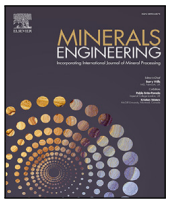
Minerals Engineering 239 (2026) 110086

Contents lists available at ScienceDirect

Minerals Engineering

journal homepage: www.elsevier.com/locate/mineng



# Gaussian process modelling of an industrial flotation bank

Johan Lindqvist [a,*], Khalid Atta [a], J. Derik le Roux [b], Andreas Johansson [a]

[a] Luleå University of Technology, 971 87, Luleå, Sweden
[b] University of Pretoria, Pretoria, Hatfield, 0028, South Africa

ARTICLE INFO

*Keywords:*
Gaussian process
Machine learning
Froth flotation
Dynamic model validation
Mineral processing

ABSTRACT

A control-oriented Gaussian process regression (GPR) model of froth flotation is developed and compared to a previously developed parametric model. The model aims to predict the behaviour of froth flotation, taking into consideration which state variables are available from measurements: air recovery, top of froth bubble size, and pulp level. The framework encodes prior knowledge of a published flotation model. Each state is modelled using a separate GP, with a custom covariance function whose form is given by the flotation model. These kernels capture the interaction between the relevant state variables and manipulated variables. The model aims to balance the complexity required to explain such a complex process with the uncertainty of its instrumentation. To evaluate the ability of the GPR model to capture the process dynamics, the GP model is assessed using an industrial data set, demonstrating its capacity to improve the performance of state prediction. The purpose of the GPR model is to enable supervisory and advanced model-based control.

## 1. Introduction

Froth flotation is a mineral separation technique that has been widely used in the mineral processing industry since its invention. While it remains one of the most prevalent methods of separation, from a control perspective, flotation faces many challenges despite its long history. The process is highly complex and nonlinear due to the interplay of mechanical, chemical, and hydrodynamic phenomena, which introduce strong coupling among process variables, making the system difficult to predict and control. The sparse instrumentation of the process further complicates this effect. Consequently, it remains difficult to develop a control-relevant model for advanced process control in industrial flotation processes. This is primarily evident in the limited success of long-term applications of supervisory control of froth flotation (Bergh and Yianatos, 2011; Oosthuizen et al., 2017).

Two recent models are of particular interest: Quintanilla et al. (2021a,b) and Oosthuizen et al. (2021), taking two different approaches to the problem. The first by Quintanilla et al. (2021b) employs a phenomenological approach, incorporating froth dynamics, an aspect often overlooked in existing froth flotation models. It was subsequently validated using a laboratory-scale flotation cell (Quintanilla et al., 2021a) and later used to demonstrate the effectiveness of economic model predictive control (MPC) to control the laboratory-scale flotation cell (Quintanilla et al., 2023, 2024). While these results are promising, scaling from a well-instrumented experimental setup to an industrial circuit remains a significant challenge. In contrast, the work of Oosthuizen et al. (2021), which was later refined in Oosthuizen (2023), places a high emphasis on the identifiability of its parameters from measurements commonly available in industrial flotation plants. With a simpler structure than the model of Quintanilla et al. (2021b), this model takes an empirical approach to modelling froth dynamics. Building on this, Venter et al. (2024) developed a reduced model based on Oosthuizen (2023), in which the desired mass and gangue are neglected, and their effects on the remaining state variables are captured through parameters. An observer was employed on a dataset collected from an industrial flotation rougher circuit to evaluate state and parameter estimation. Although the reduced model showed promising results, some dynamics remained unmodelled, with faulty predictions leading to error propagation between state variables. Consequently, Venter et al. (2024) concluded that data-driven methods might better capture the unmodelled dynamics.

The adoption of artificial intelligence (AI) systems in the mineral sector is not a new phenomenon. According to Aldrich et al. (1993), enthusiasm for the method can be traced back to at least the 1980s, when various applications of neural networks to simulate mineral processing were published. Additionally, its popularity and widespread use have grown over the past few decades (Aldrich et al., 1993, 2010; McCoy and Auret, 2019; Szmigiel et al., 2024). This development can be attributed to advances in computational and sensor capabilities and declining data storage costs (Colla, 2022). In terms of flotation, recent examples include online base-layer controller tuning (Norlund et al., 2024; Richter et al., 2025) and physics-informed neural network modelling (Nasiri et al., 2025).

* Corresponding author.
*E-mail address:* johan.lindqvist@ltu.se (J. Lindqvist).

https://doi.org/10.1016/j.mineng.2026.110086
Received 15 August 2025; Received in revised form 23 December 2025; Accepted 13 January 2026
Available online 16 January 2026

Gaussian process regression (GPR) is a probabilistic modelling approach that has its roots in geostatistics, where it was originally known as 'kriging'. It was later formalised into the framework of Bayesian nonlinear regression, as demonstrated in the work of Rasmussen and Williams (2006). A key strength of GPR is its flexible kernel structure, which allows it to encode a variety of prior beliefs. More recently, GPR was applied to the identification of dynamical systems, where it showed promise in a wide variety of systems, such as thermal grids (Simonson et al., 2020), gas–liquid separators (Kocijan and Likar, 2008), and bioreactors (Ažman and Kocijan, 2007; Morabito et al., 2022), displaying the ability of the method to represent complex and nonlinear dynamics.

In froth flotation, GPR has previously been used to model the process dynamics and performance. For example, Amankwaa-Kyeremeh et al. (2021) performed a relative neighbourhood component analysis to identify the variables most crucial for predicting rougher copper recovery using GPR. Additionally, Wang et al. (2024) utilised a weighted GPR model to predict concentrate copper grades. Recently, Wang et al. (2025) developed a GP MPC framework using simulated data from the model by Quintanilla et al. (2021a,b). They demonstrate in a simulation the capacity of GPR to be used for MPC.

In recent years, there has been growing interest in extending the capabilities of GPs by integrating prior knowledge from physics and differential equations to inform the model, e.g., Raissi et al. (2017) encode GP priors with the structure of differential operators, and Simonson et al. (2020) developed kernels inspired by differential equations of district heating pipes to simulate thermal grids. The kernels employed in this paper are heavily influenced by the methodology presented in Simonson et al. (2020).

This paper examines the use of GPR modelling of flotation cells as an alternative to parametric frameworks, with kernels formulated based on the dynamic model presented in Venter et al. (2024). The proposed kernels are trained and validated using industrial data as published in Venter et al. (2024). The predictive capacity of the GPR model is compared with that of the parametric model presented in Venter et al. (2024).

## 2. Process model

### 2.1. Flotation model

This work considers the flotation model developed by Venter et al. (2024), which is summarised below. The state variables and parameters are listed in Table 1, and the model forms the basis for constructing the GP kernels following the method of Simonson et al. (2020).

#### 2.1.1. Air recovery model

The air recovery $\alpha_n$ for cell $n$ is formulated to account for the peak in air recovery observed in flotation cells (Hadler and Cilliers, 2009):

$$\frac{\mathrm{d}\alpha_n}{\mathrm{d}t} = \frac{K_{\alpha_{J_g}}\left(J_{g_n} - K_{0,\alpha_{J_{g_n}}} - K_{\alpha_{h_f}} h_{f_n}\right)^2 + \alpha_{OS_n} - \alpha_n}{\lambda_{\mathrm{air}\ n}}, \tag{1}$$

where, $J_{g_n}$ denotes the superficial gas velocity, $\alpha_{OS_n}$ the air recovery offset, $K_{\alpha_{J_g}}$, $K_{0,\alpha_{J_{g_n}}}$, and $K_{\alpha_{h_f}}$ are empirical parameters, and $\lambda_{air_n}$ is the froth residence time given by:

$$\lambda_{air_n} = \frac{h_{f_n}}{J_{g_n}}. \tag{2}$$

froth height depends on the cell height $H_{cell}$ and the pulp level $L_n$:

$$h_{f_n} = 1000(H_{cell} - L_n). \tag{3}$$

**Table 1**
Model parameters and state variables.

| Symbol | Unit | Description |
|---|---|---|
| **Subscripts** | | |
| $t$ | – | Time index |
| $n$ | – | Subscript for flotation cell |
| **State variables** | | |
| $\alpha_n$ | – | Air recovery (see (1)) |
| $D_{BF_n}$ | mm | Top of froth bubble size (see (4)) |
| $L_n$ | m | Pulp level (see (5)) |
| **Inputs** | | |
| $J_{g_n}$ | cm $s^{-1}$ | Superficial gas velocity |
| $v_n$ | – | Valve fraction |
| **Auxiliary variables** | | |
| $h_{f_n}$ | mm | Froth height (see (3)) |
| $\lambda_{\mathrm{air}_n}$ | s | Froth residence time (see (2)) |
| $Q_{C_n}$ | $m^3$ $h^{-1}$ | Concentrate flow-rate (see (6)) |
| $Q_F$ | $m^3$ $h^{-1}$ | Feed flow-rate |
| $Q_{T_n}$ | $m^3$ $h^{-1}$ | Tailings flow-rate (see (7)) |

#### 2.1.2. Bubble size model

The bubble size of the froth has a significant impact on the drainage rate and liquid recovery (Koehler et al., 1999; Neethling and Cilliers, 2003). It can also be indicative of the stability of the froth. The top of froth bubble size $D_{BF_n}$ is described by:

$$\frac{\mathrm{d}D_{BF_n}}{\mathrm{d}t} = \frac{K_{BS_{J_g}} J_{g_n} + K_{BS_\lambda}\lambda_{\mathrm{air}_n} + D_{BF_{OS_n}} - D_{BF_n}}{\lambda_{\mathrm{air}_n}}, \tag{4}$$

where $D_{BF_{OS_n}}$ is the offset of the top of froth bubble size and $K_{BS_{J_g}}$ and $K_{BS_\lambda}$ are empirical parameters.

#### 2.1.3. Pulp level model

The pulp level is modelled using volume balance:

$$\frac{dL_n}{dt} = \frac{Q_{F_n} - Q_{C_n} - Q_{T_n}}{A_n}, \tag{5}$$

where $Q_{F_n}$ is the feed flow into the cell, $Q_{C_n}$ is the concentrate overflowing the lip of the flotation cell, $Q_{T_n}$ is the tailings flow out of the cell, and $A_n$ is the cross-sectional area of the flotation tank. The concentrate flow can be expressed as:

$$\frac{Q_{c_n}}{A_n} \approx \begin{cases} \frac{6.81 J_{g_n}^2}{k_m D_{BF_n}^2}\left(1-\alpha_n\right)\alpha_n & 0 < \alpha_n < 0.5 \\ \frac{6.81 J_{g_n}^2}{4 k_m D_{BF_n}^2} & \alpha_n \geq 0.5, \end{cases} \tag{6}$$

where the parameter $k_m$ accounts for both the fluid properties and the solids recovery. The tailings flow is given by the valve equation:

$$Q_{T_n} = C_v v_n \sqrt{L_n - L_{n+1} - \Delta_{h_n}}, \tag{7}$$

where $C_v$ is the valve parameter, $v_n$ is the valve coefficient, and $\Delta h_n$ is the height difference between adjacent cells.

#### 2.1.4. Data description

The method in this paper is evaluated on an online data set collected from a flotation plant (Steyn and Sandrock, 2021; Venter et al., 2024) shown in Fig. 1. Ground ore is fed into a 7-cell rougher bank via a surge tank, with the resulting concentrate fed into two separate concentrate hoppers. One hopper combines the concentrate flow from the first two cells and the second from the last five. This study focuses on the first two cells, since Steyn and Sandrock (2021) found that they account for the majority of the recovery of the circuit.

Air recovery is obtained by a composite measurement calculated as:

$$\alpha = \frac{w h_{f,0,n} v_{f,0,n}}{J_{g_n} A_n}, \tag{8}$$

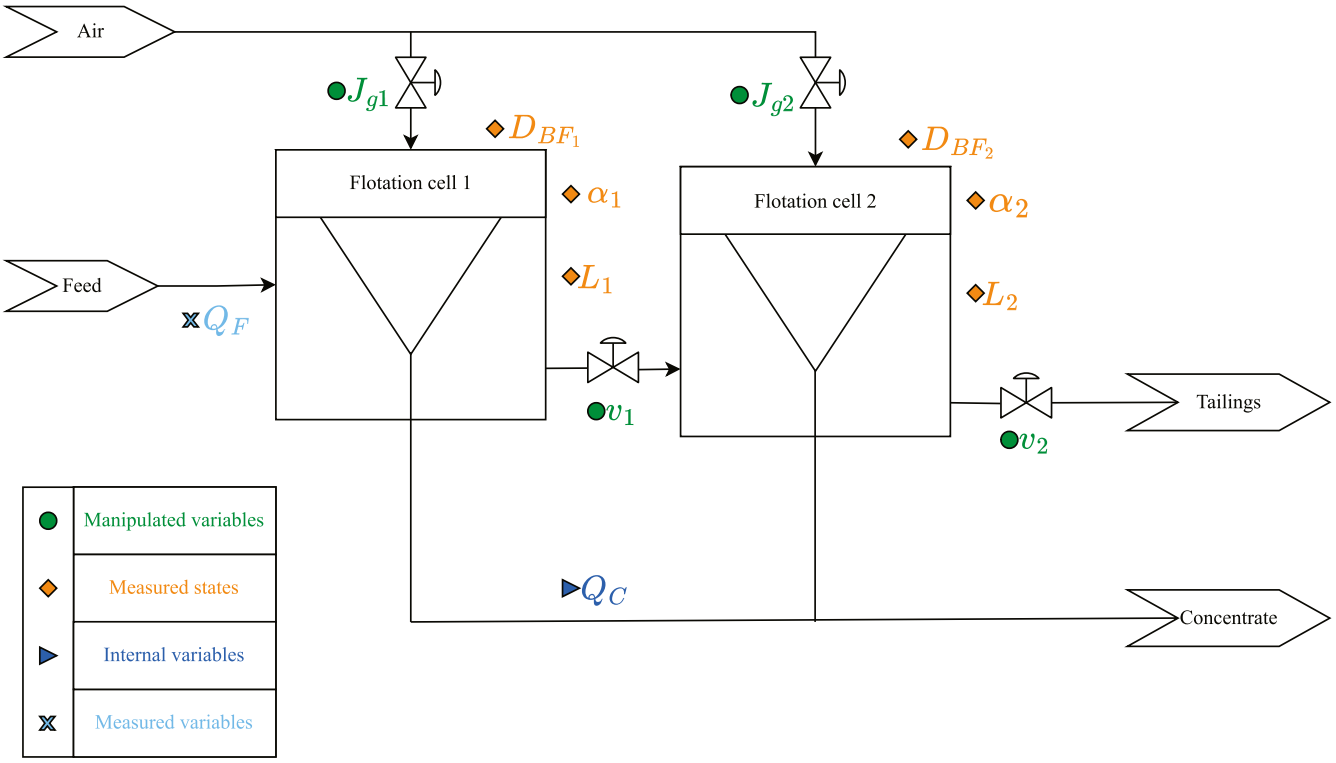



**Fig. 1.** Depiction of the first two cells of a flotation bank.

where $w$ represents the length of the launder lip, $h_{f,0,n}$ is the froth height over the lip, which is measured using a laser sensor, and $v_{f,0,n}$ is the froth overflow velocity, measured with a camera. It is important to note that all of these variables are subject to measurement noise, leading to a resulting air recovery measurement that is both noisy and uncertain. The top of froth bubble size is measured using a camera system. Given that the sample rate is set at 2 min while the circuit behaviour occurs much faster, the measurements are resampled through polynomial fit interpolation. The pulp level is determined using a ball float in conjunction with a sonar plate. For more detailed information about the dataset, readers are referred to Venter et al. (2024).

## 3. Gaussian processes

### 3.1. Gaussian process regression (GPR)

GPR is a probabilistic regression method used to model systems of the following form:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{z}) + \boldsymbol{v}, \tag{9}$$

where $\boldsymbol{y}$ are observations from the underlying function $\boldsymbol{f}(\boldsymbol{z})$, $\boldsymbol{v}$ is measurement noise, and $\boldsymbol{z}$ is the input vector or regressor. A GP is formally defined as a collection of random variables, any finite number of which have a joint Gaussian distribution (Rasmussen and Williams, 2006). This means that a GP is a multivariate Gaussian distribution completely defined by its mean function $m(\boldsymbol{z})$ and covariance function $k(\boldsymbol{z}, \boldsymbol{z}')$:

$$f(\boldsymbol{z}) \sim \mathcal{GP}(m(\boldsymbol{z}), k(\boldsymbol{z}, \boldsymbol{z}')) \tag{10a}$$

$$m(\boldsymbol{z}) = \mathbb{E}[f(\boldsymbol{z})] \tag{10b}$$

$$k(\boldsymbol{z}, \boldsymbol{z}') = \mathbb{E}\left[(f(\boldsymbol{z}) - m(\boldsymbol{z}))(f(\boldsymbol{z}') - m(\boldsymbol{z}'))\right]. \tag{10c}$$

In many cases, the data is assumed to have a zero mean, i.e., $m(\boldsymbol{z}) = 0$. A covariance function can take many forms depending on the nature of the data, allowing for a wide variety of priors in regression. The most common such function is the squared exponential covariance (SQE) function:

$$k(\boldsymbol{z}, \boldsymbol{z}') = \sigma_f^2 \exp\left(-\frac{\Delta \boldsymbol{z}^2}{2l^2}\right), \tag{11}$$

where $\Delta \boldsymbol{z} = (\boldsymbol{z} - \boldsymbol{z}')$, is the distance between two input points $\boldsymbol{z}$ and $\boldsymbol{z}'$. The behaviour of the SQE function is parameterised by the characteristic length scale $l$ and the variance of the signal $\sigma_f^2$, which determine the smoothness of the function and the variation, respectively. Since the kernel can be written as an infinite expansion of basis functions, it can model any nonlinear behaviour, provided there is sufficient data. In other words, it operates as a universal approximator.

In a GP framework, regression consists of placing a GP prior over the function $f(\boldsymbol{z})$ and subsequently formulating a joint distribution between the test points $\boldsymbol{z}^*$ and the training data $\boldsymbol{y}$:

$$p(y \mid \boldsymbol{z}, \boldsymbol{D}, \boldsymbol{\theta}) = \mathcal{N}\Big(y \Big| \boldsymbol{k}_z^\top \left(\boldsymbol{K}_N + \sigma_n^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y}, \\ K_{zz} - \boldsymbol{k}_z^\top \left(\boldsymbol{K}_N + \sigma_n^2 \boldsymbol{I}\right)^{-1} \boldsymbol{k}_z\Big). \tag{12}$$

$\sigma_n^2$ is the variance of the observation noise, and $\boldsymbol{K}_N = \boldsymbol{K}(\boldsymbol{z}, \boldsymbol{z}')$ is the covariance matrix between data points in the regression matrix $\boldsymbol{Z}$ while $\boldsymbol{k}_z = K(\boldsymbol{z}, \boldsymbol{z}^*)$ is the covariance between the training inputs $\boldsymbol{z}$ and the test input $\boldsymbol{z}_*$ and $\boldsymbol{K}_{zz} = \boldsymbol{K}(\boldsymbol{z}^*, \boldsymbol{z}^*)$ is the covariance between test inputs $z_*$. In order to perform predictions, the distribution is conditioned on observations (restricting the posterior distribution to agree with the data). This forms a posterior GP with mean and variance given by (13). These equations are used as the predictive equations for a GP model.

$$\mathbb{E}\left(y^*\right) = \mu\left(\boldsymbol{z}^*\right) = \boldsymbol{k}_z^\top \left(\boldsymbol{K}_N + \sigma_n^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y} \tag{13a}$$

$$\text{var}\left(y^*\right) = \sigma^2\left(\boldsymbol{z}^*\right) = K_{zz} - \boldsymbol{k}_z^\top \left(\boldsymbol{K}_N + \sigma_n^2 \boldsymbol{I}\right)^{-1} \boldsymbol{k}_z \tag{13b}$$

In the context of GPR, learning is achieved by identifying the hyperparameter (in the SQE case, $\boldsymbol{\theta} = [l, \sigma_f]^T$) that is most likely given the data. This can be formulated as finding the hyperparameters $\boldsymbol{\theta}$ that maximise the marginal likelihood $p(\boldsymbol{y}|\boldsymbol{Z}, \boldsymbol{\theta})$ of the GP. Due to numerical

properties, the likelihood is usually logarithmized, giving the following likelihood function to maximise:

$$\log p(f \mid \boldsymbol{Z}, \boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{y}^T(\boldsymbol{K}_N + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{y} - \frac{1}{2}\log|\boldsymbol{K}_N + \sigma_n^2\boldsymbol{I}| - \frac{N}{2}\log 2\pi. \tag{14}$$

Given the inverse of the covariance function, the computational complexity of the GPR is $O(n^3)$. A more detailed description of the derivation and implementation of GPR can be found in Rasmussen and Williams (2006).

### 3.2. Sparse GPR

In order to expand GPR to larger datasets, sparse versions of GPs have been developed, denoted as sparse Gaussian process regression (SGPR), circumventing the considerable time complexity of the classical approach. One such method is the Fully Independent Conditional Approximation (FITC) presented in the work of Snelson and Ghahramani (2005).

In FITC, each observation is assumed conditionally independent given the inducing variables $\overline{\boldsymbol{f}}$ associated with the pseudo inputs called inducing inputs $\bar{\boldsymbol{z}}_{\boldsymbol{m}}$, where the size $M$ of the inducing set is less than the real set $N$. This allows the FITC method to approximate the true GPR kernel functions, reducing the time complexity to $O(nm^2)$. Importantly, the inducing training set $\overline{\boldsymbol{D}}$ does not need to be a subset of the true observations, and the locations of the inducing inputs can be optimised alongside the hyperparameters $\boldsymbol{\theta}$, simplifying implementation. The full dataset likelihood is given by:

$$p(\boldsymbol{y} \mid \boldsymbol{Z}, \overline{\boldsymbol{Z}}, \overline{\boldsymbol{f}}) = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{K}_{NM}\boldsymbol{K}_M^{-1}\overline{\boldsymbol{f}}, \boldsymbol{\Lambda} + \sigma_n^2\boldsymbol{I}\right), \tag{15}$$

where $\boldsymbol{K}_M$ is the covariance matrix between the inducing inputs, $\boldsymbol{K}_{NM}$ denotes the covariance between training and inducing inputs, and $\boldsymbol{\Lambda} = \mathrm{diag}(K_{nn} - \boldsymbol{k}_n^{\mathrm{T}}\boldsymbol{K}_M^{-1}\boldsymbol{k}_n)$ captures the variance correction introduced by the independence assumption. Since the pseudo data will behave similarly to the real data, the inducing targets are assigned a Gaussian prior $p(\overline{\boldsymbol{f}}|\overline{\boldsymbol{Z}}) = \mathcal{N}(\overline{\boldsymbol{f}}|0, \boldsymbol{K}_M)$, which provides a compact representation of the latent process. Consequently, the prior and likelihood will then, according to Bayes' rule, give the following posterior distribution:

$$p(\overline{\boldsymbol{f}} \mid \boldsymbol{D}, \overline{\boldsymbol{Z}}) = \mathcal{N}(\overline{\boldsymbol{f}} \mid \boldsymbol{K}_M\boldsymbol{Q}_M^{-1}\boldsymbol{K}_{MN}(\boldsymbol{\Lambda} + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{y}, \boldsymbol{K}_M\boldsymbol{Q}_M^{-1}\boldsymbol{K}_M) \tag{16}$$

where $\boldsymbol{Q}_M = \boldsymbol{K}_M + \boldsymbol{K}_{MN}(\boldsymbol{\Lambda} + \sigma_n^2\boldsymbol{I})^{-1}\boldsymbol{K}_{NM}$. The optimisation of the hyperparameters $\boldsymbol{\theta}$ and the pseudo inputs $\overline{\boldsymbol{Z}}$ can be found by classical gradient descent algorithms by maximising the likelihood:

$$p(\boldsymbol{y} \mid \boldsymbol{Z}, \overline{\boldsymbol{Z}}, \boldsymbol{\theta}) = \int p(\boldsymbol{y} \mid \boldsymbol{Z}, \overline{\boldsymbol{Z}}, \overline{\boldsymbol{f}})p(\overline{\boldsymbol{f}} \mid \overline{\boldsymbol{Z}})\mathrm{d}\overline{\boldsymbol{f}} = \mathcal{N}\left(\boldsymbol{y} \mid \boldsymbol{0}, \boldsymbol{K}_{NM}\boldsymbol{K}_M^{-1}\boldsymbol{K}_{MN} + \boldsymbol{\Lambda} + \sigma_n^2\boldsymbol{I}\right). \tag{17}$$

The predictive distribution is obtained by marginalising the latent function values at the inducing inputs. Specifically, The inducing targets $\overline{\boldsymbol{f}}$ are integrated over the approximate posterior $p(\overline{\boldsymbol{f}} \mid \boldsymbol{D}, \overline{\boldsymbol{Z}})$ and the likelihood $p(y \mid \boldsymbol{z}, \overline{\boldsymbol{Z}}, \overline{\boldsymbol{f}})$ giving the predictive distribution:

$$p\left(y^* \mid \boldsymbol{z}^*, \boldsymbol{D}, \overline{\boldsymbol{Z}}\right) = \int p\left(y^* \mid \boldsymbol{z}^*, \overline{\boldsymbol{Z}}, \overline{\boldsymbol{f}}\right)p(\overline{\boldsymbol{f}} \mid \boldsymbol{D}, \overline{\boldsymbol{Z}})\mathrm{d}\overline{\boldsymbol{f}} = \mathcal{N}\left(y^* \mid \mu_*, \sigma_*^2\right). \tag{18}$$

This then gives the predictive equations:

$$\mu_* = \boldsymbol{k}_*^{\mathrm{T}}\boldsymbol{Q}_M^{-1}\boldsymbol{K}_{MN}\left(\boldsymbol{\Lambda} + \sigma_n^2\boldsymbol{I}\right)^{-1}\boldsymbol{y} \tag{19a}$$

$$\sigma_*^2 = K_{**} - \boldsymbol{k}_*^{\mathrm{T}}\left(\boldsymbol{K}_M^{-1} - \boldsymbol{Q}_M^{-1}\right)\boldsymbol{k}_* + \sigma_n^2. \tag{19b}$$

### 3.3. Gaussian process state space models

A dynamical system can be described with a discrete-time nonlinear Gaussian state space model (20), where $\boldsymbol{v}_p$ is the process noise, $\boldsymbol{v}_m$ is measurement noise, and $t$ denotes the time index:

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}\left(\boldsymbol{x}_t, \boldsymbol{u}_t\right) + \boldsymbol{v}_p \tag{20a}$$

$$\boldsymbol{y}_t = \boldsymbol{g}\left(\boldsymbol{x}_t\right) + \boldsymbol{v}_m. \tag{20b}$$

Such a system can be modelled using GPR, yielding the Gaussian process state-space model (GPSSM), which can be broadly categorised into two cases. In the first case, measurements of the state variables and outputs are available, whereas in the second case, only measurements of the output are available.

For this study, the state variables are assumed to be observable and measured. Therefore, the GPSSM is represented in a reduced form where a separate GP model is used for each state $\boldsymbol{x}_t$:

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}\left(\boldsymbol{x}_t, \boldsymbol{u}_t\right), \quad t \in \mathbb{N} \tag{21a}$$

$$\boldsymbol{f}\left(\boldsymbol{x}_t, \boldsymbol{u}_t\right) \sim \mathcal{GP}\left(\boldsymbol{m}\left(\boldsymbol{z}_t\right), \boldsymbol{k}\left(\boldsymbol{z}_t, \boldsymbol{z}_t'\right)\right) \tag{21b}$$

In the literature of GPs, the variable $\boldsymbol{x}$ is usually denoted as the input or regressor. However, since in the context of dynamical systems, this variable is reserved for the state of a system, this paper will refer to $\boldsymbol{z}$ as the regressor to the GP. Because there are both input signals and state variables that act as input to the GP, the regressor is defined as $\boldsymbol{z}_t = [\boldsymbol{x}_t^T, \boldsymbol{u}_t^T]^T$.

As expected, the industrial dataset contains considerable measurement noise. Since a GP setting assumes the inputs are noise-free, the dataset is filtered using an exponentially weighted moving average filter. This enables one-step-ahead regression and maximum-likelihood methods during training. Furthermore, since each state variable is measured, the corresponding GP can be trained independently, treating the other state variables as known inputs during training.

### 3.4. Prior knowledge

GPR provides a natural way to incorporate prior knowledge into the model through the prior. In classical GPR, this is achieved through three main mechanisms: the mean function, the covariance (kernel) function, and the selection of training data. The mean and covariance functions encode assumptions about the underlying system behaviour, while the data selection informs the posterior knowledge of the process. Furthermore, the kernel can be modified to impose different structural properties. This follows from the property that the sum of two GP kernels is itself a valid kernel, enabling the use of different covariance functions across dimensions. The resulting additive decomposition can be used to structure the components of a GP model.

## 4. GP modelling of the flotation circuit

### 4.1. Kernel structures

The flotation model in Section 2.1 is leveraged to encode the relationship between the state variables. Since it is assumed that the behaviour of the dynamics is smooth and that a flexible prior is required, the squared exponential covariance function will be used as the basis. A separate SQE function will be used to model each relation. A comparison of the impact of different kernel types, as in Simonson et al. (2020), is considered outside the scope of this study. Based on

(1), the kernel for air recovery is defined as:

$$k_{\alpha_n}(z) = \sigma_1^2 \exp\left(-\frac{1}{2}\left(\frac{\Delta J_{g_n}^2}{l_1^2} + \frac{\Delta \alpha_n^2}{l_2^2}\right)\right) + \sigma_2^2 \exp\left(-\frac{1}{2}\left(\frac{\Delta J_{g_n}^2}{l_3^2} + \frac{\Delta h_{f_n}^2}{l_4^2}\right)\right) + \sigma_3^2 \exp\left(-\frac{1}{2}\left(\frac{\Delta J_{g_n}^2}{l_5^2} + \frac{\Delta h_{f_n}^2}{l_6^2} + \frac{\Delta \alpha_n^2}{l_7^2}\right)\right) + \sigma_4^2 \exp\left(-\frac{1}{2}\left(\frac{\Delta J_{g_n}^2}{l_8^2}\right)\right), \tag{22}$$

where $\Delta$ denotes the difference between two variables. The kernel for pulp level is based on (5) and is given by:

$$k_{L_n}(z) = \sigma_1^2 \exp\left(-\frac{1}{2}\left(\frac{\Delta J_{g_n}^2}{l_1^2} + \frac{\Delta D_{BF_n}^2}{l_2^2} + \frac{\Delta \alpha_n^2}{l_3^2}\right)\right) + \sigma_2^2 \exp\left(-\frac{1}{2}\left(\frac{\Delta v_n^2}{l_4^2} + \frac{\Delta L_n^2}{l_5^2}\right)\right) + \sigma_3^2 \exp\left(-\frac{1}{2}\left(\frac{\Delta L_n^2}{l_6^2}\right)\right), \tag{23}$$

Lastly, the bubble size defined in (4) leads to the following kernel:

$$k_{D_{BF_n}}(z) = \sigma_1^2 \exp\left(-\frac{1}{2}\left(\frac{\Delta J_{g_n}^2}{l_1^2} + \frac{\Delta D_{BF_n}^2}{l_2^2} + \frac{\Delta h_{f_n}^2}{l_3^2}\right)\right) + \sigma_2^2 \exp\left(-\frac{1}{2}\left(\frac{\Delta D_{BF_n}^2}{l_4^2} + \frac{\Delta h_{f_n}^2}{l_5^2}\right)\right) + \sigma_3^2 \exp\left(-\frac{1}{2}\left(\frac{\Delta J_{g_n}^2}{l_6^2} + \frac{\Delta h_{f_n}^2}{l_7^2}\right)\right). \tag{24}$$

### 4.2. Evaluation metrics

To compare the results with those found in Venter et al. (2024), the same evaluation criteria are used. The root mean square error (RMSE) metric is a common method used to evaluate regression performance:

$$RMSE = \sqrt{RMSE_s^2 + RMSE_u^2}, \tag{25}$$

where $RMSE_s$ is the estimation error due to a modelling bias (systematic error) and $RMSE_u$ is the error as a result of random noise (unsystematic error). The ratio of these variables can be used to quantify the extent to which the systematic component exceeds the unsystematic component. As such, it can be used to indicate the performance of the prediction:

$$S/U = \frac{RMSE_s}{RMSE_u}. \tag{26}$$

The systematic and unsystematic errors are given by:

$$RMSE_s = \sqrt{N^{-1}\sum_{i=1}^{N}(\hat{P}_i - O_i)^2}, \tag{27a}$$

$$RMSE_u = \sqrt{N^{-1}\sum_{i=1}^{N}(P_i - \hat{P}_i)^2}, \tag{27b}$$

where $\hat{P}_i = a + bO_i$ is a linear best fit of the data $O_i$ given by the predictions $\hat{P}_i$.

While the prediction of the parametric model of Venter et al. (2024), is a point prediction, the GPR produces a predictive distribution. To adequately evaluate the performance of the GPR, the Mean Standardised Log Loss (MSLL) as described in Kocijan (2016) is used. MSLL is a measure that compares the predictive distribution of the GP model with a baseline Gaussian model, using the mean and variance of the training targets. Formally, the $MSLL$ is defined as:

$$MSLL = \frac{1}{2N}\sum_{i=1}^{N}\left(\ln(\sigma_i^2) + \frac{(y_i - \mu_i)^2}{\sigma_i^2}\right) - \frac{1}{2N}\sum_{i=1}^{N}\left(\ln(\sigma_y^2) + \frac{(y_i - \bar{y})^2}{\sigma_y^2}\right), \tag{28}$$

where $y_i$ are the observed test values, $\mu_i$ and $\sigma_i^2$ denote the predictive mean and variance obtained from the GPR, and $\bar{y}$ and $\sigma_y^2$ represent the mean and variance of the training targets, respectively. Negative MSLL values correspond to more accurate and better-calibrated predictive distributions, values near zero indicate performance similar to the baseline Gaussian model, and positive values indicate poorer performance than the baseline.

## 5. Results

The method was evaluated on the industrial data set published in Venter et al. (2024). Due to the abundance of steady-state data, the industrial data was split into sections containing clearer step responses and excitations. This was done to ensure sufficient excitation in the training data set to adequately capture the system's dynamics.

The model was first trained on one-step-ahead predictions on the training set and then evaluated on the test set using multi-step-ahead predictions to assess its ability to predict states. This study primarily examines the benefits of a data-driven approach to modelling compared to a parametric model. As such, the mean of the GP is most appropriate to compare. As a result, uncertainty propagation is not used as it has a negligible impact on the prediction mean. Moreover, the variance was primarily used to determine whether the model deviated from the training region, which is already covered by the predicted variance. Additionally, since predictions are made in a multi-state system, interconnections would need to be considered for the uncertainty propagation. Given the minimal improvements in predictive mean accuracy, the added information from uncertainty propagation is not considered justifiable compared to the computational cost. The horizon is set to 10 min to compare it with the result found in Venter et al. (2024).

The multi-step predictions for cells 1 and 2 are shown in Figs. 2 and 3, respectively. The GPR model prediction is compared with the parametric model prediction of Venter et al. (2024). The GPR model captures the main trends in the variables.

An RMSE analysis is presented in Table 2, which compares the RMSE values for the prediction of each state with those reported in Venter et al. (2024). Table 2 shows that the models achieve similar RMSE values and a drastic improvement in $L_n$ for the GPR model. Additionally, the model by Venter et al. (2024) appears to deviate more regularly than the GPR model. Consequently, the GPR seems capable of capturing dynamics that the parametric model does not model.

The model of Venter et al. (2024) offered better predictions for bubble size. This highlights a limitation of data-driven modelling and its reliance on a descriptive dataset. In this dataset, the bubble size had a significantly larger sample time, which necessitated interpolation. Consequently, some data points do not accurately reflect the system. Therefore, it is not surprising that a method that enforces more structure through its form, less reliant on the dataset, would have more capability to capture the underlying dynamics. These observations suggest there may be interest in a hybrid approach that incorporates both modelling paradigms, though this is a topic for future research.

In order to evaluate the ability of the GPR model to predict over longer time horizons, a 20-minute prediction was performed for both cells. The results are shown in Figs. 4–5. The model is able to capture the overall trend of the data. However, for the 2nd cell, there are instances of large deviations for bubble size and pulp level. It is important to note that the state variables display an increased confidence interval

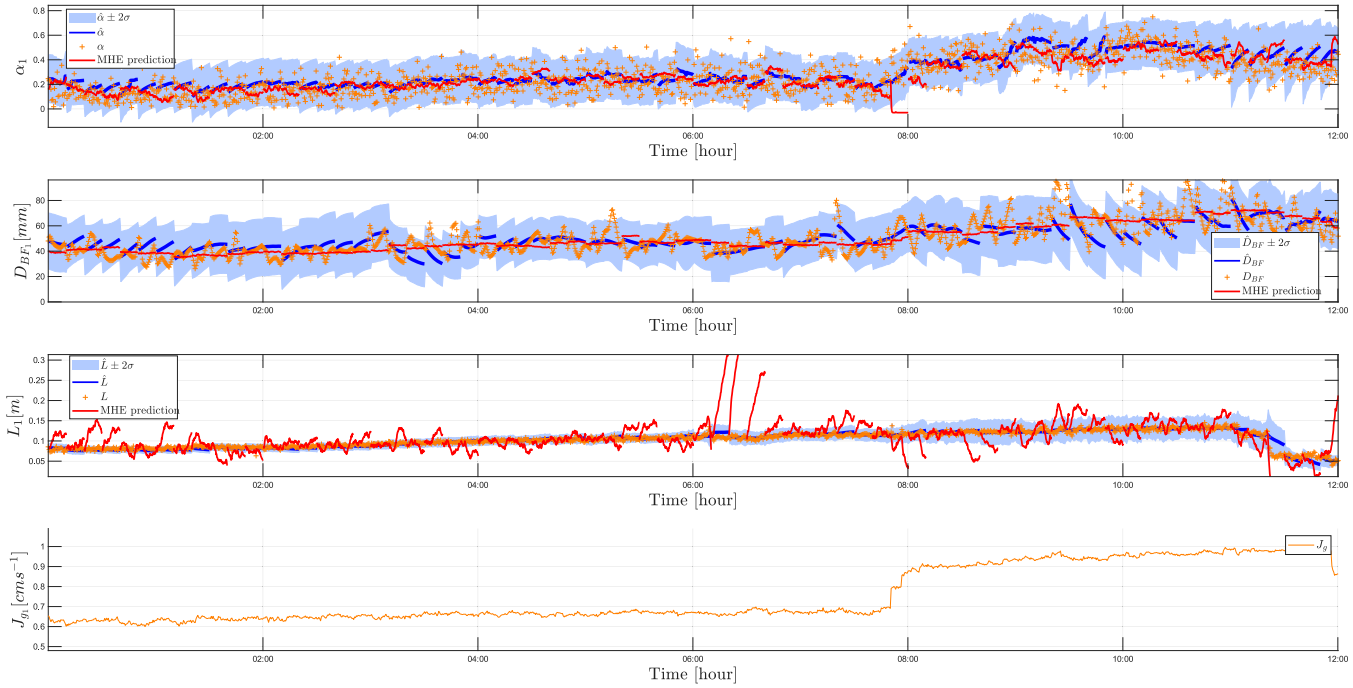



**Fig. 2.** Multi-step prediction of the air recovery $\hat{\alpha}$, pulp level $\hat{L}$, and bubble size $\hat{D}_{BF}$ of cell 1, with the input Jg shown at the bottom. Values are offset from the true values. (Legend: blue shading is the confidence interval, blue line is the GPR mean prediction, orange markers are the process data, and red is the state prediction found in Venter et al. (2024).) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

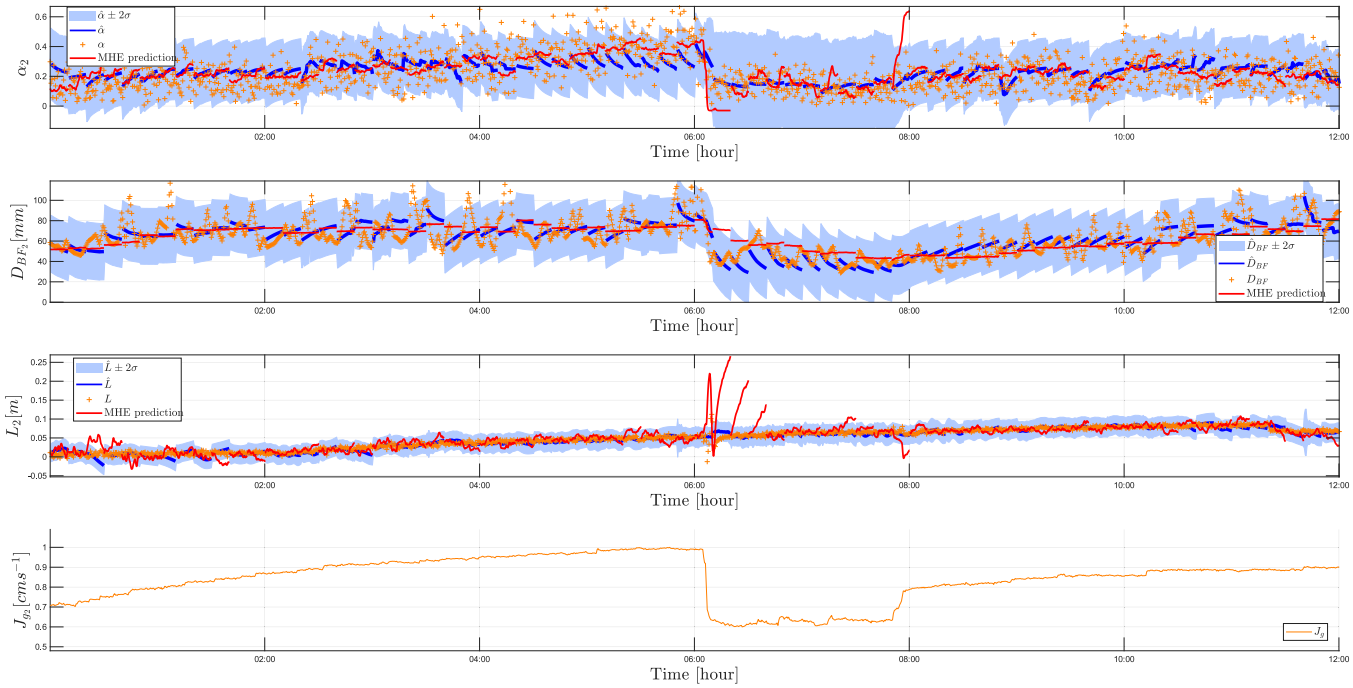



**Fig. 3.** Multi-step prediction of the air recovery $\hat{\alpha}$, pulp level $\hat{L}$, and bubble size $\hat{D}_{BF}$ of cell 2, with the input Jg shown at the bottom. Values are offset from the true values. (Legend: blue shading is the confidence interval, blue line is the GPR mean prediction, orange markers are the process data, and red is the state prediction found in Venter et al. (2024).) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

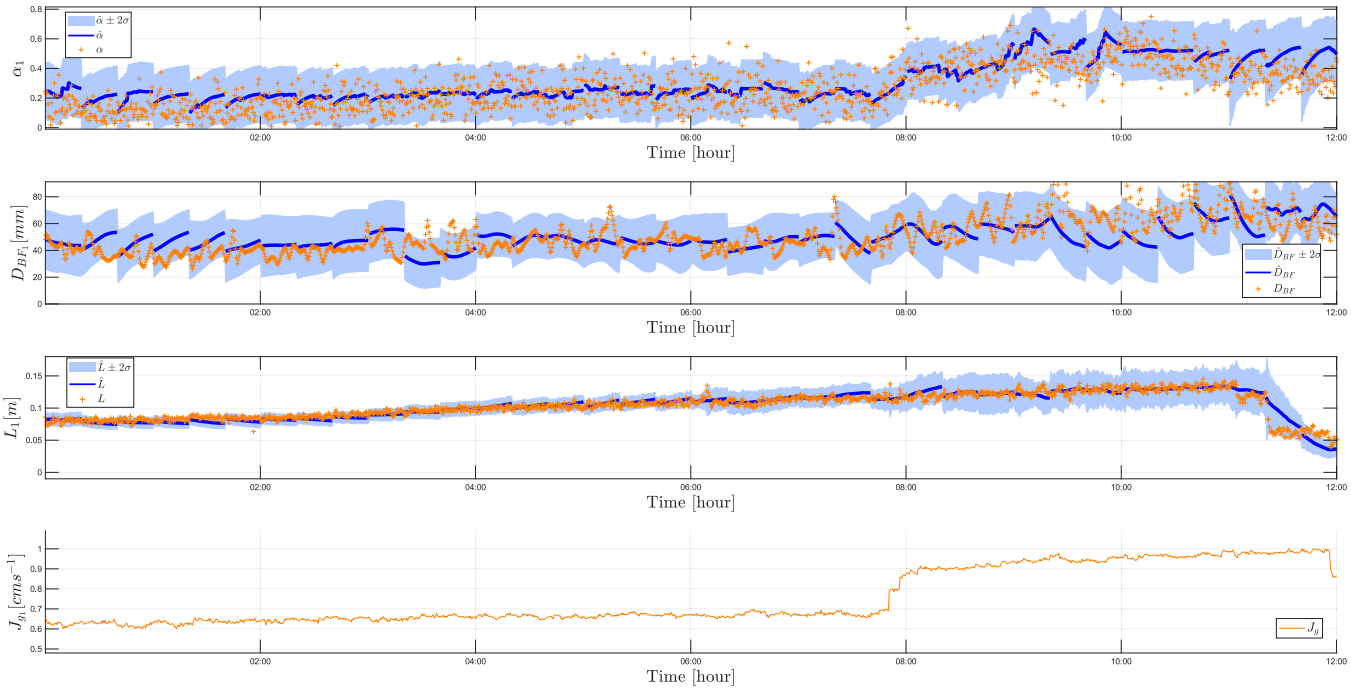



**Fig. 4.** 20 min multi-step ahead prediction of the air recovery $\hat{\alpha}$, pulp level $\hat{L}$, and bubble size $\hat{B}$ of cell 1, with the input $J_g$ shown at the bottom. Values are offset from the true values.

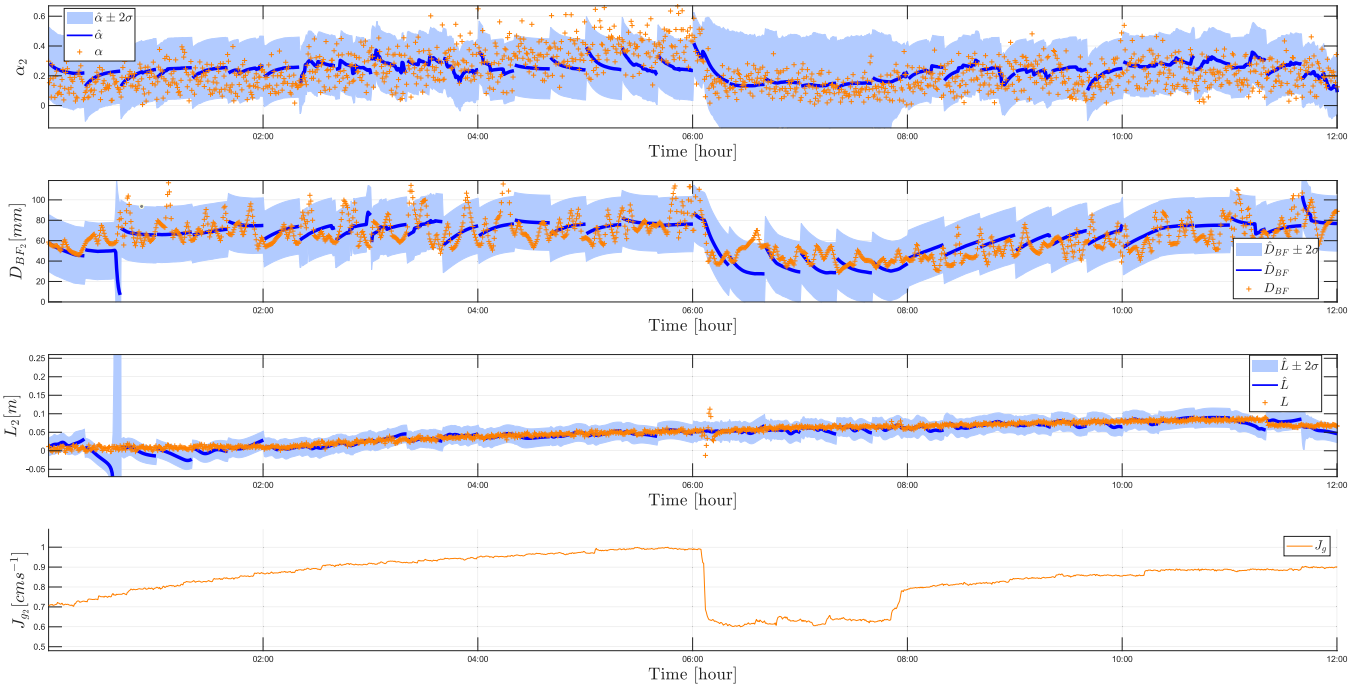



**Fig. 5.** 20 min multi-step ahead prediction of the air recovery $\hat{\alpha}$, pulp level $\hat{L}$, and bubble size $\hat{B}$ of cell 2, with the input $J_g$ shown at the bottom. Values are offset from the true values.

when deviating far from the true value, indicating a lack of trust in the prediction.

Regarding the predictive distribution of the GPR model, its MSLL values calculated using (28) are shown in Table 2. Here, the observation that the pulp level prediction performed well is evident in the MSLL value, indicating the promise of GP-based level prediction. One interesting aspect is the near-zero MSLL value for air recovery, which suggests there is only a moderate improvement in $\alpha_2$ as compared to the baseline. This is not surprising given the substantial noise in the measurement of this variable. This indicates a need for a more noise-robust technique for both air recovery and bubble size, such as the GP-FNARX method described in Frigola and Rasmussen (2013).

**Table 2**
Prediction error metrics for 10-minute prediction horizon.

| State | Venter et al. (2024) | | GPR model | | |
|---|---|---|---|---|---|
| | RMSE | S/U | RMSE | S/U | MSLL |
| $\alpha_1$ | 0.11 | 0.65 | 0.104 | 0.372 | −0.518 |
| $\alpha_2$ | 0.11 | 0.99 | 0.107 | 0.0865 | −0.126 |
| $D_{BF_1}$ | 8.70 | 0.75 | 10.1 | 0.0636 | −0.443 |
| $D_{BF_2}$ | 13.65 | 1.22 | 13.2 | 0.255 | −0.313 |
| $L_1$ | 0.06 | 0.18 | 0.0066 | 0.287 | −2.02 |
| $L_2$ | 0.03 | 0.15 | 0.0077 | 0.293 | −1.215 |

Since the purpose of the model is for use in predictive control in a feedback loop, the most important aspect of the model is to capture the direction of variation of variables. As such, the GPR model demonstrates potential for use as a predictive model, as indicated by low RMSE values and negative MSLL scores. It is regarded as suitable for MPC applications as in Oosthuizen et al. (2024) and Wang et al. (2025). In addition, the GPR approach is simpler and more attractive compared to the state estimation approach in Venter et al. (2024).

Because of the limited perturbation in the available dataset, the confidence interval ($\pm 2\sigma$) spans a large fraction of the signal range. It is clear that industrial datasets, which represent steady-state operation with a relatively low level of excitation and high levels of noise, are challenging to work with. High excitation in datasets such as step tests for system identification is expected to improve predictive capability, but such data is generally expensive to generate, as it is highly disruptive to the plant. Evaluation of the model with different data sets remains for future work.

## 6. Discussion and conclusions

The paper assessed the capacity of a GPR framework to model air recovery, the top of the froth bubble size, and pulp level in the first two cells of an industrial flotation circuit. The modelling results were compared to the parametric model of Venter et al. (2024) based on the same dataset. The GP model of the process, specifically with the flexibility of the GP prior, was able to capture dynamics that the parametric model did not. Furthermore, training and validation indicate that the GPR model can effectively handle the noisy nature of this process. However, the GPR model could not fully capture all the process dynamics, suggesting room for further improvement, particularly regarding bubble size.

Nonetheless, the GPR model demonstrates sufficient ability to handle the uncertainty of industrial data while capturing the main dynamics of the process. The model shows decent capability to predict the trend of the flotation circuit over time and could thus be used in a feedback control method for supervisory control.

## CRediT authorship contribution statement

**Johan Lindqvist:** Writing – review & editing, Writing – original draft, Methodology. **Khalid Atta:** Writing – review & editing, Supervision. **J. Derik le Roux:** Writing – review & editing, Supervision. **Andreas Johansson:** Writing – review & editing, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

This study is partially funded by the European Union's Horizon Europe research and innovation program under grant agreement No 101091885 (MINE.IO Project).

## Data availability

The data that has been used is confidential.

## References

Aldrich, C., Marais, C., Shean, B., Cilliers, J., 2010. Online monitoring and control of froth flotation systems with machine vision: A review. Int. J. Miner. Process. 96, 1–13. http://dx.doi.org/10.1016/j.minpro.2010.04.005.

Aldrich, C., Van Deventer, J., Van der Walt, T., Reuter, M., 1993. Recent advances in the simulation of mineral processing circuits using artificial intelligence.

Amankwaa-Kyeremeh, B., Zhang, J., Zanin, M., Skinner, W., Asamoah, R., 2021. Feature selection and Gaussian process prediction of rougher copper recovery. Miner. Eng. 170, 107041. http://dx.doi.org/10.1016/j.mineng.2021.107041.

Ažman, K., Kocijan, J., 2007. Application of Gaussian processes for black-box modelling of biosystems. ISA Trans. 46, 443–457. http://dx.doi.org/10.1016/j.isatra.2007.04.001.

Bergh, L.G., Yianatos, J.B., 2011. The long way toward multivariate predictive control of flotation processes. J. Process Control 21, 226–234. http://dx.doi.org/10.1016/j.jprocont.2010.11.001.

Colla, V., 2022. A big step ahead in metal science and technology through the application of artificial intelligence. IFAC-PapersOnLine 55, 1–6. http://dx.doi.org/10.1016/j.ifacol.2022.09.234.

Frigola, R., Rasmussen, C.E., 2013. Integrated pre-processing for Bayesian nonlinear system identification with Gaussian processes. In: 52nd IEEE Conference on Decision and Control. pp. 5371–5376. http://dx.doi.org/10.1109/CDC.2013.6760734.

Hadler, K., Cilliers, J., 2009. The relationship between the peak in air recovery and flotation bank performance. Miner. Eng. 22, 451–455. http://dx.doi.org/10.1016/j.mineng.2008.12.004.

Kocijan, J., 2016. Modelling and control of dynamic systems using Gaussian process models. http://dx.doi.org/10.1007/978-3-319-21021-6.

Kocijan, J., Likar, B., 2008. Gas–liquid separator modelling and simulation with Gaussian-process models. Simul. Model. Pract. Theory 16, 910–922. http://dx.doi.org/10.1016/j.simpat.2008.05.007.

Koehler, S.A., Hilgenfeldt, S., Stone, H.A., 1999. Liquid flow through aqueous foams: the node-dominated foam drainage equation. Phys. Rev. Lett. 82, 4232.

McCoy, J., Auret, L., 2019. Machine learning applications in minerals processing: A review. Miner. Eng. 132, 95–109. http://dx.doi.org/10.1016/j.mineng.2018.12.004.

Morabito, B., Pohlodek, J., Kranert, L., Espinel-Ríos, S., Findeisen, R., 2022. Efficient and simple Gaussian process supported stochastic model predictive control for bioreactors using HILO-MPC. IFAC-PapersOnLine 55, 922–927. http://dx.doi.org/10.1016/j.ifacol.2022.07.562.

Nasiri, M., Iqbal, S., Särkkä, S., 2025. Physics-informed machine learning for grade prediction in froth flotation. Miner. Eng. 227, 109297. http://dx.doi.org/10.1016/j.mineng.2025.109297.

Neethling, S., Cilliers, J., 2003. Modelling flotation froths. Int. J. Miner. Process. 72, 267–287. http://dx.doi.org/10.1016/S0301-7516(03)00104-2.

Norlund, F., Tammia, R., Hägglund, T., Soltesz, K., 2024. Linear-quadratic level control for flotation through reinforcement learning. IFAC-PapersOnLine 58, 799–804. http://dx.doi.org/10.1016/j.ifacol.2024.08.435.

Oosthuizen, D.J., 2023. A Dyndamic Flotation Model for Real-Time Control and Optimisation (Ph.D. thesis). University of Pretoria.

Oosthuizen, D.J., Craig, I.K., Jämsä-Jounela, S.L., Sun, B., 2017. On the current state of flotation modelling for process control. IFAC-PapersOnLine 50, 19–24. http://dx.doi.org/10.1016/j.ifacol.2017.12.004.

Oosthuizen, D.J., le Roux, J.D., Craig, I.K., 2021. A dynamic flotation model to infer process characteristics from online measurements. Miner. Eng. 167, 106878. http://dx.doi.org/10.1016/j.mineng.2021.106878.

Oosthuizen, D.J., le Roux, J.D., Craig, I.K., 2024. Non-linear model predictive control to improve the mineralogical efficiency of flotation circuits. IFAC-PapersOnLine 58, 1–6. http://dx.doi.org/10.1016/j.ifacol.2024.09.281, 7th IFAC Workshop on Mining, Mineral and Metal Processing MMM 2024.

Quintanilla, P., Navia, D., Neethling, S.J., Brito-Parada, P.R., 2023. Economic model predictive control for a rougher froth flotation cell using physics-based models. Miner. Eng. 196, 108050. http://dx.doi.org/10.1016/j.mineng.2023.108050.

Quintanilla, P., Navia, D., Neethling, S., Brito-Parada, P., 2024. Experimental implementation of an economic model predictive control for froth flotation. Comput. Aided Chem. Eng. 53, 1759–1764. http://dx.doi.org/10.1016/B978-0-443-28824-1.50294-5.

Quintanilla, P., Neethling, S.J., Mesa, D., Navia, D., Brito-Parada, P.R., 2021a. A dynamic flotation model for predictive control incorporating froth physics. Part II: Model calibration and validation. Miner. Eng. 173, 107190. http://dx.doi.org/10.1016/j.mineng.2021.107190.

Quintanilla, P., Neethling, S.J., Navia, D., Brito-Parada, P.R., 2021b. A dynamic flotation model for predictive control incorporating froth physics. Part I: Model development. Miner. Eng. 173, 107192. http://dx.doi.org/10.1016/j.mineng.2021.107192.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2017. Machine learning of linear differential equations using Gaussian processes. J. Comput. Phys. 348, 683–693. http://dx.doi.org/10.1016/j.jcp.2017.07.050.

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. The MIT Press.

Richter, A.V., le Roux, J.D., Craig, I.K., 2025. Bayesian optimization for automatic tuning of a MIMO controller of a flotation bank. J. Process Control 147, 103388. http://dx.doi.org/10.1016/j.jprocont.2025.103388.

Simonson, J., Atta, K.T., Birk, W., 2020. Probabilistic modeling of thermal grids using Gaussian processes. In: 2020 59th IEEE Conference on Decision and Control. CDC, pp. 36–41. http://dx.doi.org/10.1109/CDC42340.2020.9304284.

Snelson, E., Ghahramani, Z., 2005. Sparse Gaussian processes using pseudo-inputs. In: Weiss, Y., Schölkopf, B., Platt, J. (Eds.), Advances in Neural Information Processing Systems. MIT Press.

Steyn, C., Sandrock, C., 2021. Causal model of an industrial platinum flotation circuit. Control Eng. Pract. 109, 104736. http://dx.doi.org/10.1016/j.conengprac.2021.104736.

Szmigiel, A., Apel, D.B., Skrzypkowski, K., Wojtecki, L., Pu, Y., 2024. Advancements in machine learning for optimal performance in flotation processes: A review. Minerals 14, http://dx.doi.org/10.3390/min14040331.

Venter, J.L., le Roux, J.D., Craig, I.K., 2024. State and parameter estimation of a dynamic froth flotation model using industrial data. Miner. Eng. 219, 109059. http://dx.doi.org/10.1016/j.mineng.2024.109059.

Wang, Y., del Río Chanona, E.A., Quintanilla, P., 2025. Gaussian process nonlinear model predictive control for online partially observable systems: An application to froth flotation. Ind. Eng. Chem. Res. 64, 13307–13322. http://dx.doi.org/10.1021/acs.iecr.5c00660.

Wang, Z., He, D., Nie, H., 2024. Operational optimization of copper flotation process based on the weighted Gaussian process regression and index-oriented adaptive differential evolution algorithm. Chin. J. Chem. Eng. 66, 167–179. http://dx.doi.org/10.1016/j.cjche.2023.09.010.