Journal of Process Control 105 (2021) 62–77

Contents lists available at ScienceDirect

Journal of Process Control

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

A novel hybrid analysis and modeling approach applied to aluminum electrolysis process

Erlend Torje Berg Lundby [a,*], Adil Rasheed [a,b], Jan Tommy Gravdahl [a], Ivar Johan Halvorsen [b]

[a] Norwegian University of Science and Technology, Department of Engineering Cybernetics, Trondheim, Norway
[b] SINTEF Digital, Department of Mathematics and Cybernetics, Trondheim, Norway

ARTICLE INFO

Article history:
Received 29 January 2021
Received in revised form 22 April 2021
Accepted 14 June 2021
Available online 24 July 2021

Keywords:
Aluminum electrolysis
Sub-Nyquist sampling
Compressed sensing
Kalman filter
Hybrid analysis and modeling

ABSTRACT

Aluminum electrolysis cells are characterized by harsh environments where several measurements have to be done manually. Due to the operational costs related to manual sampling, the sampling rates of these measurements are low. Therefore, the information in the data can be limited, making it challenging to develop robust data-driven methods for aluminum electrolysis process. A broad array of physics-based models have been developed throughout the years to provide excellent system knowledge about the dynamics in the aluminum electrolysis cells. However, due to highly complex and interrelated sub-processes, the state-of-the-art physics-based models are insufficient to accurately express the dynamics in the cell. The combination of inadequate prediction models and low sampling rates makes estimating process variables in the aluminum electrolysis process less accurate than what is desired for optimal and safe operation of the cells. In this paper, a novel hybrid modeling approach that addresses insufficient prediction models and low sampling rates is suggested. The novel hybrid modeling approach involves manipulating a measured signal with a first principle model estimate. This manipulation, which consists of subtracting the first principle model estimate from the measurements of the signal, produces a residual that represents the unmodeled dynamics in the signal. Since the unmodeled dynamics of the measured signal is much sparser than the measured signal itself, this manipulation enables utilizing a powerful technique for estimating sparse signals from only a few measurements. The technique is called compressed sensing. The manipulated data is used as input data in a compressed sensing algorithm which produces a high fidelity estimate of the unmodeled dynamics in the original signal. Thus, the novelty in this article is two-fold. First, compressed sensing is introduced to the field of aluminum electrolysis. Second, the novel technique of sparsifying a measured signal with a first principle model in order to utilize compressed sensing on the measured data is introduced. The signal estimate of the unmodeled dynamics is integrated into an Extended Kalman filter as a pseudo measurement to improve the estimation of the system states. The novel method applies to signals with stationary periodical unmodeled dynamics. The case study in this article is conducted on simulated data of a sub-process describing the mass balance in an aluminum electrolysis cell.

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the primary production of aluminum, aluminum is extracted from alumina in the electrolytic Hall–Héroult process. The aluminum electrolysis cell consists of essential components such as anode, cathode, and electrolyte [1]. The environment in the aluminum electrolysis cell is very harsh due to the high temperature and highly corrosive electrolyte [2,3]. Sensor systems struggle to survive in the harsh environment, thereby requiring manual sampling, which is costly due to its labor-intensive nature. For this reason, there is a need to *minimize* the sampling rate without compromising on safety and efficiency of aluminum production.

Furthermore, in the industry, it is desirable to utilize advanced process control systems such as Nonlinear Model Predictive control (NMPC) to optimize the operation (safe operation, maximized production, and minimized energy consumption) [4]. But the NMPC's performance is dependent on accurate state estimation [5]. State estimation in aluminum electrolysis is a very challenging task due to low sampling rates, low observability and complex, interrelated sub-processes that are difficult to model.

* Corresponding author.
E-mail addresses: erlend.t.b.lundby@ntnu.no (E.T.B. Lundby), adil.rasheed@ntnu.no (A. Rasheed), jan.tommy.gravdahl@ntnu.no (J.T. Gravdahl), Ivar.J.Halvorsen@sintef.no (I.J. Halvorsen).

https://doi.org/10.1016/j.jprocont.2021.06.005
0959-1524/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Expensive manual measurements and the challenge of estimating state variables are the main motivations to conduct this study. Hence, we can formulate the following research needs: • minimize the number of measurements • capitalize on the data to improve predictive modeling.

Addressing the first research need has been a challenge because of the strict constraint posed by the Shanon–Nyquist sampling criterion [6], which demands a sampling frequency of twice the fastest frequency component in the signal. The Nyquist rate is far too high for what is profitable in the field of aluminum electrolysis, where sampling is very costly. Fortunately, in recent times the field of Compressed Sensing [7] has emerged as a way to bypass the Shannon Nyquist sampling criterion. Compressed sensing (also known under the terminology *compressive sensing*, *compressive sampling* and *sparse recovery*) is an emerging research area that aims to estimate a high dimensional signal from a low dimensional measurement vector. This results in an underdetermined system of linear equations that has infinitely many solutions. However, exploiting that the signal is sparse or compressible in some domain makes the problem of estimating a high dimensional signal vector from a low dimensional measurement vector possible [8]. Compressed sensing offers a framework that enables estimating signals from far fewer measurements than required by the Nyquist criterion. Therefore, introducing compressed sensing to estimate a measured signal in aluminum electrolysis makes it possible to minimize the number of expensive manual measurements taken from that signal and at the same time achieve a high-resolution estimate of the coarsely measured signal. To the best of the authors knowledge, compressed sensing has never been applied in the context of aluminum electrolysis before.

Compressed sensing has been used in a variety of applications for estimating signals and images from few measurements. In image analysis, compressed sensing has been applied in magnetic resonance imaging and hyperspectral images, in biomedical applications, compressed sensing has been used in for example electrocardiogram signals, in communication systems compressed sensing has been applied in wireless networks and multi input multi output communication, and in pattern recognition compressed sensing has been used in face recognition, speech and speaker recognition and gesture recognition to mention a few [9]. Inspired by the research field of compressed sensing, a research field called compressive system identification has evolved. Compressive system identification addresses the problem of identifying systems from only a small number of observations. In [10] compressive system identification was used to identify Auto Regressive with eXternal input (ARX) models for Linear Time-Invariant and Linear Time-Variant systems with a large number of inputs and unknown delays. In [11] the method suggested includes compressed sensing and dynamic mode decomposition with extension to actuated systems to identify reduced-order models on downsampled spatial measurements of high-dimensional systems and reconstruct full-state dynamic modes associated with the model. In [12] a method that incorporates physical knowledge into compressed sensing was developed to reduce the volume of data and number of sensors needed for modeling and monitoring the temperature field of an additive manufacturing process. Different scientific work within the field of Compressive system identification can also be found in [13–17]. In general, compressed sensing is investigated on signals or systems that are already sparse. This is rarely the case in the time-series data corresponding to the aluminum extraction process. Thus utilizing the power of compressed sensing in the current context requires manipulating the signal to make the signal sparse. This issue is addressed in this paper by sparsifying the signal of interest in the novel hybrid modeling framework.

Regarding the second research need, currently used physics-based models derived from first principles describes complex physical and chemical phenomena in the Hall–Héroult process that have contributed to increased knowledge of the process and the utilization of more efficient control strategies [18–20]. Based on well-understood physics though robust and interpretable, these models are computationally inefficient, inaccurate, and incapable of modeling the unknown/poorly understood physics on the fly. Some examples of these unknown/poorly understood physics are interactions between the sub-phenomena like magneto-hydrodynamic phenomena, the current distribution, motion induced by the bubbles formed at the anode, and reactivity and species concentration [21]. Data-driven modeling comes as an attractive alternative because of its ability to model physical phenomena, to a certain extent, without the need to know the governing physics. For predicting variables that are difficult to measure, Data-driven modeling methods have been developed for the soft-sensing purpose in a broad range of industrial processes [22]. The most commonly applied methods for the purpose of soft sensing are Artificial Neural Networks (ANN), Partial Least Square (PLS), and Support Vector Machines (SVM). These methods occupy about 75% of all data-driven methods applied in industrial processes [23]. In aluminum electrolysis, different types of data-driven methods have been applied in soft sensors. [24] suggests a Kernel Extreme Learning Machine to predict the alumina concentration in the electrolyte. Least Squares SVM models for predicting alumina concentration and electrolytic bath temperature are developed in [25]. ANN approaches have been applied in the electrolysis process to predict essential variables that are difficult to measure [26–28]. These models show good performance in modeling nonlinearities. However, they tend to be black box and non-generalizable to new scenarios.

To address the shortcomings of physics-based and data-driven modeling, a new breed of modeling called Hybrid Analysis and Modeling (HAM) is emerging [29]. It is a modeling approach that combines the interpretability, robust foundation, and understanding of a physics-based modeling approach with the accuracy, efficiency, and automatic pattern-identification capabilities of advanced data-driven modeling algorithms. Hybrid models can balance the advantages and disadvantages of first principle models and data-driven models and therefore have several advantages over these classes of models such as higher prediction accuracy, better calibration properties, enhanced extrapolation properties and better interpretability [30]. Hybrid models can be designed in many different ways to exploit the advantages of these models. In [31], a deep neural network is used to estimate process parameters utilized in a first principle model. In [32] correction terms are inferred from data and added to a model structure consisting of first principle knowledge and the learned correction terms. In [33], a general mechanistic multi-scale modeling framework for cell populations was developed. A mechanistic model describing the dynamics of the states of individual cells is combined with statistical models of the measurement data and the states and parameters of the mechanistic model that describes the dynamics of individual cells. The resulting model is a partial differential equation model that describes the time evolution of the population density among the cells.

In the proposed hybrid framework, the first principle knowledge is used in a different matter. Estimates calculated by the first principle model are subtracted from the measured values of the states, which result in residual data that can be used to estimate unmodeled dynamics of the system. While [31] and [32] suggests hybrid models that estimate all the dynamics in the system, the hybrid modeling framework in this work offers a method to estimate the unmodeled dynamics that the current model used to estimate the states in the system could not estimate.

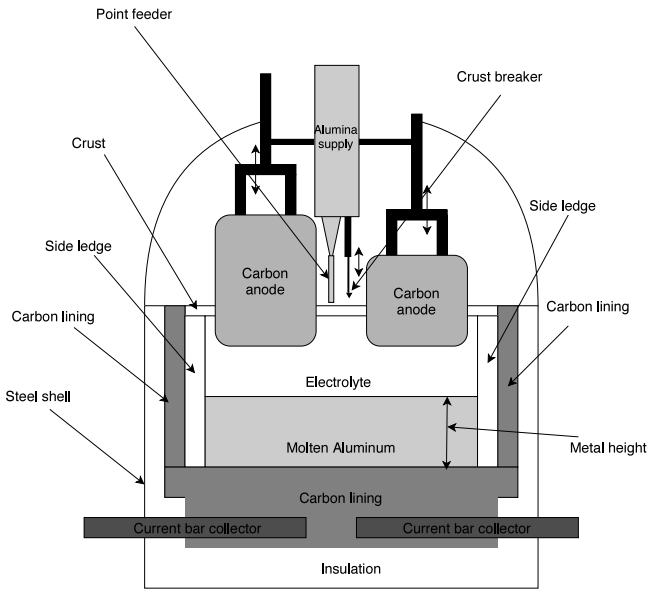


Fig. 1. Conceptual drawing of an aluminum electrolysis cell.

The purposed method can find *periodical patterns in the residual* between coarsely sampled measurements of a process variable and the predicted values of the same process variable.

Despite the aluminum electrolysis process being an interrelated process with many state variables, it is possible to consider sub-processes of the electrolysis that describe some characteristic dynamics of the process. In this article, simulated dynamics of the aluminum mass balance in the aluminum cell are considered. The dynamics of this sub-process can be expressed in terms of two state variables: the aluminum mass in the cell and the thickness of frozen electrolyte on the inside of sidewalls known as the side ledge. The side ledge works as a protective layer for the sidewall against the corrosive electrolyte and molten metal at high electrolysis temperature. Furthermore, the side ledge works as thermodynamic insulation. Thus, it is essential for both safe and efficient operations. It is practically impossible to measure the side ledge profile in operating electrolysis cells [34]. However, the side ledge can be estimated from measurements of the metal height since it affects the displacement volume for the aluminum in the cell and the metal height, see Fig. 2. However, the metal height measurements have to be sampled manually with a dipstick, observing the molten metalmark. Methods are developed for determining the metal height, temperature, and electrolyte content. However, most of these methods face technological challenges in the actual production line. The major challenges are related to the high-temperature environment, extremely corrosive bath, and the inevitable crust at the top of the bath [35], see Fig. 1.

Therefore, due to labor-intensive costs, these measurements are taken at low sampling rates. This makes it difficult to estimate the side ledge thickness from measurements. There exist models that can describe the dynamics of the side ledge thickness. However, due to many interrelated sub-processes in the aluminum electrolysis process, all dynamics in this variable cannot be expected to be captured by these models.

Thus, the main objective of the paper is to minimize the number of expensive manual measurements needed to improve state estimation of essential variables in aluminum electrolysis cells like the side ledge thickness.

To realize the objective, two research questions are formulated. These research questions are:

- How can sampled data be manipulated so that the powerful tool of compressed sensing can be utilized for estimating unmodeled dynamics from sparsely sampled data?
- How can a high fidelity signal estimate of the unmodeled dynamics be utilized in a Kalman filter to improve accuracy of the estimated system states

This study introduces the powerful tool of compressed sensing to estimate sparsely sampled signals in the process of aluminum electrolysis. Furthermore, the study presents a way of manipulating a sampled signal with a first principle model to take advantage of compressed sensing. That is, estimating a signal from far fewer measurements than what is required by the Nyquist criterion. Thus, instead of estimating the measured signal, the method suggests estimating the dynamics in the signal that the first principle model does not capture. This unmodeled dynamics in the sampled signal is much sparser than the sampled signal, and therefore it is required much fewer measurements to estimate this signal with compressed sensing. The proposed approach of sparsifying a measured signal to prepare it for compressed sensing is, to the best of the authors knowledge, not been done before. The novel hybrid modeling approach presented in this study can estimate stationary, periodical unmodeled dynamics in a sampled signal from few, randomly sampled measurements. The estimate of the unmodeled dynamics is then utilized in an Extended Kalman Filter (EKF) to improve the accuracy of estimating the side ledge thickness.

The case study in this article is conducted on a simulation model described in Section 2.1. A signal is used equivalently to a variable in this article. The term data refers to either simulated data or sampled data. Sampled data is defined as the data that is accessible for training a model or estimating a signal because the data has been measured at different time instants. Simulated data refers to the high fidelity data generated by a simulation model to generate the signals in the process and is used as the ground truth in the case study. If only the term data is used, it refers to sampled data. The structure of the article is as follows. In Section 2, the theory and methods applied in the work related to this paper is described. In Section 3, there is a detailed description of the novel hybrid modeling method and a description of how simulated data is generated for analysis. In Section 4 the results from analysis are presented and in Section 5 conclusions are made.

2. Theory

2.1. Aluminum electrolysis cell

In this study, the simulated data is generated from a simplified simulator of the aluminum electrolysis cell representing the mass balance in the cell. Aluminum is extracted in the Hall–Héroult process, which involves dissolving alumina (Al_2O_3) into an electrolytic bath mainly consisting of cryolite (Na_3AlF_6). An electric current is sent through the cell, and aluminum ions in the electrolyte are reduced to aluminum. The aluminum reduction cell consists of basic components of an electrolysis cell such as anode, cathode, and electrolyte. One or several carbon anodes are immersed into the electrolyte, also known as the bath. This is illustrated in Fig. 1. In this figure, the main components of an aluminum cell and a rough sketch of its design is presented. The carbon anodes are consumed in the process. This is expressed in the overall reaction of aluminum production in the Hall–Héroult process:

$$2Al_2O_3 + 3C \rightarrow 3CO_2(g) + 4Al(l). \quad (1)$$

The carbon in the reaction comes from the anode. The molten aluminum is considered as the cathode in the reaction.

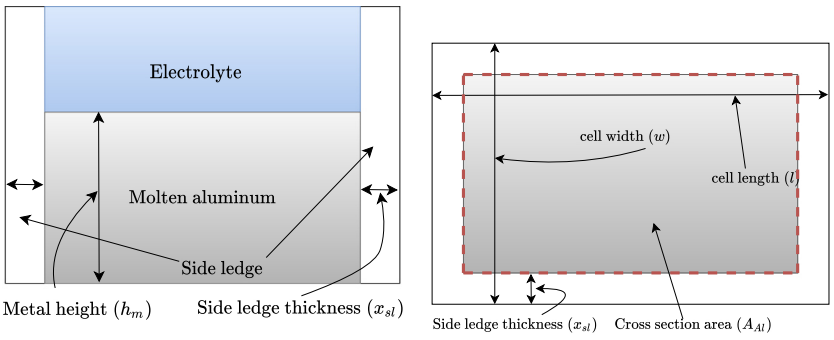


Fig. 2. Geometry of simulated aluminum cell. The side walls are assumed horizontal and the side ledge thickness is assumed uniform. In real aluminum cells, the bottom of the side walls are in general sloping walls. Furthermore, the side ledge thickness is in general not uniform.

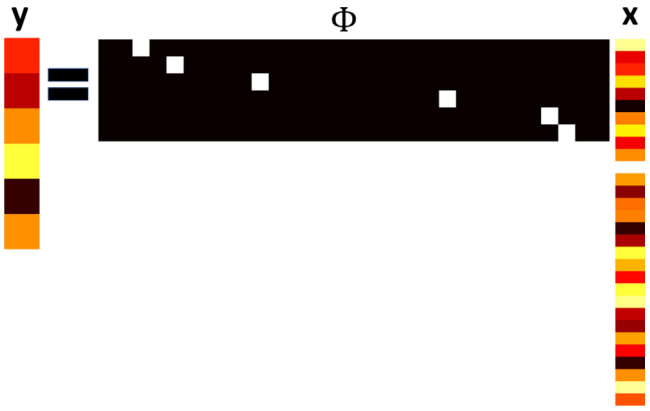
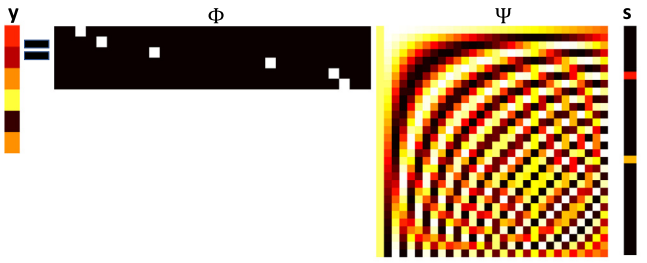
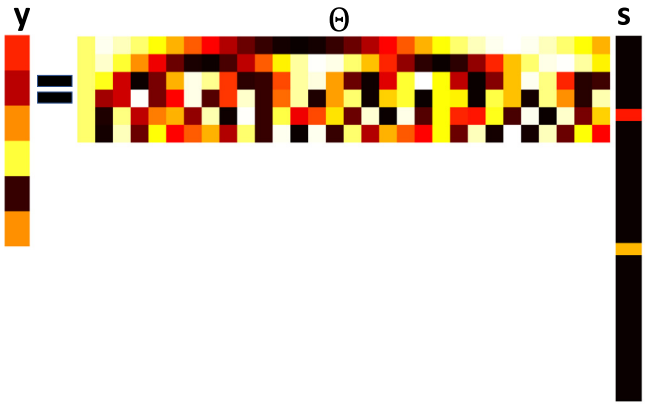


Fig. 3. Matrix illustration of the linear underdetermined system $\mathbf{y} = \Phi\mathbf{x}$. Since the system is underdetermined and the signal $\mathbf{x}$ is in general not sparse, the system can in general not be solved or have infinitely many solutions. The left vector illustrates the measurements $\mathbf{y}$, the matrix in the middle illustrates the measurement matrix Φ and the vector on the right illustrates the signal $\mathbf{x}$. The black areas in Φ are zero entries, and the white areas are ones. Φ maps values between $\mathbf{x}$ and $\mathbf{y}$. Colors in $\mathbf{x}$ and $\mathbf{y}$ correspond to numeric values and range from small values corresponding to dark and large values corresponding to light. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Represents the linear system $\mathbf{y} = \Phi\Psi\mathbf{s}$. Decompose the signal $\mathbf{x}$ into a basis matrix Ψ and basis coefficients $\mathbf{s}$



(b) Represents the linear system $\mathbf{y} = \Theta\mathbf{s}$. This is the resultin system to be solved in compressed sensing.

Fig. 4. Matrix illustration of how the signal $\mathbf{x}$ is represented in terms of a basis $\Psi\mathbf{s}$. Hence the measurements $\mathbf{y}$ are represented in terms of the measurement matrix Φ, basis matrix Ψ and basis coefficients $\mathbf{s}$. In both (a) and (b), the leftmost vector is the measurements $\mathbf{y}$ and the rightmost vector is the coefficient vector $\mathbf{s}$. In $\mathbf{s}$, the black areas are zero entries. In (a), the middle-left matrix is the measurement matrix Φ. The black areas in Φ are zero entries, and the white areas are ones. The middle-right matrix is the transform basis Ψ. In this case, Ψ is the Discrete cosine transform (DCT). In (b), the middle matrix is the matrix product $\Theta = \Phi\Psi$. Despite the resulting system in (b) being underdetermined, the system can be solved for $\mathbf{s}$ if a suitable basis is found so that $\mathbf{s}$ is sparse. That is the case in this illustration, where $\mathbf{s}$ is two-sparse. Colors in $\mathbf{y}$, Ψ and $\mathbf{s}$ correspond to numeric values and range from small values corresponding to dark and large values corresponding to light. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where Φ is the measurement matrix of size $m \times n$. Fig. 3 illustrates the linear system in matrix form. This expresses that the number of equations (m) is much smaller than the number of unknowns (n).

On the sidewalls of the cell, a ledge of frozen electrolyte called side ledge is formed. The side ledge works as a thermal insulator but most importantly as a protecting layer preventing the sidewall from corroding.

The mass balance of the aluminum cell was derived in [36]. Let m be the mass of aluminum in the cell, such that:

$$\frac{dm}{dt} = \dot{m}_{gen}(t) - \dot{m}_{out}(t), \tag{2}$$

where $\dot{m}_{gen}(t)$ is the mass rate of aluminum generated at time instant t, and $\dot{m}_{out}(t)$ is mass rate of aluminum taken out of the cell. This can be approximated to a discrete model:

$$\Delta m(k) = \Delta t(\dot{m}_{gen}(k) - \dot{m}_{out}(k)), \tag{3}$$

where $k = \{1, 2, 3, \ldots\}$ represent discrete time instants and Δt is the sampling time for when the model is updated. The mass of aluminum generated per time unit is given by Faraday's law:

$$\dot{m}_{gen}(k) = \frac{\Delta Q(k) \cdot CE \cdot M_{Al}}{F \cdot z \cdot \Delta t}, \tag{4}$$

where CE is the current efficiency, $M_{Al} = 26.98$ g/mol is the molecular mass of aluminum, $F = 96485.3329$ C/mol is Faraday's

constant and $z = 3$ is the number of electrons involved the reaction generating one aluminum atom. The charge ΔQ transferred from time instant $k-1$ to k is calculated by the time integral of the current I given by:

$$\Delta Q = \int_{k-1}^{k} I(\tau) d\tau. \tag{5}$$

Assuming that the current I is constant within a time interval Δt, generated mass rate can be formulated as:

$$\dot{m}_{gen}(k) = \frac{CE \cdot M_{Al}}{F \cdot z} I(k). \tag{6}$$

Furthermore, assuming constant current efficiency during a time interval Δt, the mass generated during a time interval Δt can be formulated as

$$m_{gen}(k) = \dot{m}_{gen}(k)\Delta t \tag{7}$$

The mass flow rate out of the cell relates to metal tapping. Assuming constant flow rate during the tapping yields:

$$m_{out}(k) = \Delta t \cdot \dot{m}_{out}(k) \tag{8}$$

The metal height h_m is a function of the average cross-section area of the cell and the volume of molten aluminum in the cell. It is assumed that the cross-section area is uniform. This follows from the assumption of horizontal side walls and uniform side ledge thickness, see Fig. 2. The metal height can be formulated as:

$$h_m = \frac{V_{Al}(m_{Al}, \rho_{Al})}{A_{Al}(l, w, x_{sl})}, \tag{9}$$

where V_{Al} is the volume of molten alumininum, m_{Al} is the mass of molten aluminum in the cell and $\rho_{Al} = 2200\ \mathrm{kg/m^3}$ is the density of molten aluminum. A_{Al} is the cross section area where the cell is occupied by aluminum, l is the cell length, w is the cell width and x_{sl} is the side ledge thickness. V_{Al} is given by:

$$V_{Al} = \frac{m_{Al}}{\rho_{Al}}. \tag{10}$$

Fig. 2(b) show a snap shot of the cross section of the electrolytic according to the simulation. Fig. 2(a) illustrates that the side ledge thickness is uniform. These drawings illustrates that the uniform cross section area of aluminum is given by:

$$A_{Al} = (l - 2x_{sb}) \cdot (b - 2x_{sb}) \tag{11}$$

The thickness of the frozen electrolyte known as side ledge x_{sl} is determined by a complex interaction between heat balance, amperage, bath composition, cell resistance, movements in the bath induced by magnetism, bubbles, and more [37]. Therefore, this variable is challenging to estimate. As stated above, this variable relates to the cross-section area A_{Al} and, consequently, to the metal height h_m. Therefore, estimating the metal height with compressed sensing techniques can reveal information about the side ledge thickness. The side ledge is crucial in preventing corrosion of the sidewalls. Thus this information is of great value. The side ledge thickness, which represents unmodeled dynamics in the cell, is simulated as a sum of cosine waves.

2.2. Extended Kalman filter

The Kalman filter is a set of equations that provide a recursive solution of the least-squares method. It supports estimates of past, present, and future states based on measurements, models, and uncertainties in the system. Although the Kalman filter was initially derived for linear systems, it has been extended to nonlinear systems through online Taylor expansions of the nonlinear system. This extension is referred to as *Extended Kalman Filter* (EKF). The EKF addresses the problem of estimating the state $\mathbf{x}$ of a nonlinear system

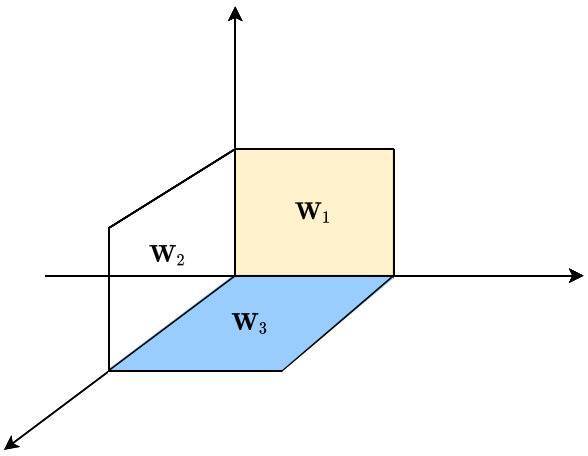


Fig. 5. Canonical example, 2-dimensional subspaces in $\mathbb{R}^3$.

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k) \tag{12}$$

$$\mathbf{y}_k = h(\mathbf{x}_k, \mathbf{v}_k), \tag{13}$$

where $f(\cdot)$ and $h(\cdot)$ are in general nonlinear functions. $\mathbf{w}_k$ represents the process noise, $\mathbf{v}_k$ represents the measurement noise, $\mathbf{y}_k$ represents the measurement, and $\mathbf{u}_k$ represent the control input. The subscript indicates at what time step the variable is sampled or estimated. The process and measurement noise are assumed to be normally distributed random variables with zero mean and covariances $\mathbf{Q}$ and $\mathbf{R}$ respectively:

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0},\ \mathbf{Q}) \tag{14}$$

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0},\ \mathbf{R}). \tag{15}$$

The Kalman filter algorithm can roughly be divided into two steps, where equations of the Kalman filter fall into the groups of either *time update* or the *measurement update*. Equations in the time update stage are responsible for projecting the current state estimate $\hat{\mathbf{x}}_k$ and error covariance estimate $\mathbf{P}_k$ forward in time to get the *a priori* state estimate $\hat{\mathbf{x}}^-_{k+1}$ and covariance estimate $\mathbf{P}^-_{k+1}$. Equations in the measurement stage are responsible for feedback from measurements to correct the *a priori* estimates, hence give the *a posteriori* estimates $\hat{\mathbf{x}}_k$ and $\mathbf{P}_k$ [38]. The *a priori* and *a posteriori* error covariance estimates are defined by:

$$\mathbf{P}^-_k = E\left[(\mathbf{x}_k - \hat{\mathbf{x}}^-_k)(\mathbf{x}_k - \hat{\mathbf{x}}^-_k)^T\right] \tag{16}$$

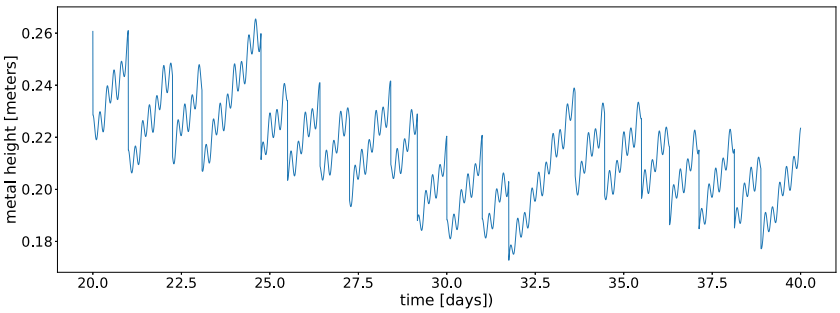
$$\mathbf{P}_k = E\left[(\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^T\right] \tag{17}$$

The algorithm is as follows:

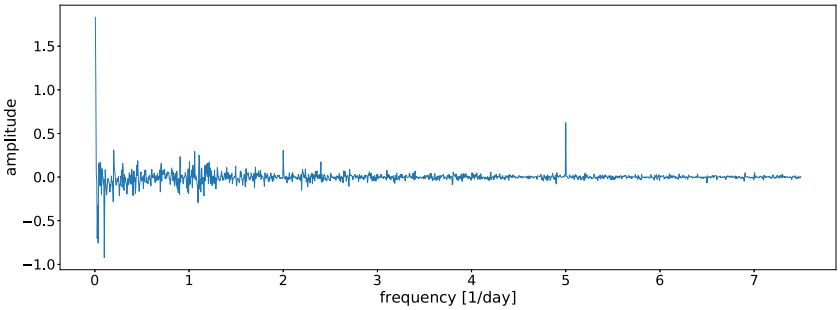
Algorithm 1: EKF

Time update:;
$\hat{\mathbf{x}}^-_{k+1} = \hat{\mathbf{x}}_k + \Delta t \cdot f(\hat{\mathbf{x}}_k, \mathbf{u}_k)$;
$\mathbf{P}^-_{k+1} = \mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^T + \mathbf{Q}_k$;
Measurement update:;
$\mathbf{K}_k = \mathbf{P}^-_k \mathbf{H}_k (\mathbf{H}_k \mathbf{P}^-_k \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$;
$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}^-_k + \mathbf{K}_k(\mathbf{y}_k - h(\hat{\mathbf{x}}^-_k, \mathbf{u}_k))$;
$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{P}^-_k$;

$\mathbf{A}_k$ is the Jacobian matrix of $f((\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k))$ with respect to $\mathbf{x}$, $\mathbf{H}_k$ is the jacobian matrix of $h(\mathbf{x}_k, \mathbf{v}_k)$ with respect to $\mathbf{x}$ and $\mathbf{K}_k$ is the Kalman gain. Δt is the sampling time and $\hat{\mathbf{x}}^-_{k+1}$ is calculated according to the *forward Euler method*.



(a) Simulated height from day 20 to day 40



(b) DCT of simulated height.

Fig. 6. In Figure (a), the simulated height h_m from day 20 to day 40 is shown. In (b), the Discrete Cosine Transform (DCT) of h_m, excluding the zero-frequency component, is shown. Frequency components of the unmodeled dynamics are those at 2 [per day] and 5 [per day]. Clearly, the transform is not sparse in the Cosine transform domain. Therefore it will be difficult to estimate the signal h_m from a limited number of measurements given the DCT as a transform basis.

2.3. Compressed sensing

The Shannon–Nyquist theorem states that the signal information is preserved if it is sampled uniformly at a rate at least two times faster than its bandwidth. In different applications, signal acquisition is prohibitive due to the cost of measuring the signal or simply because sensors do not sample the signal at rates as high as required by the Shannon–Nyquist theorem. Compressed sensing provides an alternative to Shannon–Nyquist sampling when the estimated signal is sparse or compressible [7]. Consider the signal $\mathbf{x}$ of length n represented in a transform basis Ψ such that $\mathbf{x} = \Psi\mathbf{s}$. A sparse signal $\mathbf{x}$ can in some basis Ψ be represented by $k \ll n$ nonzero coefficients in $\mathbf{s}$. A compressible signal $\mathbf{x}$ can be approximated by $k \ll n$ coefficients in $\mathbf{s}$. That is, when coefficients in $\mathbf{s}$ are sorted according to magnitude, they decay rapidly after the k'th coefficient. Compressed sensing addresses the problem of estimating a signal $\mathbf{x}$ of length n from m linear measurements $\mathbf{y}$ when $m \ll n$ by finding a solution to an underdetermined linear system:

$$\mathbf{y} = \Phi\mathbf{x}, \tag{18}$$

Now, consider the mapping of the signal $\mathbf{x}$ from time or space domain to a transform basis Ψ

$$\mathbf{x} = \Psi\mathbf{s}, \tag{19}$$

where $\mathbf{s}$ are the coefficients representing the signal $\mathbf{x}$ in Ψ. The resulting linear system is given by

$$\mathbf{y} = \Theta\mathbf{s}, \tag{20}$$

where $\Theta = \Phi\Psi$, $\Theta \in \mathbb{R}^{m\times n}$ This transformation is illustrated in Fig. 4. The rows of Φ, $\{\phi_j\}_{j=1}^{m}$, represent the measurement vectors, while the columns of Ψ, $\{\psi_j\}_{j=1}^{n}$, represent orthonormal basis vectors. If an appropriate basis Ψ is chosen so that the signal $\mathbf{x}$ is sparse in this domain, a solution for $\mathbf{s}$ of Eq. (20) can be found from far less measurements than if the original system in Eq. (18) was to be solved for $\mathbf{x}$.

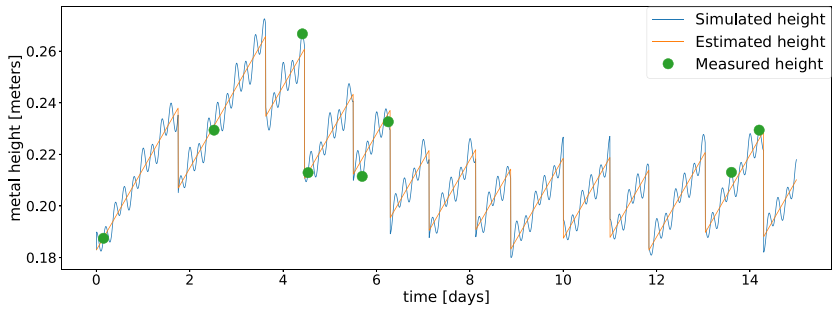
2.3.1. Low complexity structures

As mentioned, the linear system in Eq. (18) has fewer equations than unknowns, thus it is underdetermined. However, by utilizing the fact that the signal of interest $\mathbf{x} \in \mathbb{R}^n$ belongs to a low-dimensional subspace of dimension k, the system can still be solved. In other words, low complexity structures allows for recovering a signal $\mathbf{x}$ by solving the underdetermined system $\mathbf{y} = \Phi\mathbf{x}$. This touches the core of compressed sensing. All signals $\mathbf{x} \in \mathbb{R}^n$ can be expressed in a basis $\{\psi_i\}_{i=1}^{n}$ in terms of n coefficients $\{s_i\}_{i=1}^{n}$ as $\mathbf{x} = \sum_{i=1}^{n} s_i\psi_i = \Psi\mathbf{s}$. If the signal is k-sparse, it can be expressed by k nonzero coefficients s_i. This can be expressed mathematically as $\|\mathbf{s}\|_0 \leq k$, where $\|\cdot\|_0$ is the ℓ_0 pseudonorm expressing the number of nonzero coefficients. The set of all sparse signals is the union of $\binom{n}{k}$ k-dimensional subspaces spanned by k basis vectors. This gives the union-of-subspaces model and can be formulated mathematically as:

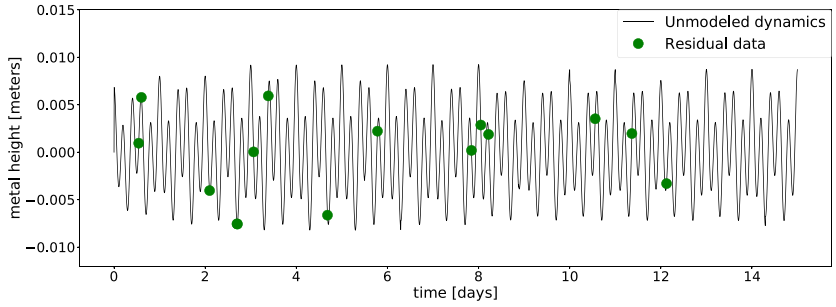
$$\mathbf{s} \in \bigcup_{S\subset[n],|S|=k} W_S =: \Sigma_k. \tag{21}$$

Here W_S is one subset of Ψ indexed by the index set $S \subset [n]$ with cardinality $|S| = k$. Hence Σ_k is the union of subspaces that correspond to vectors with at most k nonzero coefficients [39].

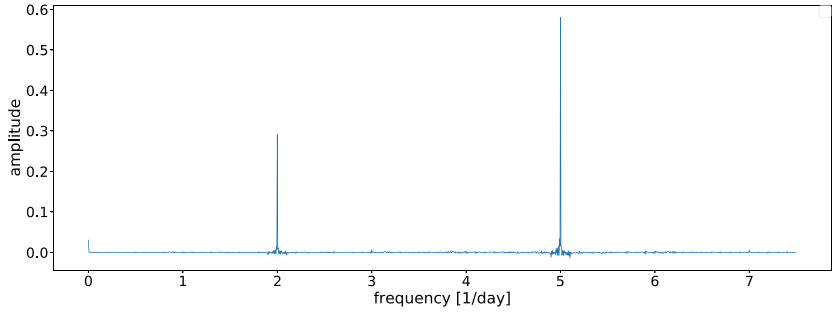
As an intuitive example, consider the canonical example in Fig. 5, where the signal lives in $\mathbb{R}^3$. The union of subspaces spanned by maximum two vectors (2-sparse vectors) is illustrated by the three 2-dimensional subspaces $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$ in $\mathbb{R}^3$. Considering the k-sparse signal in a n dimensional space, the number of possible k dimensional subspaces the solution can live in is $\binom{n}{k} \approx klog(n/k)$. This is an important quantity in compressed sensing in terms of required measurements.



(a) Simulated height and estimated height from first principle model



(b) Unmodeled dynamics. Residual between measured and estimated metal height



(c) DCT of unmodeled dynamics.

Fig. 7. In (a), the orange graph represents the estimated metal height h_{model} based on Faraday's law and knowledge of the amount of metal tapped. The blue graph represents the simulated metal height h_m. The green points represent measured metal height h_{meas}. In (b), the black graph represent the difference between simulated and estimated metal height $h_{unmod} = h_m - h_{model}$, whereas the green points represent the difference between measured and estimated metal height at the time instants when the measurements where taken $h_{residual} = h_{meas} - h_{model}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To include sets that do not necessarily form subspaces, a more general notion is needed for low-complexity structures. If the set of basis vectors is replaced with an arbitrary compact set, the signal models generated is referred to as *simple sets*:

Definition 2.1 (*Simple Set*)**.** Let $\mathcal{A} \subset \mathbb{R}^n$ be an origin-symmetric compact set, and $k \in \mathbb{R}$. Then a set $\mathcal{K} \subset \mathbb{R}^n$ of vectors on the form

$$\mathbf{x} = \sum_{i=1}^{k} c_i \mathbf{a}_i, \quad c_i \geq 0, \mathbf{a}_i \in \mathcal{A} \tag{22}$$

is called a simple set. Since elements in $\mathcal{K}$ are conic combinations of at most k elements in $\mathcal{A}$, $\mathcal{K}$ can be described as follows: $\mathcal{K} = cone_k(\mathcal{A})$. Moreover, since $\mathcal{K}$ is generated by the set $\mathcal{A}$, $\mathcal{A}$ is called an *atomic set*.

Again, consider the canonical example, where $\mathcal{A} = \{\pm \mathbf{e}_i\} \subseteq \mathbb{R}^n$. The simple set $\mathcal{K} = cone_k(\mathcal{A})$ correspond to the set $\Sigma_k(\mathbb{R}^n)$ of k-sparse vectors. Furthermore, introducing the notion of *atomic norm*, which is important in terms of compressed sensing:

Definition 2.2 (*Atomic Norm*)**.** The function

$$\|\mathbf{x}\|_{\mathcal{A}} = inf \left\{ \sum_{\mathbf{a} \in \mathcal{A}} c_{\mathbf{a}} : \mathbf{x} = \sum_{\mathbf{a} \in \mathcal{A}} c_{\mathbf{a}} \mathbf{a}, c_{\mathbf{a}} \geq 0 \ \forall \mathbf{a} \in \mathcal{A} \right\} \tag{23}$$

associated with an atomic set $\mathcal{A} \subset \mathbb{R}^n$ is called the atomic norm of $\mathcal{A}$ at $\mathbf{x}$.

A general strategy in compressed sensing is to recover or estimate simple sets through *atomic norm minimization* (ANM):

$$\min_x \quad \|\mathbf{s}\|_{\mathcal{A}} \quad \text{s.t.} \quad \mathbf{y} = \Theta \mathbf{s}. \tag{24}$$

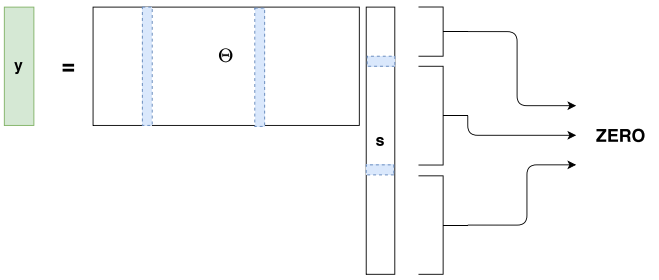


Fig. 8. The signal illustrated on matrix form. The blue columns of Θ corresponds to the blue nonzero coefficients in **s**. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3.2. Restricted isometry property

Consider the matrix $\Theta = \Phi\Psi$ in Eq. (20). In general, this matrix is rank deficient and hence loses information. However, Θ can be shown to preserve information in sparse and compressible signals if it satisfies the *restricted isometry property* (RIP).

Definition 2.3 (*RIP*)**.** A matrix **A** is said to satisfy the RIP of order k if

$$(1-\delta)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta)\|\mathbf{x}\|_2^2 \tag{25}$$

For all $\mathbf{x} \in \Sigma_k$ with $\delta > 0$.

The intuition of the RIP is that for any index set $\mathbf{S} \subset [n]$ with cardinality $|\mathbf{S}| \leq k$, the submatrix with columns of **A** indexed by **S**: $\mathbf{A}_S$ approximately acts like an isometry on the set of k-sparse vectors. Direct construction of the measurement matrix Φ such that $\Theta = \Phi\Psi$ involves verifying Eq. (25) for all $\binom{n}{k}$ sparse vectors $\mathbf{x} \in \Sigma_k$. However, RIP can be achieved with high probability by selecting Φ as a random matrix [7]. This implies that to successfully estimate the signal of interest, a reasonable sampling strategy would be to sample the variable of interest at random time instants.

2.3.3. Signal estimation techniques

Given the linear system $\mathbf{y} = \Theta\mathbf{s}$, there are infinitely many coefficient vectors **s** that is consistent with the $m \ll n$ number of measurements. Therefore, to find the correct or approximate solution **s**, it is necessary to exploit the a priori knowledge of sparsity or compressibility of the signal. This indicates minimizing the number of nonzero coefficients in **s** so that $\Theta\mathbf{s}$ still is consistent with the measurements **y**. In its most explicit form, this minimization is done through ℓ_0-minimization:

$$\min_s \quad \|\mathbf{s}\|_0 \quad \text{s.t.} \quad \mathbf{y} = \Theta\mathbf{s}. \tag{26}$$

The optimization program in Eq. (26) shows remarkable results as it is only dependent on $m = 2k$ independent measurements to recover **s** [40]. However, ℓ_0-minimization is provably NP-hard [41]. Furthermore, the solution of a ℓ_0-minimization problem can be highly sensitive to measurement noise and sparsity defects [39]. A key insight in compressed sensing is the convex relaxation of the optimization program in (26). The closest convex relaxation of (26) is the ℓ_1-minimization known as *Basis Pursuit* (BP) [42]:

$$\min_s \quad \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \mathbf{y} = \Theta\mathbf{s}. \tag{27}$$

The convex minimization problems such as the one in (27) guarantee a more stable solution that can also be solved in polynomial time. However, with this relaxation, there comes a cost in terms of increased number of required measurements, $m = \mathcal{O}(k \cdot log\,(n/k))$ as it relies on the RIP [40]. Both optimization programs in (26) and (27) are formulated for exact reconstruction of the signal **x** due to the equality constraint. To account for measurement noise, the equality constraint can be replaced by an inequality constraint such as in *Quadratic Constraint Basis Pursuit* (QCBP):

$$\min_s \quad \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \leq \eta \tag{28}$$

Here, η represents the noise level of the measurements. Therefore, it is desirable to estimate this noise level to get the most precise estimate of the signal of interest. Other convex programs are the *least-absolute shrinkage selection operator* (LASSO) [43], the *Dantzig selector* (DS) [44] and *basis pursuit denoising* (BPDN) [39]. Recovery guarantees are usually strongest for convex optimization programs. However, these programs become less practical as the problem increases in size. Thresholding algorithms represent a compromise between theoretical guarantees and efficient predictable running times. Thresholding algorithms can be divided into *hard* and *soft* thresholding. The following hard thresholding algorithm *hard thresholding pursuit* (HTP) has been involved in the scope of this article:

Algorithm 2: HTP

Result: $\mathbf{s}^i$
Input: Θ $\mathbb{R}^{m\times n}$, $\mathbf{y} \in \mathbb{R}^m$, $k \in [n]$;
Initializalize: $\mathbf{s}^0 \leftarrow 0$, $i \leftarrow 0$;
while *While condition* **do**
  instructions;
  $\mathbf{v}^{i+1} \leftarrow \mathbf{s}^n - \mu\Theta^T(\Theta\mathbf{s} - \mathbf{y})$ (Gradient descent step);
  $\mathbf{G}_{n+1} \leftarrow H_k(\mathbf{v}^{\mathbf{i+1}})$ (Support identification) ;
  $\mathbf{s}^{i+1}_{\mathbf{G}_{n+1}} \leftarrow \Theta^{\dagger}_{\mathbf{G}_{n+1}}\mathbf{y}$ (Least squares update) ;
  $i \leftarrow i + 1$;
end

Here, H_k is a hard thresholding operator, identifying the index set $G \subset [n]$ which support the k largest values of **s**, and zeroing out any values supported on $\bar{G}$. μ is a hyperparameter proportional to the gradient descent term. In the most basic case of the HTP algorithm, $\mu = 1$. In a more general case of the HTP algorithm, the hyperparameter $\mu \neq 1$. Another hard thresholding algorithm is Iterative hard thresholding (IHT) [45], which HTP is based on. The main difference between the two is that HTP converges faster than IHT [39]. Examples of soft thresholding algorithms are Smoothing proximal gradient method [46], and Fast iterative shrinkage-thresholding algorithm [47]. One of the most generic classifications of recovery algorithms split the algorithms into three different classes. Two of them are mentioned already, namely convex optimization and thresholding algorithms. Another class of recovery algorithms is iterative greedy algorithms [39]. Some of the most famous greedy methods are *Orthogonal matching pursuit* (OMP) [48] and Compressive sampling matching pursuit (CoSaMP) [49].

3. Method and data generation

3.1. Set-up for data generation and pre-processing

The cell dimensions in the simulated cell are assumed to be $l = 20$ m as the length and $w = 2$ m as the width. The initial side ledge thickness has been set to $x_{sl} = 0.08$ m and is assumed to be uniform along all sidewalls. The unmodeled dynamics are expressed as variations in the side ledge. They are assumed to be two cosine waves with frequencies 2[per day] and 5[per day] with associated amplitudes of respectively 0.02 m and 0.01 m. The fact that stationary periodical signals have been used to simulate unmodeled dynamics can be justified by the nature of the process. The aluminum electrolysis is a semi-batch process with periodical control inputs that induce periodical dynamics on the system states. Examples of these periodical control inputs are

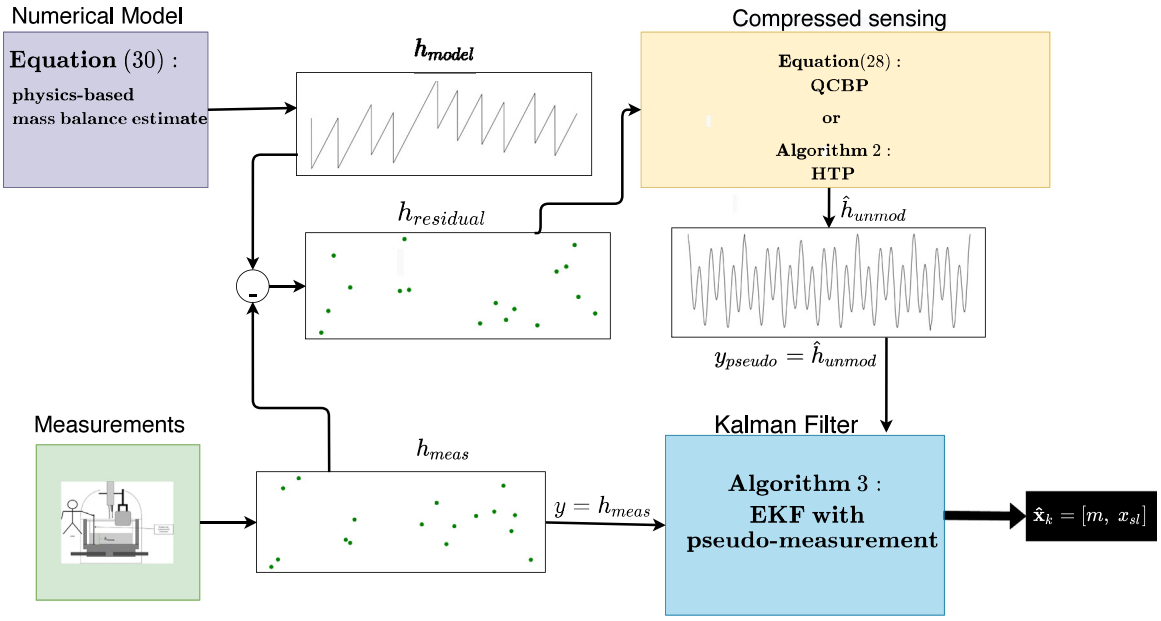


Fig. 9. Methodology: The numerical model generates high temporal resolution time-series of the height h_{model}. From the measurements, coarse resolution time-series of height h_{meas} is obtained. The difference between them when height measurements are taken is referred to as $h_{residual}$. $h_{residual}$, which has the same time resolution as h_{meas} is provided as an input to the compressed sensing algorithm to generate a signal estimate and hence high-resolution estimate h_{unmod} of the unmodeled dynamics expressed in the height signal. h_{unmod} is then provided to an EKF as a pseudo measurement to estimate the side ledge thickness. The numerical model f and the height measurement h_{meas} are also provided to the EKF to estimate the state vector $\mathbf{x}$.

alumina feed and anode change. This causes periodical dynamics in the side ledge thickness. The stationarity assumption is an approximation that can be justified for shorter periods. The initial mass of molten aluminum in the cell is set to $m_{al} = 14,700$ kg. Given the dimensions of the cell, the initial side ledge thickness, and the density of molten aluminum, this corresponds to an initial metal height of $h_m[0] = 0.183$ m. The line current is set to $I = 330$ kA and is assumed constant during the simulation. In the current work, a current efficiency of $CE = 0.95$ is used and assumed to be constant. Eq. (6) with these inputs yields the mass of aluminum produced in the cell to be 2524 kg/day ($\sim$ 2500 kg/day). The amount of aluminum mass tapped at each tapping is done according to the following control strategy:

$$m_{out} = \begin{cases} m_{ref} + k \cdot (h_{meas} - h_{ref}), & \text{if } h_{meas} \text{ not older than 5 hours} \\ m_{ref}, & \text{if } h_{meas} \text{ is older than 5 hours.} \end{cases} \quad (29)$$

where $m_{ref} = 2500$ kg and $h_{ref} = 18$ cm. h_{meas} is the measured metal height. Metal is tapped from the cell every $18 - 48$ hours. The line current I and current efficiency CE is used as input to Eq. (3) to generate a high resolution timeseries of the rate of generated aluminum mass.

As described in Section 2.3.2, having a random measurement matrix, Φ will ensure successful recovery of the signal with high probability. Therefore, measurements of the metal height are sampled at random time instants. The average sampling rate is varied for different simulations to test the limit of required data. This is also the case for the standard deviation of the measurement noise, which is varied to test the reconstruction algorithms robustness against noise. The measurement noise is assumed to be *Gaussian white*. In the simulation, the measurements are values of the simulated metal height chosen at random time instants. Each measurement has an added value drawn from a normal distribution with zero mean and a standard deviation chosen for that simulation. The simulations were conducted for a period corresponding to 100 days with a small timestep of $\Delta t = 5$ min.

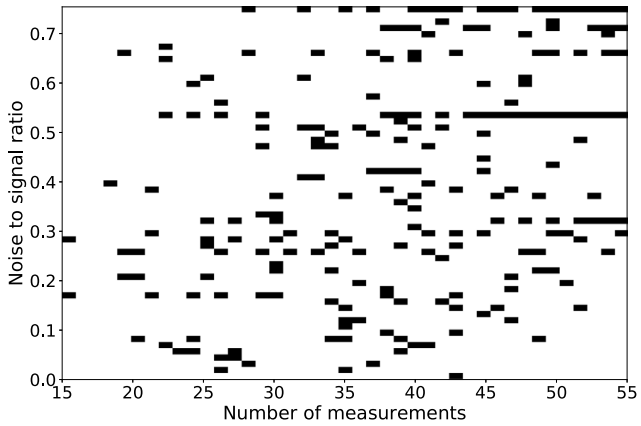
3.2. Novel hybrid framework

The proposed method is a novel hybrid approach that utilizes first principle system knowledge to manipulate a measured signal. The manipulated data is the residual between the measured signal and an estimate of the measured signal calculated by a physics-based model. This residual represents the unmodeled dynamics in the measured signal. The signal representing the unmodeled dynamics in the measured signal is much sparser than the measured signal itself. Therefore, much fewer measurements are required to estimate the unmodeled dynamics in the measured signal compared to estimating the measured signal with compressed sensing techniques. The novel method is limited to estimating stationary unmodeled dynamics. A compressed sensing technique is used to estimate the sparse residual and thereby gaining information about the periodic disturbances. This information is provided to an EKF as a pseudo measurement, leading to an increase in state estimation accuracy. Fig. 9 illustrates the novel hybrid framework developed in the work related to this article.

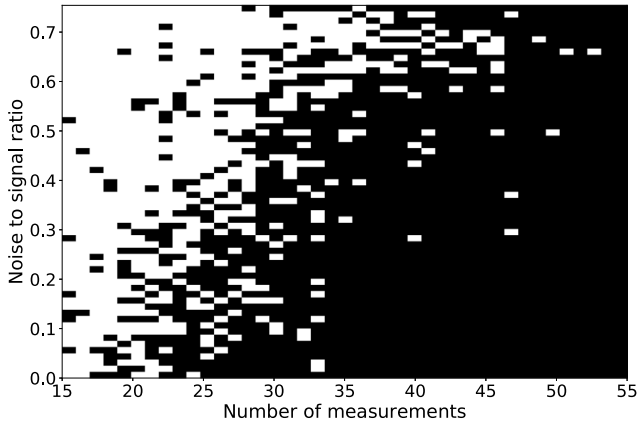
3.2.1. Sparsification of a signal

The dynamics of the metal height is in its nature a non-sparse signal in the discrete cosine transform (DCT). This is due to the discontinuities of the sawtooth shape in the signal. Fig. 6(b), which shows the DCT of the metal height signal, illustrates this point.

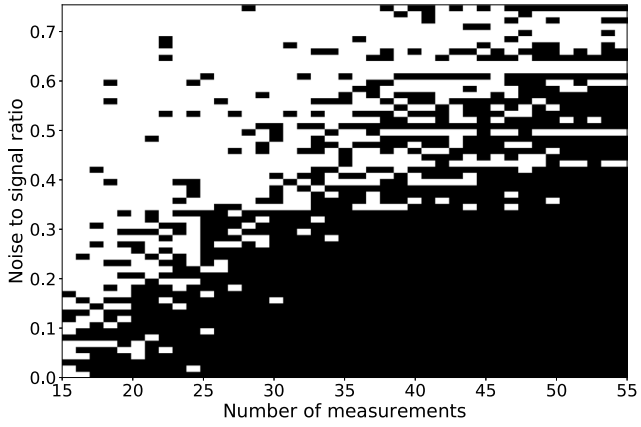
Moreover, the metal height signal is non-stationary due to the non-regular tapping of metal. The compressed sensing method used in this work estimates the frequency components of a signal based on measurements of that signal. Since the metal height signal is non-stationary, the frequency components will continuously vary. This makes it much more difficult to predict the signal in the future. Instead, by utilizing this first principle model based on Eq. (6) and the available knowledge about the amount of aluminum tapped from the cell, it is possible to sparsify the signal and remove non-stationarities from the signal.



(a) Performance plot of the HTP algorithm. Hyperparameter $\mu = 1$

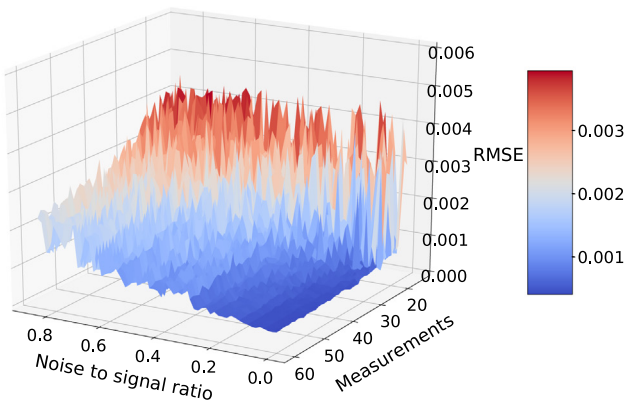


(b) Performance plot of the HTP algorithm. Hyperparameter $\mu = 700$

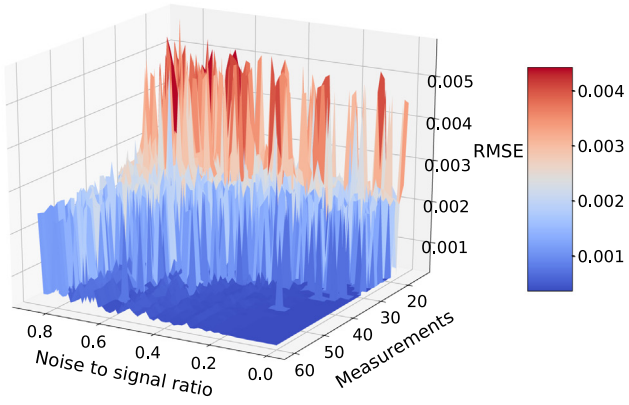


(c) Performance plot of QCBP

Fig. 10. Performance plot of the HTP algorithm and QCBP optimization program. On the horizontal axis, the number of measurements of metal height in the simulation varies, while the noise to signal ratio varies along the vertical axis. The noise to signal ratio is defined as the ratio between the standard deviation of the measurement noise and the average amplitude of the unmodeled dynamics signal for any given simulation. For a given noise to signal ratio and a given number of measurements, the color black indicates that the reconstruction algorithm found the correct support for the signal.



(a) RMSE for QCBP estimation technique



(b) RMSE for the HTP algorithm. Hyperparameter $\mu = 700$

Fig. 11. RMSE for simulations with varying measurement noise and number of measurements used in estimation of the signals.

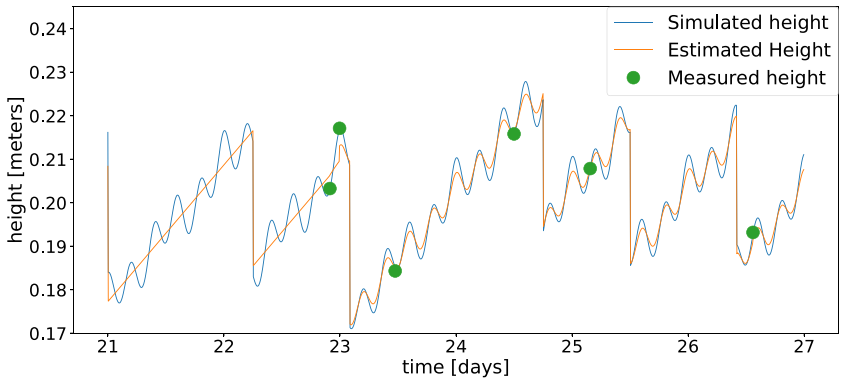
The idea is to estimate the signal h_{unmod} in Fig. 7(b) that represent the unmodeled dynamics in the metal height signal by using the data points $h_{residual}$ as input in a compressed sensing algorithm. Subtracting the estimated signal h_{model} based on the physical model from the measured metal height h_{meas} gives a signal $h_{residual}$ representing the unmodeled dynamics and measurement noise. This signal h_{unmod} is much more sparse in the Discrete Cosine transform (DCT) domain than the metal height signal h_m. This can be seen by comparing Fig. 7(c) and Fig. 6(b). Given that DCT is used as a basis, estimating the new signal h_{unmod} with compressed sensing requires much fewer datapoints than what is required for estimating the metal height h_m with compressed sensing. Furthermore, by including a first principle model in the estimation process and leaving estimation of only the unmodeled dynamics to compressed sensing, the robustness of the state-of-the-art estimate is preserved. Fig. 8 illustrates the signal representation in terms of the matrix representation from Fig. 4(b).

3.2.2. Integrated solution

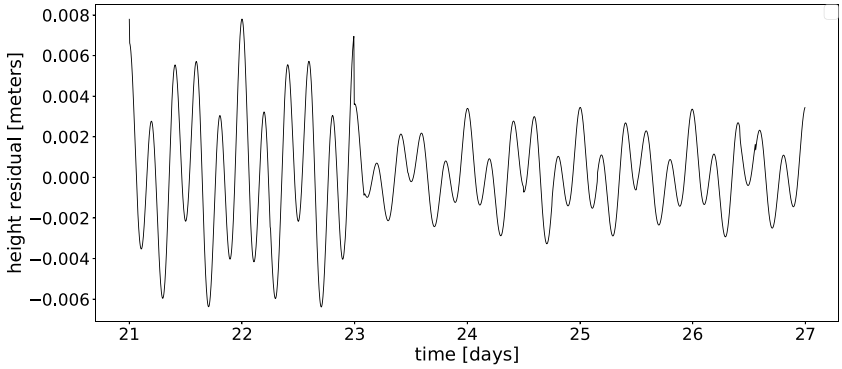
In Fig. 9, a schematic representation of the integrated solution is presented. The integrated solution is composed of a physics-based model f, a compressed sensing signal estimation algorithm, a Kalman filter, and a metal height measurement. The physics-based model is as follows:

$$f = \frac{dm}{dt} = \dot{m}_{in} - \dot{m}_{out} = \frac{CE \cdot M_{Al}}{F \cdot z} I[k] - \dot{m}_{tapped}[k]. \tag{30}$$

f estimates the mass flow in the electrolytic cell based on the inputs from the line current $I[k]$ and flow of tapped metal $\dot{m}_{tapped}[k]$



(a) Simulated and estimated height



(b) Height residual.

Fig. 12. In (a), the blue graph represents the simulated metal height h_m, whereas the orange graph represents the estimated metal height, estimated by the Kalman filter. The green points are the measured values of the metal height. In (b), the residual between simulated metal height h_m and Kalman filter-estimated metal height $\hat{h}_m$ is shown. The unmodeled dynamics in the metal height are estimated from the QCBP program. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

at each time instant k. Given an estimate of the mass of aluminum and displacement volume in the cell, a metal height estimate can be calculated. This gives the first principle estimate h_{model} visualized in Fig. 7(a). Since the dynamics of the side ledge thickness affects the displacement volume in the cell, it also affects the metal height in the cell. The lack of a model estimating the dynamics of the side ledge thickness causes the unmodeled dynamics in the height signal h_{unmod} visualized in Fig. 7(b). The compressed sensing signal estimation method estimates the unmodeled dynamics expressed in the height signal h_{unmod} using the manipulated datapoints $h_{residual}$ in Fig. 7(b). This estimate is then used as a pseudo measurement for the Kalman filter to estimate the side ledge thickness x_{sl}. In terms of a state-space model, the input vector is defined as:

$$\mathbf{u}_k = [u_{1,k},\ u_{2,k}] = \left[\frac{\dot{m}_{in,k}}{\rho},\ \frac{\dot{m}_{out,k}}{\rho}\right], \tag{31}$$

the state vector is defined as:

$$\mathbf{x}_k = [x_{1,k},\ x_{2,k}] = [m,\ x_{sl}], \tag{32}$$

and the measurement is defined as

$$y = h_{meas}. \tag{33}$$

This gives the following state-space model:

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}u_1 - u_2 + v_1\\ v_2\end{bmatrix}, \tag{34}$$

$$y = \frac{x_1}{Area(x_2)} + w. \tag{35}$$

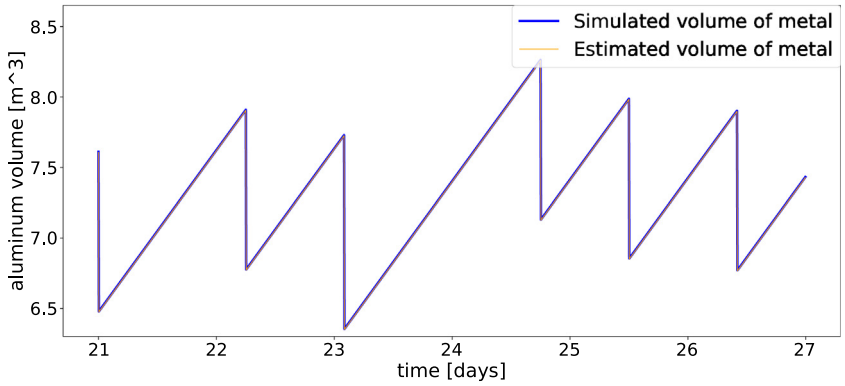
$\mathbf{v} = [v_1,\ v_2]$ is process noise, whereas w is the measurement noise.

Furthermore, the pseudo measurement representing the estimated unmodeled dynamics is defined as:

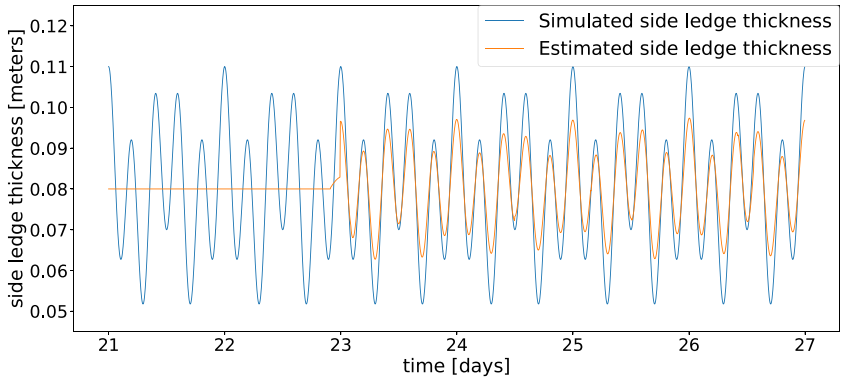
$$y_{pseudo} = \hat{h}_{unmod}. \tag{36}$$

y_{pseudo} is used as a high-frequency measurement when a signal estimate is available. The pseudo measurement has corresponding Kalman filter Covariance matrices for the pseudo measurement $\mathbf{Q}_{pseudo}$ and $\mathbf{R}_{pseudo}$. Since the physics-based model is assumed to have great abilities in estimating the mass balance of the cell, the element of $\mathbf{Q}_{pseudo}$ representing the process noise covariance for the mass estimate Q_{11} is set to zero. The element of $\mathbf{Q}_{pseudo}$ representing the process noise covariance of the side ledge thickness Q_{22} is set to one. The measurement noise covariance $\mathbf{R}_{pseudo}$ is much smaller than Q_{22}, indicating that the pseudo measurement y_{pseudo} is trusted much more than the *a priori* estimate of the side ledge thickness. Thus, the *a posteriori* estimate of the side ledge thickness will be greatly influenced by the pseudo measurement. Since $Q_{11} = 0$, the *a posteriori* estimate of the mass will not be influenced by the pseudo measurement.

Algorithm 3 describes how the pseudo measurements y_{pseudo} are incorporated into the EKF. y_{pseudo} is treated as a measurement in the EKF. In a regular measurement update, when the a posteriori estimate $\hat{\mathbf{x}}_k$ is calculated, a model estimate $h(\hat{\mathbf{x}}_k^-, \mathbf{u}_k)$ of the variable $\mathbf{x}$ is subtracted from the measurement to include both measurement and model estimate in the posterior estimate. However, since y_{pseudo} is the estimate of the unmodeled dynamics,



(a) Simulated and estimated volume



(b) Simualted and estimated side ledge thickness

Fig. 13. In (a), the simulated and estimated volume of aluminum in the cell is shown. The simulated thickness of the side ledge is shown in (b) Simulated values are in blue while estimated values are in orange. The side ledge thickness is estimated from the QCBP program. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Algorithm 3: EKF with pseudo measurement

Time update:;

$\hat{\mathbf{x}}_{k+1}^{-} = \hat{\mathbf{x}}_k + \Delta t \cdot f(\hat{\mathbf{x}}_k, \mathbf{u}_k)$;

$\mathbf{P}_{k+1}^{-} = \mathbf{A}_k \mathbf{P}_k \mathbf{A}_k^T + \mathbf{Q}_k$;

if *Meaurement* y_k *is available* **then**

  Measurement update:;

  $\mathbf{K}_k = \mathbf{P}_k^{-} \mathbf{H}_k (\mathbf{H}_k \mathbf{P}_k^{-} \mathbf{H}_k^T + \mathbf{R}_k)^{-1}$;

  $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^{-} + \mathbf{K}_k(\mathbf{y}_k - h(\hat{\mathbf{x}}_k^{-}, \mathbf{u}_k))$;

  $\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{P}_k^{-}$;

else

  $\mathbf{K}_{k,pseudo} = \mathbf{P}_{k,pseudo}^{-} \mathbf{H}_k (\mathbf{H}_k \mathbf{P}_{k,pseudo}^{-} \mathbf{H}_k^T + \mathbf{R}_{k,pseudo})^{-1}$;

  $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^{-} + \mathbf{K}_{k,pseudo} \mathbf{y}_{k,pseudo}$;

  $\mathbf{P}_{k,pseudo} = (\mathbf{I} - \mathbf{K}_{k,pseudo} \mathbf{H}_k)\mathbf{P}_{k,pseudo}^{-}$

end

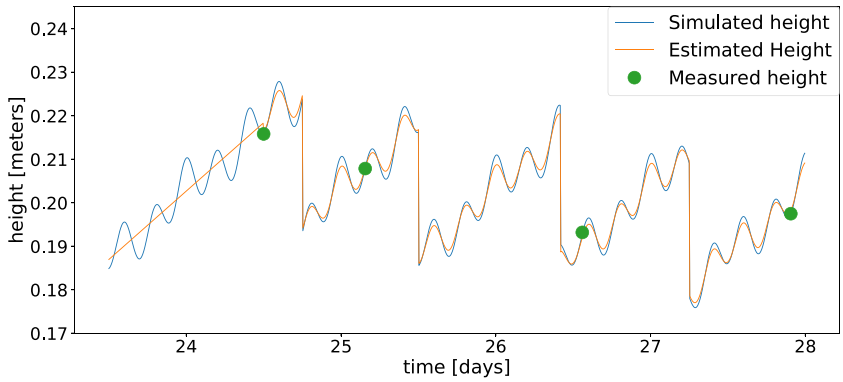
there is no model estimate of this signal. Thus, the posterior estimate after the pseudo measurement is included only depends on the a priori estimate and the pseudo-measurement, see Algorithm 3.

4. Results

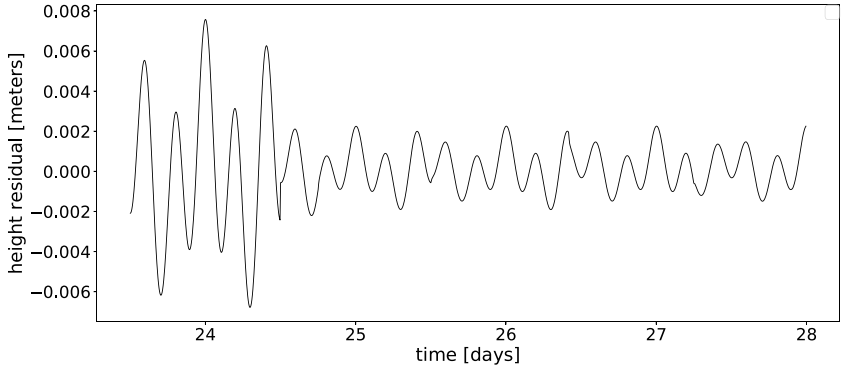
In this section, the results from a case study of the hybrid method explained in Sections 3.2.1 and 3.2.2 are presented. The data generation of the simulated data used in the case study is described in Section 3.1. The unmodeled dynamics that the hybrid modeling framework aims to estimate is a periodical, stationary signal composed of two frequency components. As stated in the introduction, it is of interest to minimize the number of measurements and at the same time improve predictive modeling. Therefore Section 4.1 investigates the number of measurements required to successfully estimate the unmodeled dynamics with the novel hybrid method. Furthermore, the robustness against measurement noise of the method is assessed. Section 4.2 illustrates how the estimate of the unmodeled dynamics provided by the novel method affects the state estimation in the Kalman filter. Throughout the study, two different compressed sensing techniques were applied. These are the QCBP optimization program and two versions of the HTP algorithm.

4.1. Noise and measurement study

The performance of the hybrid modeling approach given two different compressed sensing techniques is assessed in this section. The performance measure in Fig. 10 is a binary value stating if the correct signal support for the unmodeled dynamics was found by the compressed sensing algorithm or not. Signal support means the coefficients defining the signal in the sparse, transformed domain. The performance measure in Fig. 11 is the rooted mean square error (RMSE). The performance in both Fig. 10 and Fig. 11 is tested for several measurements and noise levels. The HTP algorithm explicitly requires that the algorithm search



(a) Simulated and estimated height

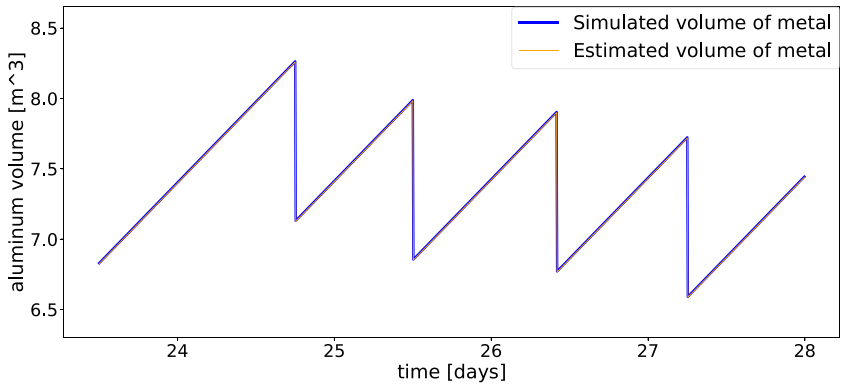


(b) Height residual.

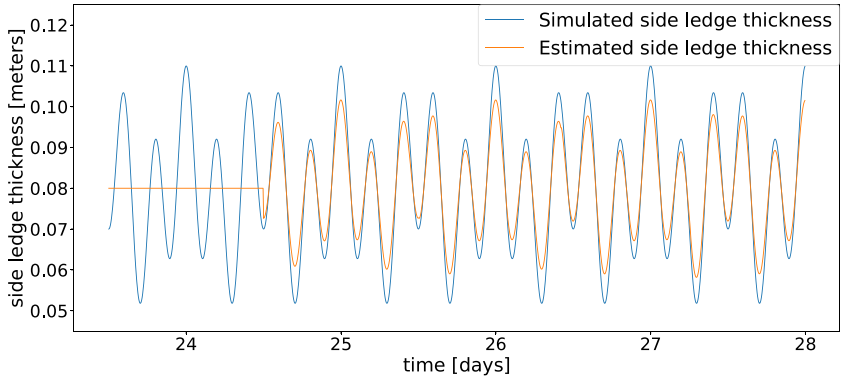
Fig. 14. In (a), the blue graph represent the simulated metal height h_m, whereas the orange graph is the estimated metal height, estimated by the Kalman filter. The green points are the measured values of the metal height. In (b), the residual between simulated metal height h_m and Kalman filter-estimated metal height $\hat{h}_m$ is shown. The unmodeled dynamics in the metal height is estimated from the HTP algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for a signal support with a given number of coefficients. Since the QCBP program is an optimization program minimizing the ℓ_1-norm of the support coefficients $\mathbf{s}$ with constraints on the quadratic error between measurement and estimated signal, it is not expected that the program will find a signal support with the exact same number of coefficients as the correct solution. In general, some coefficients that are not part of the correct solution can be expected to be included in the estimate calculated by the QCBP program. Therefore, the requirement for the QCBP program to succeed in the performance test presented in Fig. 10 is that it finds the correct support and that the largest erroneous estimated coefficient is smaller than 0.3 times the smallest of the correct coefficients.

Fig. 10 shows how QCBP and HTP estimation strategies perform to find the correct support or basis coefficients in the simulations with different levels of measurement noise and different number of measurements used in the estimation. In the HTP algorithm, tuning the hyperparameter μ turns out to be of great importance. The amplitude of the unmodeled dynamics estimated is small in magnitude ($< 0.01[m]$). Therefore, the gradient descent term $\mu\Theta^T(\Theta\mathbf{s}-\mathbf{y})$ for the basic implementation of HTP with $\mu = 1$ becomes very small after the first iteration and converges. Implementing the HTP algorithm with $\mu \gg 1$ estimates the correct support in many more cases than the basic test with $\mu = 1$. Fig. 10(b) and Fig. 10(c) show that the success of the method is dependent on both the number of measurements used and the measurement noise. The figures show a clear tendency that the number of measurements required for a successful signal estimation increases with increasing noise. Fig. 11 shows the RMSE for the estimated signals from the QCBP program in Fig. 11(a) and the RMSE for HTP algorithm with $\mu = 700$ in Fig. 11(b). Simulations with different levels of noise and number of measurements are included in the test. Fig. 10 and Fig. 11 are results from the same test. Fig. 11(a) shows that the RMSE increases proportionally with the measurement noise for the QCBP program. The explanation for this can be that the optimization program in QCBP has an inequality constraint allowing for a certain maximum quadratic error between the estimate and the measured values proportional to the standard deviation of the noise. Furthermore, comparison of Figs. 10(c) and 11(a) indicates that the RMSE does not seem to be significantly dependent on the number of measurements used to estimate the signal as long as the correct support is estimated. The same can be said about the HTP algorithm, which has consistently low values of the RMSE as long as the correct support is found. Figs. 10(b) and 11(b) show that the RMSE is significantly lower when the correct support is found compared to when the correct support is not found. The QCBP program has an inequality constraint that allows for a certain quadratic error between measured and estimated values that limits the RMSE regardless of if the correct support is found or not. The HTP algorithm does not have this limitation in RMSE in the search for correct support. Hence, in general, the RMSE will be significantly larger if HTP finds the wrong support for the signal. For the QCBP program, this difference is not that clear due to the quadratic constraint.



(a) Simulated and estimated volume



(b) Simualted and estimated side ledge thickness

Fig. 15. In (a), the simulated and estimated volume of aluminum in the cell is shown. The simulated thickness of the side ledge is shown in (b). Simulated values are in blue while estimated values are in orange. The side ledge thickness is estimated from the HTP algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. State and signal estimates

In this section, the estimated metal height, height residual as well as the state estimates of aluminum mass and side ledge thickness is presented. In Figs. 12 and 13, simulated values and estimates when using the QCBP optimization are presented, whereas in Figs. 14 and 15 simulated values and estimates when using the HTP algorithm are presented. In both cases, a noise to signal ratio equal to 0.215 was used. The time interval for the figures includes the measurement from which the compressed sensing reconstruction algorithm finds the correct support of the unmodeled dynamics. In the case presented, the HTP algorithm needed a few more measurements than the QCBP algorithm to find the correct support for the unmodeled dynamics. That is, the QCBP program needed 25 measurements of the signal to find the correct support for the signal while the HTP algorithm needed 28 measurements to find the correct support for the signal. Therefore, the time frame shown in Figs. 12 and 13 differs from the time frame in Figs. 14 and 15.

Fig. 12(a) shows the simulated and the estimated metal height. The unmodeled dynamics are estimated by the QCBP optimization program using data points $h_{residual}$, plotted as green points in Fig. 7(b). The estimate is then fed into the Kalman Filter as a pseudo measurement, where the covariance matrices $\mathbf{Q}_{pseudo}$ and $\mathbf{R}_{pseudo}$ are tuned such that only the side ledge thickness $\mathbf{x}_{sl}$ is updated. Fig. 12(b) shows the residual between simulated and estimated metal height. Fig. 12(a) shows that the estimated metal height changes character and simultaneously follows the simulated metal height as the green data point is measured at day 23. Looking at Fig. 12(b), it is clear to see the height residual $h_m - \hat{h}_m$ decreases significant at this point.

Fig. 13 shows the estimated and simulated states $\mathbf{x}$ as they are defined in the EKF. The estimate of the unmodeled dynamics is based on the same estimate as in Fig. 12.

Fig. 14 corresponds to Fig. 12 and Fig. 15 corresponds to Fig. 13. The difference is that the estimate in Figs. 14 and 15 is calculated by the HTP algorithm, while the estimates in Figs. 12 and 13 is calculated by the QCBP program. The HTP algorithm estimates a signal with a much smaller RMSE than the QCBP algorithm for this specific case. Fig. 11, shows that this is the case for the simulations where the correct support is found.

5. Conclusion

This article introduces a novel hybrid modeling method that addresses the problem of estimating stationary, periodical unmodeled dynamics from a non-sparse, non-stationary signal measured at low sampling rates. The two research questions investigated have been satisfactorily addressed. These are pointed out below:

- The first research question sought an answer regarding the potential manipulation of the coarsely sampled signal so that compressed sensing techniques could be utilized for signal estimation. It was observed that most of the non-sparsity in the measured signal was due to the linear increase caused by the aluminum production and discontinuities due to sudden drops of the amplitude in the signal resulting from a regular tapping of the molten aluminum. This

resulted in the measurement signal being non-sparse even in the frequency domain. Fortunately, the cause and effect of these linear increases and discontinuities were very well captured by the physics-based model. Simply subtracting the estimated metal height signal based on a physics-based model from the measured metal height signal yields manipulated measurements representing a new signal, namely the unmodeled dynamics of the metal height. As shown in Section 4.1, compressed sensing show promising results in estimating this residual signal from a limited number of measurements with Gaussian measurement noise.

- The second research question pertained to utilizing the estimated signal in a Kalman filter to improve the accuracy of the state estimation. The proposed method answered this by including the estimated signal as a pseudo measurement into the Kalman filter. The pseudo measurement is treated as a separate high-resolution measurement with corresponding Covariance matrices tuned according to the uncertainty about the state estimates. It was demonstrated that the state estimate of one of the variables improved significantly when a signal estimate of the unmodeled dynamics is available.

In the proposed approach, only stationary unmodeled dynamics are considered. This is because the compressed sensing techniques used in the method only consider stationary signals.

CRediT authorship contribution statement

Erlend Torje Berg Lundby: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Adil Rasheed:** Conceptualization, Methodology, validation, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Jan Tommy Gravdahl:** Validation, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Ivar Johan Halvorsen:** Validation, Writing - original draft, Writing - review & editing, Supervision, Project administration.

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.

Acknowledgments

This work was supported by the industry partners Borregaard, Norway, Elkem, Norway, Hydro, Yara, Norway and the Research Council of Norway through the project TAPI: Towards Autonomy in Process Industries, project number 294544.

References

[1] K. Grotheim, H. Kvande, Introduction To Aluminium Electrolysis-Understanding the Hall-Heroult Process, Aluminium-Verlag, Dusseldorf, Germany, 1993.
[2] H.S. Li, C.W. Jiang, Development and application of soft sensor model for heterogeneous information of aluminum reduction cells, Control Eng. Pract. 19 (10) (2011).
[3] Z. Zeng, W. Gui, X. Chen, Y. Xie, R. Wu, A mechanism knowledge-driven method for identifying the pseudo dissolution hysteresis coefficient in the industrial aluminium electrolysis process, Control Eng. Pract. 102 (2020).
[4] S. Kolås, S.O. Wasbø, A nonlinear model based control strategy for the aluminium electrolysis process, in: G. Bearne, M. Dupuis, G. Tarcy (Eds.), Essential Readings in Light Metals: Volume 2 Aluminum Reduction Technology, Springer International Publishing, Cham, 2016, pp. 825–829.
[5] S. Kolås, Estimation in Nonlinear Constrained Systems with Severe Disturbances (Ph.D. thesis), Norwegian University of Science and Technology, 2008.
[6] C.E. Shannon, Communication in the presence of noise, Proc. IEEE 86 (2) (1998) 447–457.
[7] R.G. Baraniuk, Compressive sensing, IEEE Signal Process. Mag. 24 (4) (2007).
[8] S. Foucart, H. Rauhut, An invitation to compressive sensing, in: Applied and Numerical Harmonic Analysis, 2013, no. 9780817649470.
[9] M. Rani, S.B. Dhok, R.B. Deshmukh, A systematic review of compressive sensing: Concepts, implementations and applications, IEEE Access 6 (2018) 4875–4894.
[10] B.M. Sanandaji, T.L. Vincent, M.B. Wakin, R. Tóth, K. Poolla, Compressive system identification of LTI and LTV ARX models, in: 2011 50th IEEE Conference on Decision and Control and European Control Conference, 2011, pp. 791–798.
[11] Z. Bai, E. Kaiser, J.L. Proctor, J.N. Kutz, S.L. Brunton, Dynamic mode decomposition for compressive system identification, AIAA J. 58 (2) (2020) 561–574.
[12] Y. Lu, Y. Wang, Monitoring temperature in additive manufacturing with physics-based compressive sensing, J. Manuf. Syst. 48 (2018) 60–70.
[13] R. Heckel, H. Bolcskei, Identification of sparse linear operators, IEEE Trans. Inform. Theory 59 (12) (2013).
[14] Y. Kopsinis, K. Slavakis, S. Theodoridis, Online sparse system identification and signal reconstruction using projections onto weighted $\ell 1$ balls, IEEE Trans. Signal Process. 59 (3) (2011).
[15] Y. Chen, Y. Gu, A.O. Hero, Sparse LMS for system identification, in: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2009.
[16] Y. Gu, J. Jin, S. Mei, $\ell 0$ norm constraint LMS algorithm for sparse system identification, IEEE Signal Process. Lett. 16 (9) (2009).
[17] N. Kalouptsidis, G. Mileounis, B. Babadi, V. Tarokh, Adaptive algorithms for sparse system identification, Signal Process. 91 (8) (2011).
[18] J. Antille, R. Von Kaenel, L. Bugnion, Hall-Heroult cell simulator: A tool for the operation and process control, in: Light Metals, Wiley, 2016, pp. 617–622.
[19] V. Gusberti, Modelling the mass and energy balance of aluminium reduction cells (Ph.D. thesis), University of New South Wales, 2014.
[20] I. Tabsh, M. Dupuis, Simulation of the dynamic response of aluminum reduction cells, in: Light Metals: Proceedings of Sessions, TMS Annual Meeting, Warrendale, Pennsylvania, 1997.
[21] P. Mandin, R. Wüthrich, H. Roustan, Industrial aluminium production: the Hall-Heroult process modelling, ECS Trans. 19 (26) (2009) 1–10.
[22] Z. Ge, Review on data-driven modeling and monitoring for plant-wide industrial processes, Chemometr. Intell. Lab. Syst. 171 (September) (2017) 16–25.
[23] Z. Ge, Z. Song, S.X. Ding, B. Huang, Data mining and analytics in the process industry: The role of machine learning, IEEE Access 5 (2017).
[24] S. Zhang, T. Zhang, Y. Yin, W. Xiao, Alumina concentration detection based on the kernel extreme learning machine, Sensors 17 (9) (2017).
[25] G. Yan, X. Liang, Predictive models of aluminum reduction cell based on LS-SVM, in: Proceedings - 2010 International Conference on Digital Manufacturing and Automation, ICDMA 2010, vol. 2, 2010.
[26] P.R. Chermont, F.M. Soares, R.C. De Oliveira, Simulations on the bath chemistry variables using neural networks, in: TMS Light Metals, 2016-January, 2016.
[27] A.M.F. de Souza, F.M. Soares, M.A.G. de Castro, N.F. Nagem, A.H.d.J. Bitencourt, C.d.M. Affonso, R.C.L. de Oliveira, Soft sensors in the primary aluminum production process based on neural networks using clustering methods, Sensors 19 (23) (2019).
[28] F. Frost, V. Karri, Identifying significant parameters for Hall-Heroult process using general regression neural networks, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 1821, 2000.
[29] A. Rasheed, O. San, T. Kvamsdal, Digital twin: Values, challenges and enablers from a modeling perspective, IEEE Access 8 (2020) 21980–22012.
[30] M. von Stosch, R. Oliveira, J. Peres, S. Feyo de Azevedo, Hybrid semi-parametric modeling in process systems engineering: Past, present and future, Comput. Chem. Eng. 60 (2014) 86–101.
[31] M.S.F. Bangi, J.S.I. Kwon, Deep hybrid modeling of chemical process: Application to hydraulic fracturing, Comput. Chem. Eng. 134 (2020).
[32] D. Lee, A. Jayaraman, J.S. Kwon, Development of a hybrid model for a partially known intracellular signaling pathway through correction term estimation and neural network modeling, PLoS Comput. Biol. 16 (12) (2020) 1–31.
[33] J. Hasenauer, S. Waldherr, M. Doszczak, P. Scheurich, N. Radde, F. Allgöwer, Analysis of heterogeneous cell populations: A density-based modeling and identification framework, J. Process Control 21 (10) (2011) 1417–1425.
[34] A.M. Ivanova, P.A. Arkhipov, O.Y. Tkacheva, Y.P. Zaikov, Experimental studies of the dynamic formation of the side ledge in an aluminum electrolysis cell, Russ. Metall. (Metally) 2020 (2) (2020).
[35] H. Viumdal, S. Mylvaganam, Beyond the dip stick: Level measurements in aluminum electrolysis, J. Miner. 62 (11) (2010).

[36] H. Viumdal, S. Mylvaganam, D. Di Ruscio, System identification of a non-uniformly sampled multi-rate system in aluminium electrolysis cells, Model. Identif. Control 35 (3) (2014) 127–146.
[37] L. Kiss, V. Dassylva-Raymond, Freeze thickness in the aluminum electrolysis cells, in: Light Metals 2008 - Proceedings of the Technical Sessions Presented By the TMS Aluminum Committee At the TMS 2008 Annual Meeting and Exhibition, TMS, New Orleans, LA., 2008, pp. 431–436.
[38] G. Welch, G. Bishop, An introduction to the Kalman filter, In Practice 7 (1) (2006) 1–16.
[39] N. Koep, A. Behboodi, R. Mathar, An introduction to compressed sensing, in: Applied and Numerical Harmonic Analysis, Springer International Publishing, 2019, pp. 1–65.
[40] R.G. Baraniuk, V. Cevher, M.F. Duarte, C. Hegde, Model-based compressive sensing, IEEE Trans. Inform. Theory 56 (4) (2010) 1982–2001.
[41] B.K. Natarajan, Sparse approximate solutions to linear systems, SIAM J. Comput. 24 (2) (1995) 227–234.
[42] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[43] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1) (1996) 267–288.
[44] E. Candes, T. Tao, The dantzig selector: Statistical estimation when p is much larger than n, Ann. Statist. 35 (6) (2007) 2313–2351.
[45] T. Blumensath, M.E. Davies, Iterative hard thresholding for compressed sensing, Appl. Comput. Harmon. Anal. 27 (3) (2009) 265–274.
[46] X. Chen, Q. Lin, S. Kim, J.G. Carbonell, E.P. Xing, et al., Smoothing proximal gradient method for general structured sparse regression, Ann. Appl. Stat. 6 (2) (2012) 719–752.
[47] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (1) (2009) 183–202.
[48] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inform. Theory 53 (12) (2007) 4655–4666.
[49] D. Needell, J.A. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal. 26 (3) (2009) 301–321.