on these faults is more difficult, and false alarms increase. Small drift faults, in particular, are most sensitive to the change in the smoothing factor. These results are summarised in Table 3.

Table 3 clearly shows the expected trade-off that must

Table 3. Summary of diagnostics performance parameters, averaged over all fault types and characteristics, for the different chart parameters.

λ	h	FD [%]	FI [%]	TTD [d]	FA [d]	#FAs
0.20	2	0.98	0.69	0.16	0.33	15.31
0.20	3	0.90	0.82	0.28	0.07	4.64
0.20	4	0.87	0.85	0.30	0.01	1.13
0.10	2	0.95	0.70	0.22	0.24	11.07
0.10	3	0.90	0.80	0.28	0.04	2.88
0.10	4	0.87	0.85	0.31	0.00	0.14
0.05	2	0.94	0.72	0.24	0.30	7.04
0.05	3	0.92	0.79	0.27	0.08	5.04
0.05	4	0.89	0.83	0.31	0.02	1.16

be made in the MEWMA chart design: improved detectability comes at the expense of isolability and false alarms. For a given smoothing factor, detection rates decrease by between 5 % to 11 % and isolation rates increase by 15 % to 23 % as the limit size is increased. Detection times increase by 30 % to 46 % while the number and duration of false alarms decreases by 83 % to 97 %. Smaller smoothing factors are less sensitive to the limit size.

It is worth noting that a real FD system can reasonably run several combinations of chart parameters with the strengths and weaknesses of each chart in mind. Balancing these trade-offs, the remaining analysis proceeds with $(\lambda, h) = (0.1, 3)$.

4.2 Performance of Diagnostics

Table 4 shows a comparison between bias and drift type faults, averaged across all fault parameters and sensors. The false alarms are not included as they do not differ from those presented in Table 3; false alarms are chart-dependent, not fault-dependent.

The drift faults are, as expected, harder to detect and require a longer time on average before the faults are detected. However, the isolation of drift faults is not substantially lower than that of bias faults.

The relative difficulty of detecting and isolating faults in different sensors can be seen in Figure 3 where, averaged over all fault parameters, detection and isolation statistics are shown. The most challenging faults to isolate occur in the controlled variable sensors, that is $S_{NO,2}$ and $S_{O,5}$, the lowest average detection rate among all sensors also occurs in the former. This is expected: the controller works to keep these sensor values at the set-point, obfuscating the effects of the sensor faults on the sensor itself. The proposed residual scheme relies on the use of the controller output, as mentioned previously, to reliably circumvent this issue.

Figure 3 also shows that faults in sensors TSS_4, $S_{ALK,e}$, and $S_{NO,e}$ have some of the highest detection and isolation rates and shortest detection times. This is of special importance as sensors in the effluent are important for monitoring limits related to environmental regulations. In general, sensors which have high isolation rates, such as $S_{ALK,e}$ and $S_{NO,e}$, should be subject to further careful monitoring as it is possible that they are often the target of an incorrect isolation. In the faults tested these two sensors were responsible for 26 % of incorrect isolation cases.

In order to evaluate the effects of different fault sizes and durations on detectability and isolability, the results for a single sensor ($S_{NO,2}$ - a controlled variable) are shown in Figure 4.

Considering the bias faults first: all the faults are detected, and as the size of the fault increases the time to detection decreases to a minimum of 0.026 d, or 37 min. It might be expected that the isolation rate increase with the size of the fault, however, it is important to note that this fault occurs on a controlled variable sensor. This type of fault impacts the operation of the entire process through the control system, therefore, larger faults can have a larger impact on other process variables. This can make these faults more challenging to isolate as they disrupt other variables in the process.

The chaotic behaviour of the smallest drift fault likely has similar explanation - slow drift is corrected by the controller and propagates non-linearly throughout the system. Apart from this exception, the behaviour of the drift faults is unsurprising: larger faults are detected more reliably and more rapidly, and when a fault persists for longer it is both easier to detect and easier to isolate. The minimum detection time of the drift faults is around 0.44 d, or 10 h.

Comparing the results for faults on the $S_{NO,2}$ sensor to those in Marais, Zaccaria, & Odlare (2022), where a univariate EWMA chart is used, the detection times for the bias faults are slightly longer but the time for drift fault detection has been decreased by several hours. The number of false alarms are in the same order of magnitude, and the detection rate of drift faults has increased from 56 % to 64 % to between 75 % and 100 % for faults longer than 1 d. In Marais, Zaccaria, & Odlare (2022) the results were not broken down by duration of fault so this comparison is not exhaustive. Further comparisons with the broader literature are difficult due to inconsistencies in how results are reported, varying fault sizes, and incomparable plant configurations. A cursory comparison with Luca et al. (2021, 2023) shows detection times in similar ranges with possibly better performance on the bias faults.

5 Conclusions and Recommendations

Direct modelling of sensors using GP regression in a dedicated residual scheme and monitoring with a MEWMA chart can be used for FD/FI in a WWTP. Clear comparison with the broader literature is diffi-

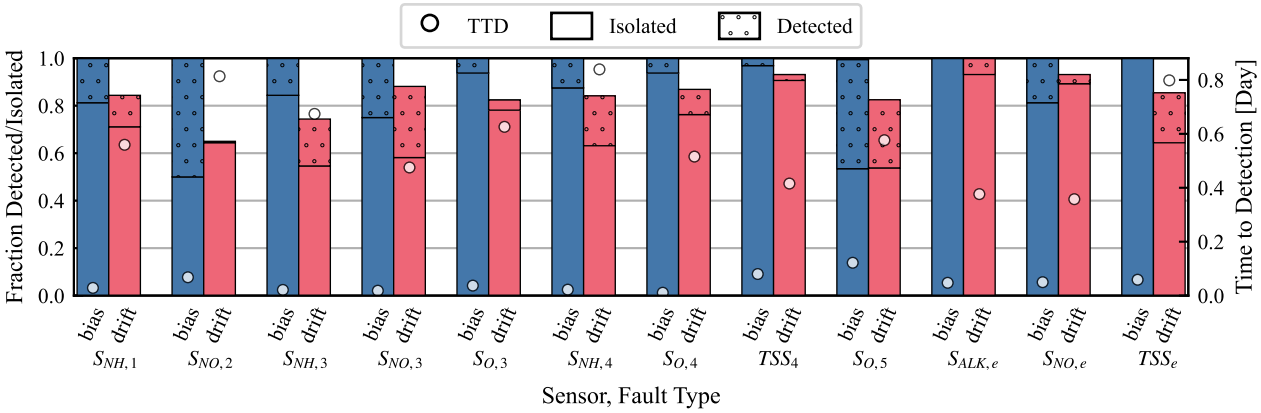

Figure 3. Detection and isolation rates, and time to detection split across bias and drift faults for each individual sensor.

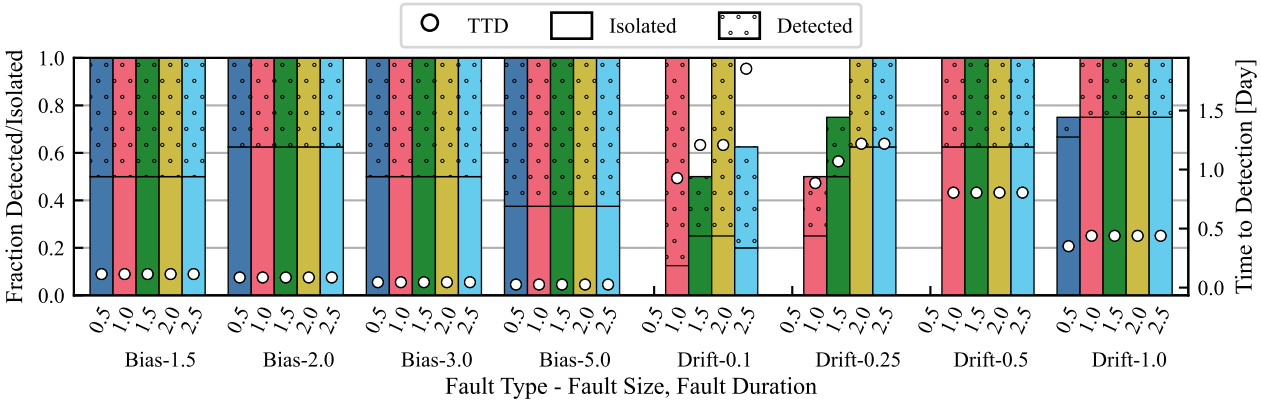

Figure 4. Detection and isolation rates, and time to detection across all faults on the $S_{NO,2}$ sensor. The bars represent a combination of fault type, size, and duration, and are grouped according to type and size.

Table 4. Detection and isolation statistics by fault type.

	FD [%]	FI [%]	TTD [d]
Bias	100	83	0.047
Drift	80	77	0.564

cult, as standardised reporting of performance evaluation parameters in studies performed in this field is lacking. In response to this difficulty, testing on a wide range of faults across many standard sensors has been reported in the hopes of facilitating future comparisons.

The method improves over a previous study using a univariate approach, and the results are comparable to other multivariate methods for FD/FI. Critically, the proposed approach is easy to visualise; a priority when developing FD/FI methods for an industry that relies heavily on operator expertise and shies away from uninterpretable automation.

The proposed approach leaves a great deal of room for further study. Without methodological changes, results across the each tested sensor can be documented, performance on out-of-distribution test data such as the BSM1 wet weather influent data can be evaluated, and more detailed FI studies carried out. The sensor models themselves can likely be simplified and made more interpretable by sharing parameters across models. Annealing the GPoE distributions, directly using the pre-GPoE sensor models in a generalised scheme, or other similar modifications to the sensor models could also yield improvements.

Acknowledgement

The authors acknowledge and express their gratitude to Dr Ulf Jeppsson and other IWA Task Group members for the availability of the BSM1 code. The work of J.A. Ivan was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

CRediT Author Contribution Statement

Conceptualization, Validation, Formal analysis, Investigation, Resources, Data curation, Writing: H.L.I., J.A.I.; Methodology: H.L.I. - FD/FI charts, J.A.I. - GP/GPoE models; Software: H.L.I. - BSM1/FD/FI, J.A.I. - GP/GPoE models; Visualization: H.L.I.;

References

Cao, Y., & Fleet, D. J. (2015). Generalized Prod-

uct of Experts for Automatic and Principled Fusion of Gaussian Process Predictions. *arXiv preprint arXiv:1410.7827*.

Chen, J., & Patton, R. J. (1999). *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Springer US. doi: 10.1007/978-1-4615-5149-2

Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q., & Wilson, A. G. (2018). Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *Advances in neural information processing systems*.

Gernaey, K. V., Jeppsson, U., Vanrolleghem, P. A., & Copp, J. B. (Eds.). (2014). *Benchmarking of Control Strategies for Wastewater Treatment Plants*. IWA Publishing. doi: 10.2166/9781780401171

Hotelling, H. (1947). Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights. In C. Eisenhart, M. Hastay, & W. Wallis (Eds.), *Techniques of Statistical Analysis* (pp. 111–184). New York: McGraw Hill.

Ivan, H. L. (2023). *Fault Detection in Wastewater Treatment : Process Supervision to Improve Wastewater Reuse* (Licentiate dissertation). Mälardalen University, Västerås, Sweden.

Jidling, C., Wahlström, N., Wills, A., & Schön, T. B. (2017). Linearly constrained Gaussian processes. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc.

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.

Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, *34*(1), 46–53. doi: 10.2307/1269551

Luca, A.-V., Simon-Várhelyi, M., Mihály, N.-B., & Cristea, V.-M. (2021). Data Driven Detection of Different Dissolved Oxygen Sensor Faults for Improving Operation of the WWTP Control System. *Processes*, *9*(9), 1633. doi: 10.3390/pr9091633

Luca, A.-V., Simon-Várhelyi, M., Mihály, N.-B., & Cristea, V.-M. (2023). Fault Type Diagnosis of the WWTP Dissolved Oxygen Sensor Based on Fisher Discriminant Analysis and Assessment of Associated Environmental and Economic Impact. *Applied Sciences*, *13*(4), 2554. doi: 10.3390/app13042554

Marais, H. L., Zaccaria, V., Ivan, J.-P. A., & Nordlander, E. (2022). Detectability of Fault Signatures in a Wastewater Treatment Process. In *The First SIMS EUROSIM Conference on Modelling and Simulation, SIMS EUROSIM 2021, and 62nd International Conference of Scandinavian Simulation Society, SIMS 2021* (pp. 418–423). Virtual Conference, Finland. doi: 10.3384/ecp21185418

Marais, H. L., Zaccaria, V., & Odlare, M. (2022). Comparing statistical process control charts for fault detection in wastewater treatment. *Water Science and Technology*, *85*, 1250–1262. doi: 10.2166/wst.2022.037

Montgomery, D. C. (2009). *Introduction to statistical quality control* (6th ed ed.). John Wiley & Sons, Incorporated.

Nagy-Kiss, A. M., & Schutz, G. (2013). Estimation and diagnosis using multi-models with application to a wastewater treatment plant. *Journal of Process Control*, *23*(10), 1528–1544. doi: 10.1016/j.jprocont.2013.09.027

Nejjari, F., Puig, V., Giancristofaro, L., & Koehler, S. (2008). Extended Luenberger Observer-Based Fault Detection for an Activated Sludge Process. *IFAC Proceedings Volumes*, *41*(2), 9725–9730. doi: 10.3182/20080706-5-KR-1001.01645

Palma, L., Coito, F., & da Silva, R. (2005). Process fault diagnosis approach based on neural observers. In *2005 IEEE Conference on Emerging Technologies and Factory Automation* (Vol. 1, pp. 4 pp.–1060). doi: 10.1109/ETFA.2005.1612642

Rasmussen, C. E., Williams, C. K. I., & Bach, F. (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Sina Tayarani-Bathaie, S., & Khorasani, K. (2015). Fault detection and isolation of gas turbine engines using a bank of neural networks. *Journal of Process Control*, *36*, 22–41. doi: 10.1016/j.jprocont.2015.08.007

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2014). Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.

VandenHul, S. P. (2002). *Decomposition of the MEWMA statistic* (Ph.D.). University of Northern Colorado, United States. (ISBN: 9780493758718)

Williams, C., & Rasmussen, C. (1995). Gaussian Processes for Regression. In *Advances in Neural Information Processing Systems* (Vol. 8). MIT Press.

Witczak, M. (2007). *Modelling and Estimation Strategies for Fault Diagnosis of Non-Linear Systems*. Berlin, Heidelberg: Springer. doi: 10.1007/978-3-540-71116-2_1