RESEARCH ARTICLE

Maximum Likelihood Estimation for an SAG Mill Model Utilizing Physical Available Measurements

ANGEL L. CEDEÑO[1,2]**, MARÍA CORONEL**[3]**, (Member, IEEE),**
RAFAEL ORELLANA[4]**, (Member, IEEE), PATRICIO VARAS**[1]**,**
RODRIGO CARVAJAL[5]**, (Member, IEEE), BORIS I. GODOY**[6]**,**
AND JUAN C. AGÜERO[1]**, (Member, IEEE)**

[1]Departamento de Electrónica, Universidad Técnica Federico Santa María, Valparaíso 2390123, Chile
[2]Advanced Center for Electrical and Electronic Engineering, Valparaíso 2390123, Chile
[3]Departamento de Electricidad, Universidad Tecnológica Metropolitana, Santiago 8330383, Chile
[4]Departamento de Ingeniería Eléctrica, Universidad de Santiago de Chile (USACH), Santiago 9170022, Chile
[5]Escuela de Ingeniería Eléctrica, Pontificia Universidad Católica de Valparaíso, Valparaíso 2340025, Chile
[6]Metropolitan College, Boston University, Boston, MA 02215, USA

Corresponding author: Angel L. Cedeño (angel.cedeno@sansano.usm.cl)

ABSTRACT In this paper, we have proposed a new paradigm for modeling of SAG mills. Typically, important parameters found in the modeling of such processes are described as state-space system model rather than unknown parameters. Here, we propose to estimate the system model using the maximum likelihood approach. Additionally, we propose using a new measurement that has not been considered in other modeling approaches. The benefits of our proposal are illustrated via numerical simulations. The results demonstrate that incorporating this new measurement within the framework of maximum likelihood estimation improves the accuracy of estimating the unknown parameters.

INDEX TERMS SAG mills, system identification, em algorithm, maximum likelihood, filtering, modeling.

I. INTRODUCTION

System identification of real-world processes is imperative for developing effective process control strategies and precise fault detection and diagnosis methods. A benefit of such identification is enhanced plant stability and performance, which can be achieved by integrating the model with advanced model-based process control, outperforming traditional control methods [7], [35], [40], [43].

In copper mining, two key processes are the grinding and flotation. These processes require a significant amount of electrical energy, particularly in the operation of the semi-autogenous (SAG) mill. An SAG mill is equipment used to mix minerals from the crushing stage with water and

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad S. Khan.

lime. Within the mill, the minerals are reduced in size through kinetic interactions with other rocks and steel balls (hence the name semi-autogenous) [41]. Several studies have been conducted to optimize SAG mill performance aimed at reducing traditional energy consumption in the mining industry, including the use of advanced control strategies and new technologies such as high-pressure grinding rolls (HPGR) and stirred mills [1], [8]. An alternative to the energy reduction using traditional sources is the inclusion of renewable energy, such as solar panels, which play a key role in the profitability of the whole copper mining process [32], [34], however, renewable energy discussion is out of the scope of the paper.

Figure 1 illustrates the open-loop SAG mill circuit used in the mining industry [28]. The mill is fed with three streams: the water flow rate (*MIW*), the ore feed rate (*MFO*), and the
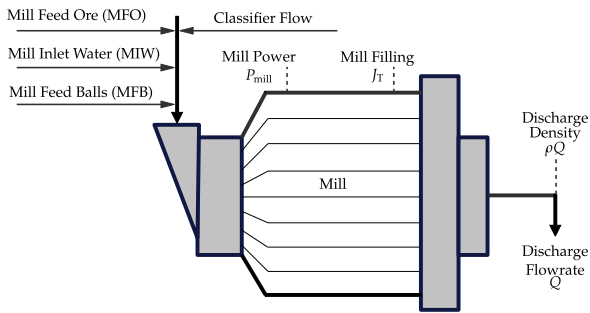
FIGURE 1. A semi-autogenous grinding mill operation.

steel balls feed rate (*MFB*), all of which combine to aid in ore breakage and form the charge in the mill, comprising a mixture of grinding media and slurry [23], [24]. Grinding media refers to the steel balls and large rocks used to break down the ore, while the slurry comprises a mixture of water and all the ore material. The fraction of the mill volume filled with charge is represented by J_T [29]. The overall process works as follows: The mill is rotated by a motor along its longitudinal axis, lifting the charge by the inner liners on the walls of the mill to a certain height. After being lifted, it cascades down, only to be lifted again by the liners as the mill rotates. If the rotational speed is high enough, the material in the charge will go airborne after reaching the top of its travel on the mill shell. The highest point where material leaves the mill shell is known as the shoulder of the charge. Airborne particles follow a parabolic path, reaching a maximum height known as the head before making contact again with the mill charge at the bottom of the mill. The cascading motion of the charge causes the ore to break down via impact, abrasion, and attrition [29]. The mill grind is the fraction of material in the mill discharge below the specified size and indicates the efficiency of the mill to break the ore down [12]. The motor (turning the mill) power draw (P_{mill}) indicates the kinetic and potential energy given to the charge [28]. The slurry is discharged through an end-discharge grate with an aperture size that limits the particle size of the discharged slurry. The slurry flow-rate at the mill discharge is given by Q, and it is assumed that the in-mill slurry density is equal to the discharge slurry density (ρ_Q).

One of the primary challenges for modeling (and control) an SAG mill process is the insufficient measurements available to estimate the necessary states and parameters in the model. In general, the number of real-time measurements available is significantly lower than the size of the state vector that needs to be measured [43]. In [4], [5], and [6], the problem of inferential modeling, state estimation, and model validation of an SAG mill is addressed using the extended Kalman filter. However, the results are obtained based on a set of measurements that are not physically measurable in the process. Along the same line of reasoning, another study is presented in [24], where states and parameters of a simplified SAG model are estimated. This study, described

in [41], focuses on the estimation of SAG mill states using a simplified model, emphasizing the use of commonly available measurements not typically utilized. For SAG mills, recent works have concentrated on the application of neural networks and machine learning techniques to estimate mill throughput or identify operational regions within the SAG mill, primarily for control purposes (see [13] and [26] and the references therein). The comparative study presented in [13] focuses on state estimation techniques for control, particularly modern predictive control. However, neural network and machine learning approaches are not suitable for identification techniques. They do not utilize the actual system model but define one based on some training and output data, generally not based on the physics properties of the actual system of interest. Other works have also focused on accurate estimation of state variables, such as load volume, but from measurements of physical variables and parameters, such as mill filling (see e.g. [19]). However, these methods include direct measurement that rely on the mill not working for a period of time, on images that require further analysis, or on precise knowledge of the mill's geometry (see [19] and the reference therein).

On the other hand, the difficulty and importance of estimating non-linear parametric systems are widely recognized in the control and system identification community, see e.g. [30], [39], and [46]. Usually, in nonlinear modeling, researchers focus on a specific class of systems, such as those characterized by Volterra kernels [27], molecular biology [22], and Hammerstein-Wiener structures [45], to mention just a few.

The maximum likelihood (ML) framework for estimating unknown parameters has been the workhorse in the system identification community when considering uncertainty in the dynamical models. The ML framework has been applied to a range of problems, including continuous-time, discrete-time, linear, and nonlinear dynamic models. To solve the ML estimation problem, the Expectation-Maximization (EM) algorithm is typically the preferred tool.

In this paper, we apply ML estimation to obtain unknown parameters in an SAG mill non-linear system model using available measurements in practice, specifically, the total SAG mill weight. The main contributions of this work are:

- We present an iterative methodology for estimating the unknown parameters of an SAG mill, leveraging both the ML method and the EM algorithm. Specifically, we employ the extended and the unscented Kalman filters and smothers, comparing their performance in addressing the ML estimation task.
- Within the EM algorithm framework, to address the inherent high nonlinearity of the model, we propose computing the auxiliary function (also known as the surrogate function) through a Taylor-based second-order approximation of the log-likelihood function.
- To validate this approach, we conduct comprehensive Monte Carlo simulations to demonstrate its robustness and accuracy in the estimations.

The remainder of the paper is as follows: In Section II, a general description of the SAG system model is presented. In Section III, the ML problem for the system of interest is stated. In Section IV, an EM-based algorithm is presented. A numerical example with simulated data is presented in Section V. Finally, in Section VI we present our conclusions.

II. MODELING AN SAG MILL

A. GENERAL SYSTEM DESCRIPTION

In this section, we provide a description of the state-space system model of an SAG mill. We consider the Hulbert SAG mill model presented in [20] and utilized in [9], [23], [24], and [31]. The SAG mill model is based on (1), which describes the mass balance using a continuous-time state-space model as follows:

$$\dot{\mathbf{x}} = \mathbf{u} - \mathbf{V}_{\text{out}} + \Phi, \tag{1}$$

where we define the following vectors

$$\mathbf{x} = \begin{bmatrix} x^w \\ x^s \\ x^f \\ x^r \\ x^b \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u^w \\ u^s \\ u^f \\ u^r \\ u^b \end{bmatrix}, \quad \mathbf{V}_{\text{out}} = \begin{bmatrix} V_{wo} \\ V_{so} \\ V_{fo} \\ 0 \\ 0 \end{bmatrix}, \quad \Phi = \begin{bmatrix} 0 \\ RC \\ FP \\ -RC \\ -BC \end{bmatrix}. \tag{2}$$

Here V_{so}, V_{fo} and V_{wo} represent the discharge of solids, fines, and water from the mill, respectively; RC denotes rock consumption, BC ball consumption, and FP fines production. The elements of the input vector, u^w, u^s, u^f, u^r, and u^b, represent the water flow-rate, solids, fines, rocks, and balls, respectively. The elements of the state vector x^w, x^s, x^f, x^r, and x^b, represent water volume, solids, rocks, and balls in the mill. The mill inflow-rate is described as follows:

$$u^w = \text{MIW}, \tag{3}$$

$$u^s = \frac{\text{MFS}}{\rho_0}(1 - \alpha_r), \tag{4}$$

$$u^f = \frac{\text{MFS}}{\rho_0}\alpha_f, \tag{5}$$

$$u_r = \frac{\text{MFS}}{\rho_0}\alpha_r \tag{6}$$

$$u^b = \frac{\text{MFB}}{\rho_b}, \tag{7}$$

where MIW represents the water flow-rate, MFS denotes the ore feed-rate to the mill, and MFB stands for the steel balls feed-rate, ρ_0 and ρ_b are the densities of the feed ore and steel balls, respectively, and α_r and α_f represent the mass fraction of rocks and fines in MFS.

Similarly, the water discharge $\mathbf{V}_{\text{out}}$, solids, and fines are given by:

$$V_{w0} = \varphi d_H x^w (\zeta_w), \quad \zeta_w = x^w/(x^s + x^w), \tag{8}$$

$$V_{s0} = \varphi d_H x^w (\zeta_s), \quad \zeta_s = x^s/(x^s + x^w), \tag{9}$$

$$V_{f0} = \varphi d_H x^w (\zeta_f), \quad \zeta_f = x^f/(x^s + x^w), \tag{10}$$

where d_H is a discharge constant. The rheology factor, φ, is used to model the slurry effect in the performance of the mill as follows [23]:

$$\varphi = \left[\max\left(0, 1 - \left(\frac{1}{\varepsilon_{sv}} - 1\right)\frac{x^s}{x^w}\right)\right]^{0.5}, \tag{11}$$

where ε_{sv} represents the maximum fraction of solids.

On the other hand, the fines production (FP) and rock and ball consumption (RC) and (BC)–in the population balance in (1)– are described as follows:

$$RC = \frac{\varphi P_{\text{mill}} x^r}{\rho_0 \kappa_r (x^r + x^s)}, \tag{12}$$

$$BC = \frac{\varphi P_{\text{mill}} x^b}{\kappa_b (\rho_0(x^r + x^s) + \rho_b x^b)}, \tag{13}$$

$$FP = \frac{P_{\text{mill}}}{\rho_0 \left\{\kappa_f \left[1 + \alpha_{\kappa_f}\left(\frac{V_{\text{LOAD}}}{v_{\text{mill}}} - v_{P_{max}}\right)\right]\right\}}. \tag{14}$$

Here P_{mill} represents the power draw of the mill (see e.g. [20] and [24]), κ_r, κ_b, κ_f denote the rock abrasion factor, steel abrasion factor, and the power needed per tonne of fines produced, respectively, α_{κ_f} represents the fractional change in power per fines produced per change in the fractional filling of the mill, v_{mill} denotes the volume of the mill, $v_{P_{max}}$ represents the fraction of the mill filled for maximum power draw, and V_{LOAD} represents the mill charge volume in terms of the volumetric states in the mill.

B. A REDUCED HULBERT MODEL

Substituting (3)–(14) in (1) we obtain the Hulbert model [20]. Notice that this model is described using five (5) state variables and several parameters, in addition to the parameters resulting from the choice of the mill's power draw P_{mill}, see e.g. [20] and [24]. Nevertheless, it is possible to obtain a reduced model, referred to as the Hulbert model, by considering the following assumptions:

- The state, x^f, which represents the volume of fines, does not affect the dynamics of the remaining state variables.
- There is instrumentation available capable of measuring input flows, thus with these estimates, the input flow equations–to the mill–can be omitted and assumed to be known [4].
- The mill's power draw, P_{mill}, can be assumed to be known. Moreover, abrasions of the rocks and steel balls (κ_r and κ_b, respectively) are inherent material properties that can be known in advance [4].

Then, the Hulbert model can be expressed as follows:

$$\dot{\mathbf{x}} = f_c(\mathbf{x}) + \mathbf{u}, \tag{15}$$

where the state and input vectors are given by

$$\dot{\mathbf{x}} = \begin{bmatrix} x^w \\ x^s \\ x^r \\ x^b \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u^w \\ u^s \\ u^r \\ u^b \end{bmatrix}. \tag{16}$$

and the nonlinear function $f_c(\mathbf{x})$ is defined as

$$f_c(\mathbf{x}) = \begin{bmatrix} -\varphi d_H x^w (\zeta_w) \\ -\varphi d_H x^w (\zeta_s) + \dfrac{\varphi P_{\text{mill}} x^r}{\rho_0 \kappa_r (x^r + x^s)} \\ -\dfrac{\varphi P_{\text{mill}} x^r}{\rho_0 \kappa_r (x^r + x^s)} \\ -\dfrac{\varphi P_{\text{mill}} x^b}{\kappa_b (\rho_0 (x^r + x^s) + \rho_b x^b)} \end{bmatrix}. \tag{17}$$

Equation (15) is an SAG system model with four state variables, and two (unknown) parameters to be estimated: the discharge constant, d_H, and the maximum fraction of solids, ε_{sv}.

C. SAG MODEL OUTPUTS

One of the challenges to analyzing an SAG mill is the limited useful information that can directly be obtained in practice, primarily because there is a lack of instrumentation available to measure internal variables within the mill. Recent studies have focused on developing methodologies for detecting and measuring internal variables, such as acoustic emissions, and residence time distribution [18], [33]. Other studies have explored the use of advanced modeling techniques to optimize SAG mill control, and to improve overall ball mill performance [36], [37]. These studies highlight the ongoing efforts to improve the analysis and control of SAG mills. For this study, we consider that the model of the output is given by:

$$\mathbf{y} = h_c(\mathbf{x}), \tag{18}$$

where, under the assumption of the measurement of the total SAG mill weight, W_c, is available. For the Hulbert model in (15) we have:

$$h_c(\mathbf{x}) = \begin{bmatrix} W_c \\ V_{\text{LOAD}} \end{bmatrix} = \begin{bmatrix} x^w + \rho_0 (x^s + x^r) + \rho_b x^b + W_{\text{mill}} \\ x^w + x^s + x^r + x^b, \end{bmatrix}, \tag{19}$$

where W_{mill} the empty mill weight, ρ_{rb} the density related with the state x^{rb}.

D. DISCRETE-TIME SYSTEM MODEL FOR THE SAG MILL

Continuous-time systems can be converted to discrete-time systems using various discretization techniques. However, in the case of non-linear continuous-time systems, it is not always feasible to have an exact representation in discrete time. Instead, an approximation method can be employed to obtain a sampled-model from the continuous-time system.

In the case of the SAG mill, a non-linear continuous-time system, its model needs to be discretized since the measurements are taken in discrete time. One common method to obtain a discrete-time model from a continuous-time system is the forward Euler approximation [11]. For instance, for a continuous variable $\mathbf{x}$, the forward Euler approximation can be written as:

$$\dot{\mathbf{x}} \approx \frac{\mathbf{x}_{t+1} - \mathbf{x}_t}{\Delta}, \tag{20}$$

where t is the sample index, and Δ is the sampled period.[1] Then, the discrete-time version of the SAG models are:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \Delta f_c(\mathbf{x}_t) + \Delta \mathbf{u}_t, \tag{21}$$

$$\mathbf{y}_t = h_c(\mathbf{x}_t). \tag{22}$$

III. MAXIMUM LIKELIHOOD ESTIMATION PROBLEM
A. PROBLEM FORMULATION

Maximum Likelihood (ML) is a method to estimate system parameters using the information provided by the observations. These observations are considered realizations of stochastic variables [25]. In general, several ML estimation algorithms have been developed for discrete-time models of dynamical systems. In this paper, we consider an SAG mill state-space model described, in general, as follows:

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \theta) + \mathbf{w}_t, \tag{23}$$

$$\mathbf{y}_t = h(\mathbf{x}_t, \mathbf{u}_t, \theta) + \mathbf{v}_t, \tag{24}$$

where $\mathbf{x}_t \in \mathbb{R}^{n_x \times 1}$ is the state variable, $\mathbf{u}_t \in \mathbb{R}^{n_u \times 1}$ and $\mathbf{y}_t \in \mathbb{R}^{n_y \times 1}$ denote the observed input signal and output signal, respectively. Furthermore, $\theta \in \mathbb{R}^{n_\theta}$ is a vector of (unknown) parameters that specifies the non-linear mappings, $f(\cdot)$ and $h(\cdot)$. Finally, $\mathbf{w}_t$ and $\mathbf{v}_t$ are independent and identically distributed random processes with zero mean and unknown variances $\mathbf{Q}$ and $\mathbf{R}$, respectively. Because of the random nature of the processes noise, $\mathbf{w}_t$, and the measurement noise, $\mathbf{v}_t$, the system model in (23) can be represented with the following stochastic description:

$$\mathbf{x}_{t+1} \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t, \beta), \tag{25}$$

$$\mathbf{y}_t \sim p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{u}_t, \beta), \tag{26}$$

where $p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t, \beta)$ is the probability density function (PDF) that describes the system model dynamics for given values of $\mathbf{x}_t$, $\mathbf{u}_t$ and β, and $p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{u}_t, \beta)$ is the PDF that describes the output measurements.

Problem 1: The problem addressed in this paper is obtaining the ML estimate, $\hat{\beta}_{ML}$, of the vector of parameters

$$\beta = [\theta^T \ \text{vec}\{\mathbf{Q}\}^T \ \text{vec}\{\mathbf{R}\}^T]^T, \tag{27}$$

utilizing N measurements of the input response, $\mathbf{u}_{1:N} = [\mathbf{u}_1 \ \dots \ \mathbf{u}_N]$, and the output response, $\mathbf{y}_{1:N} = [\mathbf{y}_1 \ \dots \ \mathbf{y}_N]$.[2]

With this in mind, the ML estimation problem can be formulated as follows:

$$\hat{\beta}_{ML} = \arg \max_{\beta} \mathcal{L}(\beta), \tag{28}$$

where $\mathcal{L}(\beta)$ is the likelihood function defined as

$$\mathcal{L}(\beta) = p(\mathbf{y}_1, \dots, \mathbf{y}_N | \mathbf{u}_{1:N}, \beta). \tag{29}$$

Notice that $p(\mathbf{y}_1, \dots, \mathbf{y}_N | \mathbf{u}_{1:N}, \beta)$ is the joint PDF of the observed output data given the input data, $\mathbf{u}_{1:N}$, and the vector

---

[1] $\mathbf{x}_i$ is the variable $\mathbf{x}$ at time $i\Delta$.
[2] $\mathbf{x}^T$ denotes the transpose of the variable $\mathbf{x}$ and vec $\{\cdot\}$ denotes the vectorization of a matrix.

of parameters, β. To compute this PDF, Bayes' rule can be used to decompose the joint density function:

$$\mathcal{L}(\beta) = p(\mathbf{y}_1) \prod_{t=2}^{N} p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{u}_{1:N}, \beta), \quad (30)$$

which ultimately transforms into (31) because of the monotonic nature of the logarithmic function. Thus, the optimization problem in (28) can be solved by minimizing the log-likelihood function, $\ell(\beta)$, as

$$\hat{\beta}_{\text{ML}} = \arg \max_{\beta} \ell(\beta),$$
$$= \arg \max_{\beta} \{\log[p(\mathbf{y}_1)]$$
$$+ \sum_{t=2}^{N} \log[p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{u}_{1:N}, \beta)]\}. \quad (31)$$

B. COMPUTING THE PREDICTION DENSITY

In order to compute the log-likelihood function in (31), we need to compute the prediction density based on the law of total probability and the Markov nature of (25) and (26) as follows:

$$p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{u}_{1:N}, \beta) = \int p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:t-1}) d\mathbf{x}_t, \quad (32)$$

Then, using Bayes' rule, we obtain:

$$p(\mathbf{x}_t | \mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t | \mathbf{y}_{1:t-1})}, \quad (33)$$

$$p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t}) = \int p(\mathbf{x}_{t+1} | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{1:t}) d\mathbf{x}_t. \quad (34)$$

Notice that (33)–(34) provide a recursive formulation to compute the prediction density $p(\mathbf{y}_t | \mathbf{y}_{1:t-1}, \mathbf{u}_{1:N}, \beta)$ as well the predicted and filtered state densities $p(\mathbf{x}_t | \mathbf{y}_{1:t-1})$, $p(\mathbf{x}_t | \mathbf{y}_{1:t})$. This recursive formulation can be used to solve the ML estimation problem in (31). In the linear and Gaussian case, the Kalman filter can be employed to solve the estimation problem. In the nonlinear case in (23), solutions based on Particle filtering (PF) and particle smoothing (PS) have been used to obtain the state estimates (see e.g. [39]). For completeness, the smoothing equation, see e.g. [38], is defined by:

$$p(\mathbf{x}_t | \mathbf{y}_{1:N}) = p(\mathbf{x}_t | \mathbf{y}_{1:t}) \int \frac{p(\mathbf{x}_{t+1} | \mathbf{y}_{1:N}) p(\mathbf{x}_{t+1} | \mathbf{x}_t)}{p(\mathbf{x}_{t+1} | \mathbf{y}_{1:t})} d\mathbf{x}_{t+1}. \quad (35)$$

Traditional methods can be used to optimize the log-likelihood function in (31). However, due to its systematic approach, the EM algorithm is the preferred tool. The EM algorithm maximizes the likelihood function by defining an auxiliary function, which can be optimized by using Quasi-Newton methods (see e.g. [15] and [16]).

IV. AN ITERATIVE ALGORITHM FOR SAG MILL MODEL ESTIMATION

The EM algorithm has gained widespread popularity for its effectiveness in identifying linear and non-linear dynamic systems in both time and frequency domains, see e.g. [3], [14], [17], and [44]. In this section, we will show the development of an EM-based estimation algorithm that can effectively solve the ML problem posed in (28).

A. THE EM-BASED ALGORITHM FORMULATION

The EM algorithm is a two-step iterative procedure to solve the ML estimation problem in the presence of latent variables, see e.g. [10]. The goal is to optimize an auxiliary function instead of the log-likelihood function. This auxiliary function is computed from the definition of the likelihood function utilizing the observed data, $\mathbf{y}_{1:N}$, and a *hidden* variable (state variable), $\mathbf{x}_{1:N+1}$. In other words, the likelihood function is determined when using the complete data.[3] Hence, the EM algorithm is given by:

$$\mathcal{Q}(\beta, \hat{\beta}^{(i)}) = \mathbb{E}\left\{\log\{p(\mathbf{x}_{1:N+1}, \mathbf{y}_{1:N} | \beta)\} | \mathbf{y}_{1:N}, \hat{\beta}^{(i)}\right\}, \quad (36)$$
$$\hat{\beta}^{(i+1)} = \arg \max_{\beta} \mathcal{Q}(\beta, \hat{\beta}^{(i)}), \quad (37)$$

where $\hat{\beta}^{(i+1)}$ is the current estimate, $p(\mathbf{x}_{1:N+1}, \mathbf{y}_{1:N} | \beta)$ is the joint PDF of $\mathbf{y}_{1:N}$ and $\mathbf{x}_{1:N+1}$, and $\mathcal{Q}(\beta, \hat{\beta}^{(i)})$ is the auxiliary function. Here, $\mathbb{E}\{\cdot|\cdot\}$ denotes the expected operator. Notice that (36) and (37) correspond to the *E-step* and *M-step* of the EM algorithm, respectively [10].

In order to develop the iterative algorithm in (36) and (37), we first need to obtain an expression for $\mathcal{Q}(\beta, \hat{\beta}^{(i)})$.

Lemma 2: Consider the vector of parameters to be estimated $\beta = [\theta^T \ \text{vec}\{\mathbf{Q}\} \ \text{vec}\{\mathbf{R}\}]^T$ for the dynamic system in (23) and (24). Then, the complete log-likelihood function, $\ell_C(\beta) = \log\{p(\mathbf{x}_{1:N+1}, \mathbf{y}_{1:N} | \beta)\}$, in (36) is given by

$$\ell_C(\beta) = -\frac{N}{2} \log\{|2\pi\mathbf{Q}|\} - \frac{N}{2} \log\{|2\pi\mathbf{R}|\}$$
$$- \frac{1}{2} \sum_{t=1}^{N} [\mathbf{x}_{t+1} - f(\mathbf{x}_t, \mathbf{u}_t, \theta)]^T \mathbf{Q}^{-1} [\mathbf{x}_{t+1} - f(\mathbf{x}_t, \mathbf{u}_t, \theta)]$$
$$- \frac{1}{2} \sum_{t=1}^{N} [\mathbf{y}_t - h(\mathbf{x}_t, \mathbf{u}_t, \theta)]^T \mathbf{R}^{-1} [\mathbf{y}_t - h(\mathbf{x}_t, \mathbf{u}_t, \theta)]$$
$$- \frac{1}{2} \log\{|2\pi\mathbf{P}_1|\} - \frac{1}{2} (\mathbf{x}_1 - \mu_1)^T \mathbf{P}_1^{-1} (\mathbf{x}_1 - \mu_1), \quad (38)$$

where $\mathbf{x}_1$ is Gaussian distributed with mean value μ_1 and variance $\mathbf{P}_1$, that is, $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{x}_1; \mu_1, \mathbf{P}_1)$.

Proof: See [2] □

Computing the auxiliary function for the EM algorithm is a difficult task because of the nonlinear functions $f(\mathbf{x}_t, \mathbf{u}_t, \theta)$ and $h(\mathbf{x}_t, \mathbf{u}_t, \theta)$ in (23) and (24), respectively. In this work, we propose to approximate the complete log-likelihood

---

[3]The complete data $\{\mathbf{x}_{1:N+1}, \mathbf{y}_{1:N+1}\}$ corresponds to the set defined by the observed data $\mathbf{y}_{1:N}$, and the unobserved data $\mathbf{x}_{1:N}$.

Algorithm 1 Expectation-Maximization Algorithm

1 **Input:** Initial guess $\beta^{(0)}$ so that $\ell(\hat{\beta}^{(0)})$ is finite.
2 Set $i = 0$.
3 **while** *Until convergence* **do**
4   **Expectation step (E-step):**
5   Compute $\tilde{\mathcal{Q}}(\beta, \hat{\beta}^{(i)})$ according to (39).
6   **Maximization step (M-step):**
7   Solve $\hat{\beta}^{(i+1)} = \arg\max_{\beta} \tilde{\mathcal{Q}}(\beta, \hat{\beta}^{(i)})$
8   Set $i = i + 1$
9 **end**
10 **Output:** The parameter vector estimate $\hat{\beta}$.

---

function in (38), by using a second-order Taylor series. From Lemma 2, the auxiliary function for our proposed EM-based algorithm can be computed using the following:

Lemma 3: Consider the dynamic system represented in the state-space form in (23)-(24), the incomplete data $\mathbf{y}_{1:N}$, and the complete data $\{\mathbf{x}_{1:N+1}, \mathbf{y}_{1:N}\}$, then the Taylor-based second-order approximation of the auxiliary function around the point $\hat{\mathbf{x}}_{1:N+1}$ is given by

$$\tilde{\mathcal{Q}}(\beta, \hat{\beta}^{(i)}) = \mathbb{E}\left\{ \ell_C(\beta) + \mathcal{G}_\beta^T \mathbf{d}_N + \frac{1}{2}\mathbf{d}_N^T \mathcal{H}_\beta \mathbf{d}_N \Big| \mathbf{y}_{1:N}, \hat{\beta}^{(i)} \right\},$$
(39)

where $\mathcal{G}_\beta = \nabla\ell_C(\beta)$ and $\mathcal{H}_\beta = \nabla^2\ell_C(\beta)$ are the gradient and the hessian of the likelihood function $\ell_C(\beta)$, respectively, evaluated at $\hat{\mathbf{x}}_{1:N+1} = \mathbb{E}\left\{\mathbf{x}_{1:N+1}|\mathbf{y}_{1:N}, \hat{\beta}^{(i)}\right\}$, and $\mathbf{d}_N$ is defined as $\mathbf{d}_N = \mathbf{x}_{1:N+1} - \hat{\mathbf{x}}_{1:N+1}$

Proof: Directly using the second-order Taylor approximation of a function $f(\mathbf{x})$ around the point $\mathbf{x}_0$, i.e. $f(\mathbf{x}) \approx f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^T \mathbf{G}_0 + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}_0(\mathbf{x} - \mathbf{x}_0)$, where $\mathbf{G}_0 = \nabla f(\mathbf{x})|_{\mathbf{x}_0}$ and $\mathbf{G}_0 = \nabla^2 f(\mathbf{x})|_{\mathbf{x}_0}$. $\square$

Finally, the resultant EM-based procedure is summarized in the Algorithm 1. Please note that to calculate the expectation in the auxiliary function (39), it is essential to implement filtering and smoothing algorithms in order to obtain the smoothed distribution $p(\mathbf{x}_{1:N+1}|\mathbf{y}_{1:N}, \hat{\beta}^{(i)})$. In the following, we present two alternative methods for computing these necessary probability density functions recursively.

B. EXTENDED KALMAN FILTER AND SMOOTHER

The extended Kalman filter (EKF) (and the corresponding extended Kalman smoother (EKS)) is an extension of the standard Kalman filter to nonlinear state-space systems considering the process and measurement of Gaussian noises. The idea of the EKF is to build a linear approximation around a state estimation by using a Taylor series expansion. This approximation allows us to rewrite the nonlinear system (23)-(24) into a linear time-varying system as follows:

$$\mathbf{x}_{t+1} = \mathbf{A}_t \mathbf{x}_t + \mathbf{B}_t + \mathbf{w}_t,$$
(40)

$$\mathbf{y}_t = \mathbf{C}_t \mathbf{x}_t + \mathbf{D}_t + \mathbf{v}_t,$$
(41)

Algorithm 2 Extended Kalman Filter

1 **Input:** The distribution of the initial condition $p(\mathbf{x}_1)$,
    e.g. $\hat{\mathbf{x}}_{0|1} = \mu_1$ and $\Sigma_{0|1} = \mathbf{P}_1$.
2 **for** $t = 1$ *to* N **do**
3   Compute the Kalman gain using:
    $\mathbf{K}_t = \Sigma_{t|t-1}\mathbf{C}_t^T(\mathbf{R} + \mathbf{C}_t\Sigma_{t|t-1}\mathbf{C}_t^T)^{-1}$.
4   **Measurement Update:**
5   Compute the filtering state $\hat{\mathbf{x}}_{t|t}$ according to
    $\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - h(\hat{\mathbf{x}}_{t|t-1}, \mathbf{u}_t))$.
6   Compute the covariance matrix $\Sigma_{t|t}$ according to
    $\Sigma_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{C}_t)\Sigma_{t|t-1}$.
7   **Time Update:**
8   Compute the predicted state $\hat{\mathbf{x}}_{t+1|t}$ according to
    $\hat{\mathbf{x}}_{t+1|t} = f(\hat{\mathbf{x}}_{t|t}, \mathbf{u}_t)$.
9   Compute the covariance matrix $\Sigma_{t+1|t}$ according
    to $\Sigma_{t+1|t} = \mathbf{Q} + \mathbf{A}_t\Sigma_{t|t}\mathbf{A}_t^T$.
10 **end**
11 **Output:** The Filtered PDFs
    $p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}, \Sigma_{t|t})$ and the predicted PDFs
    $p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_{t+1}; \hat{\mathbf{x}}_{t+1|t}, \Sigma_{t+1|t})$ for
    $t = 1, \dots, N$.

Algorithm 3 Extended Kalman Smoother

1 **Input:** The PDFs $p(\mathbf{x}_t|\mathbf{y}_{1:t}) \sim \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}, \Sigma_{t|t})$ and
    $p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) \sim \mathcal{N}(\mathbf{x}_{t+1}; \hat{\mathbf{x}}_{t+1|t}, \Sigma_{t+1|t})$ for
    $t = 1, \dots, N$ computed in Algorithm 2.
2 **for** $t = N \, to \, 1$ **do**
3   Compute the gain using $\mathbf{G}_t = \Sigma_{t|t}\mathbf{A}_t^T\Sigma_{t+1|t}^{-1}$.
4   Compute $\hat{\mathbf{x}}_{t|N}$ according to
    $\hat{\mathbf{x}}_{t|N} = \hat{\mathbf{x}}_{t|t} + \mathbf{G}_t(\hat{\mathbf{x}}_{t+1|T} - \hat{\mathbf{x}}_{t+1|t})$.
5   Compute $\Sigma_{t|N}$ according to
    $\Sigma_{t|N} = \Sigma_{t|t} + \mathbf{G}_t(\Sigma_{t+1|T} - \Sigma_{t+1|t})\mathbf{G}_t^T$
6 **end**
7 **Output:** The smoothed PDFs
    $p(\mathbf{x}_t|\mathbf{y}_{1:N}) \sim \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|N}, \Sigma_{t|N})$ for $t = 1, \dots, N$.

---

where $\mathbf{A}_t$ is the Jacobian matrix of $f(\cdot)$ with respect to $\mathbf{x}_t$ and evaluated at $\hat{\mathbf{x}}_{t|t}$, $\mathbf{C}_t$ is the Jacobian matrix of $h(\cdot)$ with respect to $\mathbf{x}_t$ and evaluated at $\hat{\mathbf{x}}_{t|t-1}$ (see Appendix), $\mathbf{B}_t = f(\hat{\mathbf{x}}_{t|t}, \mathbf{u}_t) - \mathbf{A}_t\hat{\mathbf{x}}_{t|t}$ and $\mathbf{D}_t = h(\hat{\mathbf{x}}_{t|t-1}, \mathbf{u}_t) - \mathbf{C}_t\hat{\mathbf{x}}_{t|t-1}$. The equations of the EKF are summarized in Algorithm 1. On the other hand, the EKS is an extension of the standard Rauch-Tung-Striebel Smoother. The algorithm also considers the linearized model in (40)-(41) to compute the filter gain. Notice that the smoothing algorithm starts at $t = N$ where the filtering and smoothing distribution are equal i.e. $p(\mathbf{x}_N|\mathbf{y}_{1:N})$, then continues computing the smoothing distribution in the time t based on the smoothing distribution on time $t + 1$ as is shown in Algorithm 3.

C. UNSCENTED KALMAN FILTER AND SMOOTHER

The unscented Kalman Filter [21] is a deterministic-sampling-based approach that uses samples called

sigma-points to propagate the mean and covariance of the system state (that is assumed a Gaussian random variable) through the nonlinear functions of the process or/and the output of the system. These propagated points capture the mean and covariance of the posterior state accurately to the 3rd order Taylor series expansion for any nonlinearity [42]. The key idea of the UKF is to directly approximate the mean and covariance of the posterior distribution instead of approximating the nonlinear function [21]. The unscented Kalman Filter and smoother (see Algorithms 4 and 5) are based on the unscented transformation of a random variable x into a random variable $y = g(x) + v$ where $x \sim \mathcal{N}(x; \mathbf{m}, \Sigma)$ and $v \sim \mathcal{N}(v; 0, \mathbf{P})$. *Sigma* points are defined as

$$\psi^0 = \mathbf{m}, \tag{42}$$

$$\psi^\tau = \mathbf{m} + \sqrt{n + \lambda}\left[\sqrt{\mathbf{P}}\right]_\tau, \tag{43}$$

$$\psi^{\tau+n} = \mathbf{m} - \sqrt{n + \lambda}\left[\sqrt{\mathbf{P}}\right]_\tau, \tag{44}$$

where $\tau = 1, \ldots, n$, the scaling parameter $\lambda = \alpha^2(n+\kappa) - n$, the parameters α and κ determine the propagation of the sigma points around the mean. The notation $[\mathcal{A}]_\tau$ means the τth column of the matrix $\mathcal{A}$. The notation $\sqrt{\mathcal{A}}$ means the matrix square root of the matrix $\mathcal{A}$ such that $\sqrt{\mathcal{A}}\sqrt{\mathcal{A}}^T = \mathcal{A}$. The weights associated with the unscented transformation are the sets

$$\left\{\omega^0, \omega^1, \ldots, \omega^{2n}\right\} = \{\lambda\varsigma, 0.5\varsigma, \ldots, 0.5\varsigma\}, \tag{45}$$

$$\left\{\sigma^0, \sigma^1, \ldots, \sigma^{2n}\right\} = \{\lambda\varsigma + \varrho, 0.5\varsigma, \ldots, 0.5\varsigma\}, \tag{46}$$

where $\varsigma = (n+\lambda)^{-1}$, $\varrho = 1 - \alpha^2 + \beta$. Here β is an additional algorithm parameter that can be used for incorporating prior information on the (non-Gaussian) distribution of x. Then the statistic of the transformed random variable is: mean $\mu = \sum_{\tau=0}^{2n} \omega^\tau g(\psi^\tau)$, the covariance matrix $\Sigma = \sum_{\tau=0}^{2n} \sigma^\tau [g(\psi^\tau) - \mu][g(\psi^\tau) - \mu]^T + \mathbf{P}$. Additionally, the cross covariance matrix between x and y is given by $\Sigma = \sum_{\tau=0}^{2n} \sigma^\tau [\psi^\tau - \mathbf{m}][g(\psi^\tau) - \mu]^T$.

Remark 4: Note that solving the traditional maximum likelihood estimation problem presented in (31) poses a significant challenge. Directly maximizing the log-likelihood function is often intractable due to its non-convex nature. However, the EM algorithm [10] provides a solution to this obstacle. It addresses a less complex optimization problem in each iteration, as is shown in (37), especially when leveraging the auxiliary function (Taylor-based approximation) in Lemma 3.

V. NUMERICAL EXAMPLES

In this section, we present a numerical example to illustrate the performance of the proposed EM-based algorithm. We consider the discrete–time Hulbert SAG system model in (21) and (22) in order to generate synthetic data with a sampled period $\Delta = 2[s]$.

Algorithm 4 Unscented Kalman Filter

1 Input: The PDF $p(\mathbf{x}_1)$, e.g. $\hat{\mathbf{x}}_{0|1} = \mu_1$ and $\Sigma_{0|1} = \mathbf{P}_1$, the constant α, κ, and β

2 for $t = 1$ **to** N **do**

3   **Measurement Update:**

4   Compute and store the sigma points $\psi_{t|t-1}^\tau$, the weights $\omega_{t|t-1}^\tau$ and $\sigma_{t|t-1}^\tau$ by using $\hat{\mathbf{x}}_{t|t-1}$ and $\Sigma_{t|t-1}$ for $\tau = 0, \ldots, 2n$.

5   Propagate sigma points using $h(\cdot)$ for $\tau = 0, \ldots, 2n$: $\mathcal{Y}_t^\tau = h(\psi_{t|t-1}^\tau, \mathbf{u}_t)$. Compute the gain $\mathbf{K}_t = \Gamma_t \mathbf{S}_t^{-1}$, where

$$\mu_t = \sum_{\tau=1}^{2n} \omega_{t|t-1}^\tau \mathcal{Y}_t^\tau.$$

$$\mathbf{S}_t = \sum_{\tau=1}^{2n} \sigma_{t|t-1}^\tau (\mathcal{Y}_t^\tau - \mu_t)(\mathcal{Y}_t^\tau - \mu_t)^T + \mathbf{R}.$$

$$\Gamma_t = \sum_{\tau=1}^{2n} \sigma_{t|t-1}^\tau (\psi_{t|t-1}^\tau - \hat{\mathbf{x}}_{t|t-1})(\mathcal{Y}_t^\tau - \mu_t)^T.$$

   Compute the filtering state $\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mu_t)$.

6   Compute the covariance matrix $\Sigma_{t|t} = \Sigma_{t|t-1} - \mathbf{K}_t \mathbf{S}_t \mathbf{K}_t^T$.

7   **Time Update:**

8   Compute and store the sigma points $\psi_{t|t}^\tau$, the weigths $\omega_{t|t}^\tau$ and $\sigma_{t|t}^\tau$ by using $\hat{\mathbf{x}}_{t|t}$ and $\Sigma_{t|t}$ for $\tau = 0, \ldots, 2n$.

9   Propagate sigma points using the model $f(\cdot)$ for $\tau = 0, \ldots, 2n$: $\mathcal{X}_t^\tau = f(\psi_{t|t}^\tau, \mathbf{u}_t)$.

10   Compute the predicted state $\hat{\mathbf{x}}_{t+1|t} = \sum_{\tau=1}^{2n} \omega_{t|t}^\tau \mathcal{X}_t^\tau.$

11   Compute the covariance matrix $\Sigma_{t+1|t} = \sum_{\tau=1}^{2n} \sigma_{t|t}^\tau (\mathcal{X}_t^\tau - \hat{\mathbf{x}}_{t+1|t})(\mathcal{X}_t^\tau - \hat{\mathbf{x}}_{t+1|t})^T + \mathbf{Q}.$

12 end

13 Output: The filtered PDFs $p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}, \Sigma_{t|t})$, the predicted PDFs $p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_{t+1}; \hat{\mathbf{x}}_{t+1|t}, \Sigma_{t+1|t})$, and the sigma points $\psi_{t|t}^\tau$ and $\psi_{t+1|t}^\tau$ for $t = 1, \ldots, N$.

The Hulbert model is simulated using Simulink/Matlab ő. To perform continuous-time simulations of such a model in (15) and (17), we consider the power P_{mill} given by equation (4) in [6]. The constants used to simulate the model can be summarized in Table 1.

A. SIMULATION SETUP

For comparison we consider the inputs defined in [24]:

$$\text{MIW} = \max\left(4.64 + 8\sin\left(\frac{2\pi}{12 \cdot 60}t\right), 0\right), \ (\text{m}^3/h) \tag{47}$$

$$\text{MFO} = 65.2 + 10\sin\left(\frac{2\pi}{4 \cdot 60^2}t\right), \ (\text{t}/h) \tag{48}$$

$$\text{MFB} = 5.68 + a_r\mu(t_\Delta), \ (\text{t}/\text{m}^3) \tag{49}$$

where a_r is a uniform random variable in the interval $[-1, 1]$ and $\mu(t_\Delta)$ starting at every $t_\Delta = 2$ hours. The state and

---

Algorithm 5 Unscented Kalman Smoother

---

1 **Input:** The filtered PDFs
   $p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_{t|t}, \Sigma_{t|t})$, the predicted PDFs
   $p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_{t+1}; \hat{\mathbf{x}}_{t+1|t}, \Sigma_{t+1|t})$, the sigma
   points $\psi_{t|t}^\tau$ and $\psi_{t+1|t}^\tau$ for $t = 1, \ldots, N$ computed in
   Algorithm 4.
2 **for** $t = N \, to \, 1$ **do**
3   Compute the gain $\mathbf{G}_t = \mathbf{H}_t \Sigma_{t+1|t}^{-1}$ using :

$$\mathbf{H}_t = \sum_{\tau=1}^{2M} \sigma_{t|t}^\tau \left(\psi_{t|t}^\tau - \hat{\mathbf{x}}_{t|t}\right)\left(\psi_{t+1|t}^\tau - \hat{\mathbf{x}}_{t+1|t}\right)^T.$$

   Compute the smoothing state
   $\hat{\mathbf{x}}_{t|N} = \hat{\mathbf{x}}_{t|t} + \mathbf{G}_t\left(\hat{\mathbf{x}}_{t+1|T} - \hat{\mathbf{x}}_{t+1|t}\right).$
4   Compute the covariance matrix
   $\Sigma_{t|N} = \Sigma_{t|t} + \mathbf{G}_t\left(\Sigma_{t+1|T} - \Sigma_{t+1|t}\right)\mathbf{G}_t^T.$
5 **end**
6 **Output:** The smoothed PDFs
   $p(\mathbf{x}_t|\mathbf{y}_{1:N}) \sim \mathcal{N}\left(\mathbf{x}_t; \hat{\mathbf{x}}_{t|N}, \Sigma_{t|N}\right)$ for $t = 1, \ldots, N.$

TABLE 1. Parameters used in the simulations.

Constants	Values	Unit	Description
d_H	88	$1/h$	Mill discharge constant
D_m	4.07	m	Diameter inside the mill
ϵ_{sv}	0.6	-	Maximum fraction of solids per volume of pulp at zero flow
ϕ_c	0.72	-	Critical speed fraction
g	9.8	m^2/s	Gravity constant
κ_b	90	kWh/t	Energy required for abrasion of steel balls
κ_r	6.72	kWh/t	Energy required for rock abrasion
κ_p	0.97	-	Power consumption setting parameter
L_c	0	m	Mill cone length
L_m	4.54	m	Mill cylinder length
P_{NL}	93.73	kW	Mill power at zero load
ρ_b	7.85	t/m^3	Density of steel balls
ρ_o	3.2	t/m^3	Mineral density
ρ_w	1	t/m^3	Water density
r_t	0.46	m	Mill feed opening radius
v_{mill}	59.12	m	Internal volume of the mill

measurement noises are Gaussian distributed, that is $\mathbf{w}_t \sim \mathcal{N}(\mathbf{w}_t; \mathbf{0}, \mathbf{Q})$ and $\mathbf{v}_t \sim \mathcal{N}(\mathbf{v}_t; \mathbf{0}, \mathbf{R})$, respectively. Here, we study two scenarios. Scenario 1 considers a low variance for the noise, such as:

$$\mathbf{Q} = \text{diag}\,[0.004, \; 0.0025, \; 0.1, \; 0.064]\,(\text{m}^6), \quad (50)$$

$$\mathbf{R} = 10^{-4}\text{diag}\left[1(\text{t}^2), \quad 1(\text{m}^6)\right]. \quad (51)$$

and Scenario 2, on the contrary, considers a higher value for the noise variance such as:

$$\mathbf{Q} = \text{diag}\,[0.01, \quad 0.01, \quad 0.01, \quad 0.04]\,(\text{m}^6), \quad (52)$$

$$\mathbf{R} = \text{diag}\left[200(\text{t}^4), \quad 10(\text{m}^6)\right]. \quad (53)$$

Typical initial state and input values for the system are given by:

$$\mathbf{x}_1 = \begin{bmatrix} x_1^w \\ x_1^s \\ x_1^r \\ x_1^b \end{bmatrix} = \begin{bmatrix} 8.23 \\ 1.88 \\ 4.65 \\ 4.63 \end{bmatrix}(\text{m}^3), \quad (54)$$

$$\mathbf{u} = \begin{bmatrix} u^w \\ u^s \\ u^r \\ u^b \end{bmatrix} = \begin{bmatrix} 116.6 \\ 107.7 \\ 9.6 \\ 0.7 \end{bmatrix}(\text{m}^3/h). \quad (55)$$

The initialization of the parameter to be estimated is as follows: i) The initial guest of covariance matrix $\mathbf{R}$ is obtained as the sample variance of each measurements in $\mathbf{y}_t$, that is

$$\hat{\mathbf{R}}^{(0)} = \text{diag}\left\{\text{var}\{W_c\}, \text{var}\{V_{\text{LOAD}}\}\right\}. \quad (56)$$

ii) The initial guest of covariance matrix $\mathbf{Q}$ is defined–for both Scenario 1 and 2 as

$$\hat{\mathbf{Q}}^{(0)} = \alpha_\mathbf{Q}\mathbf{Q}, \quad (57)$$

where the factor $\alpha_Q = \{0.1, 1, 10\}$, and iii) The initial guest of the initial state and the parameter vector θ are studied in two cases:

- Case factor 0.2 far from the real values:

$$\hat{\mathbf{x}}_1^{(0)} = [6.58\text{m}^3, \; 2.26\text{m}^3, \; 3.72\text{m}^3, \; 5.55\text{m}^3]^T, \quad (58)$$

$$\hat{\mathbf{P}}_1^{(0)} = \text{diag}\left\{2\text{m}^6, \; 2\text{m}^6, \; 2\text{m}^6, \; 2\text{m}^6\right\}, \quad (59)$$

$$\hat{\theta}^{(0)} = [0.48, \; 105.6]^T, \quad (60)$$

- Case factor 1.5 far from the real values:

$$\hat{\mathbf{x}}_1^{(0)} = [4.12\text{m}^3, \; 2.82\text{m}^3, \; 2.33\text{m}^3, \; 6.95\text{m}^3]^T, \quad (61)$$

$$\hat{\mathbf{P}}_1^{(0)} = \text{diag}\left\{2\text{m}^6, \; 2\text{m}^6, \; 2\text{m}^6, \; 2\text{m}^6\right\}, \quad (62)$$

$$\hat{\theta}^{(0)} = [0.3, \; 132]^T. \quad (63)$$

B. RESULTS

To compare the results obtained with the proposal presented in this paper, we have considered the conventional approach for estimating parameters in the SAG model, using filtering and smoothing techniques as described in [5], [24], and [38]. In these prior works, the authors suggest estimating system parameters by modeling them as state variables. To apply these techniques, the parameters were included as additional states and then estimated by using EKF and UKF algorithms. The estimates' mean and standard deviations are summarized in Table 2, considering different scenarios (low and high variance for the noise), and different starting points for the initial values. In general, it can be observed that there is no significant difference between using EKF and UKF. However, it is worth noting that the performance of such methods rely upon the initial values and the level of noise. A more accurate parameter estimate is typically obtained when the initial value is close to the true one. Figure 2 shows a particular case of the ones shown in

TABLE 2. Estimation of SAG mill parameters as additional states, using EKF/EKS and UKF/UKS.

$\theta = [0.6, \ 88]$		Scenario 1		Scenario 2	
		EKF/EKS	UKF/UKS	EKF/EKS	UKF/UKS
Case factor 0.20	$\alpha_\mathbf{Q} = 0.1$	0.5789 ± 0.00003 / 99.2024 ± 0.0157	0.5776 ± 0.00003 / 100.6150 ± 0.0160	0.5992 ± 0.0063 / 89.0759 ± 2.4449	0.5971 ± 0.0059 / 89.5803 ± 2.4573
	$\alpha_\mathbf{Q} = 1$	0.5690 ± 0.00002 / 106.3632 ± 0.0074	0.5833±0.00002 / 106.8906±0.0079	0.6811 ± 0.0089 / 72.6052 ± 1.3172	0.6831 ± 0.0073 / 72.1830 ± 1.3226
	$\alpha_\mathbf{Q} = 10$	0.5663 ± 0.0082 / 109.6954 ± 9.6119	0.5668 ± 0.00004 / 105.9192 ± 0.0052	0.9026 ± 0.0126 / 98.9653 ± 0.7967	0.9862 ± 0.0194 / 100.1907 ± 0.7683
Case factor 1.50	$\alpha_\mathbf{Q} = 0.1$	0.5455 ± 0.00002 / 134.3712 ± 0.0244	0.5425 ± 0.00002 / 133.7294 ± 0.0197	0.5610 ± 0.0035 / 115.8159 ± 3.2941	0.5579 ± 0.0038 / 115.1752 ± 3.1969
	$\alpha_\mathbf{Q} = 1$	0.5408 ± 0.00002 / 142.7397 ± 0.0090	0.5527 ± 0.00002 / 139.9117 ± 0.0040	0.6289 ± 0.0064 / 90.4409 ± 1.9818	0.6376 ± 0.0063 / 85.9412 ± 2.0360
	$\alpha_\mathbf{Q} = 10$	0.5454 ± 0.0540 / 134.4366 ± 13.3784	0.3 ± 0.00002 / 131.9 ± 0.0004	0.8734 ± 0.0132 / 129.1362 ± 1.3287	0.8781 ± 0.0149 / 126.5491 ± 1.2104

(a) $\hat{\theta}$ with $\alpha_Q = 1$, case factor 0.20

(b) $\hat{\theta}$ with $\alpha_Q = 1$, case factor 1.50

(c) $\hat{\theta}$ with $\alpha_Q = 1$, case factor 0.20

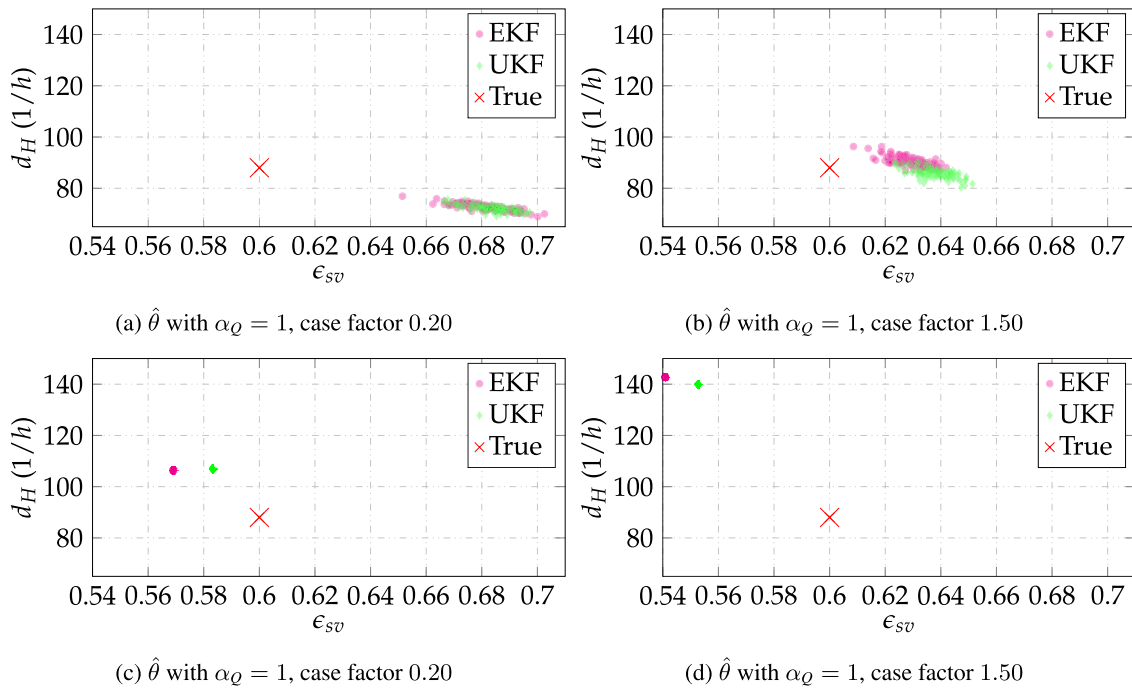(d) $\hat{\theta}$ with $\alpha_Q = 1$, case factor 1.50

FIGURE 2. Comparison of the estimates $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ considered as states, with low variance (upper plots) versus high variance (lower plots).

Table 2, considering low variance (upper plots) and high variance (lower plots), and $\alpha_Q = 1$ for both cases. Notice that the mean value of the estimated parameter is far from the true value. This discrepancy stems from the fact that the EKF, used for state and parameter estimation, employs a first-order Taylor series approximation of the nonlinear system around a state estimate. Similarly, the UKF relies on a third-order Taylor series linearization of the system. In both instances, significant estimation errors arise due to these approximations.

For the case when the unknowns are considered as parameters in the ML estimation framework, we have obtained the results shown in Table 3. Here, we can observe

that there is no significant difference between the results obtained using the EM algorithm combined with EKF/EKS and UKF/UKS in both cases, whether the noise variance is high or low. A particular case in Table 3 is depicted in Figure 3. However, the main difference can be observed when comparing the two approaches, that is, unknowns are considered as states and as parameters in the ML framework. Comparing both tables, we can infer that considering the unknowns as parameters can greatly improve the accuracy of the estimation. That is, combining a parametric estimation algorithm with filtering and smoothing techniques can help compensate for the nonlinear effects in the estimation values. With each iteration of the proposed algorithm, the parameter

estimation is progressively refined, thereby improving state estimation, and so forth.

VI. CONCLUSION

In this paper, we have proposed an ML estimation method to obtain critical parameters in the model of an SAG mill. Traditional methods consider a reduced number of measurements in order to estimate such parameters. Here, we have included an extra measurement (the total SAG mill weight) that has not been considered in previous works by other authors.

When the parameters are considered as state-space variables, a comparison between EKF and UKF is appropriate. In this case, we have carried out extensive simulations considering both cases. It has been shown that the implementation of the UKF shows no particular advantage and that the performance of both EKF and UKF depends mostly upon the chosen initial values, which have been selected as 0.2 and 1.5 of the nominal values, and the level of noise variance.

When estimating the parameters in the model as unknown parameters rather than states in a state-space model, we pose the problem in the ML framework. Here, we solve the ML estimation problem by using the EM algorithm. We have also considered two approaches when it comes to the calculation of the E-step in the EM algorithm, namely the use of the EKF/EKS and the use of UKF/UKS. The utilization of the EM algorithm greatly improved the accuracy of the estimates when compared to the approach that considers the unknowns as state variables. When it comes to the comparison of EKF and UKF, within the EM algorithm approach, no significant differences have been found, even considering higher values for the noise variances. When it comes to the selection of the initial values in the aforementioned algorithms, we have evaluated them in two cases: (i) considering a factor of 0.2 of the initial values, and (ii) considering a factor of 1.5 of the initial values. In all cases, no significant differences were found, mainly because they were not too far from the optimal value of the parameter.

APPENDIX
JACOBIAN MATRICES TO IMPLEMENT THE EXTENDED KALMAN FILTER

In this section, we obtain the Jacobian matrices to implement the Extended Kalman Filter. Let us consider the system in (21) and (22). Notice that the rheology factor in (11) can be rewritten as follows

$$\varphi\left(\mathbf{x}_t\right)=\sqrt{\max\left\{0,b_t\right\}}, \qquad b_t=1-c_{\text{sv}}\left(\frac{x_t^s}{x_t^w}\right), \quad (64)$$

where $c_{\text{sv}} = (1-\varepsilon_{\text{sv}})/\varepsilon_{\text{sv}}$. Due to the max term, the rheology factor is a non differentiable function. In this work, we propose to approximate the rheology factor using $\max\{a,b\} = \tau_{\text{r}}^{-1}\log\{\exp\{a\tau_{\text{r}}\}+\exp\{b\tau_{\text{r}}\}\}$, then, the rheology factor approximation is given by

$$\hat{\varphi}\left(\mathbf{x}_t\right)=\sqrt{\tau_{\text{r}}^{-1}\log\left\{1+\exp\left\{b_t\tau_{\text{r}}\right\}\right\}}. \quad (65)$$

where τ_{r} is positive scale parameter defined by the user. To compute the Jacobian matrix $\mathbf{A}_t$ given in (40), we define the following:

$$g_w := \frac{\partial\hat{\varphi}\left(\mathbf{x}_t\right)}{\partial x_t^w}, \qquad g_s := \frac{\partial\hat{\varphi}\left(\mathbf{x}_t\right)}{\partial x_t^s}, \quad (66)$$

then taking the partial derivative of the function $\hat{\varphi}\left(\mathbf{x}_t\right)$ w.r.t x_t^w and x_t^s we can define g_w and g_s as follows

$$g_w = \frac{-0.5(x_t^w)^{-2}x_t^s\exp\left\{b_t\tau_{\text{r}}\right\}c_{\text{sv}}}{\left(\tau_{\text{r}}^{-1}\log\left\{1+\exp\left\{b_t\tau_{\text{r}}\right\}\right\}\right)^{1/2}\left(1+\exp\left\{b_t\tau_{\text{r}}\right\}\right)}, \quad (67)$$

$$g_s = \frac{0.5(x_t^w)^{-1}\exp\left\{b_t\tau_{\text{r}}\right\}c_{\text{sv}}}{\left(\tau_{\text{r}}^{-1}\log\left\{1+\exp\left\{b_t\tau_{\text{r}}\right\}\right\}\right)^{1/2}\left(1+\exp\left\{b_t\tau_{\text{r}}\right\}\right)}. \quad (68)$$

Then, the Jacobian matrix is given by

$$\mathbf{A}_t=\begin{bmatrix} 1-J_{11}\Delta & J_{12}\Delta & 0 & 0 \\ J_{21}\Delta & 1-J_{22}\Delta & J_{23}\Delta & 0 \\ J_{31}\Delta & J_{32}\Delta & 1-J_{33}\Delta & 0 \\ J_{41}\Delta & J_{42}\Delta & J_{43}\Delta & 1-J_{44}\Delta \end{bmatrix}, \quad (69)$$

where

$$J_{11}=g_wd_Hx_t^w\Gamma_w+\hat{\varphi}\left(\mathbf{x}_t\right)d_H(2x_t^s+2x_t^w-1)\Gamma_w^2, \quad (70)$$

$$J_{12}=-g_sd_Hx_t^w\Gamma_w+\hat{\varphi}\left(\mathbf{x}_t\right)d_H\Gamma_w^2, \quad (71)$$

$$J_{21}=-g_wd_Hx_t^w\Gamma_s-\hat{\varphi}\left(\mathbf{x}_t\right)d_H\Gamma_s^2+g_w\Gamma_r, \quad (72)$$

$$J_{22}=g_sd_Hx_t^w\Gamma_s+\hat{\varphi}\left(\mathbf{x}_t\right)d_H\Gamma_w^2-g_s\Gamma_r+\hat{\varphi}\left(\mathbf{x}_t\right)\Gamma_p, \quad (73)$$

$$J_{23}=\hat{\varphi}\left(\mathbf{x}_t\right)(x_t^r)^{-1}(x_t^s)^2\Gamma_p, \quad (74)$$

$$J_{31}=-g_w\Gamma_r, \quad (75)$$

$$J_{32}=-g_s\Gamma_r+\hat{\varphi}\left(\mathbf{x}_t\right)\Gamma_p, \quad (76)$$

$$J_{33}=J_{23}, \quad (77)$$

$$J_{41}=-g_w\Gamma_b, \quad (78)$$

$$J_{42}=-g_s\Gamma_b+\rho_0\hat{\varphi}\left(\mathbf{x}_t\right)\Gamma_q, \quad (79)$$

$$J_{43}=\rho_0\hat{\varphi}\left(\mathbf{x}_t\right)\Gamma_q, \quad (80)$$

$$J_{44}=-\hat{\varphi}\left(\mathbf{x}_t\right)(x_t^b)^{-1}\rho_0(x_t^r+x_t^s)\Gamma_q. \quad (81)$$

where

$$\Gamma_r=\frac{P_t^{\text{mill}}x_t^r}{\rho_0\kappa_r(x_t^r+x_t^s)}, \qquad \Gamma_p=\frac{P_t^{\text{mill}}x_t^r}{\rho_0\kappa_r(x_t^r+x_t^s)^2}, \quad (82)$$

$$\Gamma_w=\frac{x_t^w}{x_t^s+x_t^w}, \qquad \Gamma_b=\frac{P_t^{\text{mill}}x_t^b}{\kappa_b\left[\rho_0(x_t^r+x_t^s)+\rho_bx_t^b\right]}, \quad (83)$$

$$\Gamma_s=\frac{x_t^s}{x_t^s+x_t^w}, \qquad \Gamma_q=\frac{P_t^{\text{mill}}x_t^b}{\kappa_b\left[\rho_0(x_t^r+x_t^s)+\rho_bx_t^b\right]^2}. \quad (84)$$

On the other hand, the Jacobian matrix $\mathbf{C}_t$ given in (41) is directly obtain from the output model as follows:

$$\mathbf{C}_t=\begin{bmatrix} 1 & \rho_0 & \rho_0 & \rho_b \\ 1 & 1 & 1 & 1 \end{bmatrix}. \quad (85)$$

TABLE 3. Estimation of SAG mill parameters using the proposed method, using EKF/EKS and UKF/UKS, but considering the unknowns as parameters.

$\theta = [0.6,\ 88]$		Scenario 1		Scenario 2	
		EKF/EKS-EM	UKF/UKS-EM	EKF/EKS-EM	UKF/UKS-EM
Case factor 0.20	$\alpha_{\mathbf{Q}} = 0.1$	0.6006±0.0001 / 87.7677±0.0550	0.5976±0.0002 / 88.5060±0.0791	0.6003±0.0028 / 87.7930±1.5398	0.5972±0.0032 / 88.4451±1.4964
	$\alpha_{\mathbf{Q}} = 1$	0.6005±0.0001 / 87.7663±0.0326	0.5993±0.0002 / 88.0788±0.0453	0.5993±0.0028 / 88.3064±1.5242	0.5968±0.0029 / 88.7077±1.3728
	$\alpha_{\mathbf{Q}} = 10$	0.6004±0.0004 / 87.8262±0.1029	0.5974±0.0017 / 88.3976±0.6149	0.5981±0.0029 / 88.8739±1.6109	0.5973±0.0037 / 89.3307±2.0342
Case factor 1.50	$\alpha_{\mathbf{Q}} = 0.1$	0.6005±0.0001 / 87.7849±0.0273	0.5972±0.0001 / 88.6554±0.0382	0.5941±0.0032 / 90.7119±1.8797	0.5860±0.0048 / 93.8975±2.7506
	$\alpha_{\mathbf{Q}} = 1$	0.6003±0.00004 / 87.8852±0.0107	0.5990±0.0002 / 87.9955±0.0366	0.5952±0.0032 / 90.2764±1.8617	0.5915±0.0040 / 91.1640±2.0677
	$\alpha_{\mathbf{Q}} = 10$	0.6002±0.0005 / 87.9477±0.0884	0.5989±0.0001 / 87.9289±0.0140	0.5960±0.0033 / 89.8591±1.8688	0.5906±0.0053 / 91.6230±2.7246

(a) $\hat\theta$ with $\alpha_Q = 1$, case factor 0.20

(b) $\hat\theta$ with $\alpha_Q = 1$, case factor 1.50

(c) $\hat\theta$ with $\alpha_Q = 1$, case factor 0.20

(d) $\hat\theta$ with $\alpha_Q = 1$, case factor 1.50

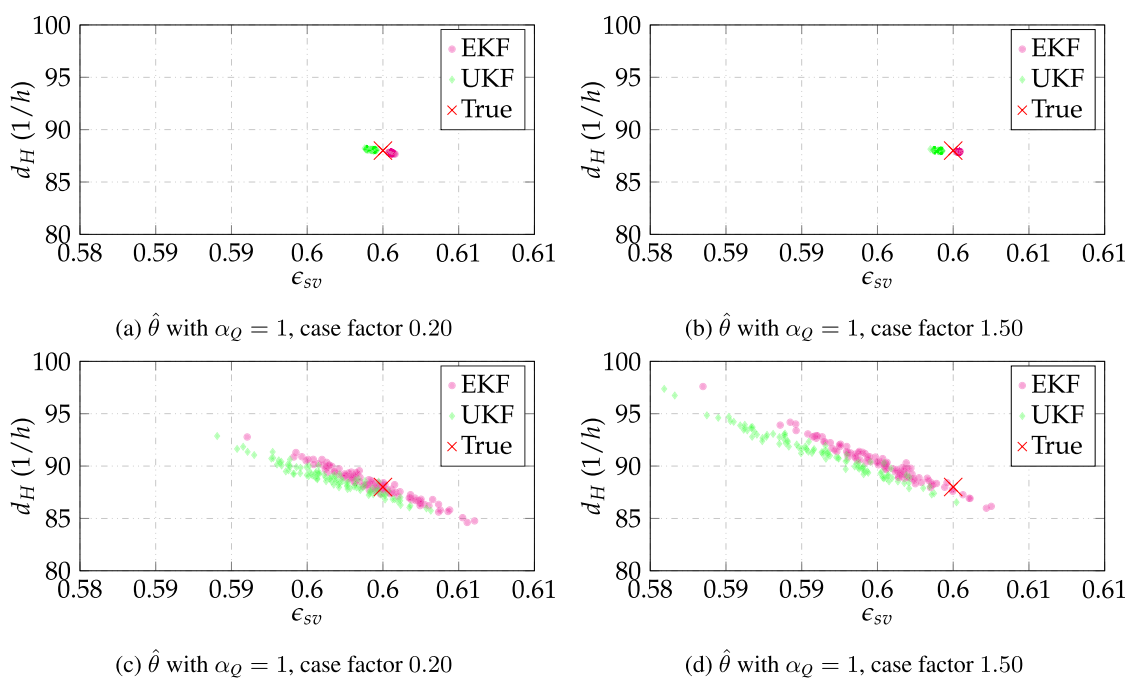**FIGURE 3.** Comparison of estimates $\hat\theta = (\hat\theta_1, \hat\theta_2)$ considered as unknown parameters, with low variance (upper plots) versus high variance (lower plots).

REFERENCES

[1] A.-Z.-M. Abouzeid and D. W. Fuerstenau, "Grinding of mineral mixtures in high-pressure grinding rolls," *Int. J. Mineral Process.*, vol. 93, no. 1, pp. 59–65, Sep. 2009.

[2] J. C. Agüero and G. C. Goodwin, "Approximate EM algorithms for parameter and state estimation in nonlinear stochastic models," in *Proc. 44th IEEE Conf. Decis. Control (CDC)*, 2005, pp. 1–15.

[3] J. C. Agüero, W. Tang, J. I. Yuz, R. Delgado, and G. C. Goodwin, "Dual time–frequency domain system identification," *Automatica*, vol. 48, no. 12, pp. 3031–3041, Dec. 2012.

[4] T. A. Apelt, S. P. Asprey, and N. F. Thornhill, "Inferential measurement of SAG mill parameters III: Inferential models," *Minerals Eng.*, vol. 15, no. 12, pp. 1055–1071, Dec. 2002.

[5] T. A. Apelt, S. P. Asprey, and N. F. Thornhill, "Inferential measurement of SAG mill parameters II: State estimation," *Minerals Eng.*, vol. 15, no. 12, pp. 1043–1053, Dec. 2002.

[6] T. A. Apelt and N. F. Thornhill, "Inferential measurement of sag mill parameters IV: Inferential model validation," *Minerals Eng.*, vol. 22, no. 12, pp. 1032–1044, Oct. 2009.

[7] J. J. Burchell, J. D. Le Roux, and I. K. Craig, "Nonlinear model predictive control for improved water recovery and throughput stability for tailings reprocessing," *Control Eng. Pract.*, vol. 131, Feb. 2023, Art. no. 105385.

[8] S. C. Chelgani, M. Parian, P. S. Parapari, Y. Ghorbani, and J. Rosenkranz, "A comparative study on the effects of dry and wet grinding on mineral flotation separation—A review," *J. Mater. Res. Technol.*, vol. 8, no. 5, pp. 5004–5011, Sep. 2019.

[9] L. C. Coetzee, I. K. Craig, and E. C. Kerrigan, "Robust nonlinear model predictive control of a run-of-mine ore milling circuit," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 1, pp. 222–229, Jan. 2010.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 39, no. 1, pp. 1–22, Sep. 1977.

[11] A. Dontchev and W. Hager, "The Euler approximation in state constrained optimal control," *Math. Comput.*, vol. 70, no. 233, pp. 173–203, Jan. 2001.

[12] D. Foszcz, D. Krawczykowski, T. Gawenda, E. Kasinska-Pilut, and W. Pawlos, "Analysis of process of grinding efficiency in ball and rod mills with various feed parameters," in *Proc. IOP Conf. Ser., Mater. Sci. Eng.*, Sep. 2018, vol. 427, no. 1, p. 12031.

[13] Z. Ghasemi, F. Neumann, M. Zanin, J. Karageorgos, and L. Chen, "A comparative study of prediction methods for semi-autonomous grinding mill throughput," *Minerals Eng.*, vol. 205, Jan. 2024, Art. no. 108458.

[14] S. Gibson, A. Wills, and B. Ninness, "Maximum-likelihood parameter estimation of bilinear systems," *IEEE Trans. Autom. Control*, vol. 50, no. 10, pp. 1581–1596, Oct. 2005.

[15] B. I. Godoy, J. C. Agüero, R. Carvajal, G. C. Goodwin, and J. I. Yuz, "Identification of sparse FIR systems using a general quantisation scheme," *Int. J. Control*, vol. 87, no. 4, pp. 874–886, Apr. 2014.

[16] B. I. Godoy, G. C. Goodwin, J. C. Agüero, D. Marelli, and T. Wigren, "On identification of FIR systems having quantized output data," *Automatica*, vol. 47, no. 9, pp. 1905–1915, Sep. 2011.

[17] R. B. Gopaluni, "A particle filter approach to identification of nonlinear processes under missing observations," *Can. J. Chem. Eng.*, vol. 86, no. 6, pp. 1081–1092, Dec. 2008.

[18] A. Hassanzadeh, "Measurement and modeling of residence time distribution of overflow ball mill in continuous closed circuit," *Geosystem Eng.*, vol. 20, no. 5, pp. 251–260, Sep. 2017.

[19] M. M. Hilden, M. S. Powell, and M. Yahyaei, "An improved method for grinding mill filling measurement and the estimation of load volume and mass," *Minerals Eng.*, vol. 160, Jan. 2021, Art. no. 106638.

[20] D. G. Hulbert, "Simulation of milling circuits: Part 1 & 2," Mintek, Johannesburg, South Africa, Tech. Rep., 2005.

[21] S. J. Julier and J. K. Uhlmann, "New extension of the Kalman filter to nonlinear systems," *Proc. SPIE*, vol. 3068, pp. 182–193, Apr. 1997.

[22] M. J. Korenberg, "Applications of nonlinear system identification in molecular biology," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2006, pp. 256–259.

[23] J. D. Le Roux, I. K. Craig, D. G. Hulbert, and A. L. Hinde, "Analysis and validation of a run-of-mine ore grinding mill circuit model for process control," *Minerals Eng.*, vols. 43–44, pp. 121–134, Apr. 2013.

[24] J. D. Le Roux, A. Steinboeck, A. Kugi, and I. K. Craig, "An EKF observer to estimate semi-autogenous grinding mill hold-ups," *J. Process Control*, vol. 51, pp. 27–41, Mar. 2017.

[25] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[26] P. Lopez, I. Reyes, N. Risso, M. Momayez, and J. Zhang, "Machine learning algorithms for semi-autogenous grinding mill operational regions' identification," *Minerals*, vol. 13, no. 11, p. 1360, Oct. 2023.

[27] W. Taam, "Nonlinear system analysis and identification from random data," *Technometrics*, vol. 33, no. 4, pp. 482–483, Nov. 1991.

[28] S. Morrell, "Predicting the overall specific energy requirement of crushing, high pressure grinding roll and tumbling mill circuits," *Minerals Eng.*, vol. 22, no. 6, pp. 544–549, May 2009.

[29] T. J. Napier-Munn, S. Morrell, R. D. Morrison, T. Kojovic, "Mineral comminution circuits: their operation and optimization," Julius Kruttschnitt Mineral Res. Centre, Monograph Ser. Mining'Mineral Process., Univ. Queensland. Australia, 1996.

[30] B. Ninness, "Some system identification challenges and approaches," in *Proc. 15th IFAC Symp. Syst. Identificat.*, 2009, vol. 42, no. 10, pp. 1–20.

[31] L. E. Olivier, B. Huang, and I. K. Craig, "Dual particle filters for state and parameter estimation with application to a run-of-mine ore mill," *J. Process Control*, vol. 22, no. 4, pp. 710–717, Apr. 2012.

[32] J. M. Ortiz, W. Kracht, G. Pamparana, and J. Haas, "Optimization of a SAG mill energy system: Integrating rock hardness, solar irradiation, climate change, and demand-side management," *Math. Geosci.*, vol. 52, no. 3, pp. 355–379, Apr. 2020.

[33] K. B. Owusu, M. Zanin, W. Skinner, and R. K. Asamoah, "AG/SAG mill acoustic emissions characterisation under different operating conditions," *Minerals Eng.*, vol. 171, Sep. 2021, Art. no. 107098.

[34] G. Pamparana, W. Kracht, J. Haas, J. M. Ortiz, W. Nowak, and R. Palma-Behnke, "Studying the integration of solar energy into the operation of a semi-autogenous grinding mill. Part I: Framework, model development and effect of solar irradiance forecasting," *Minerals Eng.*, vol. 137, pp. 68–77, Jun. 2019.

[35] A. Pomerleau, D. Hodouin, A. Desbiens, and É. Gagnon, "A survey of grinding circuit control methods: From decentralized PID controllers to multivariable predictive controllers," *Powder Technol.*, vol. 108, nos. 2–3, pp. 103–115, Mar. 2000.

[36] V. A. Rodriguez, R. M. de Carvalho, and L. M. Tavares, "Insights into advanced ball mill modelling through discrete element simulations," *Minerals Eng.*, vol. 127, pp. 48–60, Oct. 2018.

[37] M. Ruel, "Fuzzy logic control on a SAG mill," in *Proc. 16th IFAC Symposium Control*, 2013, vol. 46, no. 16, pp. 282–287.

[38] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[39] T. B. Schön, A. Wills, and B. Ninness, "System identification of nonlinear state-space models," *Automatica*, vol. 47, no. 1, pp. 39–49, Jan. 2011.

[40] D. A. Silva and L. A. Tapia, "Experiences and lessons with advanced control systems for the SAG mill control in minera los pelambres," in *Proc. 2nd IFAC Workshop Automation Mining*, 2009, vol. 42, no. 23, pp. 25–30.

[41] P. Varas, R. Carvajal, and J. C. Agüero, "State estimation for SAG mills utilizing a simplified model with an alternative measurement," in *Proc. IEEE CHILEAN Conf. Electr., Electron. Eng., Inf. Commun. Technol. (CHILECON)*, Nov. 2019, pp. 1–7.

[42] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. IEEE Adapt. Syst. Signal Process., Commun., Control Symp.*, Oct. 2000, pp. 153–158.

[43] D. Wei and I. K. Craig, "Grinding mill circuits—A survey of control and economic concerns," in *Proc. 17th IFAC World Congr.*, 2008, vol. 41, no. 2, pp. 1000–1005.

[44] A. Wills, B. Ninness, and S. Gibson, "Maximum likelihood estimation of state space models from frequency domain data," in *Proc. Eur. Control Conf. (ECC)*, Jul. 2007, pp. 1545–1552.

[45] A. Wills, T. B. Schön, L. Ljung, and B. Ninness, "Identification of Hammerstein—Wiener models," *Automatica*, vol. 49, no. 1, pp. 70–81, 2013.

[46] C. Yuwen, B. Sun, and S. Liu, "A dynamic model for a class of semi-autogenous mill systems," *IEEE Access*, vol. 8, pp. 98460–98470, 2020.

ANGEL L. CEDEÑO was born in Manabí, Ecuador. He received the bachelor's degree in electronics and control engineering from Escuela Politécnica Nacional, Quito, Ecuador, and the M.Sc. and Ph.D. degrees in electronic engineering from Universidad Técnica Federico Santa María, Chile, in 2022 and 2023, respectively. Currently, he was a Postdoctoral Researcher with the Advanced Center for Electrical and Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile. His research interests include system identification, state estimation, control, and optimization-based predictive control for embedded systems.

MARÍA CORONEL (Member, IEEE) was born in San Cristóbal, Venezuela. She received the Electronic Engineering degree from Universidad Nacional Experimental del Tàchira, Venezuela, the M.Sc. degree in automation and instrumentation from Universidad de Los Andes, Venezuela, and the M.Sc. and Ph.D. degrees in electronic engineering from Universidad Técnica Federico Santa María, Chile, in 2021 and 2022, respectively. She is currently an Academic with the Department of Electricity, Universidad Tecnológica Metropolitana. Her research interests include continuous-time system models, systems identification, stochastic processes, and control and instrumentation.

RAFAEL ORELLANA (Member, IEEE) was born in San Cristóbal, Venezuela. He received the Electronic Engineering degree from Universidad Nacional Experimental del Táchira, Venezuela, the M.Sc. degree in automation and instrumentation from Universidad de Los Andes, Venezuela, and the M.Sc. and Ph.D. degrees in electronic engineering from Universidad Técnica Federico Santa María, Chile, in 2020 and 2021, respectively. He is currently an Academic with the Department of Electrical Engineering, Universidad de Santiago de Chile. His research interests include system identification, statistical signal processing, control, and instrumentation.

PATRICIO VARAS was born in Chile. He received the Ingeniería Civil Electrónica and Master of Engineering degrees from Universidad Técnica Federico Santa María, Chile. His research interests include system identification, bayesian inference, and stochastic signal processing.

RODRIGO CARVAJAL (Member, IEEE) was born in Valparaíso, Chile. He received the Ingeniería Civil Electrónica and M.Sc. degrees from Pontificia Universidad Católica de Valparaíso, Chile, in 2003 and 2007, respectively, and the Ph.D. degree from The University of Newcastle, Australia, in 2013. He is currently with the School of Electrical Engineering, Pontificia Universidad Católica de Valparaíso. His research interests include communication systems, statistical signal processing, and systems identification. He received the Postgraduate Research Excellence Prize for Electrical, Computer, and Telecommunication Engineering, in 2011.

BORIS I. GODOY was born in Valparaíso, Chile, in 1975. He received the Ingeniería Civil Electrónica and M.Sc. degrees in electronics engineering from Universidad Técnica Federico Santa María, Chile, in 2001, and the Ph.D. degree in electrical engineering from The University of Newcastle, Australia, in 2008. He is currently pursuing the master's degree in data analytics with Boston University, USA. Since then, he has held several academic and industry positions in Australia. He was a Researcher with the Department of Automatic Control, Lund University, Lund, Sweden, until 2023. His research interests include system identification, time-varying systems, process control, and dynamical systems in general.

JUAN C. AGÜERO (Member, IEEE) was born in Osorno, Chile. He received the Ingeniería Civil Electrónica and Master of Engineering degrees from Universidad Técnica Federico Santa María, Chile, in 2000, and the Ph.D. degree from The University of Newcastle, Australia, in 2006. He gained industrial experience at El Teniente, Codelco, Chile, in the copper mining industry, from 1997 to 1998. He was a Research Academic with The University of Newcastle, until 2014. He is currently an Associate Professor with Universidad Técnica Federico Santa María. His research interests include system identification, control, and statistical signal processing. He was an Associate Editor of *Automatica*, from 2017 to 2022.

• • •