Identifying Complex Dynamics of Power Grid Frequency

Xinyi Wen
Institute for Automation and Applied Informatics (IAI)
Karlsruhe Institute of Technology
Germany
xinyi.wen@kit.edu

Ulrich Oberhofer
Institute for Automation and Applied Informatics (IAI)
Karlsruhe Institute of Technology
Germany
ulrich.oberhofer@kit.edu

Leonardo Rydin Gorjão
Faculty of Science and Technology
Norwegian University of Life Sciences
Norway
leonardo.rydin.gorjao@nmbu.no

G.Cigdem Yalcin
Department of Physics
Istanbul University
Turkey
gcyalcin@istanbul.edu.tr

Veit Hagenmeyer
Institute for Automation and Applied Informatics (IAI)
Karlsruhe Institute of Technology
Germany
veit.hagenmeyer@kit.edu

Benjamin Schäfer
Institute for Automation and Applied Informatics (IAI)
Karlsruhe Institute of Technology
Germany
benjamin.schaefer@kit.edu

ABSTRACT

The energy system is undergoing rapid changes to integrate a growing number of intermittent renewable generators and facilitate the broader transition toward sustainability. This increases the complexity of the energy system in many aspects, including the power grid and its dynamics. As millions of consumers and thousands of (volatile) generators are connected to the same synchronous grid, no straightforward bottom-up models describing the dynamics are available on a continental scale comprising all of these necessary details. Hence, to identify this unknown power grid dynamics, we propose to leverage the Sparse Identification of Nonlinear Dynamics (SINDy) method. Thereby, we unveil the governing equations underlying the dynamical system directly from data measurements. Investigating the power grids of Iceland, Ireland and the Balearic islands as sample systems, we observe structurally similar dynamics with remarkable differences in both quantitative and qualitative behavior. Overall, we demonstrate how complex, i.e. non-linear, noisy, and time-dependent, dynamics can be identified straightforwardly.

CCS CONCEPTS

• **Applied computing** → **Mathematics and statistics**; **Engineering**; • **Mathematics of computing** → **Mathematical analysis**; • **Hardware** → **Power and energy**; • **Computing methodologies** → *Machine learning*; *Artificial intelligence*.

KEYWORDS

Power grid frequency, symbolic regression, sparse identification of non-linear dynamics (SINDy), time-dependent dynamics, noisy dynamics

This work is licensed under a Creative Commons Attribution International 4.0 License.

E-Energy '24, June 04–07, 2024, Singapore, Singapore

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0480-2/24/06
https://doi.org/10.1145/3632775.3661944

ACM Reference Format:
Xinyi Wen, Ulrich Oberhofer, Leonardo Rydin Gorjão, G.Cigdem Yalcin, Veit Hagenmeyer, and Benjamin Schäfer. 2024. Identifying Complex Dynamics of Power Grid Frequency. In *The 15th ACM International Conference on Future and Sustainable Energy Systems (E-Energy '24), June 04–07, 2024, Singapore, Singapore.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3632775.3661944

1 INTRODUCTION

Energy systems are presently experiencing a swift shift toward a more environmentally sustainable future. The stable functioning of our society relies heavily on the electrical power system and the maintenance of a stable power grid, transporting power from distributed generators to consumers. The power grid frequency, which indicates the number of oscillations per second on alternating current, is a crucial indicator of the equilibrium between electricity supply and demand, serving as a vital factor in the reliability of energy systems [21]. The single value of frequency is used throughout the grid and represents the center-of-inertia (COI) frequency[31]. Power grid frequency typically remains within a few percentage points of its reference value of 50 Hz or 60 Hz, respectively [6, 26]. Deviations from these respective standard frequencies indicate an imbalance of the system and can cause additional costs for grid operators as control systems need to be activated. For example, frequency values below the reference indicate an abundance of demand and hence require additional generation to be deployed. Furthermore, frequency deviations can impact the performance of electronic devices, industrial processes, and other critical power system components. Hence, understanding, modeling, and forecasting power grid frequency dynamics and statistics is crucial [26].

To transition towards a more sustainable energy system, flexible dispatchable fossil generators are being replaced by renewable sources, such as wind and solar power. In contrast to traditional power sources, renewables introduce a fluctuating and uncertain (non-deterministic) power supply, and are often coupled via power electronics, lowering the overall inertia in the power grid. The non-deterministic generation and the reduced inertia increase volatility in the power system and raise the need for additional control and balancing capacity. However, the interplay of the numerous different actors in the same cyber-physical system makes the overall

dynamics intransparent and almost impossible to estimate using detailed models of all components (bottom-up), especially since access to all these devices and their data is rarely possible [14, 23, 32].

Power grids are complex systems: They are driven by both stochastic and deterministic influences, i.e. we have to cope with potentially time-dependent and noisy dynamics. Furthermore, while many control laws are linear by design [21], their interplay might lead to non-linear effective equations [24, 36]. The interplay between stochastic (random) and deterministic influences creates a challenging environment where traditional linear models fall short. Therefore, we require an accessible, flexible, and data-driven approach, to derive linear and nonlinear differential models of the power-grid frequency. Accurate modeling of these governing equations is crucial for optimizing grid stability, e.g. to identify and potentially adjust the amplitude of power imbalances and control. Neural-network-based approaches can process high-dimensional data and enhance data utility. However, they are often regarded as black-box methods due to their complex internal workings and lack of interpretability [11, 19].

In the current study, we propose to identify the complex dynamics of the power grid frequency by utilizing the Sparse Identification of Nonlinear Dynamics (SINDy) method [2]. The key approach of SINDy is using sparse regression to estimate the most relevant parameters of the governing equations for dynamical systems. This technique offers a systematic and efficient means of uncovering the underlying dynamics and relationships within the power grid frequency, allowing for a more nuanced understanding and improved parameter estimation within the data-driven framework.

Since the first publication of Brunton et al. [2] the concept of SINDy has been applied and adapted on application in various areas of research. In [1] and [4] SINDy has been applied for stochastic dynamics, in [5] a sparse framework is implemented for creating data-driven physics-informed models. Other examples, where SINDy is successfully used to describe nonlinear dynamics can be seen in [22] for biological networks as well as for nonlinear dynamics with control [3]. In the field of power systems, only few studies explored SINDy, e.g. for state estimation in [17, 18, 30]. However, many previously considered (power) systems are comparatively simple dynamic systems. The main contribution of the present paper is the focus on uniquely complex time series from power systems, both synthetic and empirical. In contrast to dynamic systems such as the Lorenz attractor investigated in [2], the time series investigated in the present paper are characterized not only by non-linearity but also include strong noise and time-dependent dynamics [24]. While previous studies [20, 29] only considered time-dependent coefficients for SINDy, we advance the state of the art by including time as an explicit variable in the model.

This article is structured as follows. We first introduce a mathematical model that we use as a description of the frequency dynamics and to generate validation data sets, Sec. 2. Subsequently, in Sec. 3, as well as in App. A.1, we validate the SINDy model on these synthetic time series. In Sec. 4, we analyze data from three different power grids, noting distinct differences in the inferred dynamics, additionally, in App. A.2, we present an evaluation of the performance of our models. We close with a discussion and outlook in Sec. 5.

2 DATA-DRIVEN MODEL

The basis for modeling the grid frequency is the aggregated swing equation [37]. It is given as a linear stochastic differential equation:

$$\frac{d\theta}{dt} = \omega, \quad \frac{d\omega}{dt} = -c_\omega \omega - c_\theta \theta + \Delta P(t) + \epsilon \xi(t). \tag{1}$$

The expression θ represents the bulk angle of the voltage signal which in the following we abbreviate as angle, and its bulk angular velocity ω is given by $\omega = 2\pi(f - f_{\text{ref}})$, with the reference frequency $f_{\text{ref}} = 50$ Hz. The power mismatch, i.e. the imbalance between generation and demand, is given as ΔP. In order to describe the mismatch between continuous demand and step-wise constant generation, ΔP is often denoted as $\Delta P(t) = P_0 + P_1 t$ [34]. The symbols ϵ, and ξ represent the noise amplitude and Gaussian white noise function, respectively, which models the unforeseen influences, such as demand or generation fluctuations. The terms $-c_\omega\omega$ and $-c_\theta\theta$ combine the effects of damping and control. Specifically, $-c_\theta\theta$ models an integral component, as often included in secondary control, while $-c_\omega\omega$ includes any proportional control effects (from either primary or secondary control). For simplicity, we refer to these parameters as primary control parameter c_ω and secondary control parameter c_θ. Typically, the magnitude of the primary control parameter is much larger than the secondary control parameter, i.e., $c_\omega \gg c_\theta$. Detailed properties and explanations of the swing equation and its approximations, parameters, and modeling methods are extensively explained in [25] and [24]. We remark that it is a useful model for high-voltage transmission systems.

In the present paper, we investigate empirical data that has been recorded with electrical data recorders spanning three synchronous areas [15, 26]. In particular, we consider approximately 90 days of Balearic frequency data from 2019, along with several weeks of data from Ireland and Iceland obtained at the end of 2021. To conduct a precise analysis, we eliminate any gaps in the data, utilizing a continuous dataset. For the implementation of the models we use the Python implementation of SINDy [9, 16]. Code for the reproducibility of the models is available on GitHub [35].

In contrast to the model-based approach presented so far, SINDy is a procedure for extracting interpretable and generalizable dynamical systems models from time series data. The SINDy algorithm works by constructing a library of potential functions of the state variables, such as polynomials, trigonometric functions, or any functions that might be appropriate for the system. It then uses sparse regression to determine the smallest subset of these functions that most accurately predicts the time derivatives of the state data. The resulting model is a system of ordinary differential equations (ODEs) that approximates the dynamics of the system, see [2, 10] for details, thus the stochastic elements in equation 1 cannot be directly assessed.

The optimization problem is formulated as:

$$\Xi = \underset{\hat{\Xi}}{\operatorname{argmin}} \left\| \Theta(X)\hat{\Xi} - \frac{dX}{dt} \right\|_2^2 + \lambda \left\| \hat{\Xi} \right\|_2 . \tag{2}$$

The term Ξ represents the matrix of coefficients to be determined, $\Theta(X)$ is a library of functions applied to the state variables X, $\frac{dX}{dt}$ is the time derivative of the state variables, and we state $\frac{dX}{dt} = f(X) \approx \Theta(X)\Xi$. The term $\lambda \left\| \hat{\Xi} \right\|_2$ denotes the L2 regularization [12],

controlled by the parameter λ, which promotes sparsity in the coefficients Ξ. In our model, we use $\lambda = 0.05$ and set the threshold for the coefficients as $1{\cdot}10^{-10}$.

3 VALIDATING SINDY ON SYNTHETIC POWER GRID DYNAMICS

Before we infer unknown dynamics, we first demonstrate that SINDy correctly infers complex power grid dynamics by utilizing synthetic datasets based on realistic parameters, effectively validating the method. For this purpose, we use a stochastic data-driven model for equation (1), which is created using parameter estimation of a Fokker-Planck equation. This model, denoted as 1D-L-KM (1-dimensional linear Kramers-Moyal model), is extensively detailed in [25] and [24]. In these studies, an aggregated swing equation is modeled, see equation (1), which is used to generate synthetic time series with the help of the Euler-Maruyama algorithm. For the validation of SINDy for power frequency dynamics, we consider three different adaptions of the 1D-L-KM model.

The three synthetic data sets are characterized as follows:

(1) Base case: linear parameters c_ω and c_θ, step function as power mismatch ΔP, no noise ($\epsilon = 0$), see Sec. 3,
(2) Added noise: linear parameters c_ω and c_θ, step function as power mismatch ΔP, with Gaussian noise added to the dynamics ($\epsilon > 0$), see Appendix A.1 for details.
(3) Added time-dependency: linear parameters c_ω and c_θ, Gaussian noise ($\epsilon > 0$), and time-dependent power mismatch $\Delta P(t)$, see Appendix 3 for details.

The mean frequency exhibits jumps every hour due to market dispatch [28]. These jumps arise from a mismatch between step-wise generation and a continuous load curve against the step-wise generation schedules, ensuring system stability and efficiency [33, 34]. Hence, we systematically partition our dataset into discrete chunks, each spanning a time scale of one hour.

To include this external driving, we introduce time, denoted as T, as an additional variable in our model. This addition transforms a non-autonomous system into an autonomous one. The power mismatch ΔP ramps upwards for some hours and downwards for other hours, i.e. the time-dependent features will have different signs at different hours of the day. To accommodate this variability and capture the overall magnitude of time's influence, we report the mean of the absolute values of these coefficients.

Linear noise-free validation model

To validate the SINDy approach for complex power grid data, we initially generate one noise-less dataset from the 1D-L-KM model. The equation is shown below:

$$\frac{d\theta}{dt} = \omega, \quad \frac{d\omega}{dt} = c_\omega\omega + c_\theta\theta + P_0, \quad \frac{dT}{dt} = 1 \tag{3}$$

where P_0 is a piece-wise constant step function for 1-hour intervals. For creating synthetic time series we use values that are calculated in [24] by estimation of the Kramers-Moyal coefficients [27] of empirical time series from the Balearic grid. Specifically, the estimated values that are used as ground truth in our validation of SINDy for frequency data for c_ω and c_θ are $-2.95{\cdot}10^{-2}\mathrm{s}^{-1}$ and $-4.52{\cdot}10^{-5}\mathrm{s}^{-2}$ respectively. For reasons of clarity, we omit the physical units of all variables and parameters used in the equations of the present paper.

We consider a function library that consists of polynomials up to a degree of 2. Terms proportional to T, T^2, θT, and ωT fluctuate between positive and negative values, likely due to fluctuations in generation and demand. Thus, we represent these with a ± and indicate their magnitude, which is small to the other polynomials. The governing equations from SINDy are obtained as follows:

$$\begin{aligned}\frac{d\theta}{dt} = {} & -5.17{\cdot}10^{-3}\mathrm{s}^{-1} + 2.35{\cdot}10^{-5}\mathrm{s}^{-1}\cdot\theta \\ & + 9.34{\cdot}10^{-1}\cdot\omega \pm 9.67{\cdot}10^{-5}\mathrm{s}^{-2}\cdot T \\ & - 2.57{\cdot}10^{-8}\mathrm{s}^{-1}\cdot\theta^2 - 1.69{\cdot}10^{-5}\cdot\theta\omega \\ & \pm 3.97{\cdot}10^{-7}\mathrm{s}^{-2}\cdot\theta T - 5.33{\cdot}10^{-4}\mathrm{s}\cdot\omega^2 \\ & \pm 2.46{\cdot}10^{-4}\mathrm{s}^{-1}\cdot\omega T \pm 7.63{\cdot}10^{-13}\mathrm{s}^{-3}\cdot T^2 \end{aligned} \tag{4}$$

$$\begin{aligned}\frac{d\omega}{dt} = {} & 1.11{\cdot}10^{-2}\mathrm{s}^{-2} - 4.59{\cdot}10^{-5}\mathrm{s}^{-2}\cdot\theta \\ & - 2.79{\cdot}10^{-2}\mathrm{s}^{-1}\cdot\omega \pm 2.46{\cdot}10^{-6}\mathrm{s}^{-3}\cdot T \\ & + 7.23{\cdot}10^{-10}\mathrm{s}^{-2}\cdot\theta^2 + 4.75{\cdot}10^{-7}\mathrm{s}^{-1}\cdot\theta\omega \\ & \pm 1.02{\cdot}10^{-8}\mathrm{s}^{-3}\cdot\theta T + 1.50{\cdot}10^{-5}\cdot\omega^2 \\ & \pm 6.32{\cdot}10^{-6}\mathrm{s}^{-2}\cdot\omega T \pm 1.98{\cdot}10^{-15}\mathrm{s}^{-4}\cdot T^2 \end{aligned} \tag{5}$$

$$\begin{aligned}\frac{dT}{dt} = {} & 1.00 - 9.77{\cdot}10^{-12}\cdot\theta \\ & - 3.17{\cdot}10^{-9}\mathrm{s}\cdot\omega \pm 5.70{\cdot}10^{-13}\mathrm{s}^{-1}\cdot T \\ & + 2.03{\cdot}10^{-14}\cdot\theta^2 + 1.32{\cdot}10^{-11}\mathrm{s}\cdot\theta\omega \\ & \pm 2.37{\cdot}10^{-15}\mathrm{s}^{-1}\cdot\theta T + 3.77{\cdot}10^{-10}\mathrm{s}^2\cdot\omega^2 \\ & \pm 1.46{\cdot}10^{-12}\cdot\omega T \pm 1.35{\cdot}10^{-20}\mathrm{s}^{-2}\cdot T^2 \end{aligned} \tag{6}$$

The system of equations derived from the SINDy method reveals a notable correlation between the variables θ and ω. Specifically, in the equation for $\frac{d\theta}{dt}$, approximates as $\frac{d\theta}{dt} \approx \omega$, as expected. The coefficients c_ω and c_θ closely approximate the true values, with c_ω being approximately $-2.79{\cdot}10^{-2}\mathrm{s}^{-1}$ and c_θ as $-5.19{\cdot}10^{-5}\mathrm{s}^{-2}$. In addition, the constant term in the equation for $\frac{d\omega}{dt}$, specifically $1.11{\cdot}10^{-2}\mathrm{s}^{-2}$, corresponds to the mean value of P_0 as defined in our equation (3).

Moreover, the SINDy algorithm identifies additional terms in the equation, and coefficients associated with these discovered terms are comparatively small. Importantly, these coefficients exhibit magnitudes at least one order of magnitude smaller than the coefficients related to the principal terms in the equations. Note that the typical range for $|\omega|$ is smaller than 1, while for $|\theta|$ it is < 100. This suggests that while these terms contribute to the overall model, their influence is relatively subtle compared to the dominant effects represented by c_ω and c_θ.

Therefore, we conclude that the derived governing equations from SINDy successfully capture the system dynamics in the absence of noise within the framework of a one-dimensional linear model. Unfortunately, we cannot easily derive "sparse" models as some of the relevant terms (c_θ in particular) are small.

We also validate our SINDy algorithm on noisy and time-dependent systems. The case including noise and time-dependency is shown in the following, for the method where only noise is added

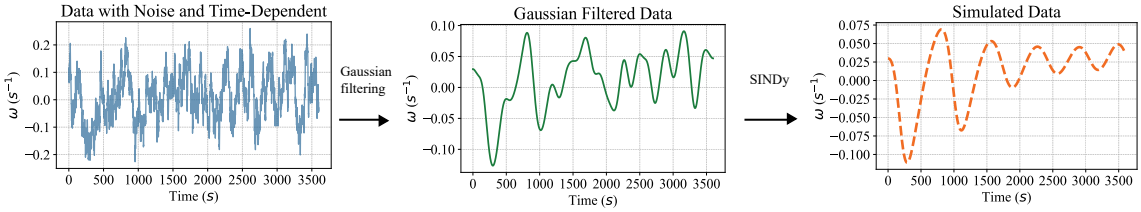

Figure 1: Schematic of the process for discovering the dynamics of power grid frequency: Data collection is followed by applying Gaussian filtering as the SINDy algorithm primarily focuses on deterministic dynamics and is sensitive to noise, then using our SINDy algorithm to derive the governing equation.

to the system see Appendix A.1. The noisy data is first smoothed with a Gaussian filter [13] to ensure stable results, the concept is demonstrated in Figure 1. Overall, we show that SINDy is applicable for identifying the relevant parameters of noise-less time series.

Linear validation model with noise and time-dependent driving

In order to increase the complexity of the validation, we examine the performance of SINDy on the linear model including noise and time-dependent driving. This evaluation allows us to assess how well SINDy performs in capturing the dynamics of a nonlinear system subjected to stochastic elements. The following equation is presented:

$$\frac{d\omega}{dt} = c_\omega \omega + c_\theta \theta + P_0 + P_1 t + \epsilon \xi(t). \tag{7}$$

Before applying SINDy, a critical denoising step is undertaken to enhance the accuracy and stability of the algorithm. In this context, the Gaussian filter method is employed for denoising.

The kernel width parameter σ adjusts the smoothness of the filtered signal, with lower values resulting in less smoothing and higher values producing more smoothing. To ascertain the most appropriate σ value for the Gaussian filter, We calculate the Mean Squared Error (MSE) between the original ω and the simulated ω for various σ values. The MSE indicates the accuracy of the simulations, with lower values suggesting a closer match to the original data. This analysis aids in determining the optimal parameter for the Gaussian filter, ensuring an effective denoising process and, consequently, a more accurate and stable SINDy model, especially for the analysis of the empirical data in Sec. 4. Figure 3 in the Appendix presents that the MSE value is the smallest for $\sigma = 60$. Therefore, $\sigma = 60$ is used for denoising the synthetic data and empirical data.

The expression capturing the governing equation within the SINDy framework is given by ($\frac{d\theta}{dt} \approx \omega$, $\frac{dT}{dt} \approx 1$):

$$\begin{aligned}\frac{d\omega}{dt} = &-4.06{\cdot}10^{-5}\mathrm{s}^{-2} - 3.05{\cdot}10^{-5}\mathrm{s}^{-2} \cdot \theta \\ &- 1.02{\cdot}10^{-3}\mathrm{s}^{-1} \cdot \omega \pm O(10^{-6})\mathrm{s}^{-3} \cdot T \\ &+ O(10^{-8})\mathrm{s}^{-2} \cdot \theta^2 + O(10^{-6})\mathrm{s}^{-1} \cdot \theta\omega \\ &\pm O(10^{-8})\mathrm{s}^{-3} \cdot \theta T - 1.72{\cdot}10^{-3} \cdot \omega^2 \\ &\pm O(10^{-6})\mathrm{s}^{-2} \cdot \omega T \pm O(10^{-10})\mathrm{s}^{-4} \cdot T^2\end{aligned}$$

Again, we mostly reproduce the expected dynamics: The coefficient for ω in $\frac{d\theta}{dt}$ and the correct sign of c_ω and c_θ in $\frac{d\omega}{dt}$ are accurately captured. Further, the coefficient c_ω is estimated to be $-1.02{\cdot}10^{-3}\mathrm{s}^{-1}$, and c_θ is estimated as $-3.05{\cdot}10^{-5}\mathrm{s}^{-2}$.

Additionally, we show the presence and significance of the ω^2 term in the $\frac{d\omega}{dt}$ equation. In the linear model with time-dependent data, the coefficient associated with ω^2 is specifically identified as $-1.72{\cdot}10^{-3}$. The fact that this coefficient is of the same order as the coefficient associated with the linear term (ω) indicates that the quadratic term is not negligible and plays a substantial role in influencing the dynamics of $\frac{d\omega}{dt}$. This may be indicative of more complex, possibly nonlinear behaviors in energy systems.

4 SINDY ON EMPIRICAL DATASET

Now that we have demonstrated that our SINDy approach is in principle capable of inferring complex power grid dynamics, including noise and explicit time-dependency within our linear stochastic differential equation, see 3 and Appendix A.1, we utilize it to infer the dynamics of empirical data from the Balearic islands, Ireland and Iceland with a one-second resolution. We analyze the model's predictive performance across various model complexities by evaluating the root mean square error (RMSE) and the proportion of non-convergent intervals of the model simulation. We select the second-order polynomial model due to its lowest RMSE and less unstable intervals. Hence, we apply the second-order model to all empirical datasets, see Appendix A.2 for details.

The mean value of the ω term in the $\frac{d\theta}{dt}$ equation consistently converges close to 1 across all three regions, with a relatively narrow standard deviation. Conversely, the mean coefficients associated with other terms within the equation are relatively small, significantly less than 10^{-4}, which indicates that the contribution of these terms is weaker than the ω-term. This aligns with our stochastic differential equation, i.e. the inferred dynamics is coherent with the underlying mathematical model, i.e. we again obtain $\frac{d\theta}{dt} \approx \omega$.

Therefore, we focus on the $\frac{d\omega}{dt}$ equation. In figure 2, we present a comparison of mean coefficients derived from the SINDy models applied to the Balearic, Iceland and Ireland regions. This reflects the variability of each coefficient, with longer bars (standard deviation) indicating greater uncertainty or variability in the model across different datasets. We can see that the coefficient for θ ranges from $10^{-4}\mathrm{s}^{-2}$ to $10^{-5}\mathrm{s}^{-2}$ and the coefficient for ω falls within the interval of $10^{-2}\mathrm{s}^{-1}$ to $10^{-3}\mathrm{s}^{-1}$, aligning notably well with the values obtained during our validation process. Hence, if we were simply interested in estimating primary (c_ω) and secondary control (c_θ) amplitudes, we would not have needed SINDy.

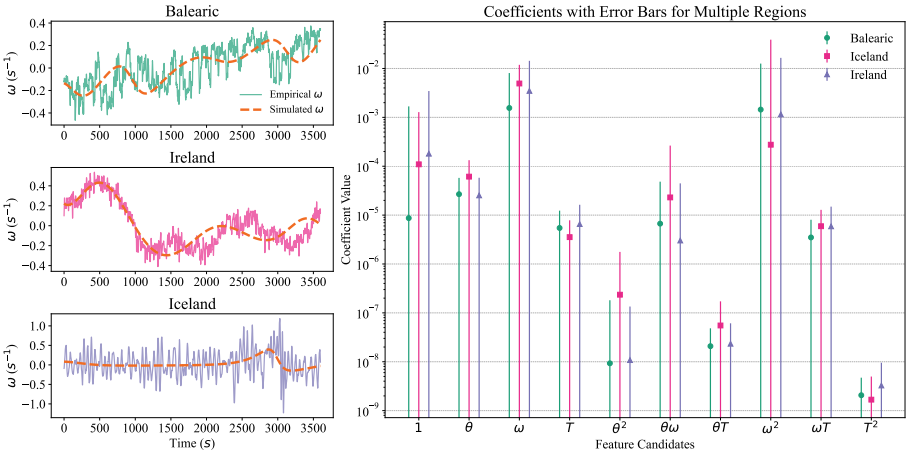

Figure 2: The empirical data, presented in a one-hour example, exhibits power grid frequency dynamics observed in the real world, whereas the simulations generated by our SINDy model are shown on the left-hand side. On the right-hand side, it shows that the Balearic islands, Iceland and Ireland display structurally similar models with substantial quantitative differences. We display the coefficients on a logarithmic scale with mean values (large symbol) and standard deviation (bar).

More interesting is the analysis of the remaining terms: The coefficients of $\theta\omega$ and ω^2 stand out due to their relatively high magnitude and significant standard deviation across the Balearic, Iceland and Ireland regions. This indicates that non-linear relationships could play a critical role in the dynamics of the power grid. These are not included in previous models [24, 25]. In addition, we note substantial impacts of time-dependency: Within a one-hour interval, the contribution of the T-coefficient reaches an amplitude ranging from $10^{-3}\mathrm{s}^{-3}$ to $10^{-2}\mathrm{s}^{-3}$ compared to the $10^{-5}\mathrm{s}^{-2}$ to $10^{-4}\mathrm{s}^{-2}$ range of the 1-coefficient. The constant value in the governing equations represents P_0, the range of P_0 spans from $10^{-9}\mathrm{s}^{-2}$ to $10^{-2}\mathrm{s}^{-2}$. This means that the power mismatch ΔP should be modeled as a time-dependent function, potentially including non-linear or mixed (ωT) terms as well.

5 DISCUSSION AND CONCLUSION

In this study, we demonstrate the enhanced capability of a Sparse Identification of Nonlinear Dynamics (SINDy) algorithm to accurately capture the complex behaviors of power grid frequency dynamics in energy systems. To validate this method we use the 1D-L-KM model across three different synthetic datasets.

Each dataset is generated with different conditions to test the robustness of SINDy and partitioned into chunks corresponding to hourly intervals. Time T is introduced as an explicit variable in the modeling process, and the focus on absolute values of time-related coefficients underlines the algorithm's sensitivity. Upon the successful validation against synthetic datasets, the SINDy algorithm is applied to real-world datasets obtained from Balearic, Ireland and Iceland. This study identifies key dynamic relationships and potential weaknesses in the face of real-world data complexity.

The derived system of equations from SINDy reproduces some of the expected relationships between the variables angle θ and angular velocity ω. In particular, we closely approximate the coefficients c_ω and c_θ, particularly in the noise-free datasets. When noise and time dependence are introduced, SINDy's estimation exhibits slight deviations from the exact coefficient values but maintains the correct sign and order of magnitude. SINDy also reproduces the observed dynamics for the empirical data sets, allowing us to interpret the estimated coefficients in further detail. SINDy identifies several nonlinear terms included in ω^2 and $\theta\omega$. Since ω^2 also emerges in the validation set with noise, it is likely not based on a physical process, while $\theta\omega$ might be a nonlinearity present in the empirical system. Note that these are potential non-linear term to reproduce the observed dynamics, while the actual underlying equations might differ.

Concluding, we present a version of SINDy that can infer complex (noisy, non-linear, time-dependent) dynamics in power systems. Thereby, we offer a perspective to refine existing mathematical models. The data-driven approach allows flexibility to update models during the energy transition or to infer models for previously unmodelled systems.

Having established the feasibility of symbolic approaches for power systems, we open up the research field for numerous future activities. On the one hand, we could consider different optimization algorithms or further modify the available basis functions in the symbolic regression (library of functions of SINDy). Additionally, incorporating stochastic elements into the SINDy framework could significantly improve its applicability, especially in scenarios where randomness and uncertainty are present. Bayesian inference techniques could prove helpful for this [7]. Furthermore, we can extend the SINDy algorithm across to other data sets, including other power grids, higher time resolutions of power grid frequency data or different power systems such as new power hardware [8]. In future work, we plan to address the limitations of assuming Gaussian white noise in our models, especially given the complex, correlated, and non-Gaussian nature of disturbances in power systems, such as those introduced by renewable energy sources. In addition, our future work will focus on incorporating more sophisticated noise-robust techniques or inferring the full stochastic dynamics.

ACKNOWLEDGMENTS

We sincerely acknowledge the funding from the Helmholtz Association and the Networking Fund through Helmholtz AI and under grant no. VH-NG-1727.

REFERENCES

[1] Lorenzo Boninsegna, Feliks Nüske, and Cecilia Clementi. 2018. Sparse learning of stochastic dynamical equations. *The Journal of Chemical Physics* 148, 24 (2018), 241723. https://doi.org/10.1063/1.5018409

[2] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. 2016. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* 113, 15 (2016), 3932–3937. https://doi.org/10.1073/pnas.1517384113

[3] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. 2016. Sparse Identification of Nonlinear Dynamics with Control (SINDYc). *IFAC-PapersOnLine* 49, 18 (2016), 710–715. https://doi.org/10.1016/j.ifacol.2016.10.249

[4] Jared L. Callaham, Jean-Christophe Loiseau, Georgios Rigas, and Steven L. Brunton. 2021. Nonlinear stochastic modelling with Langevin regression. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 477, 2250 (2021), 20210092. https://doi.org/10.1098/rspa.2021.0092

[5] Kathleen Champion, Peng Zheng, Aleksandr Y. Aravkin, Steven L. Brunton, and J. Nathan Kutz. 2020. A Unified Sparse Optimization Framework to Learn Parsimonious Physics-Informed Models From Data. *IEEE Access* 8 (2020), 169259–169271. https://doi.org/10.1109/ACCESS.2020.3023625

[6] Council of European Union. 2016. COMMISSION REGULATION (EU) 2016/631 of 14 April 2016. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0631&rid=1.

[7] Kevin Course and Prasanth B. Nair. 2023. State estimation of a physical system with unknown governing equations. *Nature* 622, 7982 (2023), 261–267. https://doi.org/10.1038/s41586-023-06574-8

[8] Giovanni De Carne, Georg Lauss, Mazheruddin H Syed, Antonello Monti, Andrea Benigni, Shahab Karrari, Panos Kotsampopoulos, and Md Omar Faruque. 2022. On modeling depths of power electronic circuits for real-time simulation–a comparative analysis for power systems. *IEEE Open Access Journal of Power and Energy* 9 (2022), 76–87. https://doi.org/10.1109/OAJPE.2022.3148777

[9] Brian de Silva, Kathleen Champion, Markus Quade, Jean-Christophe Loiseau, J. Kutz, and Steven Brunton. 2020. PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data. *Journal of Open Source Software* 5, 49 (2020), 2104. https://doi.org/10.21105/joss.02104

[10] Urban Fasela, J. Nathan Kutz, Bingni W. Brunton, and Steven L. Brunton. 2022. Ensemble-SINDy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 478, 2260 (2022), 20210904. https://doi.org/10.1098/rspa.2021.0904

[11] Ting-Ting Gao and Gang Yan. 2023. Data-driven inference of complex system dynamics: A mini-review. *Europhysics Letters* (2023).

[12] Arthur E. Hoerl and Robert W. Kennard. 1970. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* 12, 1 (1970), 69–82. https://doi.org/10.2307/1267352 Publisher: [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality].

[13] Kazufumi Ito and Kaiqi Xiong. 2000. Gaussian filters for nonlinear filtering problems. *IEEE Trans. Automat. Control* 45, 5 (2000), 910–927. https://doi.org/10.1109/9.855552 Conference Name: IEEE Transactions on Automatic Control.

[14] Bilkisu Jimada-Ojuolape and Jiashen Teh. 2020. Impact of the Integration of Information and Communication Technology on Power System Reliability: A Review. *IEEE Access* 8 (2020), 24600–24615. https://doi.org/10.1109/ACCESS.2020.2970598

[15] Richard Jumar, Heiko Maaß, Benjamin Schäfer, Leonardo Rydin Gorjão, and Veit Hagenmeyer. 2021. Database of Power Grid Frequency Measurements. arXiv:2006.01771 [physics] http://arxiv.org/abs/2006.01771

[16] Alan A. Kaptanoglu, Brian M. de Silva, Urban Fasel, Kadierdan Kaheman, Andy J. Goldschmidt, Jared Callaham, Charles B. Delahunt, Zachary G. Nicolaou, Kathleen Champion, Jean-Christophe Loiseau, J. Nathan Kutz, and Steven L. Brunton. 2022. PySINDy: A comprehensive Python package for robust sparse system identification. *Journal of Open Source Software* 7, 69 (2022), 3994. https://doi.org/10.21105/joss.03994

[17] Javad Khazaei and Ali Hosseinipour. 2023. Data-Driven Feedback Linearization Control of Distributed Energy Resources using Sparse Regression. *IEEE Transactions on Smart Grid* (2023), 1–1. https://doi.org/10.1109/TSG.2023.3298133

[18] Subhash Lakshminarayana, Saurav Sthapit, Hamidreza Jahangir, Carsten Maple, and H Vincent Poor. 2022. Data-driven detection and identification of IoT-enabled load-altering attacks in power grids. *IET Smart Grid* 5, 3 (2022), 203–218.

[19] Tania B Lopez-Garcia, Alberto Coronado-Mendoza, and José A Domínguez-Navarro. 2020. Artificial neural networks in microgrids: A review. *Engineering Applications of Artificial Intelligence* 95 (2020), 103894.

[20] Jeremy D. Lore, Sebastian De Pascuale, M. Paul Laiu, Ben Russo, Jae-Sun Park, Jin Myung Park, Steven L. Brunton, J Nathan Kutz, and Alan Ali Kaptanoglu. 2023. Time-dependent SOLPS-ITER simulations of the tokamak plasma boundary for model predictive control using SINDy *. *Nuclear Fusion* 63, 4 (2023), 046015. https://doi.org/10.1088/1741-4326/acbe0e

[21] Jan Machowski, Zbigniew Lubosny, Janusz W Bialek, and James R Bumby. 2020. *Power System Dynamics: Stability and Control*. John Wiley & Sons.

[22] Niall M. Mangan, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. 2016. Inferring Biological Networks by Sparse Identification of Nonlinear Dynamics. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications* 2, 1 (2016), 52–63. https://doi.org/10.1109/TMBMC.2016.2633265

[23] Shima Mohebbi, Qiong Zhang, E. Christian Wells, Tingting Zhao, Hung Nguyen, Mingyang Li, Noha Abdel-Mottaleb, Shihab Uddin, Qing Lu, Mathews J. Wakhungu, Zhiqiang Wu, Yu Zhang, Anwesh Tuladhar, and Xinming Ou. 2020. Cyber-physical-social interdependencies and organizational resilience: A review of water, transportation, and cyber infrastructure systems and processes. *Sustainable Cities and Society* 62 (2020), 102327. https://doi.org/10.1016/j.scs.2020.102327

[24] Ulrich Oberhofer, Leonardo Rydin Gorjão, G. Cigdem Yalcin, Oliver Kamps, Veit Hagenmeyer, and Benjamin Schäfer. 2023. Non-linear, bivariate stochastic modelling of power-grid frequency applied to islands. In *2023 IEEE Belgrade PowerTech*. 1–1. https://doi.org/10.1109/PowerTech55446.2023.10202986

[25] Leonardo Rydin Gorjão, Mehrnaz Anvari, Holger Kantz, Christian Beck, Dirk Witthaut, Marc Timme, and Benjamin Schäfer. 2020. Data-Driven Model of the Power-Grid Frequency Dynamics. *IEEE Access* 8 (2020), 43082–43097. https://doi.org/10.1109/ACCESS.2020.2967834

[26] Leonardo Rydin Gorjão, Richard Jumar, Heiko Maass, Veit Hagenmeyer, G. Cigdem Yalcin, Johannes Kruse, Marc Timme, Christian Beck, Dirk Witthaut, and Benjamin Schäfer. 2020. Open database analysis of scaling and spatio-temporal properties of power grid frequencies. *Nature Communications* 11, 1 (2020), 6362. https://doi.org/10.1038/s41467-020-19732-7

[27] Leonardo Rydin Gorjão and Francisco Meirinhos. 2019. kramersmoyal: Kramers–Moyal coefficients for stochastic processes. *Journal of Open Source Software* 4, 44 (2019), 1693. https://doi.org/10.21105/joss.01693

[28] Benjamin Schäfer, Marc Timme, and Dirk Witthaut. 2018. Isolating the Impact of Trading on Grid Frequency Fluctuations. In *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. 1–5. https://doi.org/10.1109/ISGTEurope.2018.8571793

[29] Oliver Schön, Ricarda-Samantha Götte, and Julia Timmermann. 2022. Multi-Objective Physics-Guided Recurrent Neural Networks for Identifying Non-Autonomous Dynamical Systems*. *IFAC-PapersOnLine* 55, 12 (2022), 19–24. https://doi.org/10.1016/j.ifacol.2022.07.282

[30] Alex M. Stanković, Aleksandar A. Sarić, Andrija T. Sarić, and Mark K. Transtrum. 2020. Data-driven Symbolic Regression for Identification of Nonlinear Dynamics in Power Systems. In *2020 IEEE Power & Energy Society General Meeting (PESGM)* (2020). 1–5. https://doi.org/10.1109/PESGM41954.2020.9281935

[31] Andreas Ulbig, Theodor S Borsche, and Göran Andersson. 2014. Impact of low rotational inertia on power system stability and operation. *IFAC Proceedings Volumes* 47, 3 (2014), 7290–7297.

[32] Wei Wei, Danman Wu, Qiuwei Wu, Miadreza Shafie-Khah, and João P. S. Catalão. 2019. Interdependence between transportation system and power distribution system: a comprehensive review on models and applications. *Journal of Modern Power Systems and Clean Energy* 7, 3 (2019), 433–448. https://doi.org/10.1007/s40565-019-0516-7

[33] Tobias Weißbach, Simon Remppis, and Hendrik Lens. 2018. Impact of Current Market Developments in Europe on Deterministic Grid Frequency Deviations and Frequency Restauration Reserve Demand. In *2018 15th International Conference on the European Energy Market (EEM)*. 1–6. https://doi.org/10.1109/EEM.2018.8469210

[34] Tobias Weissbach and Ernst Welfonder. 2009. High frequency deviations within the European Power System: Origins and proposals for improvement. In *2009 IEEE/PES Power Systems Conference and Exposition*. 1–6. https://doi.org/10.1109/PSCE.2009.4840180

[35] Xinyi Wen. 2024. Identify Complex Dynamics of Power Grid Frequency. https://github.com/KIT-IAI-DRACOS/Identify-Complex-Dynamics-of-Power-Grid-Frequency.

[36] Xinyi Wen, Mehrnaz Anvari, Leonardo Rydin Gorjão, G Cigdem Yalcin, Veit Hagenmeyer, and Benjamin Schäfer. 2023. Non-standard power grid frequency statistics in Asia, Australia, and Europe. *arXiv preprint arXiv:2308.16842* (2023).

[37] Dirk Witthaut, Frank Hellmann, Jürgen Kurths, Stefan Kettemann, Hildegard Meyer-Ortmanns, and Marc Timme. 2022. Collective nonlinear dynamics and self-organization in decentralized power grids. *Reviews of Modern Physics* 94, 1 (2022), 015005. https://doi.org/10.1103/RevModPhys.94.015005

A APPENDIX

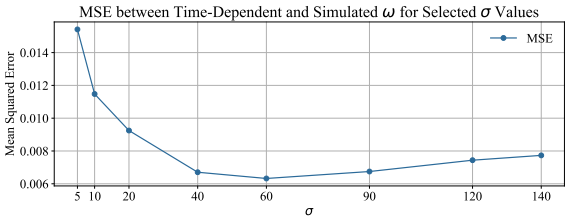

Figure 3: As the smoothing parameter σ increases in the Gaussian filtering process, an upward trend in Mean Squared Error (MSE) is observed. This suggests a potential divergence in simulation accuracy with a higher degree of smoothing.

A.1 Linear validation model including noise

To assess the impact of stochastic elements on the performance of SINDy, we generate a dataset that includes noise from the 1D-L-KM model. The equation governing this noise-inclusive dataset is expressed as follows:

$$\frac{d\omega}{dt} = c_\omega \omega + c_\theta \theta + P_0 + \epsilon \xi(t). \tag{8}$$

For denoising the data, the same filtering method is used as in 3, the evaluation of the kernel width parameter is shown in Figure 3. The expression for the governing equation derived from the SINDy method is as follows:

$$\begin{aligned}
\frac{d\theta}{dt} &= 1.64{\cdot}10^{-5}\mathrm{s}^{-1} - 1.65{\cdot}10^{-5}\mathrm{s}^{-1} \cdot \theta \\
&\quad + 9.99{\cdot}10^{-1} \cdot \omega \pm O(10^{-6})\mathrm{s}^{-2} \cdot T \\
&\quad + O(10^{-9})\mathrm{s}^{-1} \cdot \theta^2 + O(10^{-6}) \cdot \theta\omega \\
&\quad \pm O(10^{-8})\mathrm{s}^{-2} \cdot \theta T - 1.54{\cdot}10^{-4}\mathrm{s} \cdot \omega^2 \\
&\quad \pm O(10^{-6})\mathrm{s}^{-1} \cdot \omega T \pm O(10^{-10})\mathrm{s}^{-3} \cdot T^2 \\
\frac{d\omega}{dt} &= 3.21{\cdot}10^{-5}\mathrm{s}^{-2} - 3.36{\cdot}10^{-5}\mathrm{s}^{-2} \cdot \theta \\
&\quad - 1.28{\cdot}10^{-3}\mathrm{s}^{-1} \cdot \omega \pm O(10^{-6})\mathrm{s}^{-3} \cdot T \\
&\quad + O(10^{-9})\mathrm{s}^{-2} \cdot \theta^2 + O(10^{-6})\mathrm{s}^{-1} \cdot \theta\omega \\
&\quad \pm O(10^{-8})\mathrm{s}^{-3} \cdot \theta T - 2.28{\cdot}10^{-4} \cdot \omega^2 \\
&\quad \pm O(10^{-6})\mathrm{s}^{-2} \cdot \omega T \pm O(10^{-10})\mathrm{s}^{-4} \cdot T^2 \\
\frac{dT}{dt} &= 1.00
\end{aligned}$$

In the presence of noise, while the coefficient of ω in $\frac{d\theta}{dt}$ is accurately estimated, there is a slight deviation in the estimated values of c_ω and c_θ in $\frac{d\omega}{dt}$ from their true coefficients. Specifically, in the noisy data, the coefficient c_ω is estimated to be $-1.28{\cdot}10^{-3}\mathrm{s}^{-1}$, and c_θ is estimated as $-3.36{\cdot}10^{-5}\mathrm{s}^{-2}$.

The sign of both c_ω and c_θ aligns with their true coefficients, indicating the correct direction of influence. However, both magnitudes are smaller, particularly for c_θ. This suggests that noise and followed filtering leads to an underestimation of the deterministic dynamics and hence less precise determination of the coefficients' true values.

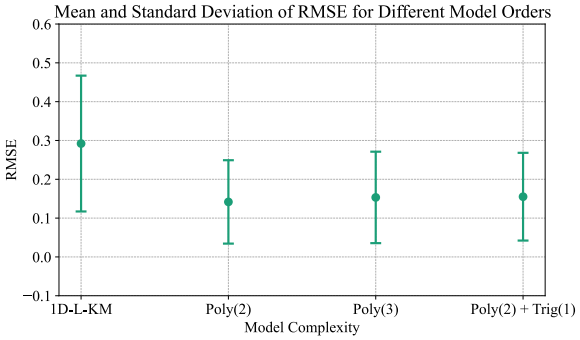

Figure 4: Compared to the RMSE value of 1D-L-KM model, the higher-order polynomial models show increased mean RMSE. The combined Order 2 polynomial and trigonometric model exhibits the highest mean of RMSE.

A.2 Evaluation of model performance

To choose the best library functions and polynomial order for our model, we calculate the root mean square error (RMSE) between the simulated and empirical angular velocity from the Balearic grid across various model complexities, using the 1D-L-KM model as a reference. We use all full-hour intervals available, each comprising one hour of data points with a resolution of one second. We notice that for certain intervals, the simulated ω fails to converge, indicating instability. In addition, we only consider the RMSE values below 0.5 due to the values of ω. Consequently, we also calculate the share of unstable intervals.

Model Complexity	Unstable Intervals Share	Mean of RMSE
1D-L-KM	0	0.292
Poly(2)	0.399	0.142
Poly(3)	0.463	0.153
Poly(2) + Trig(1)	0.332	0.155

Table 1: Comparison of Models Performance

Figure 4 and Table 1 present a comparison of the mean of RMSE values and share of unstable intervals respectively, illustrating the predictive accuracy of different model complexities. The SINDy model with second-order polynomials (Poly(2)) demonstrates the smallest mean of RMSE, indicating a strong predictive performance close to the reference. In contrast, the second-order model augmented with first-order trigonometric functions (Poly(2)+Trig(1)) exhibits the highest mean of RMSE, indicating a potential decrease in predictive performance possibly due to overfitting. However, it also demonstrates the least share of unstable intervals. This may suggest that the inclusion of trigonometric functions could help in capturing periodic patterns.

Considering the balance of complexity and precision, the second-order model is selected for its superior RMSE and share of unstable intervals.