Physics-Informed Machine Learning for Power Grid Frequency Modeling

Johannes Kruse [1,2] Eike Cramer [3] Benjamin Schäfer [4] and Dirk Witthaut [1,2,*]

[1] *Forschungszentrum Jülich, Institute for Energy and Climate Research (IEK-STE), Jülich 52428, Germany*
[2] *Institute for Theoretical Physics, University of Cologne, Köln 50937, Germany*
[3] *Forschungszentrum Jülich, Institute for Energy and Climate Research—Energy Systems Engineering (IEK-10), Jülich 52428, Germany*
[4] *Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen 76344, Germany*

The operation of power systems is affected by diverse technical, economic, and social factors. Social behavior determines load patterns, electricity markets regulate the generation, and weather-dependent renewables introduce power fluctuations. Thus, power system dynamics must be regarded as a nonautonomous system whose parameters vary strongly with time. However, the external driving factors are usually only available on coarse scales and the actual dependencies of the dynamic system parameters are generally unknown. Here, we propose a physics-informed machine learning model that bridges the gap between large-scale drivers and short-term dynamics of the power system. Integrating stochastic differential equations and artificial neural networks, we construct a probabilistic model of the power grid frequency dynamics in continental Europe. Its probabilistic prediction outperforms the daily average profile, which is an important benchmark, on a time horizon of 15 min. Using the integrated model, we identify and explain the parameters of the dynamical system from the data, which reveal their strong time-dependence and their relation to external drivers such as wind power feed-in and fast generation ramps. Finally, we generate synthetic time series from the model, which successfully reproduce central characteristics of the grid frequency such as their heavy-tailed distribution. All in all, our work emphasizes the importance of modeling power system dynamics as a stochastic nonautonomous system with both intrinsic dynamics and external drivers.

I. INTRODUCTION

Mitigation of climate change requires a comprehensive transformation of our economy and lifestyle, in particular the way we generate and utilize electric power [1,2]. Power plants based on fossil fuels must be replaced by renewable sources such as wind and solar power, which are volatile and uncertain [3]. Various sectors are being integrated, for instance through electric heatpumps [4], introducing numerous new interdependencies and increasing system complexity. The electric power system is at the heart of this transformation. Hence, understanding risks and guaranteeing stability of the electric power system is critical amidst far-reaching challenges [5].

Power system operation is determined by various technical, economic, and social influences and perturbations. Power generation from renewable sources is essentially determined by the weather [6,7], while the dispatch of conventional power plants is determined on various electricity markets [8]. Moreover, the load depends on the decisions and actions of millions of consumers [9]. As the power grid does not store electric energy, generation and load must be balanced at all times. On long time scales of hours, this is achieved by trading on electricity markets [10]. On short time scales of seconds and minutes, several layers of control reserves balance the grid, e.g., to counteract unforeseen perturbations and forecasting errors [11]. The activation of these reserves is mainly controlled by the grid frequency, which directly monitors the power imbalance: a scarcity of generation leads to a drop of the frequency, which is easily monitored anywhere in the grid. The stability of this load-frequency control system is challenged by the energy transformation, as the effective inertia of the grid decreases, making the frequency more susceptible to perturbations [12].

The realistic modeling of frequency dynamics in large-scale power systems is profitable but complex due to its nonautonomous character. Stochastic dynamical models have successfully reproduced central characteristics of frequency measurements such as their nonstandard distributions [13–15]. Such models can be used to generate synthetic frequency time series, which are, for example, employed to optimize electric devices [16]. Moreover, they can be used to explore dynamics under different operating conditions, e.g., with an increased wind power generation [17]. However, multiple technical, economic, and social influences and perturbations shape power system dynamics, as explained above. As a consequence, the power system must be regarded as a stochastic nonautonomous dynamical system, which makes grid frequency modeling a daunting task.

In this context, the data-driven representation of external drivers can greatly facilitate realistic models, but data assimilation is challenging due to insufficient data sources. Integrating actual load time series can improve stochastic models of grid frequency dynamics in continental Europe [14]. The assimilation of load and generation data enabled an accurate reproduction of grid frequency recordings for the Gran Canaria island [17]. However, load and generation time series are typically only available at hourly time scales [18], while frequency dynamics happen at much smaller time scales, thus requiring a careful adoption of these external drivers. For large-scale power systems such as the continental European grid, the data are often incomplete with missing or unrealistic data points [19]. The physical model is uncertain as, for example, control schemes vary among local control zones and detailed setups are not publicly available [20]. Finally, important dynamical parameters such as the inertia cannot be calculated exactly due to scarce time series on the power plant level [21].

Overall, the dynamics of the power grid constitute a complex system: an intrinsic, potentially nonlinear dynamics [13,22] is enriched by external drivers [23] and stochastic influences. This complexity is somewhat simplified as some processes take effect on different time horizons, leading to a separation of time scales. The superposition of time-varying parameters on a longer time scale then leads to an effective superstatistics [24], observed in earlier contributions [25].

In this work, we propose physics-informed machine learning (PIML) to approach these challenges. Compared to numerical simulations, PIML models can perform better in solving inverse problems with noisy, insufficient data and imperfect physical models [26,27]. Moreover, they can better generalize from small amounts of data than common machine learning methods [26]. Notably, inverse problems are particularly important for power system control to estimate hidden states or dynamical parameters from measurements [28]. The inverse problem of inferring system

parameters from input-output data is known as system identification [29] and PIML models offer a promising tool for such applications [30].

In particular, we develop a PIML model for the load-frequency dynamics of electric power systems, which includes proportional and integral controllers, stochastic noise, and external techno-economic driving factors. The internal dynamics is described by a set of stochastic differential equations, which admit a semianalytic solution. The external driving is manifested through specific system parameters, which depend on a variety of techno-economic features such as the generation mix. This dependency is deduced via an artificial feed-forward neural network (FFNN), which is trained on data of the continental European power system in a maximum likelihood approach. Finally, we interpret our model with SHapley Additive exPlanation (SHAP) values [23,31] to extract the dependency between dynamical parameters and techno-economic features. All in all, the model bridges the gap between the large-scale behavior of interdependent energy systems and markets and the short-term dynamics of the power system.

The article is organized as follows. In Sec. II we introduce the physics-informed machine learning model for the grid frequency dynamics and discuss its implementation. In Sec. III we present and evaluate three model applications: probabilistic prediction, system identification and explanation, and generation of synthetic time series. Finally, we discuss our results as well as possible future directions in Sec. IV.

II. AN INTEGRATED MODEL FOR POWER SYSTEM LOAD-FREQUENCY DYNAMICS

Here, we present the details of how we constructed a physics-informed model of the power grid frequency, including a stochastic description of short-term dynamics and the interaction with power system operation and markets on longer time scales. Furthermore, we introduce our approach on modeling these interactions with an artificial neural network prediction based on techno-economic features. The detailed implementation of our model pipeline, as well as all input data and the results are available from Zenodo [32,33].

A. Short-term dynamics and control of the grid frequency

Our starting point is a stochastic model for the dynamics of the grid frequency on short time scales between 1 s and 1 h [cf. Fig. 1(a)]. The rate of change of the frequency at time t is determined by the balance of power generation and load as well as the load-frequency control system (details are provided in Appendix A 1). Denoting the deviation from the reference as $\omega(t) = 2\pi(f(t) - f_{\text{ref}})$, we have

the equation of motion

$$M\frac{d\omega}{dt} = P_{\text{im}}(t) + P_{\text{control}}(t) + P_{\text{noise}}(t), \qquad (1)$$

where M is the aggregated inertia constant. The power imbalance on the right-hand side has been decomposed into three contributions. First, the term $P_{\text{im}}(t)$ denotes sustained power imbalances, for instance due to a recurring mismatch of the load and the scheduled generation of dispatchable power pants [cf. Fig. 1(b)]. Their generation changes in a stepwise manner, as electricity is traded on the spot markets in blocks of 15, 30, or 60 min [8]. In contrast, the load changes continuously, which leads to deterministic power imbalances [34]. For time intervals of an hour, we can approximate the time dependence of these imbalances by an affine linear function

$$P_{\text{im}}(t) = M \underbrace{\left(\sum_{n=1}^{4} q_n \Theta\left(t - n\frac{\Delta t}{4}\right) + rt \right)}_{=:P(t)}, \qquad (2)$$

where $\Theta(t)$ is the Heaviside function. Parameters $q_1, \ldots, q_4$ model the power steps of scheduled generation after quarters of the hour $\Delta t = 1$ h, which is the shortest trading block. Parameter r represents the continuous drift of the load [cf. Fig. 1(c)]. Notably, q_n and r are strongly correlated as a rising load typically leads to an increase of scheduled generation. However, irregularities such as forecast errors can also lead to a sustained mismatch of load and scheduled generation. This requires an independent modeling of power steps q_n and drift r.

Second, the term $P_{\text{control}}(t)$ denotes the balance of primary and secondary load-frequency control, which can be modeled by a proportional-integral law as

$$P_{\text{control}}(t) = -\frac{M}{\tau}\omega(t) - \frac{M}{\kappa^2}\underbrace{\int_{t_i}^{t} \omega(t')dt'}_{=:\theta(t)} \qquad (3)$$

with time constants τ and κ. Parameter τ quantifies the effective primary control time scale and κ can be interpreted as the intrinsic time scale of secondary control. In contrast, the effective time scale of secondary control is approximated by κ^2/τ [13], as the frequency decays with this time constant in the overdamped case [14]. We note that some simplifications are necessary to keep the model tractable. For instance, Eq. (3) neglects the existence of a small deadband in the proportional control law.

Finally, we assume that the remaining power imbalance fluctuations $P_{\text{noise}}(t)$ are faster than 1 s and thus uncorrelated on the time scale of our short-term dynamical model.

Therefore, we modeled such fluctuations as

$$P_{\text{noise}}(t) = MD\xi(t), \qquad (4)$$

where $\xi(t)$ is white noise with a standard normal distribution and D quantifies the strength of fast imbalance fluctuations. We used a Gaussian distribution because frequency increments on time scales of seconds are approximately Gaussian in continental Europe [35].

Because of the presence of noise, the equation of motion (1) must be interpreted as a stochastic differential equation (SDE) with the explicit form

$$\frac{d\theta}{dt} = \omega,$$
$$\frac{d\omega}{dt} = P(t) - \omega\tau^{-1} - \theta\kappa^{-2} + D\xi(t). \qquad (5)$$

Note that this model only contains effective parameters that were rescaled by the inertia M. For example, the actual primary control strength is M/τ [Eq. (3)]. However, the whole model is invariant under a scaling of M (cf. Appendix A2) such that it is only possible to estimate the ratio of parameters and the inertia.

Applying Itô's calculus [36], the SDE can be recast into a Fokker-Planck equation (FPE) of the probability density function $\mathcal{P}(\theta, \omega; t)$:

$$\frac{\partial}{\partial t}\mathcal{P}(\theta, \omega; t) = \left[-\frac{\partial}{\partial \omega}(P(t) - \tau^{-1}\omega - \kappa^{-2}\theta) \right.$$
$$\left. -\frac{\partial}{\partial \theta}\omega + \frac{D^2}{2}\frac{\partial^2}{\partial \omega^2} \right]\mathcal{P}(\theta, \omega; t). \qquad (6)$$

As we show in Appendix A3, the FPE is solved by a multivariate Gaussian distribution

$$\mathcal{P}(\mathbf{x}; t) = \frac{1}{2\pi|\Sigma|}\exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top\Sigma^{-1}(\mathbf{x} - \mu) \right) \qquad (7)$$

with $\mathbf{x}^\top = (\theta, \omega)$ and time-dependent parameters

$$\mu(t) = \begin{pmatrix} \mu_\theta(t) \\ \mu_\omega(t) \end{pmatrix}, \qquad \Sigma(t) = \begin{pmatrix} \sigma_\theta^2(t) & \sigma_{\theta\omega}(t) \\ \sigma_{\theta\omega}(t) & \sigma_\omega^2(t) \end{pmatrix},$$

if the parameters satisfy the ordinary differential equations

$$\frac{d}{dt}\mu_\theta = \mu_\omega, \qquad (8a)$$

$$\frac{d}{dt}\mu_\omega = P(t) - \tau^{-1}\mu_\omega - \kappa^{-2}\mu_\theta, \qquad (8b)$$

$$\frac{d}{dt}\sigma_\theta^2 = 2\sigma_{\theta\omega}, \qquad (8c)$$

$$\frac{d}{dt}\sigma_{\theta\omega} = \sigma_\omega^2 - \tau^{-1}\sigma_{\theta\omega} - \kappa^{-2}\sigma_\theta^2, \qquad (8d)$$

$$\frac{d}{dt}\sigma_\omega^2 = -2\tau^{-1}\sigma_\omega^2 - 2\kappa^{-2}\sigma_{\theta\omega} + D^2. \qquad (8e)$$
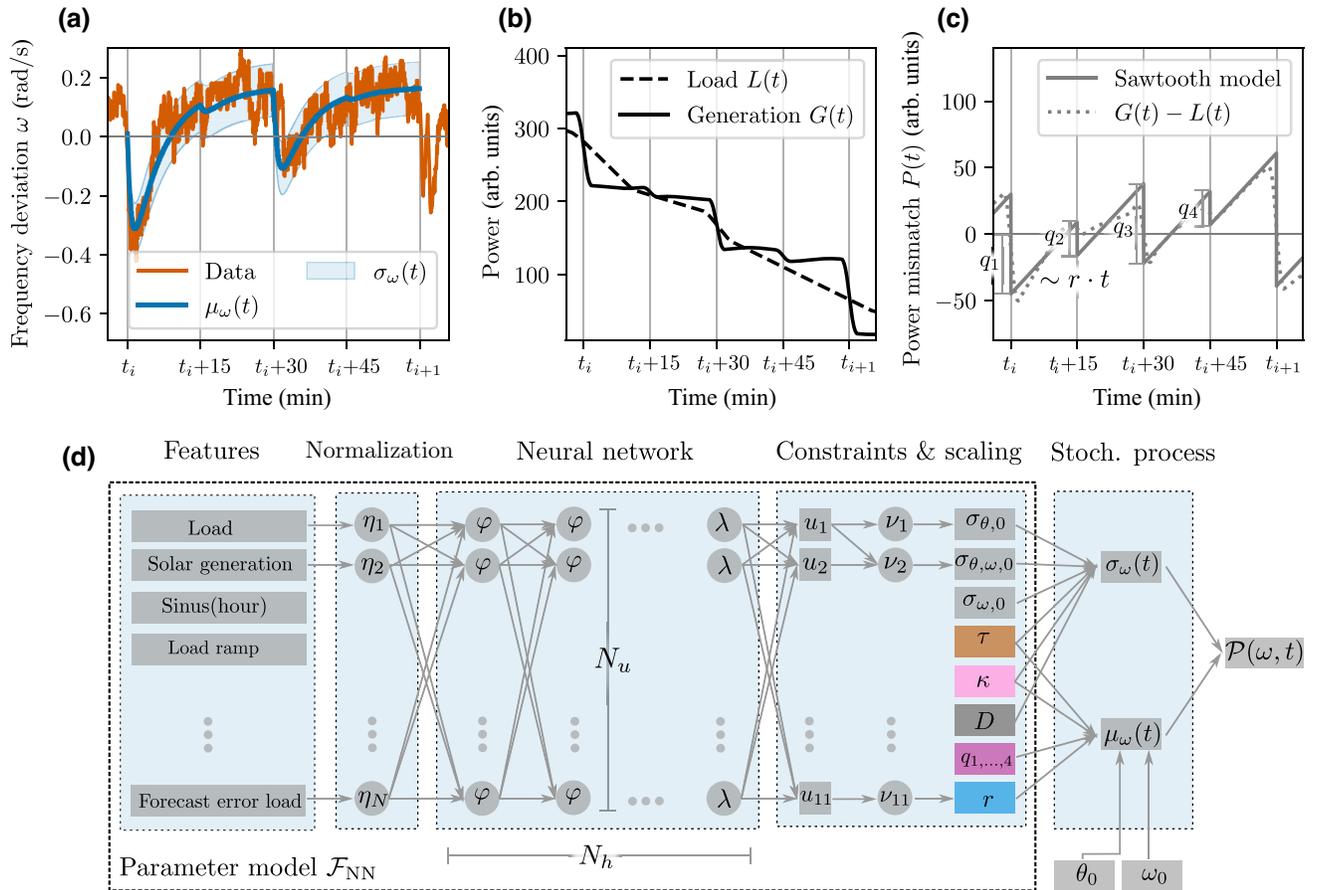
FIG. 1. A physics-informed machine learning model for the power grid frequency. (a) In each interval $[t_i, t_{i+1}]$, we modeled the grid frequency with a stochastic process that provides a normal distribution with time-dependent mean $\mu_\omega(t)$ and standard deviation $\sigma_\omega(t)$. The panel depicts our model prediction of the continental European grid frequency at $t_i = 22:00$ on October 11, 2018 in comparison to the recorded data. (b) Deviations from the reference frequency are partly driven by deterministic power imbalances $P_{\text{im}}(t) = MP(t)$. These result from a different time evolution of generation $G(t)$ and load $L(t)$ due to the market-based dispatch of power generation, which is illustrated in this panel using self-engineered synthetic data. (c) We approximate the deterministic mismatch with a sawtooth function $P(t)$. (d) The full probabilistic model $\mathcal{P}(\omega, t)$ incorporates deterministic imbalances $P_{\text{im}}(t)$, additional stochastic imbalance fluctuations $P_{\text{noise}}(t)$, and the load-frequency control $P_{\text{control}}(t)$. The parameter model $\mathcal{F}_{\text{NN}}$ predicts the model parameters, i.e., the imbalance and control parameters (color highlight), as well as the initial covariances, from N techno-economic features by using a feed-forward neural network (FFNN). The feature sets differ in the day-ahead and the *ex post* models, which are two different model types defined in Sec. II C.

Here, μ and σ are the mean and standard deviation of the angle and the frequency deviation, while $\sigma_{\theta\omega}$ represents their covariance.

In Appendix A 5, we provide a semianalytical solution for the moment equations (8), which enables a numerically tractable calculation of the likelihood $\mathcal{P}(\omega; t)$.

B. Power system operation and interdependencies

The SDE for the frequency dynamics in Eq. (5) contains several parameters, describing the load-frequency control system (τ, κ) or the power imbalances (r, $q_{1,...,4}$, D). The parameters are not constant, but change in time due to different operating conditions of the power system.

For instance, the market-based scheduling of dispatchable power plants causes deterministic power imbalances [34, 37], which we represent in our model via the imbalance parameters $q_{1,...,4}$ and r [Figs. 1(b) and 1(c)]. As the load changes throughout the day, the imbalance parameters will also vary over time. Another example is the (effective) primary control time scale τ, which can change in time due to a variation of inertia M caused by a varying share of conventional power plants (cf. Sec. II A). In general, the operating conditions mainly change due to a different dispatch on electricity markets. The largest volume of electricity in Europe is still traded in hourly blocks [38]. Therefore, we assume that our model parameters change on time scales of 1 h or longer, but are approximately constant within each hour.

We thus propose a model that links different temporal and technological scales, integrating the fast dynamics of the frequency-control system, the stochastic noise, and the impact of slowly changing operating conditions. In every hourly interval, the frequency is modeled by SDE (5). The system parameters τ, κ, D, r, $q_{1,\ldots,4}$ are constant within the hour, but change from interval to interval in a stepwise manner. The operating conditions are characterized with different techno-economic features, as detailed below, and their impact on the model parameters is predicted with an FFNN that is trained such that the stochastic dynamics best fits the recorded time series. More precisely, the FFNN constitutes a parameter model $\mathcal{F}_{NN} : \mathbf{X} \mapsto \boldsymbol{\vartheta}$, where $\mathbf{X}$ summarizes the values of techno-economic features [Fig. 1(d)]. Vector $\boldsymbol{\vartheta}$ includes the system parameters τ, κ, D, r, $q_{1,\ldots,4}$ as well as the initial covariances $\sigma_{\omega\theta,0}$, $\sigma_{\omega,0}^2$, and $\sigma_{\theta,0}^2$ at time $t = t_i$, while the initial means $\mu_{\omega,0}$ and $\mu_{\theta,0}$ are directly obtained from the data. Integrating the parameter model and the stochastic dynamical model, we obtained a PIML model that predicts a probability distribution $\mathcal{P}(\omega; t)$ for 1 h from techno-economic features.

C. Techno-economic features affecting power system dynamics and operation

As input features for the PIML model, we used several operational time series of the continental European power system from Ref. [23]. The data set includes 64 techno-economic features with an hourly resolution, which are based on data from the ENTSO-E Transparency Platform [18].

We created two different PIML models using subsets with N features from the original dataset. The day-ahead model offers a probabilistic day-ahead forecast of the grid frequency (Sec. III A). It uses only day-ahead available features ($N = 13$): day-ahead forecasts of load and renewable generation, day-ahead electricity prices and scheduled generation, as well as their respective ramps (time derivative of a feature) and the hour of the day (cf. Table I).

In contrast, the *ex post* model uses only *ex post* available features ($N = 51$): actual load and generation per type, their ramps and forecast errors (day-ahead minus actual), as well as day-ahead electricity prices (to keep a price feature in the input set). As these features describe the actual system operation, we used the *ex post* model for parameter inference (Sec. III B) and the generation of synthetic time series (Sec. III C), but we also compared it to the day-ahead forecasting model in Sec. III A. A full list of features is provided in Table I. Details on the preprocessing for the continental European power system are provided in Ref. [23]. Notably, the number of available features can vary in other power systems, as, for example, certain generation types might not be installed.

The grid frequency recordings used in this work were taken from Ref. [39], which provides preprocessed frequency data from the German transmission system operator TransnetBW [40].

D. Artificial neural network model

The integrated model provides a probabilistic prediction of the power grid frequency for every hourly interval $[t_i, t_{i+1}]$. The architecture of the model is depicted in Fig. 1(d).

As input, we used the techno-economic features $\mathbf{X}(i) = (X_1(i), \ldots, X_N(i))^T$ for each time interval $[t_i, t_{i+1}]$, where N depends on the model type (cf. Sec. II C). In the first step, each feature X_k was normalized using functions $\eta_k(X_k) = (X_k - \langle X_k \rangle)/\sigma_k$ to improve numeric stability, where $\langle \cdot \rangle$ denotes the average and σ_k the standard deviation of the feature.

The normalized features were fed into an FFNN of N_h hidden layers with N_u units and activation functions φ. The last layer comprises a linear activation λ, as we aim to predict real-valued parameters $\boldsymbol{\vartheta}$.

The following layer rescales the output of the FFNN and implements several constraints. The rescaling was implemented to improve training efficiency and stability. After random initialization, the outputs u_j of the FFNN typically have the same scale, but the physical parameters do not. Such a mismatch will yield large initial errors along certain parameter axes, leading to inhomogeneous loss landscapes that can make optimization inefficient and more difficult [41]. This difficulty can be mitigated by a suitable rescaling. Furthermore, several output variables must respect physical constraints. For instance, the time constants τ and κ must be positive and respect the inequality $\kappa \geq 2\tau$ to avoid an unphysical oscillation behavior of the solution. Rescaling and constraints were implemented with parameter-specific functions $\nu_j(u_j)$, which are described in Appendix B.

After rescaling, the output $\boldsymbol{\vartheta}(i)$ of the parameter model $\mathcal{F}_{NN}$ was used to compute a probabilistic prediction of the grid frequency for the entire time interval $[t_i, t_{i+1}]$ based on Eq. (7). Vector $\boldsymbol{\vartheta}(i)$ contains the system parameters as well as the covariances at $t = t_i$, while the means $\mu_{\omega,0}$ and $\mu_{\theta,0}$ are directly taken from data. For training and forecasting applications, we used the actual value of the frequency $\mu_{\omega,0}(i) = \omega(t_i)$ and estimated $\mu_{\theta,0}(i) = \int_{t_i - 60\,s}^{t_i} \omega(t') \, dt'$. For the generation of synthetic time series, we predicted intervals sequentially in time and estimated $\mu_{\omega,0}$ and $\mu_{\theta,0}$ from the preceding prediction and not from the data.

E. Training, testing, and interpretation

Our complete data set comprises 26 859 data points from 2015 to 2019. Each data point corresponds to one interval $[t_i, t_{i+1}]$ and comprises a feature vector $\mathbf{X}(i)$ and a

TABLE I.   External features used in the *ex post* and day-ahead models. The data are available online [33], data acquisition and preprocessing are described in Ref. [23].

Ex post	Generation and load (MW)	Load, total generation, synchronous generation, biomass generation, coal gas generation, gas generation, geothermal generation, hard coal generation, lignite generation, nuclear generation, oil generation, other generation, other renewable generation, pumped hydro generation, reservoir hydro generation, run-off-river hydro generation, solar generation, waste generation, wind offshore generation, wind onshore generation
	Ramps (MW/h)	Load ramp, total generation ramp, biomass ramp, coal gas ramp, gas ramp, geothermal ramp, hard coal ramp, lignite ramp, nuclear ramp, offshore wind ramp, onshore wind ramp, oil ramp, other ramp, other renewables ramp, pumped hydro ramp, reservoir hydro ramp, run-off-river hydro ramp, solar ramp, waste ramp
	Forecast errors of load and generation (MW)	Forecast error load, forecast error total generation, forecast error solar, forecast error offshore wind, forecast error onshore wind
	Forecast errors of ramps (MW/h)	Forecast error load ramp, forecast error generation ramp, forecast error solar ramp, forecast error offshore wind ramp, forecast error onshore wind ramp
	Other	Price day-ahead (euro/MWh)
Day-ahead	Generation and load (MW)	Load day-ahead, solar day-ahead, offshore wind day-ahead, onshore wind day-ahead
	Ramps (MW/h)	Load ramp day-ahead, generation ramp day-ahead, solar ramp day-ahead, offshore wind ramp day-ahead, onshore wind ramp day-ahead
	Other	Price day-ahead (euro/MWh), price ramp day-ahead (euro/MWh/h), $\cos(\pi/12 \times \text{hour})$, $\sin(\pi/12 \times \text{hour})$

frequency time series $\omega(i) = (\omega(t_i), \ldots, \omega(t_i + \Delta t))^T$ with $\Delta t = 60$ min.

We quantified the ability of the model to predict the stochastic frequency dynamics by the negative loglikelihood. The model performance can vary with the prediction length $t_{\max} \leq \Delta t$ in each interval. Therefore, we define the loglikelihood for subsets $\mathcal{I} = [t_i, t_i + t_{\max}]$:

$$\mathcal{C}(\omega(i), \vartheta(i)) = -\sum_{t \in \mathcal{I}} \log \mathcal{P}(\omega; t | \vartheta(i)) \qquad (9)$$

with $P(\omega; t | \vartheta)$ the marginal of PDF (7) evaluated at the measured data $\omega(t)$. During the training process, we utilized the full intervals with $t_{\max} = \Delta t$. During model evaluation, we also investigated the performance for different prediction lengths $t_{\max}$. Notably, the loglikelihood is a negatively oriented metric, i.e., smaller values represent a better performance.

To train the PIML model, we initialized the FFNN weights using the Glorot uniform initializer [42]. Using data from 2015 to 2017, we trained the weights with stochastic gradient descent using the ADAM optimizer with a fixed learning rate [41]. As a loss function, we chose the negative loglikelihood (9), summed over all hourly intervals in the training set. The model hyperparameters were optimized using random search on data from 2018 (as a validation set) and with parameter choices defined in Table II. In particular, we trained the model for 100 epochs and applied early stopping based on the validation loss.

Then, we retrained the best model on data from 2015 to 2018 and evaluated the performance in terms of the negative loglikelihood on data from 2019 as a test set.

We benchmarked the developed model by comparing its performance to the daily profile of the grid frequency, which is defined as follows. For a fixed time of the day t_d, we collected all frequency values recorded on all days in the training set and calculated their average $\mu_p(t_d)$ and the corresponding standard deviation $\sigma_p(t_d)$. Our daily profile model $\mathcal{P}_p$ returns a normal distribution $\mathcal{P}_p(\omega; t) = \mathcal{N}(\mu_p(t_d), \sigma_p(t_d))$ based on the time of the day $t_d(t)$ of time step t. For example, the predicted mean $\mu_p(t_d)$ for January 11, 2019 at 11:00 equals the average of frequency values at 11:00 over all days in the training set. In addition to the daily profile, we applied the constant model as a benchmark, which simply provides a normal distribution using

TABLE II.   Parameter choices during hyperparameter optimization. Here, Sig(x) is the sigmoid function and $\tanh(x)$ the hyperbolic tangent.

	Possible values
Learning rate	$10^{-4}, 10^{-3}$
Dropout rate	0, 0.15, 0.3
N_u	64, 128
N_h	1, 3, 5
Activation $\varphi(x)$	Sig(x), $\tanh(x)$
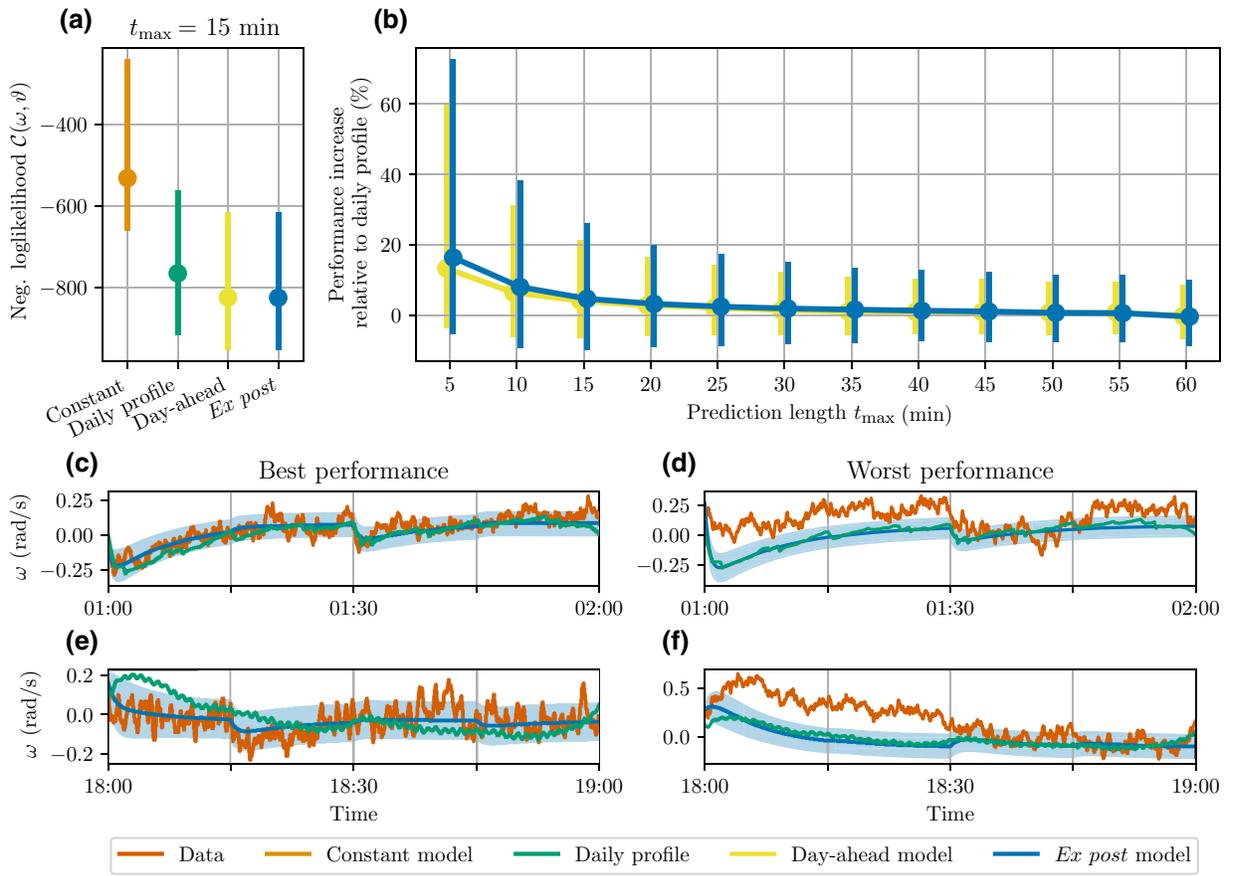
FIG. 2. Probabilistic machine learning model outperforms elementary benchmarks on a prediction horizon of 15 min. (a) We quantified the performance of our probabilistic frequency prediction with the negative loglikelihood loss, which implicates better performance if values are lower. For a prediction length of $t_{max} = 15$ min, the day-ahead and *ex post* models showed better performance than our benchmarks, namely, the constant model and the daily profile predictor. (b) The performance increase of our models relative to the daily profile shrank with growing prediction length t_{max}. The dots indicate the median and the bars represent the 25% and 75% quantiles. (c)–(f) Prediction examples with best and worst performance at 01:00 and 18:00 illustrate the strengths and limitations of our model.

the global mean and variance of the whole frequency time series.

Finally, we interpreted our parameter model $\mathcal{F}_{NN} : \mathbf{X} \mapsto \boldsymbol{\vartheta}$ with Shapely additive explanation (SHAP) values [31], which attribute the prediction of a single parameter $\vartheta_j(i)$ to the impact of different features $X_k(i)$. Aggregating individual SHAP values offers a tool to inspect feature importance and dependencies extracted by the FFNN. In particular, we used KernelSHAP [43], which approximates SHAP values for any machine learning model.

Our FFNN model is implemented with tensorflow [44] and tensorflow probability [45] and we used keras tuner for hyperparameter optimization [46].

III. MODEL APPLICATION AND EVALUATION

We demonstrate and evaluate three applications of our PIML model. First, it provides a probabilistic prediction of the grid frequency trajectory in each time interval, which

we evaluate in terms of the performance and compare it to elementary benchmarks (Sec. III A). Second, the model infers time-dependent imbalance and control parameters based on the data, i.e., we can use it for system identification. We analyze their time dependence, compare our estimates with values from the literature, and explain their dependency on techno-economic features with SHAP values (Sec. III B). Third, our model provides a tool for generating synthetic frequency time series by drawing samples from the stochastic process. Such synthetic time series should reproduce central stochastic characteristics of the grid frequency, which we evaluate in Sec. III C.

A. Probabilistic prediction of the grid frequency

Our physics-informed model provides a probabilistic prediction for unseen samples of the grid frequency. Its performance depends both on the length of the prediction and on the available set of features (Fig. 2).
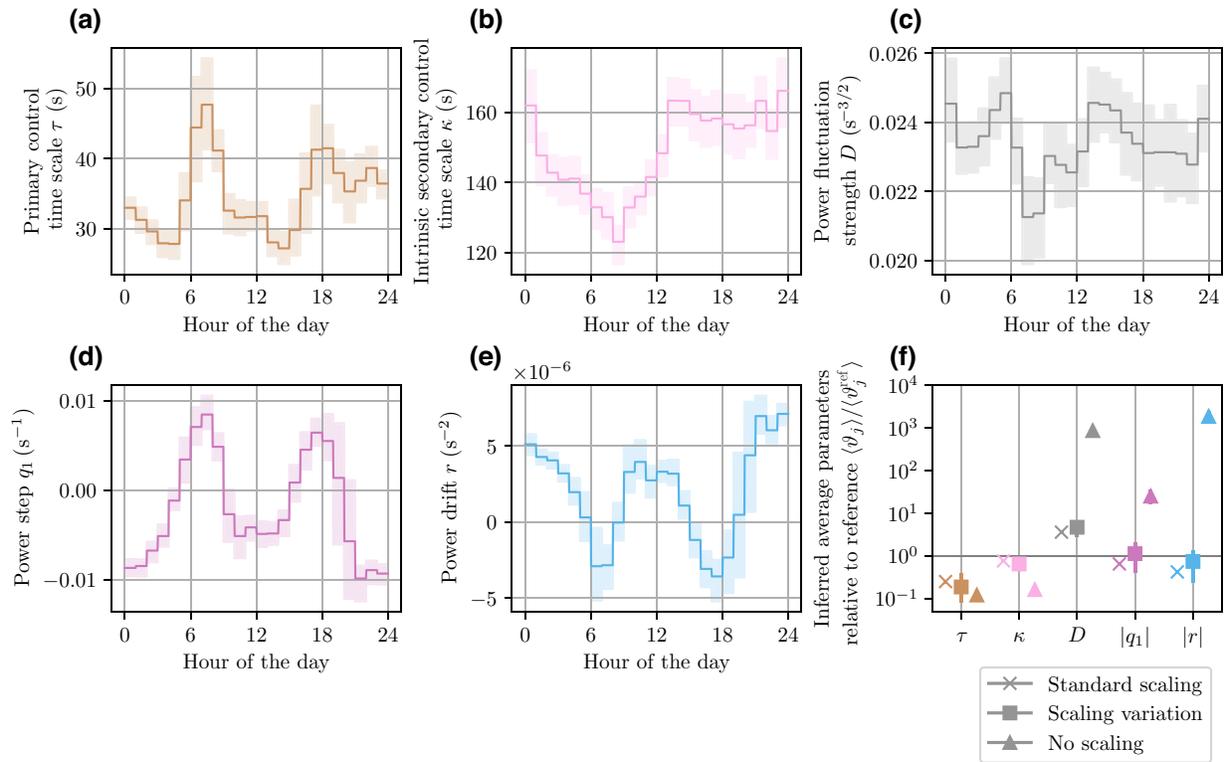
FIG. 3. Physics-informed machine learning model infers dynamical system parameters. (a)–(e) The inferred parameters showed strong daily patterns, which underlines the importance of time-dependent dynamical properties in our model. The panels depict the daily means (solid lines) and the range between the 25% and 75% quantile (area). (f) The ratio between the average inferred parameters and the reference value ϑ_j^{ref} obtained from time-independent stochastic inference (cf. Table III). In addition to the standard scaling used in the main model and a model with no scaling, we tested small variations of the scaling parameters (defined in Table IV) with ten random weight initializations for each combination of scaling parameters. The ensemble means (markers) and the data range (error bar) indicate the relation between the inferred parameters and the reference, which showed particularly large deviations if no scaling is applied.

The machine-learning model outperforms elementary benchmarks on a prediction horizon of 15 min [Fig. 2(a)]. The *ex post* model, defined in Sec. II C, yielded a lower median loss than the daily profile and the constant model for predicting the first $t_{\max} = 15$ min of each interval. Restricting the feature set to day-ahead available data (cf. Sec. II C) yielded a similar performance, which enables us to forecast future frequency deviations better than the daily profile. With growing prediction length $t_{\max}$, the performance increase of our PIML models (relative to the daily profile) deteriorated [Fig. 2(b)]. This is probably due to an accumulation of errors as the model starts to deviate from the initial condition, which was fixed by the data.

The prediction examples in Figs. 2(c)–2(f) illustrate the strengths and limitations of our model. The intervals with the best model performance at 01:00 and 18:00 demonstrate how our model outperforms the daily profile in the first 15 min [Fig. 2(c)] or even the full hour [Fig. 2(e)]. A remarkable aspect is observed when inspecting the intervals with the worst performance: as the physics-informed

model fails to capture the dynamics, so does the daily profile, and both models offer a similar prediction [Figs. 2(d) and 2(f)]. In these cases, our PIML model reproduced a daily average behavior, possibly missing enough detailed information in the techno-economic feature set.

In the following sections, we explain dynamical parameters based on techno-economic features, among others. The *ex post* model better suits for explanation as it also includes actually measured features, such as forecast errors. Therefore, we only focus on the *ex post* model and predict the full interval (with $t_{\max} = 1$ h) in following sections.

B. System identification and explanation

1. Inference and variation of system parameters

In addition to probabilistic prediction, our PIML model provides a tool to infer dynamical system parameters from frequency measurements and techno-economic features (Fig. 3). In contrast to time-independent models [14,47], our parameter model $\mathcal{F}_{\text{NN}}$ extracts time-dependent system parameters $\vartheta(i)$, which mirror the local dynamical

TABLE III. Reference values of the system parameters $\vartheta_j^{\mathrm{ref}}$ obtained in a time-independent model [14]. We note that this study employed the actual grid frequency instead of the angular velocity. Thus, we rescaled the results according to $D \to 2\pi D$, $q_1 \to 2\pi q_1, r \to 2\pi r, \theta \to 2\pi\theta$, while τ and κ stayed the same. Furthermore, this study did not directly provide a result for r, but its value was implicitly defined through the constraint $\langle P_{\mathrm{im}}(t)\rangle_t = 0$, which yields $r = -2q/t_{\max}$.

Parameter	Value
τ	120 s
κ	183 s
D	0.007 s$^{-3/2}$
q_1	0.01 s^{-2}
r	0.000 009 s^{-3}

properties of load-frequency control (cf. Fig. 1). These parameters change in a stepwise manner for each hourly interval i (cf. Sec. II B). Note that we only estimated effective parameters that also contain the impact of the inertia (cf. Sec. II), which we discuss later in Sec. IV. When it comes to the deterministic imbalance parameters $q_{1,...,4}$, we only discuss the results for q_1, the power step at the beginning of the hour as an application example.

The inferred parameters strongly changed during the day, which illustrates the importance of time-dependent dynamical modeling [Figs. 3(a)–3(e)]. The daily profile of the primary control time scale showed variations of 37%, while the intrinsic secondary control time scale varied by 11%. The strength of high-frequency imbalance fluctuations D varied by 6% and the deterministic imbalance parameters q_1 and r changed by 80% and 120%.

The imbalance parameters showed distinct patterns that reveal physically meaningful impact factors on the grid frequency. We inferred upward power steps q_1 and negative drifts r in the morning around 06:00 and in the evening around 18:00, while the opposite behavior was estimated around noon and during the night. This successfully models the deterministic imbalances between scheduled generation and continuous load: around noon and during the night, the load is decreasing, thus causing downward power steps and positive drifts [cf. Figs. 1(b) and 1(c)].

The inferred parameters agree well with estimates from the literature. Typical values of the system parameters were obtained in a time-independent model using the Kramer-Moyal expansion by Rydin Gorjão et al. [14] and are summarized in Table III. Figure 3(f) depicts the ratio between the time average of our absolute parameter estimates and the reference value from the literature $\vartheta_j^{\mathrm{ref}}$. Our model (referred to as "standard scaling") infers parameter values within the same order of magnitude as the reference values, i.e., the ratio to the reference is near one.

The adequate inference of dynamical parameters is greatly facilitated by our implementation of appropriate scaling steps. In our PIML model, the last model layer rescales the inferred parameter to a similar order of magnitude (see, Sec. II D). In particular, we used the scaling constants defined in Appendix B throughout the paper (referred to as "standard scaling"). To assess the impact of the rescaling, we defined two other scaling configurations. If no scaling is applied (triangle markers), the estimated parameters deviated from the reference by multiple orders of magnitude. However, small variations of our standard scaling coefficients, which are defined in Appendix B, did not strongly change the parameter estimates (square markers). This indicates that scaling is important to obtain reasonable parameter estimates with fixed hyperparameters (number of epochs, learning rate, etc.), but the results seem to be independent of the exact choice of the scaling. As already discussed in Sec. II D, the rescaling probably leads to better results due to the large difference in scale between the model parameters, which renders the FFNN training inefficient and unstable [41].

2. Techno-economic drivers of dynamical system properties

Using SHAP values [31], we explain the dependencies between the techno-economic features and the dynamical parameters ϑ_j, which the model extracted from the data (Fig. 4). Focusing on the deterministic mismatch parameters q_1 and r and the primary control time scale τ, we analyze feature importance quantified by the mean absolute SHAP value [Fig. 4(a)], as well as dependencies [Figs. 4(b)–4(d)] that display SHAP values for different feature values.

The power step q_1 was mostly determined by generation and load ramps [Figs. 4(a) and 4(b)]. The load ramp strongly correlates with the direction of scheduled generation steps as electricity markets aim at covering the load. It thus had a positive impact on the step q_1. In addition, fast generation types are important features. It is known that especially fast generation ramps drive the power step and thus the rate of change of frequency (RoCoF) at the beginning of the market intervals [23]. Accordingly, the PIML model yielded a high positive impact of hydro power ramps, which are among the fastest in the European power system.

Drift r of the deterministic mismatch mirrors continuous changes of the load [Figs. 1(b) and 1(c)]. Consistently, load ramps obtained the highest feature importance for r with positive load ramps leading to negative slopes r [Figs. 4(a) and 4(c)]. Solar ramps were also ranked highly, but their dependency showed the opposite behavior. This mirrors the fact that in addition to the load, solar power also shapes the slow evolution of the deterministic mismatch $P_{\mathrm{im}}(t) \sim rt$ [37]: the load and aggregated solar power typically change slowly and continuously on a time scale of hours, with the load having a negative impact and the solar
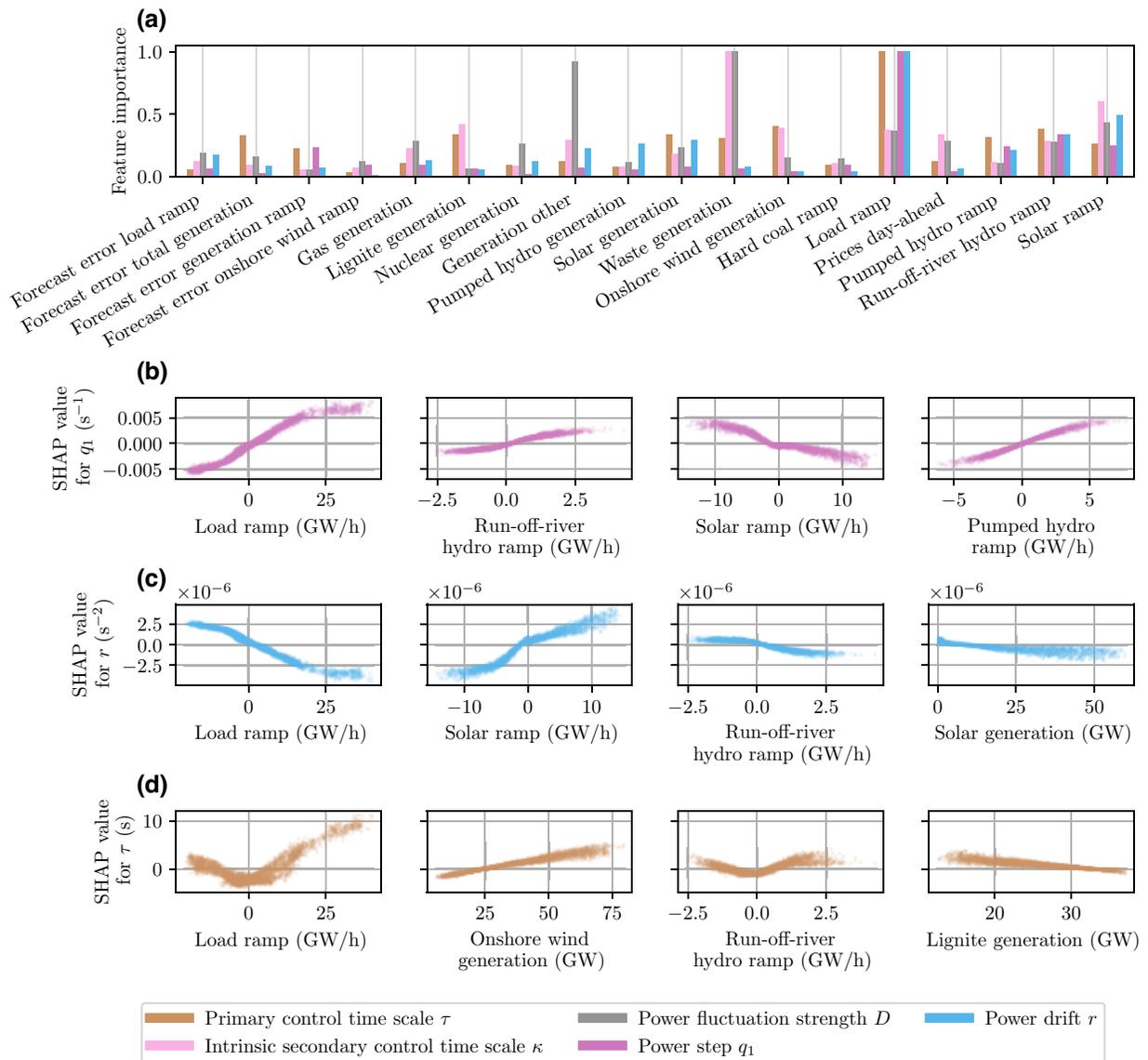
FIG. 4. SHAP values reveal dependencies between techno-economic features and dynamical system parameters. (a) The mean absolute SHAP values quantify the overall importance of a feature, which varies strongly among the different dynamical system parameters. We show the eight most important features for each model parameter. (b)–(d) The relation between feature values and SHAP values reveals the dependencies that were extracted by our model. The inferred dependencies generally represent physically meaningful effects that agree with domain knowledge (see the main text).

power having a positive impact on the power imbalance. This perfectly manifests in the opposite effects of load and solar ramps on the mismatch slope r, which were identified by our PIML model [Fig. 4(c)].

The effective time scale of primary control τ was determined by load and generation ramps, wind power, and lignite power generation [Figs. 4(a) and 4(d)]. Most interestingly, increasing wind power generation led to a larger time scale of primary control. In an Ornstein-Uhlenbeck process, τ quantifies the time to revert back to the mean after a disturbance. A large wind power feed-in can cause large stochastic imbalances and thus effectively reduce

the mean-reverting time τ. This would be consistent with our SHAP results and with previous studies that showed an increased variability of short-term frequency dynamics with increasing wind power feed-in [48]. Notably, this dependency cannot be caused by the rescaling of dynamical parameters with the inertia. If the actual primary control strength M/τ was constant (cf. Sec. II), increasing wind power, and thus decreasing inertia M, would correspond to decreasing values of τ. However, we observed the opposite: the dependency showed increasing values of τ [Fig. 4(d)], thus pointing to other causes such as an increased variability.

C. Generation of synthetic grid frequency time series

A third major application of probabilistic machine learning models is the generation of synthetic time series. Scenario generation, i.e., the generation of multiple synthetic samples from the model, is important for simulation or optimization models [49,50]. Given a data set of techno-economic features, synthetic time series are obtained as follows. For every interval i, we applied the FFNN to predict the system parameters $\vartheta(i)$. We then integrated the original SDE (5) using a standard Euler-Maruyama method. To ensure continuity, we used the final values of ω and θ from one interval as initial states for the following interval. As a test case, we generated a synthetic trajectory from April 06 to April 13 in our test set, for which the first hours are shown in Fig. 5(a).

Power grid frequency trajectories exhibit several highly characteristic stochastic properties [25]: the distribution of frequency deviations $\omega(t)$ is heavy tailed [Fig. 5(b)], i.e., large deviations are much more likely than expected from conventional normal statistics. Furthermore, the autocorrelation function peaks at multiples of a quarter hour, the smallest interval of electricity trading in Europe, which are most strongly pronounced after one hour [Fig. 5(d)]. All these characteristic patterns were well reproduced by our PIML model.

A discrepancy of the data and model is observed in the statistics of the frequency increments $\Delta\omega_T(t) = \omega(t+T) - \omega(t)$ for $T = 1$ s, as shown in Fig. 6. The increments are correlated for time scales of up to 8 s, which indicates correlated power fluctuations $P_{\text{noise}}(t)$. In contrast, the PIML model assumes Gaussian white noise [cf. Eq. (4)], and thus cannot reproduce the correlation of the increments for $T = 1$ s. At the same time, the PIML model strongly overestimates the variance of the increments. Increments with a longer time lag of $T = 60$ s are much better reproduced though.

We conclude that the assumption of uncorrelated power fluctuations $P_{\text{noise}}(t)$ is not fully justified. The model compensates for the missing correlations by increasing the noise strength, such that the frequency diffusion on longer timescales fits the data. We note, however, that the assumption of uncorrelated fluctuations is needed to derive the Fokker-Planck equation and its analytic solution. Going beyond this approximation would render model training close to impossible.

All in all, our results reveal important aspects of load-frequency dynamics and control. The success of our model suggests that the non-normal statistics is a direct consequence of the nonautonomous character of the power system. We recall that a *time-invariant* linear Gaussian process always leads to a Gaussian distribution of the observables. The changing system parameters in the power grid naturally induce heavy tails in the frequency distribution, without the need for heavy-tailed power fluctuations or nonlinear evolution equations; cf. the discussion in Ref. [15,25].
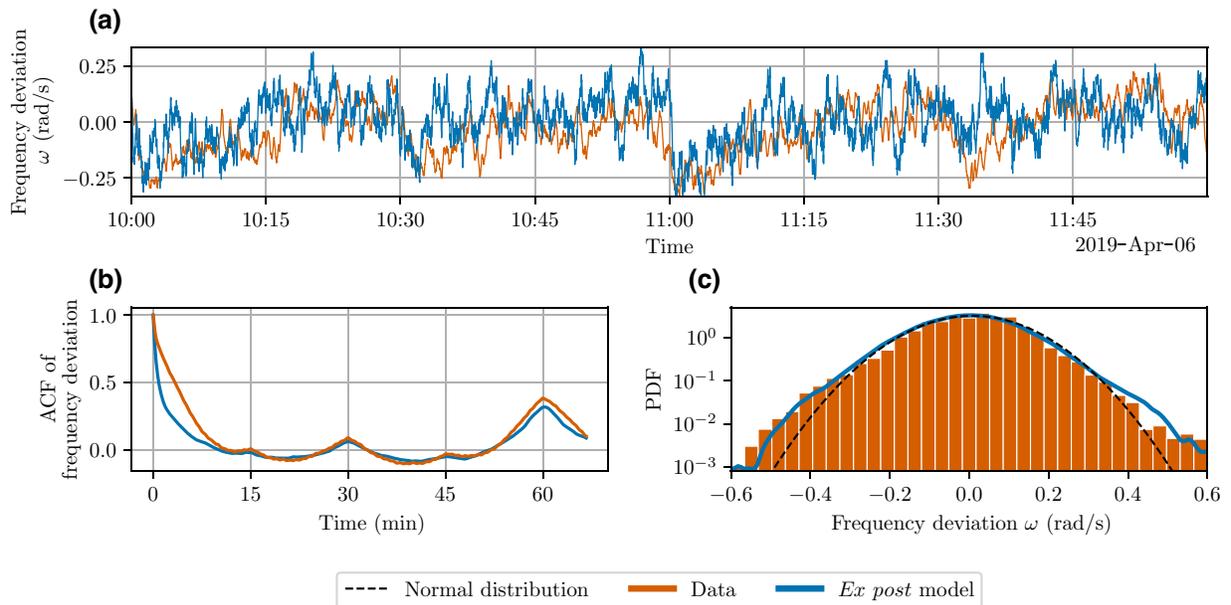
FIG. 5. Physics-informed machine learning model generates synthetic grid frequency time series with characteristic properties. (a) We sampled a synthetic frequency trajectory from our probabilistic model for the test period April 06 to April 13 in our test set, which was the longest period without missing or corrupted data points. The panel depicts the start of this period, showing good agreement between real and synthetic time series. (b),(c) The autocorrelation function (ACF) and the probability density function (PDF) of the frequency deviation $\omega(t)$ are well reproduced by our synthetic time series.
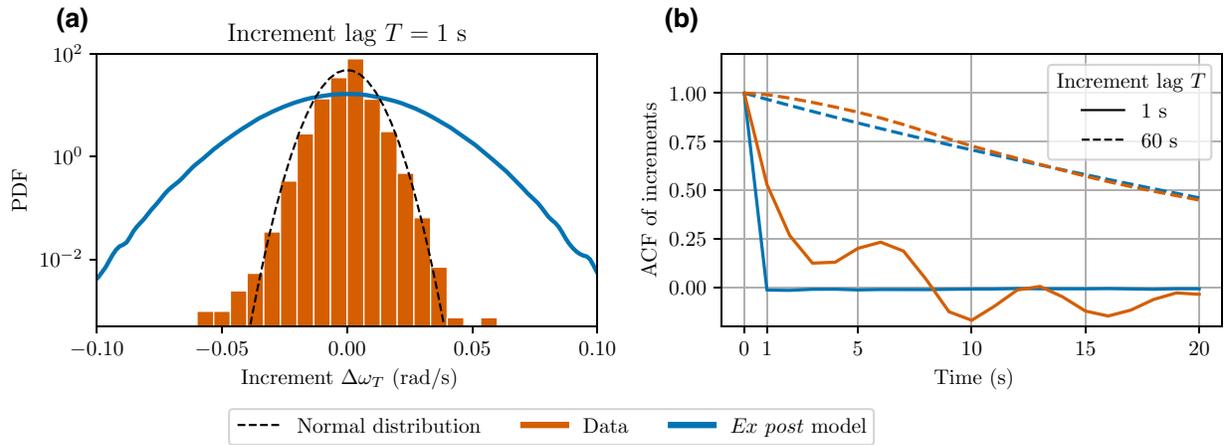
FIG. 6. Discrepancy of the data and model at the shortest time scales. (a) The PIML model strongly overestimated the variability of small-scale frequency increments $\Delta\omega_T(t) = \omega(t + T) - \omega(t)$ ($T = 1$ s), as indicated by their PDF. (b) The ACF of the increments indicates the presence of correlations on time scales of several seconds in the power fluctuations P_{noise}. In contrast, we have to assume uncorrelated noise in the PIML model to keep the training process feasible. To still reproduce the variability of the grid frequency, the model had to assume a larger variance in the increment PDF.

IV. DISCUSSION

We have developed a model of power system operation that integrates both the internal system dynamics and the external techno-economic features. The integration has been achieved by the combination of an explicit simulation model in terms of stochastic differential equations, and an artificial neural network to link the external influences to the system parameters. We thus obtained a generic physics-informed machine learning model of power system dynamics and control.

Using grid frequency recordings from the continental European power grid as a test case, we demonstrated three applications of our physics-informed model. First, we provided a probabilistic prediction of the grid frequency in intervals of 1 h. In the first 15 min, our model outperformed the daily average profile of the grid frequency, which is already a good predictor in continental Europe [23]. This was also possible when using only day-ahead available techno-economic features, thus providing the possibility to forecast grid frequency dynamics 15 min ahead. Previous grid frequency predictors only used historic frequency data as inputs [51–53]. Approaches that integrated external features previously focused on aggregated frequency deviations [23,54], which is also a cause of the data quality. Techno-economic features are typically available only on aggregated time scales of 15 min or 1 h, while the frequency fluctuates on a time scale of seconds (and even shorter time scales). We bridge the gap between large-scale techno-economic features and short-term frequency dynamics by using a physics-informed model. It connects the time-aggregated features with dynamical parameters of a stochastic process that well describes the short-term grid frequency fluctuations.

Second, our model provides a tool for system identification and explanation. The model inferred the effective system parameters for every hour from frequency measurements and techno-economic input features. The parameters were rescaled by the inertia (cf. Sec. II), but the actual system parameters can be obtained by incorporating inertia time series, which however can only be approximated for large-scale power systems [21,55]. The inferred parameters varied strongly in time, which indicates the importance of modeling the grid frequency as a nonautonomous system with time-dependent parameters. Explaining the inferred parameters with SHAP values further revealed their dependency on techno-economic drivers. For example, the primary control time scale increased with rising feed-in of wind power, which is harder to control and thus effectively leads to longer relaxation times. Our tool therefore extracts and explains physically meaningful system parameters and their time-dependent drivers.

Third, we used our model for the generation of synthetic grid frequency time series. The synthetic data well approximated the heavy-tailed distribution of frequency deviations and the recurrent patterns in its autocorrelation. In contrast to previous stochastic models [13–15], the synthetic time series also reproduced the actual frequency trajectory with its local time-dependent characteristics. Notably, the model requires very little system-specific information as inputs, but learns them directly from the data. Hence, the model is highly flexible and can be easily transferred to other grids.

A key feature of the load-frequency dynamics is the presence of three distinct time scales. The power balance fluctuates on time scales of seconds and below, which was represented by the term $P_{\mathrm{noise}}(t)$ in the model. The data

suggest that fluctuations are correlated on time scales of a few seconds, but these correlations had to be approximated to keep the model tractable. Load-frequency control acts on the time scale of minutes to restore the power balance and bring the frequency back to its reference value. This dynamics has been modeled by stochastic differential equations, where the high-frequency fluctuations enter as noise. The power system further evolves on time scales of hours due to changes in the dispatch of power plants, the trading on electricity markets, and the demand. This evolution is included in the model via changing system parameters, including the droop constants of the control system and the systematic power balance P_{im}. This evolution has been modeled by a neural network in terms of large-scale energy system features.

Notably, the load-frequency control system itself comprises different processes with different time constants. The inertia of the synchronous machines determines the response to a perturbation or a change in the dispatch. Though we cannot infer the numerical value of the inertia, the response is very well reproduced by the model. This "momentary reserve" is the fastest contribution to the load-frequency control system, and its role in future renewable power systems is an important topic in power engineering [12]. Primary control restores the power balance, while secondary control brings the grid frequency back to its reference value. We assumed that primary control is faster than secondary control ($\tau < \kappa$) to ensure an overdamped dynamics. The numerical results yield values around $\tau \approx 35$ s and $\kappa \approx 145$ s, which consistently satisfy our assumptions and agree with reference values [14].

In the context of power system dynamics and control, physics-informed machine learning methods have become popular during the past years [56,57]. Classical physics-informed neural networks (PINNs) are commonly applied to the differential equations directly [58], which we circumvented by providing a semianalytical solution for our system. However, a (semi)analytical solution is not possible anymore when including nonlinearities such as deadbands. In the future, our model can be modified to leverage classical PINNs to also treat nonlinearities and more generic power system dynamics. Previous applications of PINNs to power system dynamics have successfully addressed autonomous dynamics [57,59]. We contribute to these developments by proposing a model that explicitly models nonautonomous dynamics, which may greatly advance the application of physics-informed machine learning in the energy sector.

ACKNOWLEDGMENTS

APPENDIX A: A STOCHASTIC MODEL OF GRID FREQUENCY DYNAMICS AND CONTROL

1. The aggregated swing equation

Grid frequency deviations $\Delta f(t) = (f(t) - f_{ref})$ from the reference $f_{ref} = 50$ or 60 Hz reflect power imbalances ΔP in the grid, which have to be compensated via different sources. Firstly, the rotational energy of synchronous machines provides momentary reserve power P_{rot} through changes in the rotation speed [60]:

$$P_{rot} = \frac{2HS_B}{f_{ref}}\frac{d\Delta f}{dt} = 2\pi M \frac{d\Delta f}{dt}. \quad (A1)$$

Here, H denotes the average inertia constant of a synchronous machine, which is typically around 6 s for conventional generators [13] and S_B is the total rated power of all generators. Loosely speaking, H equals the kinetic energy of the rotating machine, rotating at f_{ref} divided by its rated power. Parameter M then denotes the aggregated inertia of the grid.

Secondly, damping power is provided via primary control, also referred to as frequency containment reserve, and frequency-sensitive loads [60],

$$P_{prim} = K_1 \Delta f, \quad (A2)$$

where $K_1 = K_T + K_L$ is the inverse droop coefficient, which comprises the effect of control K_T and load damping K_L ($[K_1] = W/Hz$). The control effect K_T is typically 1 or 2 orders of magnitude larger than K_L [11]. Their values are often provided in the *per unit* (pu) system with $K_1^{pu} = f_{ref}K_1/P_0$, as parameter K_T depends on the steady-state load P_0 within the specific system. For example, in Great Britain a typical value of $K_1^{pu} = 12.5$ is reported [13].

Thirdly, secondary control, also referred to as frequency restoration reserve, restores the frequency back to its reference f_{ref}. Secondary control is typically implemented as an integral controller (but other implementations exist) [61]:

$$P_{sec} = K_2 \bar{\theta} \quad (A3)$$

with the integrated frequency deviation

$$\bar{\theta}(t) = \int_{t_i}^{t} \Delta f(t')\, dt'. \quad (A4)$$

Parameter K_2 is the secondary control gain ($[K_2] = $ W), which reads $K_2^{\mathrm{pu}} = f_{\mathrm{ref}} K_2 / P_0$ in the pu system. For example, in Great Britain a typical value of $K_2^{\mathrm{pu}} = 0.05/\mathrm{s}$ is reported [13]. In interconnected power grids, secondary control may also be used to reduce unscheduled flows between different control areas [62]. These control actions are applied reciprocally in two areas and thus have only minor effects on the overall frequency dynamics.

Finally, this yields the aggregated swing equation, which is simply the power balance of all contributions:

$$\frac{d\bar{\theta}}{dt} = \Delta f, \tag{A5}$$

$$2\pi M \frac{d\Delta f}{dt} = -K_1 \Delta f - K_2 \bar{\theta} + \Delta P(t). \tag{A6}$$

Here, $\Delta P(t)$ denotes the imbalance of power generation and load, excluding the contribution of the control system (A2)–(A3).

2. Stochastic differential equations

Following Rydin Gorjão *et al.* [14], we modeled the power imbalances $\Delta P(t) = P_{\mathrm{im}}(t) + \bar{D}\xi(t)$ as a sum of "deterministic" power imbalances $P_{\mathrm{im}}(t)$ and stochastic deviations $\bar{D}\xi(t)$. In the large European power grids, deterministic power imbalances repeatedly arise due to a different ramping of dispatchable generators and the load or due to forecasting errors [63]. Stochastic deviations are modeled as Gaussian noise $\xi(t)$ defined as the derivative of a Wiener process W_t that has independent, normally distributed increments dW_t with zero mean and variance $\langle dW_t^2 \rangle = dt$.

To estimate model parameters, we have to rescale Eq. (A6) by M as all parameters are otherwise undefined up to a multiplicative factor. In addition, we transition from frequencies and integrated frequencies to angular velocities and angles. We define

$$\tau^{-1} = \frac{K_1}{2\pi M}, \tag{A7}$$

$$\kappa^{-2} = \frac{K_2}{2\pi M}, \tag{A8}$$

$$D = \frac{\bar{D}}{M}, \tag{A9}$$

$$P(t) = \frac{P_{\mathrm{im}}(t)}{M}, \tag{A10}$$

$$\theta = 2\pi\bar{\theta}. \tag{A11}$$

Based on these definitions, we obtain a stochastic model for the angular grid frequency deviation $\omega = 2\pi \Delta f$:

$$d\theta = \omega \, dt, \tag{A12}$$

$$d\omega = (P(t) - \tau^{-1}\omega - \kappa^{-2}\theta)dt + D \, dW_t. \tag{A13}$$

Collecting both stochastic variables into a vector $\mathbf{X} = (\theta, \omega)^T$, we can write our model as a two-dimensional matrix equation

$$d\mathbf{X} = \mathbf{a}(\mathbf{X}, t)dt + \mathbf{D} \, d\mathbf{W}_t, \tag{A14}$$

using the drift vector $\mathbf{a}(\mathbf{X}, t) = (a_\theta, a_\omega)^T$ with $a_\theta = \omega$, $a_\omega = P(t) - \tau^{-1}\omega - \kappa^{-2}\theta$, the diffusion matrix $\mathbf{D} = \mathrm{diag}(D, 0)$, and a two-dimensional Wiener process $\mathbf{W}_t$.

The stochastic differential equation (A14) can be recast into a Fokker-Planck equation for the joint probability density function $\mathcal{P}(\theta, \omega; t)$ that describes the distribution of the two random variables θ and ω at time t [64]. Using Itô calculus, one obtains

$$\frac{\partial}{\partial t}\mathcal{P}(\theta, \omega; t) = \left[-\frac{\partial}{\partial \omega}(P(t) - \tau^{-1}\omega - \kappa^{-2}\theta) \right.$$
$$\left. -\frac{\partial}{\partial \theta}\omega + \frac{D^2}{2}\frac{\partial^2}{\partial \omega^2} \right]\mathcal{P}(\theta, \omega; t). \tag{A15}$$

3. Solution of the Fokker-Planck equation

In this section, we prove that the Fokker-Planck equation (A15) is solved by a multivariate normal distribution with PDF

$$\mathcal{P}(\mathbf{x}; t) = \frac{1}{2\pi |\Sigma|} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu) \right) \tag{A16}$$

with $\mathbf{x}^\top = (\theta, \omega)$ and time-dependent parameters

$$\mu(t) = \begin{pmatrix} \mu_\theta(t) \\ \mu_\omega(t) \end{pmatrix}, \qquad \Sigma(t) = \begin{pmatrix} \sigma_\theta^2(t) & \sigma_{\theta\omega}(t) \\ \sigma_{\theta\omega}(t) & \sigma_\omega^2(t) \end{pmatrix}$$

if the parameters satisfy the ordinary differential equations

$$\frac{d}{dt}\mu_\theta = \mu_\omega, \tag{A17a}$$

$$\frac{d}{dt}\mu_\omega = P(t) - \tau^{-1}\mu_\omega - \kappa^{-2}\mu_\theta, \tag{A17b}$$

$$\frac{d}{dt}\sigma_\theta^2 = 2\sigma_{\theta\omega}, \tag{A17c}$$

$$\frac{d}{dt}\sigma_{\theta\omega} = \sigma_\omega^2 - \tau^{-1}\sigma_{\theta\omega} - \kappa^{-2}\sigma_\theta^2, \tag{A17d}$$

$$\frac{d}{dt}\sigma_\omega^2 = -2\tau^{-1}\sigma_\omega^2 - 2\kappa^{-2}\sigma_{\theta\omega} + D^2. \tag{A17e}$$

We prove this result using the characteristic function, which is defined via the Fourier transform

$$\phi(\mathbf{u}; t) = \int e^{i\mathbf{u}^\top \mathbf{x}} \mathcal{P}(\mathbf{x}; t)d^2\mathbf{x}. \tag{A18}$$

In terms of the characteristic function, the FPE reads

$$
\frac{\partial}{\partial t}\phi(\mathbf{u};t) = \mathcal{L}\phi(\mathbf{u};t)
$$

$$
= is\left[\left(P(t) + i\tau^{-1}\frac{\partial}{\partial s} + i\kappa^{-2}\frac{\partial}{\partial r}\right)\right.
$$

$$
\left. + r\frac{\partial}{\partial s} - \frac{D^2}{2}s^2\right]\phi(\mathbf{u};t), \quad \text{(A19)}
$$

where we have defined $\mathbf{u}^{\top} = (r,s)$. The characteristic function of the normal distribution (A16) reads

$$
\phi(\mathbf{u};t) = \exp\left(i\mathbf{u}^{\top}\mu - \tfrac{1}{2}\mathbf{u}^{\top}\Sigma\mathbf{u}\right),
$$

$$
\phi(r,s;t) = \exp\left(i(r\mu_{\theta} + s\mu_{\omega})\right. \quad \text{(A20)}
$$

$$
\left. - \tfrac{1}{2}(r^2\sigma_{\theta}^2 + s^2\sigma_{\omega}^2 + 2rs\sigma_{\theta\omega})\right).
$$

We now show that the normal distribution (A20) with parameters (A17) satisfies the Fokker-Planck equation (A19). We first evaluate the right-hand side of the FPE,

$$
\mathcal{L}\phi(\mathbf{u};t) = is\left[\left(P(t) + i\tau^{-1}\frac{\partial}{\partial s} + i\kappa^{-2}\frac{\partial}{\partial r}\right) + r\frac{\partial}{\partial s} - \frac{D^2}{2}s^2\right]
$$

$$
\times \exp\left(i(r\mu_{\theta} + s\mu_{\omega})\right.
$$

$$
\left. - \frac{1}{2}(r^2\sigma_{\theta}^2 + s^2\sigma_{\omega}^2 + 2rs\sigma_{\theta\omega})\right),
$$

$$
= \left[isP(t) - \frac{D^2}{2}s^2 - (\tau^{-1}s - r)\right.
$$

$$
\times (i\mu_{\omega} - s\sigma_{\omega}^2 - r\sigma_{\theta\omega})
$$

$$
\left. - \kappa^2 s(i\mu_{\theta} - r\sigma_{\theta}^2 - s\sigma_{\theta\omega})\right]\phi(\mathbf{u};t). \quad \text{(A21)}
$$

Now we proceed with the left-hand side,

$$
\frac{\partial}{\partial t}\phi(\mathbf{u};t) = \left[i\left(s\frac{d\mu_{\omega}}{dt} + r\frac{d\mu_{\theta}}{dt}\right)\right.
$$

$$
\left. - \frac{1}{2}\left(r^2\frac{d\sigma_{\theta}^2}{dt} + s^2\frac{d\sigma_{\omega}^2}{dt} + 2rs\frac{d\sigma_{\theta\omega}}{dt}\right)\right]\phi(\mathbf{u};t). \quad \text{(A22)}
$$

Inserting Eqs. (A17) then yields

$$
\frac{\partial}{\partial t}\phi(\mathbf{u};t) = \left[isP(t) - \frac{D^2}{2}s^2 - (\tau^{-1}s - r)\right.
$$

$$
(i\mu_{\omega} - s\sigma_{\omega}^2 - r\sigma_{\theta\omega})
$$

$$
\left. - \kappa^2 s(i\mu_{\theta} - r\sigma_{\theta}^2 - s\sigma_{\theta\omega})\right]\phi(\mathbf{u};t), \quad \text{(A23)}
$$

which coincides with the right-hand side (A21).

4. Moment equations

The ordinary differential equations for parameters (A17) can also be obtained in a more direct way, once we know that the PDF remains Gaussian at all times. In fact, we can exploit the fact that the parameters of a Gaussian PDF equal the mean and the (co)variances. The dynamics of the mean and the (co)variances are determined by the moment equations, which we extracted using Itô's lemma. For any twice differentiable scalar function $g(\mathbf{X})$ of the random variable $\mathbf{X}$ in Eq. (A14), Itô's lemma reads [36]

$$
dg = \left[(\nabla g)^T\mu + \tfrac{1}{2}\mathrm{Tr}(\mathbf{D}^T\mathbf{H}_g\mathbf{D})\right]dt + (\nabla g)^T\mathbf{D}d\mathbf{W},
$$

$$
\text{(A24)}
$$

where ∇g is the gradient and $\mathbf{H}_g$ is the Hessian matrix of function $g(\mathbf{X})$. This yielded in our case

$$
dg = \left[\frac{\partial g}{\partial \omega}\mu_{\omega} + \frac{\partial g}{\partial \theta}\omega + \frac{D^2}{2}\frac{\partial^2 g}{\partial \omega^2}\right]dt + \frac{\partial g}{\partial \omega}D\,dW. \quad \text{(A25)}
$$

To apply this to moment functions, we further assumed that $d\langle g\rangle = \langle dg\rangle$. For the first moments (averages) $g = \langle\theta\rangle$ and $g = \langle\omega\rangle$, we obtained

$$
\frac{d\langle\theta\rangle}{dt} = \langle\omega\rangle, \quad \text{(A26)}
$$

$$
\frac{d\langle\omega\rangle}{dt} = P(t) - \tau^{-1}\langle\omega\rangle - \kappa^{-2}\langle\theta\rangle. \quad \text{(A27)}
$$

The second moments $g = \langle\theta^2\rangle$, $g = \langle\omega^2\rangle$ and the mixed moment $g = \langle\omega\theta\rangle$ yielded

$$
\frac{d\langle\theta^2\rangle}{dt} = 2\langle\theta\omega\rangle, \quad \text{(A28)}
$$

$$
\frac{d\langle\omega^2\rangle}{dt} = 2P(t)\langle\omega\rangle - 2\tau^{-1}\langle\omega^2\rangle - 2\kappa^{-2}\langle\omega\theta\rangle + D^2,
$$

$$
\text{(A29)}
$$

$$
\frac{d\langle\omega\theta\rangle}{dt} = P(t)\langle\theta\rangle - \tau^{-1}\langle\omega\theta\rangle - \kappa^{-2}\langle\theta^2\rangle + \langle\omega^2\rangle. \quad \text{(A30)}
$$

In this derivation, we used $\langle\omega\,dW\rangle = 0$ and $\langle\theta\,dW\rangle = 0$. Identifying $\mu_{\theta} = \langle\theta\rangle$, $\mu_{\omega} = \langle\omega\rangle$, $\sigma_{\omega}^2 = \langle\omega^2\rangle - \langle\omega\rangle^2$, $\sigma_{\theta}^2 = \langle\theta^2\rangle - \langle\theta\rangle^2$, and $\sigma_{\theta,\omega} = \langle\omega\theta\rangle - \langle\omega\rangle\langle\theta\rangle$ then reproduces Eqs. (A17).

5. Solution of the moment equations

We now provide a semianalytic solution for the ordinary differential equations (A17) describing the evolution of the parameters μ_{θ}, μ_{ω}, σ_{ω}^2, σ_{θ}^2, and $\sigma_{\theta,\omega}$. We first note that the equations for the deterministic part (the means) and the stochastic part [the (co)variances] decouple; hence, they

can be treated separately. We collected the deterministic equations using the vector $\mathbf{y}_d = (\mu_\theta, \mu_\omega)^T$:

$$\frac{d\mathbf{y}_d}{dt} = \mathbf{A}_d \mathbf{y}_d + \mathbf{b}_d(t), \tag{A31}$$

$$\mathbf{A}_d = \begin{pmatrix} 0 & 1 \\ -\kappa^{-2} & -\tau^{-1} \end{pmatrix}, \tag{A32}$$

$$\mathbf{b}_d(t) = \begin{pmatrix} 0 \\ P(t) \end{pmatrix}. \tag{A33}$$

With $\mathbf{y}_s = (\sigma_\theta^2, \sigma_{\theta,\omega}, \sigma_\omega^2)^T$, the stochastic part yielded

$$\frac{d\mathbf{y}_s}{dt} = \mathbf{A}_s \mathbf{y}_s + \mathbf{b}_s, \tag{A34}$$

$$\mathbf{A}_s = \begin{pmatrix} 0 & 2 & 0 \\ -\kappa^{-2} & -\tau^{-1} & 1 \\ 0 & -2\kappa^{-2} & -2\tau^{-1} \end{pmatrix}, \tag{A35}$$

$$\mathbf{b}_d = \begin{pmatrix} 0 \\ 0 \\ D^2 \end{pmatrix}. \tag{A36}$$

These equations are linear, ordinary differential equations (ODEs) for which several solution methods exist.

a. Solution of the homogeneous equations

The general solution of the homogeneous ODE $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$ with time-independent coefficients $\mathbf{A}$ is given by

$$\mathbf{y}_h(t) = \mathbf{U}(t)\mathbf{U}^{-1}(0)\mathbf{y}_0. \tag{A37}$$

The columns of matrix $\mathbf{U}(t)$ span the solution space of the homogeneous ODE. The column vectors are given by $\mathbf{u}_i = \mathbf{v}_i e^{\lambda_i t}$, where the $\mathbf{v}_i$ are the eigenvectors of $\mathbf{A}$ and the λ_i are the corresponding eigenvalues. Using any computer algebra program, we calculated the eigenvalues and eigenvectors for matrix $\mathbf{A}_d$ from the deterministic part:

$$\lambda_{d,1} = -\frac{1}{2\tau}\sqrt{1 - \frac{4\tau^2}{\kappa^2}} - \frac{1}{2\tau}, \tag{A38}$$

$$\lambda_{d,1} = \frac{1}{2\tau}\sqrt{1 - \frac{4\tau^2}{\kappa^2}} - \frac{1}{2\tau}, \tag{A39}$$

$$\mathbf{v}_{d,1} = (\lambda_{d,2}\kappa^2, 1)^T, \tag{A40}$$

$$\mathbf{v}_{d,2} = (\lambda_{d,1}\kappa^2, 1)^T. \tag{A41}$$

For matrix $\mathbf{A}_s$ from the stochastic part, we obtained

$$\lambda_{s,1} = -\tau^{-1}, \tag{A42}$$

$$\lambda_{s,2} = -\frac{1}{\tau}\sqrt{1 - \frac{4\tau^2}{\kappa^2}} - \frac{1}{\tau}, \tag{A43}$$

$$\lambda_{s,3} = \frac{1}{\tau}\sqrt{1 - \frac{4\tau^2}{\kappa^2}} - \frac{1}{\tau}, \tag{A44}$$

$$\mathbf{v}_{s,1} = \left(\kappa^2, -\frac{\kappa^2}{2\tau}, 1\right)^T, \tag{A45}$$

$$\mathbf{v}_{s,2} = \left(-\kappa^2 - \lambda_{s,3}\frac{\kappa^4}{2\tau}, \frac{\lambda_{s,3}\kappa^2}{2}, 1\right)^T, \tag{A46}$$

$$\mathbf{v}_{s,3} = \left(-\kappa^2 - \lambda_{s,2}\frac{\kappa^4}{2\tau}, \frac{\lambda_{s,2}\kappa^2}{2}, 1\right)^T. \tag{A47}$$

To specify the homogeneous ODE solutions for $\mu_\omega(t)$ and $\sigma_\omega^2(t)$, we needed the inverses of matrix $\mathbf{U}(t)$, i.e., matrices $\mathbf{U}_d^{-1}(t)$ and $\mathbf{U}_s^{-1}(t)$, which we again obtained through a computer algebra program. The ω components of the solution for the homogeneous system (A37) then read

$$\mu_{\omega,h}(t) = \frac{1}{\lambda_{d,1} - \lambda_{d,2}}[\kappa^{-2}\theta_0(e^{\lambda_{d,2}t} - e^{\lambda_{d,1}t}) + \omega_0(\lambda_{d,1}e^{\lambda_{d,1}t} - \lambda_{d,2}e^{\lambda_{d,2}t})], \tag{A48}$$

$$\begin{aligned}
\sigma_{\omega,h}^2(t) = &\left[\sigma_{\theta,0}^2(e^{\lambda_{s,2}t} + e^{\lambda_{s,3}t} - 2e^{\lambda_{s,1}t})\frac{\tau^2}{\kappa^2}\right. \\
&+ \sigma_{\omega,\theta,0}\left(-2\tau e^{\lambda_{s,1}t} + \frac{8\tau^2}{\kappa^2}\frac{\lambda_{s,2}\kappa^2(4\tau)^{-1} + 1}{\lambda_{s,2} - \lambda_{s,3}}e^{\lambda_{s,2}t}\right. \\
&\left.+ \frac{8\tau^2}{\kappa^2}\frac{\lambda_{s,3}\kappa^2(4\tau)^{-1} + 1}{\lambda_{s,3} - \lambda_{s,2}}e^{\lambda_{s,3}t}\right) \\
&+ \sigma_{\omega,0}^2\left(-2\tau e^{\lambda_{s,1}t} + \frac{2\lambda_{s,2}\tau^2 - \lambda_{s,2}\kappa^2 - 2\tau}{\lambda_{s,3} - \lambda_{s,2}}e^{\lambda_{s,2}t}\right. \\
&\left.\left.+ \frac{2\lambda_{s,3}\tau^2 - \lambda_{s,3}\kappa^2 - 2\tau}{\lambda_{s,2} - \lambda_{s,3}}e^{\lambda_{s,3}t}\right)\right]\frac{1}{\kappa^2 - 4\tau^2},
\end{aligned} \tag{A49}$$

where θ_0, ω_0 denote the initial conditions of the averages and $\sigma_{\theta,0}^2, \sigma_{\theta,\omega,0}, \sigma_{\omega,0}^2$ represent the initial conditions of the covariances.

b. Semianalytical solution of inhomogeneous equations

The general solution of the inhomogeneous ODE $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y} + \mathbf{b}(t)$ with time-independent coefficient $\mathbf{A}$ is given by

the sum of the homogeneous solution and an inhomogeneous contribution $\mathbf{y}_{\text{in}}(t)$:

$$\mathbf{y}_{\text{ODE}}(t) = \mathbf{y}_h(t) + \mathbf{y}_{\text{in}}(t),$$

$$= \mathbf{y}_h(t) + \mathbf{U}(t) \int_0^t \mathbf{U}^{-1}(t')\mathbf{b}(t')dt'. \qquad \text{(A50)}$$

We provide a semianalytical solution, which leaves the integration of inhomogeneity $b_d(t)$ to a numerical routine. This enables us to flexibly insert a different power function $P(t)$. The ω components of the inhomogeneous contributions yield

$$\mu_{\omega,\text{in}}(t) = \frac{1}{\lambda_{d,1} - \lambda_{d,2}} \left[\lambda_{d,1} e^{\lambda_{d,1}t} \int_0^t P(t') e^{-\lambda_{d,1}t'} dt' - \lambda_{d,2} e^{\lambda_{d,2}t} \int_0^t P(t') e^{-\lambda_{d,2}t'} dt' \right], \qquad \text{(A51)}$$

$$\sigma_{\omega,\text{in}}^2(t) = \frac{D^2}{\kappa^2 - 4\tau^2} \left[\frac{2\tau^2}{\lambda_{s,1}}(1 - e^{\lambda_{s,1}t}) + \frac{2\tau/\lambda_{s,2} - 2\tau^2 + \kappa^2}{\lambda_{s,3} - \lambda_{s,2}}(1 - e^{\lambda_{s,2}t}) \right.$$

$$\left. + \frac{2\tau/\lambda_{s,3} - 2\tau^2 + \kappa^2}{\lambda_{s,2} - \lambda_{s,3}}(1 - e^{\lambda_{s,3}t}) \right]. \qquad \text{(A52)}$$

The deterministic part $\mu_{\omega,\text{in}}(t)$ contains an integral over the deterministic power imbalance $P(t)$, which we can compute numerically for any power function. However, the factors $e^{-\lambda t}$ in Eq. (A51) can become very large as $\lambda < 0$ for stable systems, thus causing numerical problems. To use this semianalytical solution during neural network training, one has to mitigate these numerical problems. We did this by restricting the parameter space to ensure that λ does not become too large (Appendix B).

Note that we only require the marginal probability density $\mathcal{P}(\omega; t) = \int \mathcal{P}(\theta, \omega; t) \, d\theta$ to model the grid frequency dynamics; hence, we only needed a closed-form solution for μ_ω and σ_ω.

APPENDIX B: PARAMETER SCALING AND CONSTRAINTS

The developed PIML model includes a layer that rescales the parameters and ensures some physical constrains [Fig. 1(d)]. The outputs u_j of the FFNN do not necessarily fulfil the physical constraints of parameters ϑ_j (cf. Table IV), as the linear activation of the output takes arbitrary real values, while τ, for example, only takes positive real values. Moreover, the physical parameters ϑ_j vary strongly in scale (cf. Table III), but outputs u_j of the initialized FFNN typically have the same scale due to uniform random initialization of the weights [65]. This will yield large initial errors along certain parameter axes, thus leading to inhomogeneous loss landscapes that can make optimization inefficient and more difficult [41].

Therefore, we added a constraint and scaling layer that applies functions v_j to the FFNN outputs. The results then represent the parameter estimates $\vartheta_j = v_j(u_j)$. First, functions v_j enforce the physical constraints. For example, a softplus function $\mathcal{S}_+(u) = \log(\exp(u) + 1) \in (0, \infty)$ enforces positivity, and the sigmoid function $\text{Sig}(u)$

$= (1 + \exp(-u))^{-1} \in (0, 1)$ was used to ensure that $\tau \leq \kappa/2$ holds. Numerical imprecision can lead to a violation of these constraints, so we added a safety factor $\delta = 0.999$ in some cases. Second, factors s_j, which mirror the typical scale of parameters ϑ_j, are applied to make the optimization more efficient. Third, minimum values are added in certain cases to ensure numerical stability during optimization. For example, a very small standard deviation $\sigma_{\omega,0}$ can lead to probability densities beyond float precision. All in all, we defined the following functions using minimum values, scaling factors, and constraints from Table IV:

$$v_1(u_1) = \mathcal{S}_+(u_1)s_1 + \underline{v}_1,$$

$$v_2(u_2) = \delta \tanh(u_1)v_1(u_1)v_3(u_3),$$

$$v_3(u_3) = \mathcal{S}_+(u_3)s_3 + \underline{v}_3,$$

$$v_4(u_4) = \left(\frac{2\underline{v}_4}{v_5(u_5)} + \delta \text{Sig}(u_4) \left(1 - \frac{2\underline{v}_4}{v_5(u_5)} \right) \right) \frac{v_5(u_5)}{2},$$

$$v_5(u_5) = \mathcal{S}_+(u_5)s_5 + \underline{v}_5,$$

$$v_6(u_6) = \mathcal{S}_+(u_6)s_6 + \underline{v}_6,$$

$$v_j(u_j) = u_j s_j \quad \text{for } j \in \{7, 8, 9, 10\},$$

$$v_{11}(u_{11}) = u_{11}s_{11}.$$

To test the impact of the scaling with s_j, we varied the scaling parameters according to Table IV. In particular, we trained the PIML model for each combination of the individual scaling choices listed in the table. For each scaling tuple, we additionally simulated ten different random initializations of the FFNN weights. Finally, we trained ten initializations using the standard scaling defined in Table IV and no scaling with $s_j = 1$ (cf. Sec. III B).

TABLE IV. Properties of dynamical system parameters that are predicted by the FFNN [cf. Fig. 1(d)]. The parameters are subject to several physical constraints, which are summarized in the third row. The output of the FFNN is rescaled by constant factors listed in the fourth row to improve training efficiency (referred to as the standard scaling). To test the impact of the scaling, we varied scaling s_j according to the parameter choices in the fifth row. Furthermore, we ensure that the dynamical system parameter exceeds a minimum value listed in the sixth row.

Parameter ϑ_j	ϑ_1	ϑ_2	ϑ_3	ϑ_4	ϑ_5	ϑ_6	$\vartheta_{7,...,10}$	ϑ_{11}
Name	$\sigma_{\theta,0}$	$\sigma_{\theta,\omega,0}$	$\sigma_{\omega,0}$	τ	κ	D	$q_{1,...,4}$	r
Physical constraints	$\sigma_{\theta,0} > 0$	$|\sigma_{\theta,\omega,0}| \leq \sigma_{\theta,0}\sigma_{\omega,0}$	$\sigma_{\omega,0} > 0$	$0 < \tau \leq \kappa/2$	$\kappa > 0$	$D > 0$	$q_{1,...,4} \in \mathbb{R}$	$r \in \mathbb{R}$
Standard scaling s_j	1	$\cdots$	0.1	$\cdots$	100	0.01	10^{-3}	10^{-6}
Scaling variation	$\{1\}$	$\cdots$	$\{0.1\}$	$\cdots$	$\{1000, 100\}$	$\{0.01, 0.1\}$	$\{10^{-3}, 10^{-2}\}$	$\{10^{-5}, 10^{-6}\}$
Minimum v_j	10^{-3}	$\cdots$	10^{-3}	10	30	10^{-4}	$\cdots$	$\cdots$
Range of $v_j(u_j)$	(v_1, ∞)	$(-\sigma_{\theta,0}\sigma_{\omega,0}, \sigma_{\theta,0}\sigma_{\omega,0})$	(v_3, ∞)	$(v_4, \kappa/2)$	(v_5, ∞)	(v_6, ∞)	$(-\infty, \infty)$	$(-\infty, \infty)$

[1] J. Rockström, O. Gaffney, J. Rogelj, M. Meinshausen, N. Nakicenovic, and H. J. Schellnhuber, A roadmap for rapid decarbonization, Science **355**, 1269 (2017).

[2] J. Rogelj, G. Luderer, R. C. Pietzcker, E. Kriegler, M. Schaeffer, V. Krey, and K. Riahi, Energy system transformations for limiting end-of-century warming to below 1.5 °C, Nat. Clim. Chang. **5**, 519 (2015).

[3] I. Staffell and S. Pfenninger, The increasing impact of weather on electricity supply and demand, Energy **145**, 65 (2018).

[4] A. Orths, C. L. Anderson, T. Brown, J. Mulhern, D. Pudjianto, B. Ernst, O. Mark, J. McCalley, and G. Strbac, Flexibility from energy systems integration: Supporting synergies among sectors, IEEE Power Energy Mag. **17**, 67 (2019).

[5] D. Witthaut, F. Hellmann, J. Kurths, S. Kettemann, H. Meyer-Ortmanns, and M. Timme, Collective nonlinear dynamics and self-organization in decentralized power grids, Rev. Mod. Phys. **94**, 015005 (2022).

[6] D. Heide, L. von Bremen, M. Greiner, C. Hoffmann, M. Speckmann, and S. Bofinger, Seasonal optimal mix of wind and solar power in a future, highly renewable Europe, Renew. Energy **35**, 2483 (2010).

[7] S. Collins, P. Deane, B. Ó. Gallachóir, S. Pfenninger, and I. Staffell, Impacts of inter-annual wind and solar variations on the European power system, Joule **2**, 2076 (2018).

[8] J. Lin and F. H. Magnago, *Electricity Markets: Theories and Applications* (Wiley-IEEE Press, 2017), 1st ed.

[9] M. Anvari, E. Proedrou, B. Schäfer, C. Beck, H. Kantz, and M. Timme, Data-driven load profiles and the dynamics of residential electricity consumption, Nat. Commun. **13**, 4593 (2022).

[10] C. Han, H. Hilger, E. Mix, P. C. Böttcher, M. Reyers, C. Beck, D. Witthaut, and L. Rydin Gorjão, Complexity and Persistence of Price Time Series of the European Electricity Spot Market, PRX Energy **1**, 013002 (2022).

[11] J. Machowski, J. Bialek, and J. Bumby, *Power System Dynamics: Stability and Control* (Wiley, 2008), 2nd ed.

[12] F. Milano, F. Dörfler, G. Hug, D. J. Hill, and G. Verbic, in *2018 Power Systems Computation Conference (PSCC)* (IEEE, Dublin, Ireland, 2018), p. 1.

[13] P. Vorobev, D. M. Greenwood, J. H. Bell, J. W. Bialek, P. C. Taylor, and K. Turitsyn, Deadbands, droop, and inertia impact on power system frequency distribution, IEEE Trans. Power Syst. **34**, 3098 (2019).

[14] L. Rydin Gorjão, M. Anvari, H. Kantz, C. Beck, D. Witthaut, M. Timme, and B. Schäfer, Data-driven model of the power-grid frequency dynamics, IEEE Access **8**, 43082 (2020).

[15] D. Kraljic, Towards realistic statistical models of the grid frequency, IEEE Trans. Power Syst. **38**, 256 (2022).

[16] F. Guo and D. Schlipf, A spectral model of grid frequency for assessing the impact of inertia response on wind turbine dynamics, Energies **14**, 2492 (2021).

[17] M. Martínez-Barbeito, D. Gomila, and P. Colet, Dynamical model for power grid frequency fluctuations: Application to islands with high penetration of wind generation, IEEE Trans. Sustainable Energy **14**, 1436 (20223).

[18] ENTSO-E Transparency Platform: https://transparency.entsoe.eu/ (2020).

[19] L. Hirth, J. Mühlenpfordt, and M. Bulkeley, The ENTSO-E Transparency Platform—A review of Europe's most ambitious electricity data platform, Appl. Energy **225**, 1054 (2018).

[20] ENTSO-E, Report on Deterministic Frequency Deviations, https://consultations.entsoe.eu/system-development/deterministic_frequency_deviations_report/user_uploads/report_deterministic_frequency_deviations_final-draft-for-consultation.pdf (2019).

[21] ENTSO-E, Inertia and rate of change of frequency (RoCoF), https://eepublicdownloads.azureedge.net/clean-documents/SOC%20documents/Inertia%20and%20RoCoF˙v17˙clean.pdf (2020).

[22] U. Oberhofer, L. Rydin Gorjão, G. C. Yalcin, O. Kamps, V. Hagenmeyer, and B. Schäfer, Non-linear, bivariate stochastic modelling of power-grid frequency applied to islands, Preprint ArXiv:2301.04551 (2023).

[23] J. Kruse, B. Schäfer, and D. Witthaut, Revealing drivers and risks for power grid frequency stability with explainable AI, Patterns **2**, 100365 (2021).

[24] C. Beck and E. G. Cohen, Superstatistics, Phys. A: Stat. Mech. Appl. **322**, 267 (2003).

[25] B. Schäfer, C. Beck, K. Aihara, D. Witthaut, and M. Timme, Non-Gaussian power grid frequency fluctuations characterized by Lévy-stable laws and superstatistics, Nat. Energy **3**, 119 (2018).

[26] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Physics-informed machine learning, Nat. Rev. Phys. **3**, 422 (2021).

[27] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[28] J. Zhao, A. Gómez-Expósito, M. Netto, L. Mili, A. Abur, V. Terzija, I. Kamwa, B. Pal, A. K. Singh, J. Qi, Z. Huang, and A. P. S. Meliopoulos, Power system dynamic state estimation: Motivations, definitions, methodologies, and future work, IEEE Trans. Power Syst. **34**, 3188 (2019).

[29] L. Ljung, Perspectives on system identification, Annu. Rev. Control **34**, 1 (2010).

[30] J. Stiasny, G. S. Misyris, and S. Chatzivasileiadis, in *2021 IEEE Madrid PowerTech* (Madrid IEEE Power & Energy Society, 2021), p. 1.

[31] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intel. **2**, 56 (2020).

[32] J. Kruse, E. Cramer, B. Schäfer, and D. Witthaut, Supplementary code: "Physics-informed machine learning for power grid frequency modeling", https://doi.org/10.5281/zenodo.8014136 (2023).

[33] J. Kruse, E. Cramer, B. Schäfer, and D. Witthaut, Supplementary data: "Physics-informed machine learning for power grid frequency modeling", https://doi.org/10.5281/zenodo.8014065 (2023).

[34] T. Weissbach and E. Welfonder, in *2009 IEEE/PES Power Systems Conference and Exposition* (IEEE, Seattle, WA, USA, 2009), p. 1.

[35] L. Rydin Gorjão, R. Jumar, H. Maass, V. Hagenmeyer, G. C. Yalcin, J. Kruse, M. Timme, C. Beck, D. Witthaut, and B. Schäfer, Open database analysis of scaling and spatio-temporal properties of power grid frequencies, Nat. Commun. **11**, 6362 (2020).

[36] B. Øksendal, *Stochastic Differential Equations*, Universitext (Springer, 2003).

[37] J. Kruse, B. Schäfer, and D. Witthaut, in *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)* (IEEE, Aachen Germany, 2021), p. 133.

[38] EPEX SPOT, Press release: Intraday trading increases by 27.0% on year, https://www.epexspot.com/en/news/intraday-trading-increases-270-year (2020).

[39] J. Kruse, B. Schäfer, and D. Witthaut, Pre-Processed Power Grid Frequency Time Series, https://doi.org/10.5281/zenodo.5105820 (2020).

[40] TransnetBW GmbH, Regelenergie Bedarf + Abruf, https://www.transnetbw.de/de/strommarkt/systemdienstleistungen/regelenergie-bedarf-und-abruf (2020).

[41] P. Mehta, M. Bukov, C.-H. Wang, A. G. R. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, A high-bias, low-variance introduction to machine learning for physicists, Phys. Rep. A **810**, 1 (2019).

[42] X. Glorot and Y. Bengio, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (JMLR Workshop and Conference Proceedings, Sardinia, Italy, 2010), p. 249.

[43] S. M. Lundberg and S.-I. Lee, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17 (Curran Associates Inc., Long Beach California USA, 2017), p. 4768.

[44] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, *et al.*, in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (USENIX Association, Savannah, GA, USA, 2016), p. 265.

[45] J. V. Dillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, and R. A. Saurous, TensorFlow distributions, Preprint ArXiv:1711.10604 (2017).

[46] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, and L. Invernizzi, *et al.*, Keras tuner, https://github.com/keras-team/keras-tuner (2019).

[47] M. Anvari, L. Rydin Gorjão, M. Timme, D. Witthaut, B. Schäfer, and H. Kantz, Stochastic properties of the frequency dynamics in real and synthetic power grids, Phys. Rev. Res. **2**, 013339 (2020).

[48] H. Haehne, J. Schottler, M. Waechter, J. Peinke, and O. Kamps, The footprint of atmospheric turbulence in power grid frequency measurements, EPL (Europhys. Lett.) **121**, 30001 (2018).

[49] E. Cramer, L. Rydin Gorjão, A. Mitsos, B. Schäfer, D. Witthaut, and M. Dahmen, Validation methods for energy time series scenarios from deep generative models, IEEE Access **10**, 8194 (2022).

[50] E. Cramer, L. Paeleke, A. Mitsos, and M. Dahmen, Normalizing flow-based day-ahead wind power scenario generation for profitable and reliable delivery commitments by wind farm operators, Comput. Chem. Eng. **166**, 107923 (2022).

[51] J. Kruse, B. Schäfer, and D. Witthaut, Predictability of power grid frequency, IEEE Access **8**, 149435 (2020).

[52] W. Bang and J. W. Yoon, in *2019 International Conference on Information and Communication Technology Convergence (ICTC)* (Korea Institute of Electrical and Electronics Engineers (IEEE), Jeju Island, 2019), p. 1218.

[53] J. Dong, X. Ma, S. Djouadi, H. Li, and Y. Liu, Frequency prediction of power systems in FNET based on state-space approach and uncertain basis functions, IEEE Trans. Power Syst. **29**, 2602 (2014).

[54] S. Kaur, S. Agrawal, and Y. P. Verma, in *2013 5th International Conference and Computational Intelligence and Communication Networks* (2013), p. 311.

[55] S. Homan, N. Mac Dowell, and S. Brown, Grid frequency volatility in future low inertia scenarios: Challenges and mitigation options, Appl. Energy **290**, 116723 (2021).

[56] S. Chatzivasileiadis, A. Venzke, J. Stiasny, and G. Misyris, Machine learning in power systems: Is it time to trust it?, IEEE Power Energy Mag. **20**, 32 (2022).

[57] G. S. Misyris, A. Venzke, and S. Chatzivasileiadis, in *2020 IEEE Power Energy Society General Meeting (PESGM)* (Institute of Electrical and Electronics Engineers (IEEE), Montreal Canada, 2020), p. 1.

[58] M. Raissi, P. Perdikaris, and G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. **378**, 686 (2019).

[59] B. Huang and J. Wang, Applications of physics-informed neural networks in power systems–a review, IEEE Trans. Power Syst. **38**, 572 (2022).

[60] A. Ulbig, T. S. Borsche, and G. Andersson, Impact of low rotational inertia on power system stability and operation, IFAC Proc. Vol. **47**, 7290 (2014).

[61] E. Weitenberg, Y. Jiang, C. Zhao, E. Mallada, C. De Persis, and F. Dörfler, Robust decentralized secondary frequency control in power systems: Merits and tradeoffs, IEEE Trans. Automat. Contr. **64**, 3967 (2019).

[62] P. C. Böttcher, A. Otto, S. Kettemann, and C. Agert, Time delay effects in the control of synchronous electricity grids, Chaos **30**, 013122 (2020).

[63] L. Hirth and I. Ziegenhagen, Balancing power and variable renewables: Three links, Renew. Sustainable Energy Rev. **50**, 1035 (2015).

[64] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, 2009), 4th ed, Vol. 3.

[65] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer New York, 2016), 2nd ed.