Chaos
An Interdisciplinary Journal of Nonlinear Science

RESEARCH ARTICLE | MARCH 24 2025

Analyzing deterministic and stochastic influences on the power grid frequency dynamics with explainable artificial intelligence

Tim Drewnick ; Xinyi Wen ; Ulrich Oberhofer ; Leonardo Rydin Gorjão ; Christian Beck ; Veit Hagenmeyer ; Benjamin Schäfer

Chaos 35, 033153 (2025)
https://doi.org/10.1063/5.0239371

Analyzing deterministic and stochastic influences on the power grid frequency dynamics with explainable artificial intelligence

Cite as: Chaos 35, 033153 (2025); doi: 10.1063/5.0239371
Submitted: 19 September 2024 · Accepted: 7 March 2025 ·
Published Online: 24 March 2025

Tim Drewnick,[1] Xinyi Wen,[1] Ulrich Oberhofer,[1] Leonardo Rydin Gorjão,[2] Christian Beck,[3] Veit Hagenmeyer,[1] and Benjamin Schäfer[1,a)]

AFFILIATIONS

[1]Institute for Automation and Applied Informatics (IAI), Karlsruhe Institute of Technology (KIT), Karlsruhe 76344, Germany
[2]Faculty of Science and Technology, Norwegian University of Life Sciences, Ås 1433, Norway
[3]Queen Mary University of London, School of Mathematical Sciences, London E1 4NS, United Kingdom

[a)]**Author to whom correspondence should be addressed:** benjamin.schaefer@kit.edu

ABSTRACT

Power grids are essential for our society, connecting consumers and generators. Their frequency stability is impacted by supply and demand changes, including deterministic and stochastic dynamics, e.g., from market activities or fluctuating renewables. The first two Kramers–Moyal coefficients allow for a description of both the deterministic (via drift) and stochastic (via diffusion) aspects of these dynamics. Such a description and understanding could be critical to stabilizing power systems. However, how drift and diffusion differ between synchronous areas, how they vary over time, and how the generation mix influences them, remains unclear. Analyzing temporal patterns in drift and diffusion for frequency data from Australia (AUS) and Continental Europe (CE), we reveal a positive correlation between drift and diffusion. In addition, we utilize both gradient-boosted trees and neural network models to train drift and diffusion models for AUS and CE. Shapley additive explanations make these black-box models transparent and allow us to identify the total generation and load to influence the drift, while calendar features seem critical for the diffusion coefficient estimates.

Climate change is forcing us to look for more sustainable technologies in all aspects of our lives, including the power grid. This leads to a shift toward more renewable energy sources, which, in turn, challenges the power grid frequency stability as generation and demand are no longer easily balanced. Understanding the impact of different power generation methods on the power grid stability is, therefore, essential in ensuring a reliably functioning power grid. This article investigates these impacts by analyzing the stochastic and deterministic dynamics of the power grid frequency. In particular, we estimate drift and diffusion coefficients from frequency time series of the Continental European and the Australian power grid and analyze periodic and seasonal behavior. In order to determine the influence of techno-economic features on the frequency stability, we train a gradient-boosted tree model and a neural network for the ex-post prediction of drift and diffusion and explain the results using SHapley Additive exPlanations (SHAP).

I. INTRODUCTION

The amount of renewable energy in power grids worldwide is rising every year.[1] This increase is a direct measure to reduce carbon emissions and combat climate change. Consequently, power grids experience continuous changes such as integrating renewable energy sources like solar and wind generators. Supply and demand always need to be balanced, and this balance is easily monitored by the power grid frequency,[2] which is also used to inform control actions. Hence, a reduction in the relative amount of rotating mass in power systems directly impacts the frequency stability and control[3] by increasing the variability and unpredictability of the power grid dynamics and therefore challenging the maintenance of stability. Stability is essential in ensuring a reliable supply of power, one of our society's pillars.[4] Understanding how different factors, such as individual power generation methods, influence power grid stability is crucial for designing efficient balancing mechanisms and transitioning to a highly sustainable grid.

In the present article, we first provide some information about power grids and our data-driven methods (Sec. II). Next, we perform an initial data analysis of the extracted drift and diffusion coefficients (Sec. III) and discuss our machine learning results (Sec. IV), before closing with a discussion. The interested reader will find further details in Appendixes A–G and our code.[5]

II. BACKGROUND AND METHODS

A. Electrical power grids

An electrical power grid is a complex network consisting of suppliers and consumers of electrical power. Suppliers are entities responsible for generating and providing electricity. There are mainly two categories for power plants. Dispatchable power plants have a predictable power output and their output can be planned far in advance and adjusted on short notice. They include coal, oil, gas, or nuclear power plants, among others. Non-dispatchable power plants are less predictable and mostly made up of volatile renewable energy sources. Consumers range from individual households to large corporations and entire countries. They use energy with varying demand patterns, further complicating the balance between supply and demand. These factors result in a highly dynamic system for the power grid. Increasing the power supply leads to a higher frequency while decreasing power generation lowers the frequency.[6] Usually, these deviations are within a few percentage points of the reference frequency, but a large enough change from that value can lead to blackouts or infrastructure damage.[7] The rotational energy in the turbine of power plants helps to combat these deviations. If power is missing in the power grid, the kinetic energy of the connected turbines is converted into electrical energy, slowing down the generator and reducing the grid frequency. During an energy surplus, the generators speed up, increasing the grid frequency. This inertial response only covers the first seconds of deviation and is then complemented by a cascade of control actions (primary, secondary, etc.).[8] Renewable energy power plants do not provide this benefit as they are coupled via power electronics to the power grid. Instead, they mainly depend on the weather, which is hard to predict[9] precisely. Thus, their unpredictability changes the amount of energy in the power grid, resulting in possibly too much or too little power that needs to be compensated for by the other power plants. This tends to introduce more variability in the power grid frequency. Thus, understanding the behavior of power grid frequency, especially the deterministic and stochastic dynamics, becomes increasingly crucial. See also Ref. 8 or 10 for further details.

B. Detrending frequency data

The power grid frequency data used in this study are publicly available and include both frequency and feature data for Australia (AUS) and Continental Europe (CE). For Australia, the data cover the period from January 1, 2021, to March 31, 2024, while for Continental Europe, they span from March 13, 2017, to December 31, 2019. Power grid frequency dynamics are affected by many different influences, such as control mechanisms, dispatch schedules, and stochastic perturbations. We focus on the stochastic dynamics and deterministic influences on a small time scale and neglect the longer-timescale trend and remove it from the time series. For this purpose, a one-dimensional Gaussian filter is applied to the time series data. The resulting time series containing the slower dynamics of the original dataset is then subtracted from the original data, and we obtain a detrended time series. Details are given in Appendix B.

C. Drift and diffusion

A common method for analyzing time series involves examining the probability density function (PDF) and its corresponding Fokker–Planck equation.[11] This approach assumes Markovian dynamics for the time series such that a Fokker–Planck equation is locally applicable. The Fokker–Planck equation, which describes the time evolution of the PDF $p(x,t)$, is given by

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}[\mu(x,t)p(x,t)] + \frac{\partial^2}{\partial x^2}[D(x,t)p(x,t)], \quad (1)$$

where the partial differential equation describes the evolution of the space-and-time-dependent PDF of a physical observable under the influence of deterministic and stochastic forces. The Fokker–Planck equation contains two coefficients, the drift (or first Kramers–Moyal) coefficient $\mu(x,t)$ and the diffusion (or second Kramers–Moyal) coefficient $D(x,t)$. The drift represents the deterministic influence on the observable, while the diffusion represents the strength of the stochastic dynamics. We assume both these coefficients are dependent on space and time. These coefficients can be estimated from time series data using kernel density estimation.[12] We apply this method to grid frequency time series in the CE and AUS power grids.

For estimating and analyzing drift and diffusion, we consider 1-h intervals of frequency data with no gaps in their availability. The CE frequency data are at a resolution of 1 s[13] and the AUS frequency data are at a resolution of 4 s,[14] resulting in 3600 and, respectively, 900 frequency data points per 1-h interval. This approach balances the large amount of required data points for Machine Learning (ML) models with the decrease in accuracy when using fewer values for the calculations. Figure 1 shows the resulting Kramers–Moyal coefficients for four randomly chosen hours.

Next, we transform the Kramers–Moyal coefficients into one-dimensional drift and diffusion amplitudes. For the drift amplitude, we calculate the negative slope of the first-order Kramers–Moyal coefficient at 50 Hz, assuming a local linear behavior. Similarly, the diffusion amplitude is the value of the second-order Kramers–Moyal coefficient at 50 Hz. In the following, we refer to these amplitude values as drift, respectively, diffusion for shorter notation. The drift describes the deterministic changes in the frequency, given by the power grid's (droop) control and damping forces, which push it back to its reference frequency. Diffusion is the counterpart that quantifies random fluctuations in the power grid frequency that typically cause it to deviate from the reference value.[6] Understanding these parameters can help to operate the grid in a stable state.

D. Machine learning models

Machine Learning (ML) and Artificial Intelligence (AI) are becoming increasingly popular to analyze power systems. In previous works, explainable Artificial Intelligence (XAI) has been

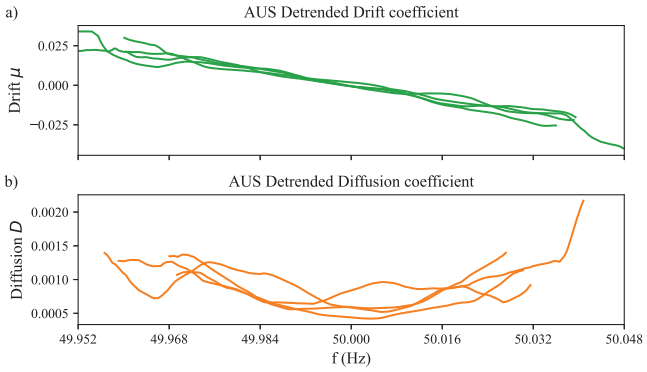

FIG. 1. The first and second Kramers–Moyal coefficients, drift and diffusion, respectively, of the detrended AUS frequency data for four randomly chosen hours. (a) To calculate the drift parameter, we use the negative slope of the drift coefficient at 50 Hz. (b) We take the diffusion coefficient value at 50 Hz for the diffusion parameter.

proposed for power grid studies, motivating our use of Gradient Boosted Trees (GBTs) in the present work as well.[15–17] In the following, we employ two different frameworks for GBTs as comparative models which allow us to check the models for consistency. We use a total of six different machine-learning models: Three of these are GBTs from the LightGBM and XGBoost frameworks. The two XGBoost models differ in the selection of the loss function, in one case the Squared Error function (SE) is used as loss and in the other the Absolute Error (AE). The remaining three models are a Random Forest (RF), a Multi-Layer Perceptron (MLP), and a Linear Regression (LR). We apply these models to a selection of techno-economic electricity grid features to estimate drift and diffusion in the respective time intervals.

In order to train these models we use a random search with 1000 trials for each model. This is shown to provide equal or better results compared to a grid search.[18] Next, we evaluate each model using the following well-established metrics: Mean Squared Error (MSE), Mean Average Error (MAE), Mean Absolute Percentage Error (MAPE), and the R^2-score. Minimizing these metrics serves as the indicator of sufficient quality for meaningful interpretation.

E. Shapley additive explanations

We use SHAP[19] values to explain the GBTs and RF machine learning models used in this study. SHAP is a framework that assigns an importance value to each input feature in a prediction, indicating the impact of each feature on the prediction. SHAP defines the explanation as $g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j$, where the explanation model g is a weighted linear combination of binary feature presence indicators z', designed to represent the contribution of the features to the output of the machine learning model. $z' \in \{0,1\}^M$ is a vector of the features used, M is the maximum number of features used, and $\phi_j \in \mathbb{R}$ are the Shapley values, representing the contribution of j-th feature to the prediction. The term ϕ_0 represents the baseline value, which corresponds to the model's prediction when no features are included.[20] We are particularly interested in the Shapley values where z' is a vector of ones, meaning all features are selected. Then, the expression simplifies to $g(x') = \phi_0 + \sum_{j=1}^{M} \phi_j$, where x' is a vector of ones. We get the explanation model g from our machine learning model, and our goal is to calculate the Shapley values ϕ_j. SHAP values approximate Shapley values for ML models.[19]

In this article, we estimate and analyze SHAP values from tree-based machine learning models. While SHAP can be applied to other types of models, such as using KernelSHAP, we focus on TreeSHAP.[21]

While SHAP generates feature importance for every feature, exercising caution when analyzing these results is important. SHAP calculates how a model assigns feature importance to every feature, but this assigned importance does not have to align with actual correlations or causation in the physical world. To provide a baseline for irrelevant features, we introduce a random noise feature. The noise feature should not impact the prediction as it has no connection to the actual processes happening in the real world. Therefore, if the model assigns the random noise a relatively high feature importance, it suggests that all features with a lower feature importance than the random noise feature are virtually useless for the quality and interpretation of the model. Alternatively, removing some highly impactful features may bring further insights. It can help to evaluate how that feature impacts the model, whether other features can replace it, or if the model quality deteriorates. We will employ both strategies to disentangle meaningful from meaningless important features from the data.

III. ANALYZING STOCHASTIC AND DETERMINISTIC ASPECTS

Before applying machine learning models to forecast drift and diffusion dynamics depending on techno-economical features, we analyze the values for drift and diffusion that are estimated from the data.

A. Distribution of drift and diffusion

The drift in the original data is approximately between six and eight times lower than in the detrended data while exhibiting a significantly higher variance, as shown in Table I. The original data still include various trends that appear in intervals longer than 1 min. Usually, these trends are primarily deterministic and add to the predictability of the frequency because specific supply and demand patterns repeat daily, weekly, or on even longer scales.[22,23] However, they still affect the frequency and cause deviations from the reference value. We calculate the drift on a 1-h interval of frequency data. Longer-scale patterns cannot be captured in the 1-h interval of frequency data we use, instead contributing to perceived noise. Therefore, the drift measuring systematic changes in power grid frequency increases if we consider the detrended data.

This general tendency is easily understandable. The temporary drift can be positive and negative for the original data. These contributions of both signs partially cancel out and lead to an average value close to 0. On the other hand, for the detrended data, the drift coefficient is always positive, and it leads to a relaxation of the grid toward its stable operating point.

TABLE I. Mean and variance of AUS and CE original and detrended drift. The original drift is much smaller and has a higher variance, possibly due to additional noise in the data.

Parameter	Mean	Variance	Mean	Variance
	Drift		Diffusion	
AUS original	1.15×10^{-2}	9.14×10^{-4}	1.00×10^{-3}	4.61×10^{-8}
AUS detrended	9.49×10^{-2}	4.59×10^{-4}	6.02×10^{-4}	2.00×10^{-8}
CE original	1.62×10^{-3}	1.85×10^{-5}	3.41×10^{-5}	3.25×10^{-10}
CE detrended	1.06×10^{-2}	6.25×10^{-6}	3.10×10^{-5}	1.28×10^{-10}

Similarly, to the drift, detrending the data also affect mean diffusion. In Table I, we see a decrease in diffusion when detrending, albeit much smaller than the drift change. For AUS data, the same explanation applies as with the drift. Detrending removes long-term trends that are interpreted as noise on a short timescale, resulting in lower diffusion. However, this logic does not apply to the CE diffusion. The observed reduction is only 9% compared to nearly a 40% reduction for AUS data. This slight difference indicates that short-term fluctuations in the grid dominate the CE diffusion.

Generally, from a mathematical point of view, our detrended model is essentially an Ornstein–Uhlenbeck process with $\mu(t)$ and $D(t)$ time-dependent on a larger time scale.

B. Patterns

After observing some distribution properties of the estimated drift and diffusion, our analysis shifts to examining trends and periodic behavior within the drift and diffusion time series. In order to better capture the underlying patterns, we will now focus exclusively on the detrended data. This approach, as detailed in Appendix D, removes long-term trends and allows us to observe the relevant power system dynamics more clearly.

We examine the weekly (7-day), monthly (28-day), seasonal (three-month), and yearly patterns in the detrended frequency, drift, and diffusion time series data for AUS and CE.

Starting with the 7-day intervals, the AUS diffusion exhibits daily peaks lasting around half the day, from late afternoon until the morning (see Fig. 2). In contrast, the drift behavior does not follow clear patterns. Therefore, temporal patterns in the frequency mostly arise due to patterns in the diffusion. A high diffusion results in increased deviations of the frequency from the reference value.

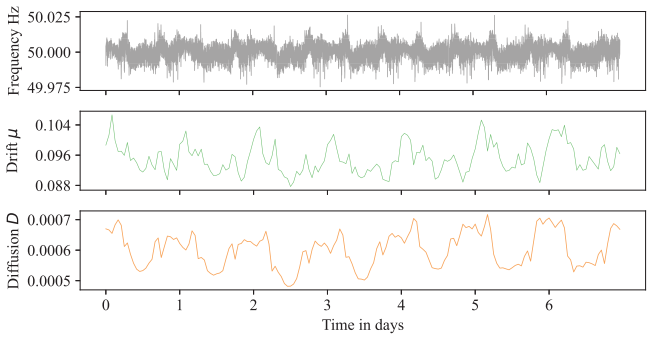

FIG. 2. Frequency, drift, and diffusion of the detrended AUS data, averaged over every 7-day interval.

Our results show that the drift in CE follows a clear pattern over a 7-day interval, peaking daily at midnight and rising slowly throughout the day, with higher values observed on weekdays compared to weekends. CE exhibits relatively stable diffusion on weekdays, with only minor variations throughout the day. For the 28-day interval, the AUS data reveal a weekly cycle in diffusion, which peaks progressively from Monday to Sunday before resetting. The CE data continue to show consistent weekly patterns without new insights. Yearly analysis uncovers a prominent seasonal pattern in CE, where diffusion is higher in the warmer months and lower in the colder months, with drift following a similar but less pronounced trend. For further details, please refer to Appendix F.

IV. MACHINE LEARNING MODELS FOR DRIFT AND DIFFUSION

A. Predictions

To provide a baseline comparison for our machine-learning-based model, we introduced a non-machine-learning model, the mean prediction. This baseline calculates the average value for each hour using the entire dataset. In contrast, the machine-learning models are trained and evaluated using an 80:20 train–test split, coupled with fivefold cross-validation during the grid search process. We examine the performance of the models described in Sec. II D on the test set and further predict the drift and diffusion using the whole dataset to verify the presence of any long-term trends in the data that the model should pick up. In Sec. IV B, we then only analyze the best-performing models, which are the GBT models.

In all four subplots of Fig. 3, the ML models predict the average drift and diffusion in most of the cases very close to the average target value. The mean prediction as well as the linear regression are less accurate in terms of the average value. With the exception of the Australian drift, the first and third quartiles are also well matched by the GBT as well as by the RF models and the MLP. The values at the edges of the distributions of drift and diffusion are for the most part well approximated by the GBT models. Boxplots displaying the full distribution of drift and diffusion (including fliers) are shown in Fig. 8 in Appendix E.

The error values of each model presented in Table II further help identify the most trustworthy model. For the majority of the models, the MAE is within 10%–15%, indicating similar performance across all models, even for the different parameters. However,

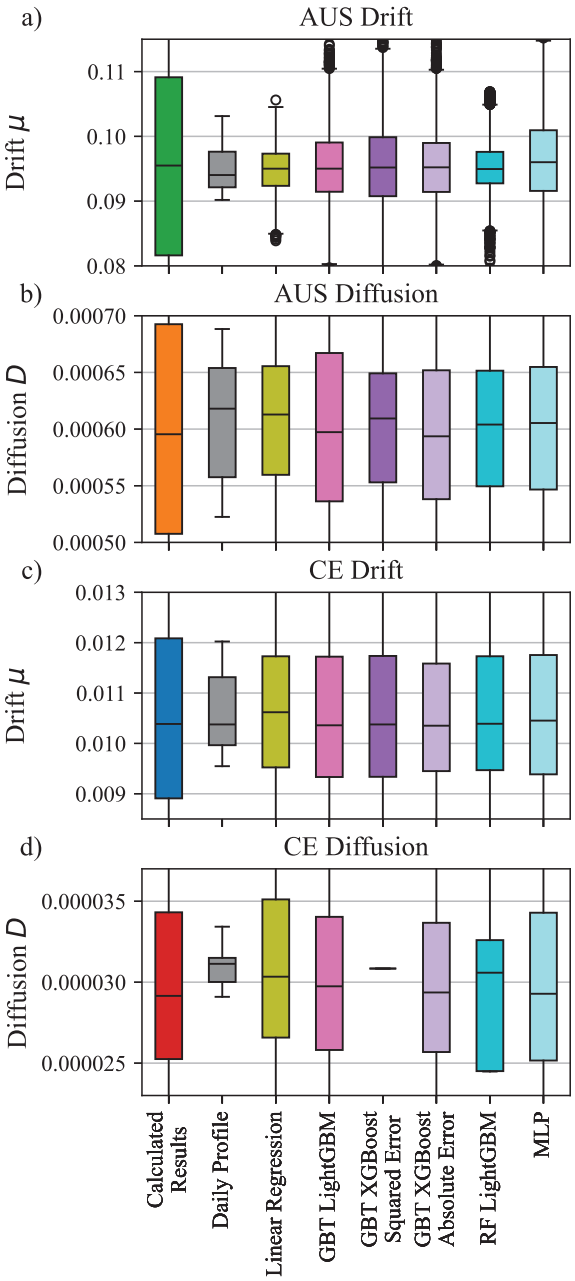

FIG. 3. Predictions of all models compared to our calculated values for (a) Australian drift, (b) Australian diffusion, (c) CE drift, (d) CE diffusion. The diffusion models are closer to the target than the drift models. See also Fig. 8 for a plot including fliers.

since the values for drift and diffusion (Table I) are close to zero, the MAPE can be misleading.[24] On the other hand, the R^2 score varies significantly between models with different parameters and among models for the same parameter. Models with a higher R^2 score also have a lower MSE in our predictions for the same parameter. Combining these two error measures allows us to choose the best-performing models for drift and diffusion in each area. Overall, the diffusion models perform better compared to the drift models, while the CE models are significantly more accurate than the AUS models. In each area, the GBT models outperform the RF and MLP models, as well as the linear regression and the mean prediction baseline model, where the latter two are considered as baseline models in our analysis. These performance differences are also present when using the models to predict the drift and diffusion over the whole time series, with the GBT models best predicting long-term trends. In Table II and Fig. 8, the GBT XGBoost model using a squared error loss function on the CE diffusion data is a strong outlier. This model did not train correctly despite undergoing the same hyperparameter optimizations using a random search and we report the result for consistency.

B. Feature importance

After using the models to predict values for drift and diffusion in dependence on various electricity grid features, we analyze the feature importance for the drift and diffusion parameters for CE and Australia using SHAP. Since the GBT models outperformed all other models, we will present the results for only these models, especially, we will discuss the results of the GBT LightGBM model, which are presented in Fig. 4. Note that SHAP values itself do not provide any insight but require additional assumptions and domain knowledge for contextual explanations, as we discuss them in the following.

1. Problems of Australian drift

Across all used machine learning models, the predictions for Australian drift consistently have an R^2 score of less than 0.16 (Table II). This low R^2 score means that every model explains less than 16% of the variance in the data. Further, compared to the CE drift predictions, the model has a higher MAPE and a significantly lower R^2 score. Considering the SHAP values of the CE drift predictions, the two time features cosine of the hour and sine of the hour (for an explanation of the calendar and time features, see Appendix A) are the two most important features for all GBT models. As the performance of the ML models is worse than for the other cases, we will not further analyze the importance of the features for the Australian drift.

2. Australian diffusion

The best performing model for the Australian diffusion is the GBT LightGBM model which has an R^2-score of 40.5%, with a good performance in the overall trend prediction. Figure 4(c) reveals the solar (Utility) feature as having the highest importance for the Australian diffusion. The model indicates an inverse relation between the amount of solar generation and diffusion. It models a decrease in diffusion when solar generation increases. This contradicts what we expect, as solar generation methods have no inertia—a stabilizing property of traditional synchronous generators that resists frequency changes.[8] Solar generation is also somewhat unpredictable, which should increase the diffusion. However, the model might not capture the actual causation in this case. We observe a clear peak

TABLE II. Mean prediction and machine learning (ML) model errors for all models using detrended AUS and CE data to predict drift and diffusion. Diffusion estimations are better than drift estimations in the same area, while CE models outperform AUS models when comparing the same parameter. The model with the lowest error (highest R_2 score) is highlighted in bold. For CE drift multiple models perform (almost) equally well. The GBT models, i.e., the LightGBM model, perform best in most of the cases, while the linear regression, RF, and MLP models perform worse. The mean prediction performs the worst on all error metrics, thereby motivating the use of ML models. SE, squared error; AE, absolute error.

Model	MSE	MAE	MAPE	R^2
		AUS drift		
Mean prediction	4.48×10^{-4}	1.65×10^{-2}	5.45×10^{-1}	2.5%
Linear regression	4.42×10^{-4}	1.64×10^{-2}	2.70×10^{-1}	2.8%
GBT LightGBM	3.86×10^{-4}	1.55×10^{-2}	2.18×10^{-1}	15.0%
GBT XGBoost SE	3.83×10^{-4}	1.55×10^{-2}	2.17×10^{-1}	15.7%
GBT XGBoost AE	3.91×10^{-4}	1.56×10^{-2}	2.23×10^{-1}	13.8%
RF LightGBM	4.04×10^{-4}	1.59×10^{-2}	2.33×10^{-1}	10.9%
MLP	4.15×10^{-4}	1.61×10^{-2}	2.43×10^{-1}	8.6%
		AUS diffusion		
Mean prediction	1.73×10^{-8}	1.02×10^{-4}	5.25	13.7%
Linear regression	1.56×10^{-8}	9.75×10^{-5}	1.74×10^{-1}	18.5%
GBT LightGBM	1.14×10^{-8}	8.41×10^{-5}	1.47×10^{-1}	40.5%
GBT XGBoost SE	1.51×10^{-8}	9.59×10^{-5}	1.71×10^{-1}	21.3%
GBT XGBoost AE	1.18×10^{-8}	8.44×10^{-5}	1.46×10^{-1}	38.7%
RF LightGBM	1.50×10^{-8}	9.57×10^{-5}	1.71×10^{-1}	21.7%
MLP	1.48×10^{-8}	9.47×10^{-5}	1.68×10^{-1}	22.7%
		CE drift		
Mean prediction	5.64×10^{-6}	1.81×10^{-3}	1.85×10^{-1}	9.7%
Linear regression	3.57×10^{-6}	1.43×10^{-3}	1.41×10^{-1}	39.7%
GBT LightGBM	3.28×10^{-6}	1.36×10^{-3}	1.35×10^{-1}	44.6%
GBT XGBoost SE	3.25×10^{-6}	1.36×10^{-3}	1.35×10^{-1}	45.0%
GBT XGBoost AE	3.26×10^{-6}	1.35×10^{-3}	1.34×10^{-1}	44.9%
RF LightGBM	3.34×10^{-6}	1.38×10^{-3}	1.37×10^{-1}	43.6%
MLP	3.39×10^{-6}	1.39×10^{-3}	1.38×10^{-1}	42.7%
		CE diffusion		
Mean prediction	1.27×10^{-10}	6.18×10^{-6}	1.96×10^{-1}	1.0%
Linear regression	6.83×10^{-11}	4.00×10^{-6}	1.22×10^{-2}	34.5%
GBT LightGBM	4.25×10^{-11}	3.02×10^{-6}	8.84×10^{-2}	59.3%
GBT XGBoost SE	1.04×10^{-10}	6.04×10^{-6}	1.97×10^{-1}	-3.63×10^{-4}
GBT XGBoost AE	4.87×10^{-11}	2.45×10^{-6}	6.59×10^{-2}	53.3%
RF LightGBM	6.46×10^{-11}	3.95×10^{-6}	1.19×10^{-1}	38.1%
MLP	7.29×10^{-11}	4.24×10^{-6}	1.29×10^{-1}	30.1%

in diffusion at night, which is when solar generation is low. The model captures this correlation perfectly but does not offer any causal explanations for which further domain knowledge would be necessary.

In contrast, the inverse relationship is expected for coal power generation. High coal generation increases the inertia in the system and thus leads to a lower diffusion. The GBT LightGBM model identifies coal power generation as the second highest feature. Coal is also the largest contributor to Australian power generation, accounting for over 45% of the energy produced.[25]

The high importance attributed to the weekday_sin and month_sin features by SHAP indicates the presence of long-term patterns. We observe a weekly repeating pattern of increased diffusion in the second half of the week, confirmed by the weekday_sin feature. The monthly feature indicates that winter months have a lower diffusion, while summer months have a higher diffusion. This pattern is vaguely visible in the yearly pattern plot in Appendix F.

Further, the cosine of the hour is also among the features with the highest importance. The feature importance plots reveal that high feature values lead to increased diffusion for the model while a low value decreases it. The feature hour_cos has a high value at the beginning and end of every day, meaning the model predicts low diffusion during the day and a high diffusion at night. This is precisely the day-night difference observed in the 7-day patterns.

To emphasize the findings on the feature importance of the LightGBM model, we compare them with the results from the other GBT models. Nine of the ten most important features are the same

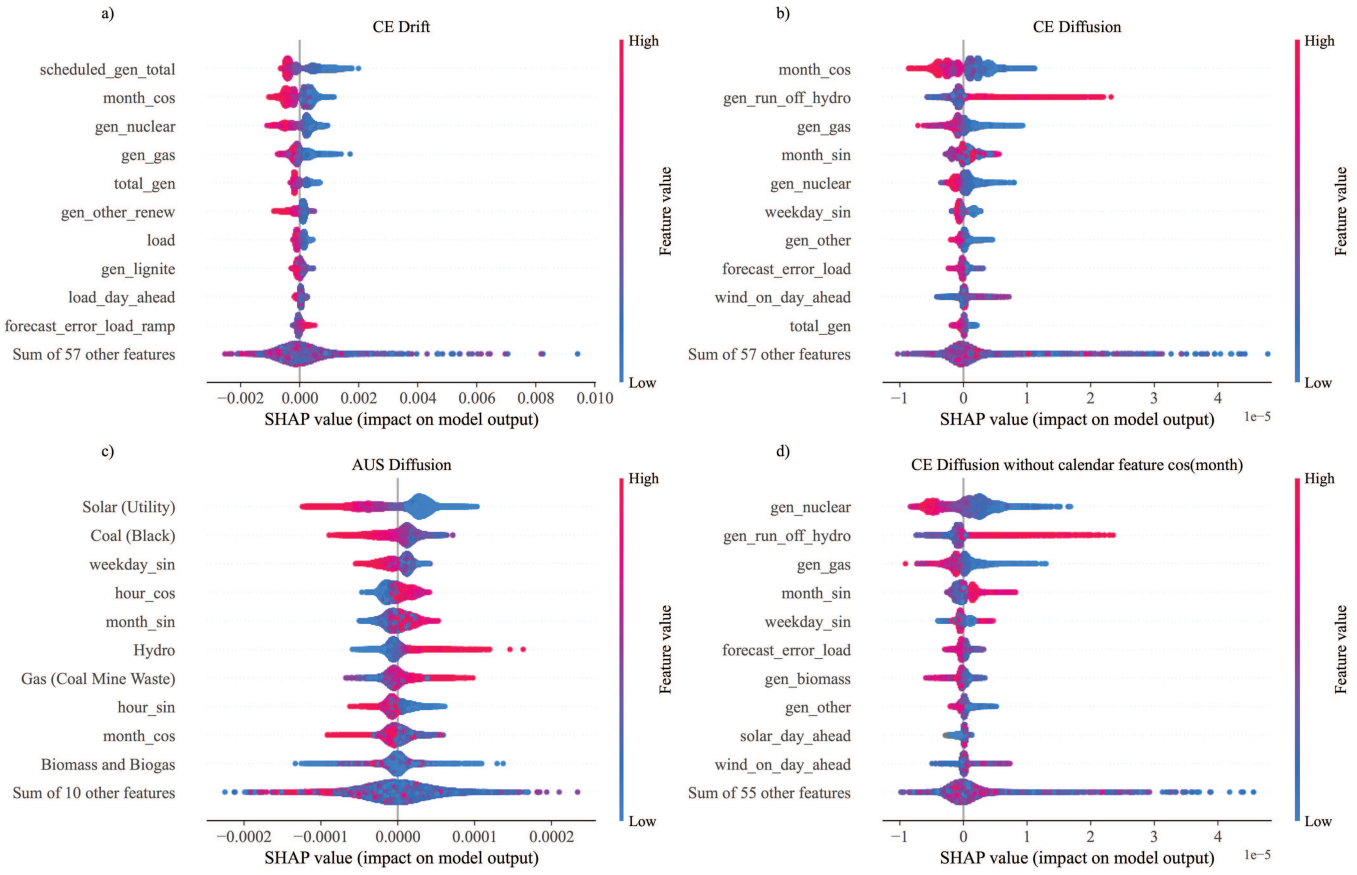

FIG. 4. Feature importance of the top 10 features for the best performing model (GBT LightGBM) for (a) CE drift with random noise, (b) CE diffusion with random noise, (c) AUS diffusion with random noise, (d) CE diffusion without the previously most important feature.

in the two best-performing models, the LightGBM and the absolute error XGBoost model, and each feature has the same qualitative impact according to SHAP. This improves the trustworthiness of the model and the derived explanations.

3. Continental European drift and diffusion

For each parameter, we consider the best-performing models in order to analyze the feature importance. The GBT LightGBM performs best for CE diffusion, while for the drift, it shares the best performance with the GBT XGBoost models. As already described in the case of the Australian data, we see major similarities in feature importance between the different models. Therefore, in the following, we discuss the results of the GBT LightGBM model

Appendix G reveals a high correlation between the detrended drift and diffusion for Continental Europe. This correlation means that when the drift is high, the diffusion tends to be high as well, and vice versa. Therefore, the generally made assumption that a high drift results in a low diffusion is not applicable here. Instead, they depend on each other: low diffusion indicates a small deviation from the reference frequency, requiring minimal correction to maintain grid stability. Conversely, high diffusion leads to greater frequency deviations, necessitating a higher drift to keep the grid stable.

According to Fig. 4, the most important feature in the model for the CE drift is the totally scheduled generation. This is defined as the planned amount of electricity that the power plants are to produce in a certain period of time and is stated in MW. It is, therefore, strongly correlated with the actual total generation. Figure 5(a) presents the partial dependency of the drift given on the total scheduled generation. There is a significant split in the average total scheduled generation. A total scheduled generation above the average leads to a reduced drift, but further increases in the generation only lead to a slight reduction. For a below-average total scheduled generation, we observe an approximately linear impact on the drift. The values for a total scheduled generation close to zero should be disregarded, as this is unrealistic for a grid as large as Continental Europe. During an above-average total generation, the system's inertia is also higher, resulting in low deviations from the reference frequency in case of a disturbance. Small deviations require only a minor correction, resulting in a low drift. Further increases in generation typically involve more renewable energy sources, which introduce variability and potential disturbances. However,

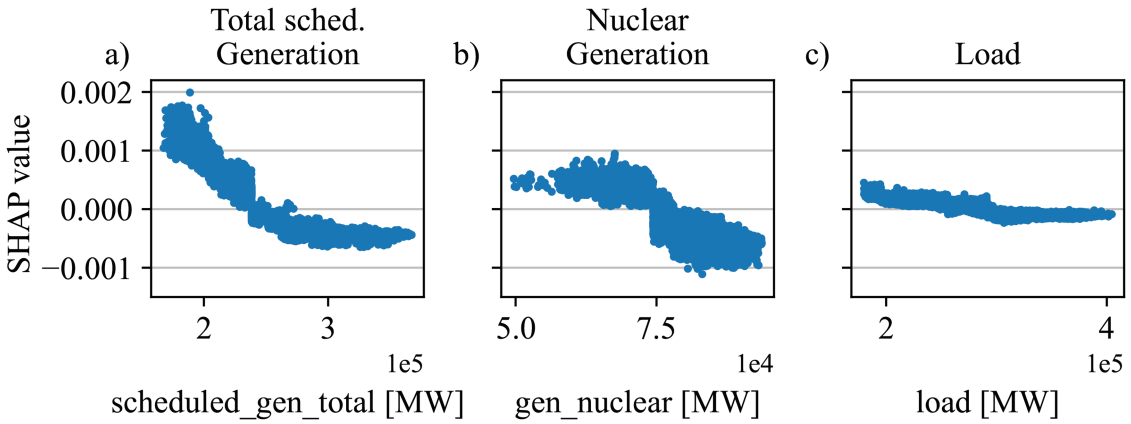

FIG. 5. Partial dependency of the Continental European drift on the total scheduled generation (a), nuclear generation (b), and load (c) for the GBT LightGBM model.

the increased inertia can counteract this variability, keeping the impact on the drift stable even at high generation levels. For a low total scheduled generation, most of the power is provided by the base production from plants such as nuclear or coal, which help maintain grid stability.

In both drift and diffusion, nuclear generation has a high feature importance with a negative correlation. A high, above-average nuclear generation leads to a lower drift and diffusion [Figs. 4(a) and 4(b)]. We see this statement confirmed in the partial dependency plots in Figs. 5(b) and 6(d). As diffusion and drift are correlated, influencing one impacts the other. A large increase in nuclear generation increases the power grid's inertia, allowing for an easier correction of deviations. This decreases the diffusion and also leads to a lower required drift. Additionally, nuclear power plants, employed in large numbers e.g., in France, are used to correct frequency deviations by providing rapid power adjustments.[26] Therefore, they often significantly increase generation when corrective measures are needed, which are connected to a reduced drift. The additional nuclear power plants also add to grid inertia, keeping the grid stable and thus reducing diffusion.

Showing a similar qualitative behavior for the drift as the total scheduled generation also the total generation is among the most important features for the drift, and at the same time for the diffusion. The impact on drift and diffusion follows the same explanations as for the total scheduled generation. Further, load, respectively, total energy consumption is among the ten features with the highest importance for the drift. We also observe this relationship in Appendix G, which is confirmed by the models using

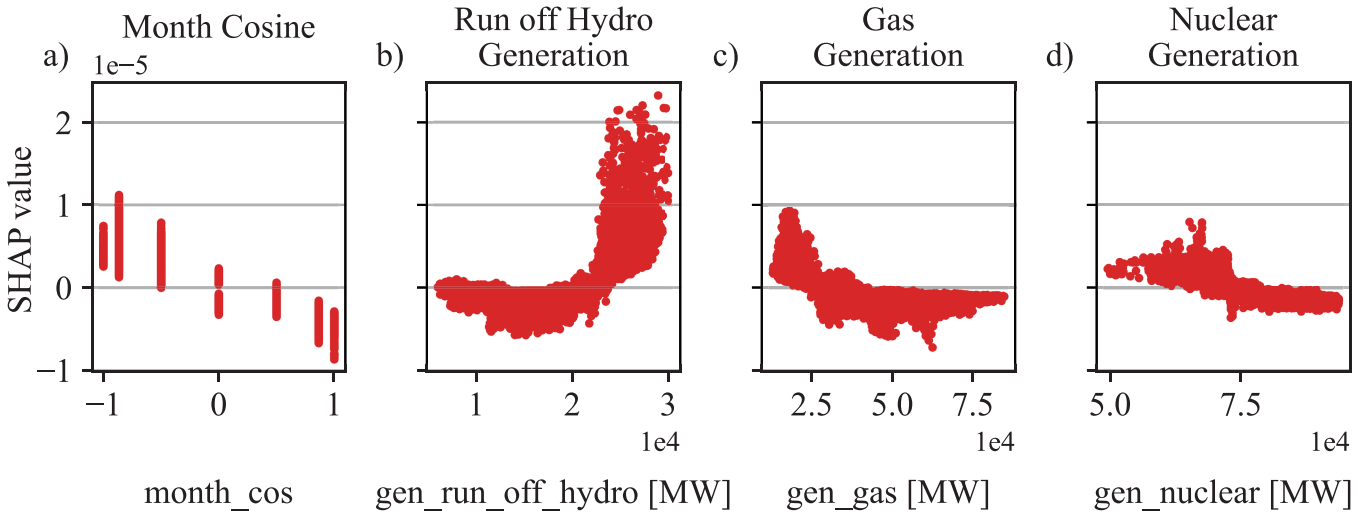

FIG. 6. Partial dependency of the Continental European diffusion on the month cosine (a), run off hydro generation (b), gas generation (c), and nuclear generation (d) for the GBT LightGBM model.

TABLE III. ML model scores for the LightGBM model and the XGBoost gradient boosted tree model with an absolute error loss function using detrended CE data to predict the diffusion. The month_cos feature, which has the highest importance in these models, is removed. The difference in performance is not significant.

Model	MSE	MAE	MAPE	R^2	MSE	MAE	MAPE	R^2
	All features				Month_cos removed			
XGBoost AE	4.87×10^{-11}	2.45×10^{-6}	6.59×10^{-2}	53.3%	4.88×10^{-11}	2.48×10^{-6}	6.70×10^{-2}	53.2%
LightGBM	4.25×10^{-11}	3.02×10^{-6}	8.84×10^{-2}	59.3%	4.17×10^{-11}	3.04×10^{-6}	8.98×10^{-2}	60.1%

load as one of the more important features. Figure 5(c) shows the partial dependency between the load and drift, exhibiting a negative linear relation for below-average loads. High loads correlate with a constant negative influence on the drift.

In Sec. III B, we showcase daily peaks at midnight, with a slow rise in drift during the day and a general increase on weekends. Compared to the Australian models, the models for CE drift do not rely on the added time features to produce this trend. However, the monthly time feature is still highly important, strongly contributing to the yearly trends. The models relate summer months with a decrease in drift and winter months with an increase. For the diffusion model, calendar features are even more important. Up to three of the six most important features in the diffusion models are such calendar features [Fig. 4(b)]. This may partly be the reason for the high performance of the diffusion models. In Australia, the diffusion model also uses multiple calendar features in its top ten and performs much better than the Australian drift models. These features are noise-free and represent daily, weekly, or yearly intervals. The high dependence on the calendar features reveals that the diffusion itself depends on time and is likely more robust to the various other features. The features in the dataset include the generation from different power plants, their ramps, day-ahead predictions, and prices. However, they do not contain any consumer behavior patterns. The higher dependence on calendar features in the Continental European diffusion models, compared to the drift model, indicates that power generation primarily influences drift. In contrast, diffusion has a relatively stronger dependence on its consumption, which is not included in the features and, therefore, is replaced with calendar-based variables. Figure 6(a) presents one such time feature, the cosine of the month, where larger values correspond to months close to the beginning or end of the year. The partial dependency shows that winter months experience a lower diffusion and summer months a higher diffusion. This aligns with the seasonal patterns previously identified in Sec. III B.

Gas power generation is among the top three features of the diffusion models and has a negative non-linear relationship with diffusion. Calculating the mean value from the data set on gas generation, which is around 39 GW and considering the partial dependency in Fig. 6(c) shows that the diffusion is lowered or almost unaffected for a large part of the gas power generation. Only low amounts of gas power generation lead to an increase in diffusion.

Among the ten most important features for both CE drift and diffusion, the run-off hydro feature in the diffusion models is the only one that is positively correlated to the diffusion [Fig. 4(b)]. Run-of-the-river hydroelectricity generates power by diverting part of a flowing water stream through a turbine, thereby generating electricity.[27] This means its power generation heavily depends on the current river conditions. Figure 6(b) reveals that the partial dependence of the diffusion on the run-off hydro feature is quadratic: both low and high generation values increase the diffusion, while average generation decreases it. This is likely due to the unpredictable nature of river flows, which causes unplanned changes in power generation, thereby affecting the balance of power supply and demand and increasing grid instability during times of unpredictable river flow changes.

Overall, almost all the top ten features for the selected GBT LightGBM drift model are negatively related to the drift. However, their impact on the drift around their average value is often positive or near zero. This negative relationship is also the case for eight out of ten of the top features in the GBT LightGBM model for diffusion. Only big changes in load or generation negatively impact the grid stability. The usual operation of power plants keeps the drift high and the grid stable.

4. CE diffusion without its most important feature

As shown in Fig. 4(b), the cosine of the month has the highest impact on the CE diffusion models according to SHAP. In order to gain further insights, we can remove this feature from the dataset, and compare its performance with the calculated feature importance.

Table III presents the updated model scores of the best performing models for CE diffusion when using all features except the cosine of the month compared to the scores achieved when using all features. All values stay quite similar to before. Therefore, dropping this feature does not change the model's performance significantly. The ML models each use the same hyperparameters as the models with all features. A reason for the consistent performance can be that some of the features that are used as model inputs are strongly correlated.

Comparing the recalculated feature importance values for the GBT LightGBM model in Fig. 4(d) with those from the complete feature set reveals that from the remaining top nine features, eight remain and stay in a similar position as before. It is worth noting that nuclear generation has moved from fourth to first position among the most important features. As nuclear availability is usually the highest in the winter months than during the summer, the influence of nuclear generation is consistent with the influence of the monthly feature on the model. This means that the nuclear generation can replace the monthly feature and achieve the same model result. The other calendar features such as month_sin can also supplement and replace the impact of month_cos. The impact of other features on

the model, therefore, can maintain stable performance despite the omission of the most important feature.

V. CONCLUSION AND OUTLOOK

The generation mix of power grids is evolving due to climate change and technological advancements. Renewable generators are becoming more relevant for power grids every year.[1] With fewer synchronous generators and hence lower inertia,[28] the operation of the power system becomes more challenging. In this article, we presented an analysis of drift and diffusion coefficients of the power grid frequency, as an indicator for balance and stability.[8] The drift describes the deterministic changes in the frequency given by the (droop) control and damping forces in the grid, which push it back to the reference value. Diffusion quantifies random fluctuations in the power grid frequency, causing it to deviate from the reference frequency. Utilizing both simple correlations as well as machine learning models (GBT, RF, and MLP), we identified important relations between deterministic and stochastic dynamics and calendars as well as techno-economical features in the Australian and Continental European power grids.

Naïve analysis of drift and diffusion coefficients stresses the importance of distinguishing between correlation and causation: A simple correlation analysis revealed that drift and diffusion coefficients are positively correlated, i.e., large stochastic deviations coincide with a larger control effort (Sec. III). However, the large drift (and thereby control values) likely does not cause large diffusion. Instead, we conjecture that noisy and unsteady operating situations (large diffusion coefficients) lead to an increased control effort (large drift coefficients). Vice versa, when the grid is operating stably, e.g., as generation and load are very predictable, both drift and diffusion coefficients are naturally small.

The careful distinction between correlation and causation also continues when interpreting the machine learning models for drift and diffusion. Neither the used ML models (GBTs, RF, and MLP) nor the interpretability method SHAP guarantee that the found relations between features and the target values are based on actual causal relations in the data or the real world. Consistent with earlier analysis in power systems,[15–17] we find that tree-based methods, specifically GBTs, are performing well on the tabular data we are working with. Interpreting these black-box models is critical for understanding the relationships between generation-mix and power system dynamics.[29]

Using SHAP,[19] we identify important features. For Australia, we find that coal power generation is an important feature coinciding with low diffusion. For the Continental European data set, the total load is one of the important features for the drift, confirming the previously observed negative correlation: High load is associated with a low drift. We can speculate that this drift is measuring the relative control effort, i.e., when the system is heavily loaded, dispatchable generation and control are already active. Furthermore, fluctuations tend to average out during high-load settings, when the load is aggregated over a high number of consumers and the overall system inertia is likely the highest.

In addition to the load, nuclear power generation is highly impactful on both drift and diffusion in Continental Europe. The typical operation of these power plants increases both drift and diffusion. Removing the most important feature for the diffusion model, the cosine of the month, from the ML model, even makes nuclear generation the most important feature and, furthermore, leads to almost no change in performance. This indicates that the two features strongly correlate and that the impact of the calendar feature can be sufficiently explained by nuclear generation and the remaining features. In general, it can be seen that calendar features are very important for the result of the diffusion model, indicating that the magnitude of diffusion at a given time is almost deterministic. Hence, we reproduce results that SHAP has to be carefully interpreted when it comes to feature selections.[30] Furthermore, nuclear power is changing slowly and hence calendar features contain similar information.

Concluding, we have demonstrated that the Kramers–Moyal expansion and estimation of drift and diffusion coefficients yield interesting insights into the deterministic and stochastic dynamics of power systems. We argue that combining these stochastic estimates with external information, machine learning, and explainability methods is useful for understanding these complex systems. Thereby, relations beyond linear correlation can be quantified, and stabilizing or destabilizing effects are identified.

We briefly lay out some future research opportunities. To bring the drift predictions to the same performance level as the diffusion predictions, additional data might be necessary and alternative models could be considered. On the data side, 1 s or even higher resolution data of the frequency data for Australia would be desirable as well as higher than hourly resolution of the external features, such as load and generation. Additional interesting features include the activated control power, such as the Frequency Containment Reserve (FCR), which is already available for limited time periods for some countries.[31] On the modeling side, one could consider attention-based methods, specialized on tabular data[32] or include known causal knowledge in the model building, e.g., via Shapley flows.[33] Another interesting direction would be to explore the prediction of model coefficients, as proposed by recent studies,[34] which could provide deeper insights into the underlying dynamics.

ACKNOWLEDGMENTS

We gratefully acknowledge funding from the Helmholtz Association and the Networking Fund through Helmholtz AI under Grant No. VH-NG-1727. The authors acknowledge support by the state Baden-Württemberg through bwHPC. C.B. gratefully acknowledges funding by a KIT International Excellence Fellowship as well as by the UKRI Science and Technology Facilities Council under Project No. APP50058.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

T.D., X.W., and U.O. contributed equally to this paper.

Tim Drewnick: Investigation (equal); Software (equal); Visualization (equal); Writing – original draft (equal); Writing – review

& editing (equal). **Xinyi Wen:** Methodology (equal); Supervision (equal); Writing – review & editing (equal). **Ulrich Oberhofer:** Software (equal); Supervision (supporting); Visualization (equal); Writing – review & editing (equal). **Leonardo Rydin Gorjão:** Writing – review & editing (equal). **Christian Beck:** Writing – review & editing (equal). **Veit Hagenmeyer:** Writing – review & editing (equal). **Benjamin Schäfer:** Conceptualization (equal); Funding acquisition (equal); Project administration (equal); Writing – review & editing (equal).

DATA AVAILABILITY

All data and code used in this article (the Australian frequency data are available from AEMO,[14] the Australian feature data from OpenNEM,[35] the Continental European frequency data from the German TSO TransnetBW GmbH,[13] and the Continental European feature data from the ENTSO-E Transparency Platform)[36] are openly available in GitHub, Ref. 5.

APPENDIX A: DATA PROCESSING

The raw data should undergo cleaning procedures to ensure errors do not deteriorate the calculations' and training's performance. We use two cleaning procedures for Australian frequency data. First, we identify absolute increments between two measurements that are beyond a threshold of 0.05 Hz. Second, we find windows of 15 or more constant measurements in the data. We replace both with NaN values. Next, we forward-fill these and previously present NaN values in the data up to a limit of six values. Finally, we delete all hours of frequency data with missing data points, as the drift and diffusion calculations cannot handle gaps in the data. These cleaning procedures[37] are based on those used in Ref. 38. The Continental European data are available pre-cleaned.[39] This pre-cleaning expands upon our methods by removing values that diverge by more than 1 Hz from the reference frequency.

We average all feature data to 1-h intervals and then remove hours with many missing features. We add features capturing time information for Australian and Continental European feature data. These calendar features are the sine and cosine of the feature's month, weekday, and hour, which we calculate in the following way:

$$f_{\sin,p}(x) = \sin\left(\frac{x}{p} * 2\pi\right),$$

$$f_{\cos,p}(x) = \cos\left(\frac{x}{p} * 2\pi\right),$$

where x is the xth month, weekday, or hour and p is the length of the feature period. In the text, we refer to these features as month/weekday/hour_sin/cos. These two variations allow the model to pick up on many different time-based relationships in the target value.

APPENDIX B: DETRENDING THE FREQUENCY DATA

We use the frequency data in two different ways to make most calculations. The drift and diffusion calculations are based on the findings in Ref. 6, which used the frequency data only after detrending to calculate the drift and diffusion. We use the detrended and unmodified frequency data for our analysis.

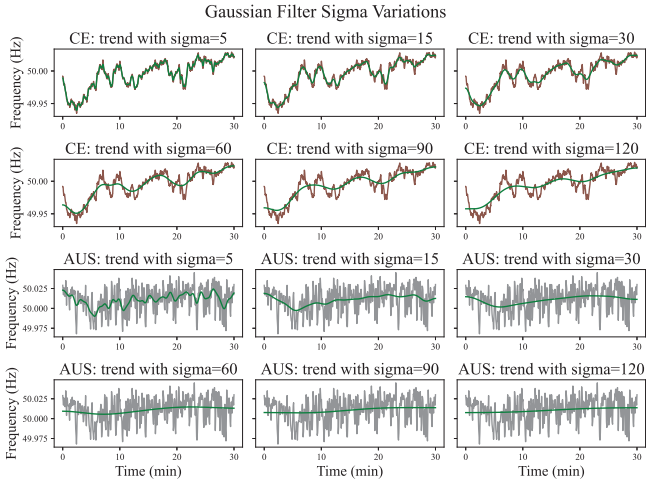

FIG. 7. Selection of six plots each for CE and AUS frequency and their trend using six different values for sigma, the length of the Gaussian filter used to calculate the trend.

Detrending data aim to capture and remove slow changes in the power grid frequency to be left with stochastic motion. We detrend our data using a one-dimensional Gaussian filter. For this filter, we must select a length that captures the trend without following the stochastic motion. To achieve this, we plot the trend and frequency for six different Gaussian filter lengths as shown in Fig. 7. The resolution of the frequency of CE and AUS data differs by a factor of four: a 1-s resolution for CE and a 4 s resolution for AUS data. We aim to use a Gaussian filter that is also different by the same factor to keep the actual time of the filter interval the same. The plots for each Gaussian filter length reveal that a 60 s filter captures the trend for the CE data well while not following too much stochastic motion. This filter length aligns with the one used in Ref. 6. For the Australian data, which generally seems noisier, a value of 15 achieves this. We now define the filter length as 60 s, which also fulfills the goal of keeping the overall time for the filter the same in both datasets. To calculate the detrended data, we remove the trend given by the Gaussian filter from the frequency: $f_d = f_o - f_t$, where f_d is the detrended frequency, f_o is the original frequency, and f_t is the trend. Then, we use this detrended data to calculate the drift and diffusion.

APPENDIX C: IMPACT OF DATA RESOLUTION

Because the frequency datasets for the Australian and Continental European power grid have two different resolutions, it is interesting to know the impact of such a change. Continental European frequency data are available with a time resolution of 1 s, while the Australian frequency data have a time resolution of 4 s. kramersmoyal python library[12] notes that we must normalize the result by dividing it with the time resolution delta t. To estimate the accuracy of the comparisons between the two different grids, we test

TABLE IV. Normalization impact of average drift and diffusion for CE data. Lower resolution increases the calculated value despite normalization.

Parameter	1s resolution	4s resolution	Factor
CE original drift	1.62×10^{-3}	4.68×10^{-3}	2.89
CE detrended drift	1.06×10^{-2}	1.25×10^{-2}	1.18
CE original diffusion	3.41×10^{-5}	7.60×10^{-5}	2.23
CE detrended diffusion	3.10×10^{-5}	6.70×10^{-5}	2.16

how the results change when we modify the data resolution of the same power grid. We calculate the drift and diffusion of the CE data by only using a resolution of 4 s and normalizing the result with the new delta t value. The resulting drift and diffusion increase by the factors shown in Table IV.

This comparison shows that the initial resolution of the data greatly impacts the calculation result with the Kramers Moyal library. We assume that the calculations are more accurate when using data with a higher resolution, which is why we keep the 1 s resolution for Continental European data throughout this article. Based on the presented differences the drift and diffusion for AUS are difficult to compare with those from CE or those of other papers using 1s resolution. However, we can still analyze the drift and diffusion patterns as the actual size of the drift and diffusion is less important than individual, possibly repeating patterns present no matter by which factor we scale the result.

APPENDIX D: DISCARDING ORIGINAL DATA

Using both the detrended and original datasets duplicates the number of models we need to interpret and introduces unneeded redundancy. To allow a more focused analysis, we first remove the original dataset and continue working with only the detrended data. This decision is based on the observation that the patterns in drift and diffusion are much clearer in the detrended data, which enhances the interpretation of these patterns. Additionally, the drift values in the original dataset can be negative, but their significance is unclear. This distribution over positive and negative values is present throughout all hours of the day and is not correlated with variations in power demand. Finally, the ML models using the drift and diffusion calculated on original data perform significantly worse than those using the detrended values.

APPENDIX E: BOXPLOTS OF THE MODEL PREDICTIONS DISPLAYING THE ENTIRE DISTRIBUTION OF DRIFT AND DIFFUSION

In Fig. 3 we show the predictions of the ML and the baseline models, including the distribution around the average values drift and diffusion. In Fig. 8 the boxplots displaying the full distribution of drift and diffusion including fliers are displayed.

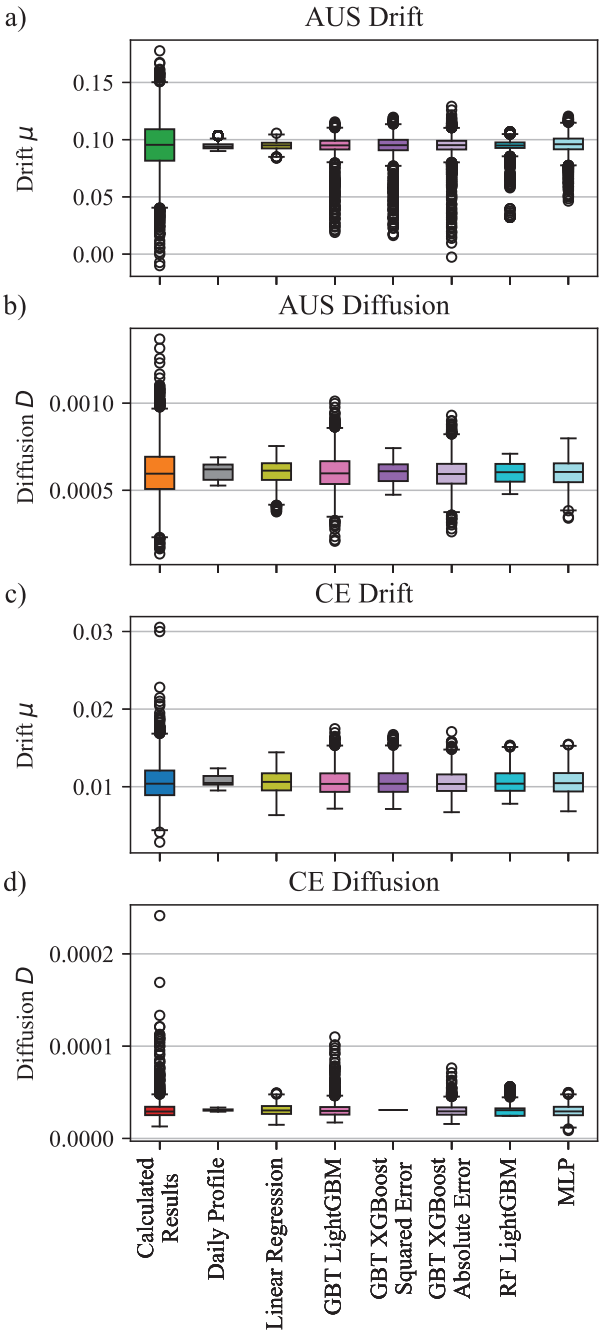

FIG. 8. Full predictions of all models compared to our calculated values for (a) Australian drift, (b) Australian diffusion, (c) CE drift, (d) CE diffusion.

APPENDIX F: AVERAGED DRIFT, DIFFUSION, AND FREQUENCY OVER VARIOUS INTERVALS

The drift in CE follows a clear pattern (cf. Fig. 9). It peaks daily at midnight and rises slowly during the day. The drift is higher on weekdays than on the weekends. Comparing the frequency with the drift reveals that the drift rises whenever the frequency deviates strongly from the reference value. This correlation may be counter-intuitive initially because the drift measures how closely

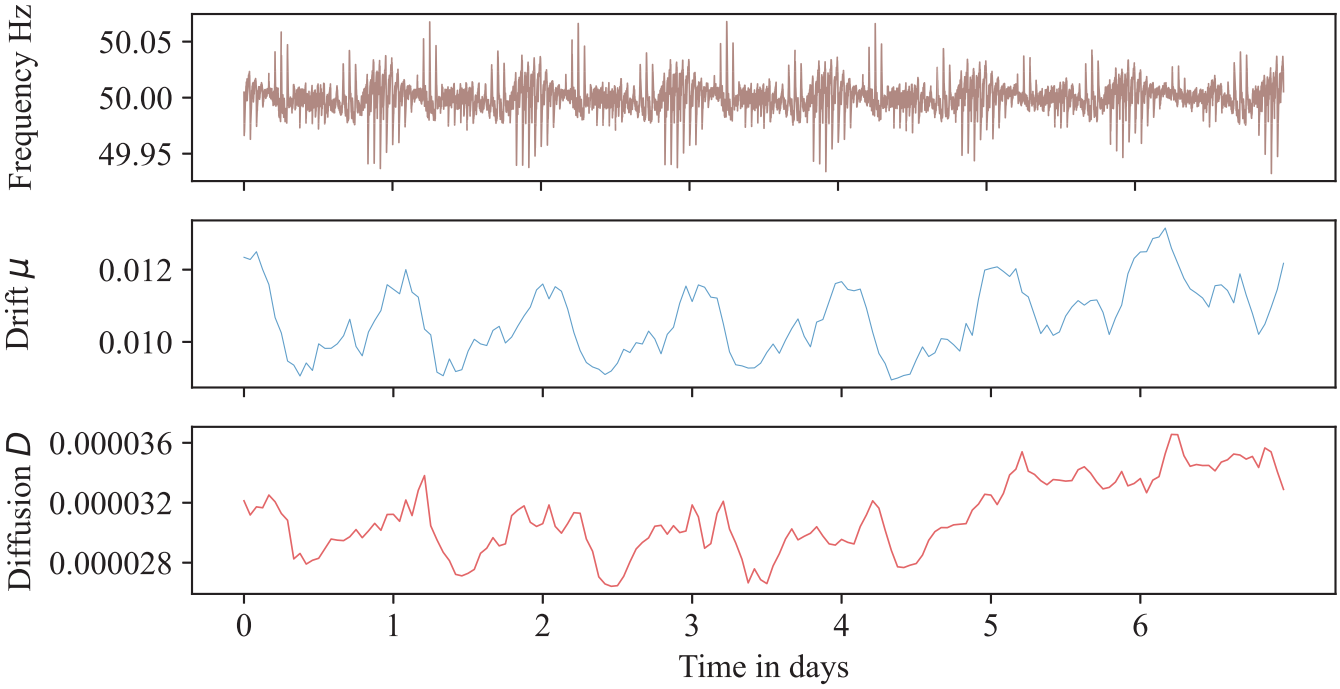

FIG. 9. Frequency, drift, and diffusion of the detrended CE data, averaged over every 7-day interval.

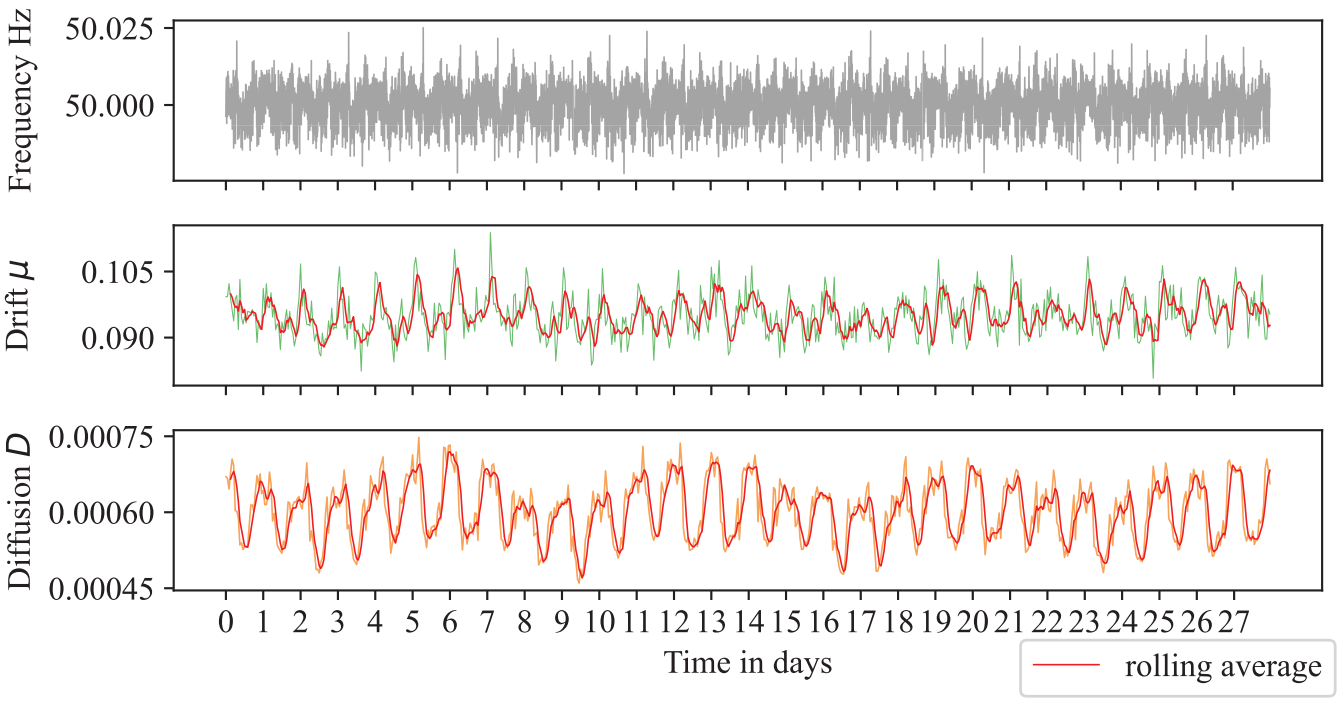

FIG. 10. Frequency, drift, and diffusion of the detrended AUS data, averaged over every 28-day interval.

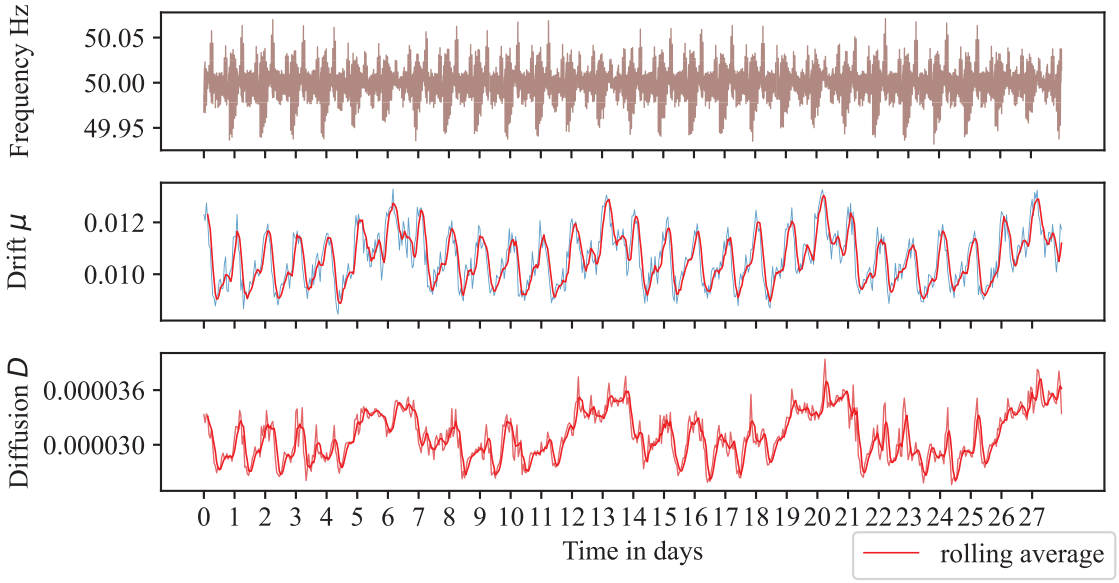

FIG. 11. Frequency, drift, and diffusion of the detrended CE data, averaged over every 28-day interval.

the frequency is maintained to the reference value. But without deviations from this value, there is also no need for a strong drift. Therefore, the drift is only large when the frequency deviates significantly. These large frequency deviations happen primarily in the hours before midnight when solar power generation ends, and other power sources are required to support the energy demand. This could be due to a change in the power generation mix while the demand stays high. Matching the decrease in solar power generation with an increase in the generation of other power plants is difficult as the solar power output can vary, and power plants cannot

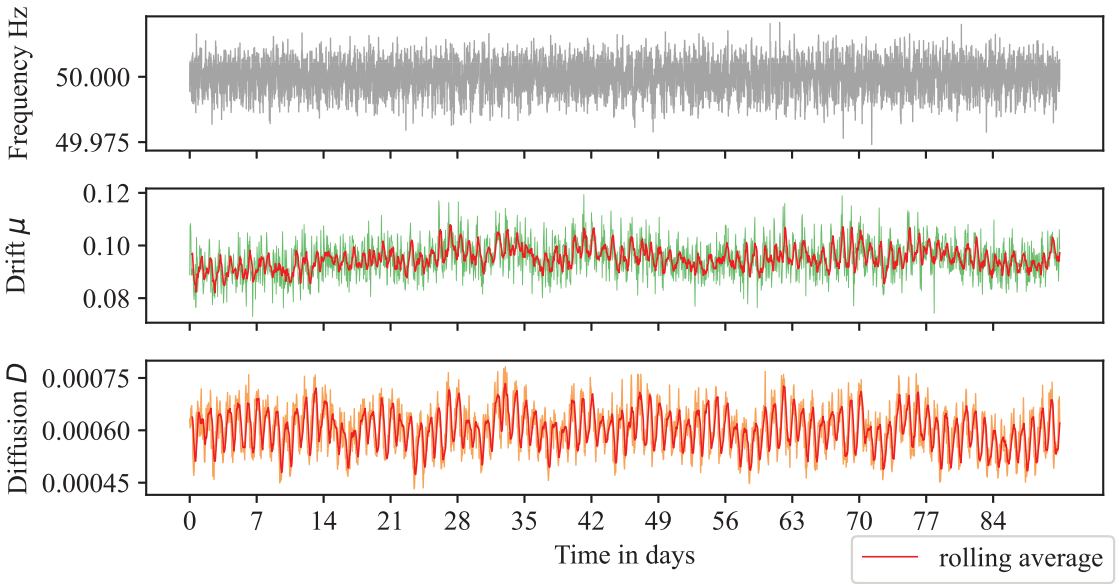

FIG. 12. Frequency, drift, and diffusion of the detrended AUS data, averaged over every 91-day interval.

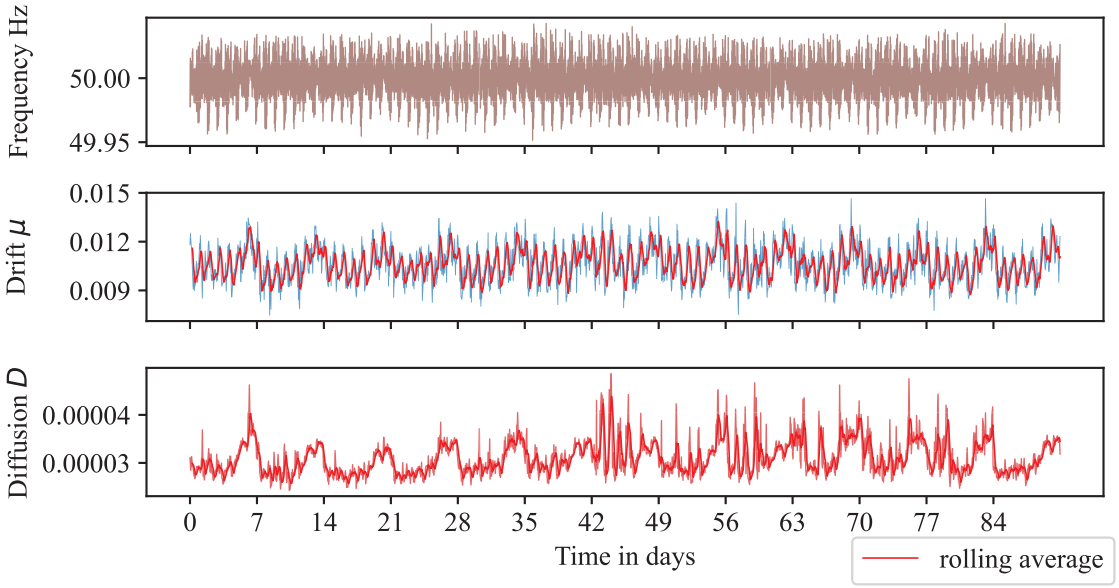

FIG. 13. Frequency, drift, and diffusion of the detrended CE data, averaged over every 91-day interval.

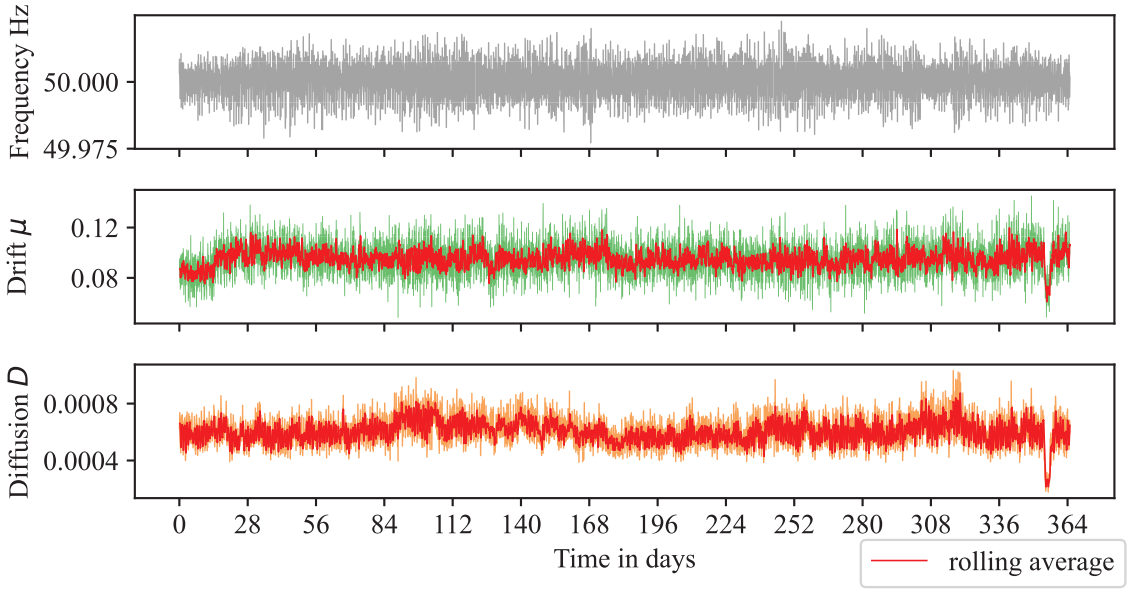

FIG. 14. Frequency, drift, and diffusion of the detrended AUS data, averaged over every 1-year interval beginning in January. The diffusion dominates the drift, which is confirmed by higher deviations in the frequency at times of high diffusion.

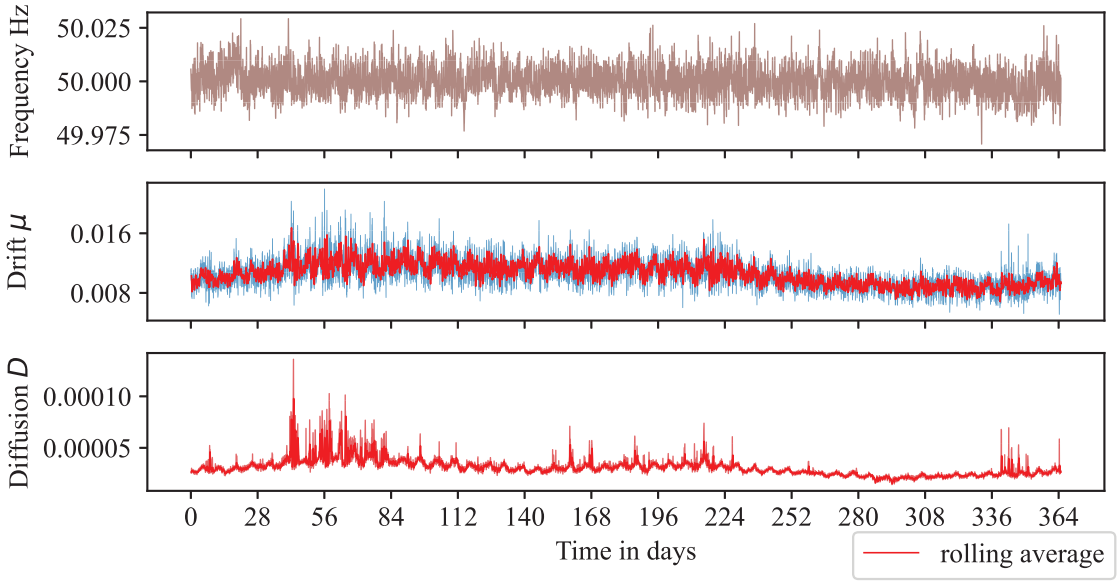

FIG. 15. Frequency, drift, and diffusion of the detrended CE data, averaged over every 1-year interval beginning in mid-March. Summer months show a higher-than-average drift and diffusion; in winter, the drift and diffusion are below average.

reliably increase their generation at the same rate as the solar power decreases its generation.

Continental Europe has a relatively stable diffusion, slightly below its average throughout the weekdays (cf. Fig. 9). There is a small peak in the morning followed by an immediate dip before noon and it remains almost constant for the rest of the day. Contrary to that, the weekend has a much higher diffusion, which does not dip below the highest diffusion values on weekdays. The heightened diffusion is reflected in the frequency as well. During the weekends, the deviation from the reference value is smaller due to the higher drift but also more noisy due to the diffusion.

In addition to the daily peaks that are already visible in the 7-day interval, the 28-day interval for AUS data shows a slight rise in diffusion every week (cf. Fig. 10). The peaks in diffusion continuously get higher from Monday to Sunday and are reset to their lower value the following week. This cyclic pattern indicates a mechanism that resets the diffusion weekly. No new insights compared to the 7-day intervals are possible for the 28-day CE data (cf. Fig. 11). The drift and diffusion continue to follow similar weekly patterns. The 3-month intervals also do not reveal new patterns (cf. Figs. 12 and 13). However, looking at the data at yearly intervals reveals new insights. The Australian interval begins in January, while the Continental European begins in March due to the nature of the dataset.

The AUS drift again does not follow any recognizable yearly patterns and stays relatively stable (cf. Fig. 14). In contrast, diffusion has multiple visible intervals of higher or lower values. This combination means that the diffusion patterns again dominate the drift, which is also visible in the frequency. During the intervals of higher diffusion, the frequency also tends to deviate further from the reference frequency.

CE has a very prominent yearly pattern of drift and diffusion (cf. Fig. 15. The diffusion is above average during April, May, June, July, September, and October, whereas November to March experienced below-average diffusion. In August, diffusion slightly dips to around the average. In contrast, the drift is constantly slightly above average from April to October and slightly below average from November to March. This deviation in drift is not as substantial as in the diffusion. The frequency during these months is also narrower, which can be attributed to the increased drift.

The months with higher diffusion and drift correspond to the CE summer months. One possible explanation for the heightened diffusion is the increase in solar power generation during these months due to longer sunlight hours. At the same time, there is also a lower average energy consumption.[40] Increased solar generation further decreases the relative amount of control as fewer coal or nuclear generators for a base power supply are needed. This imbalance reduces the grid stability, leading to a higher diffusion. Consequently, a higher drift is also necessary to keep the frequency stable and combat this increased diffusion during these months.

APPENDIX G: CORRELATION BETWEEN DRIFT AND DIFFUSION

When analyzing data, we observed an increased drift and diffusion at night on the daily scale as well as an increase during the summer months on the yearly scale, see Appendix F. These are both times during which the load is lower. Hence, we might speculate that there could be a correlation between drift, diffusion, and the load, respectively.

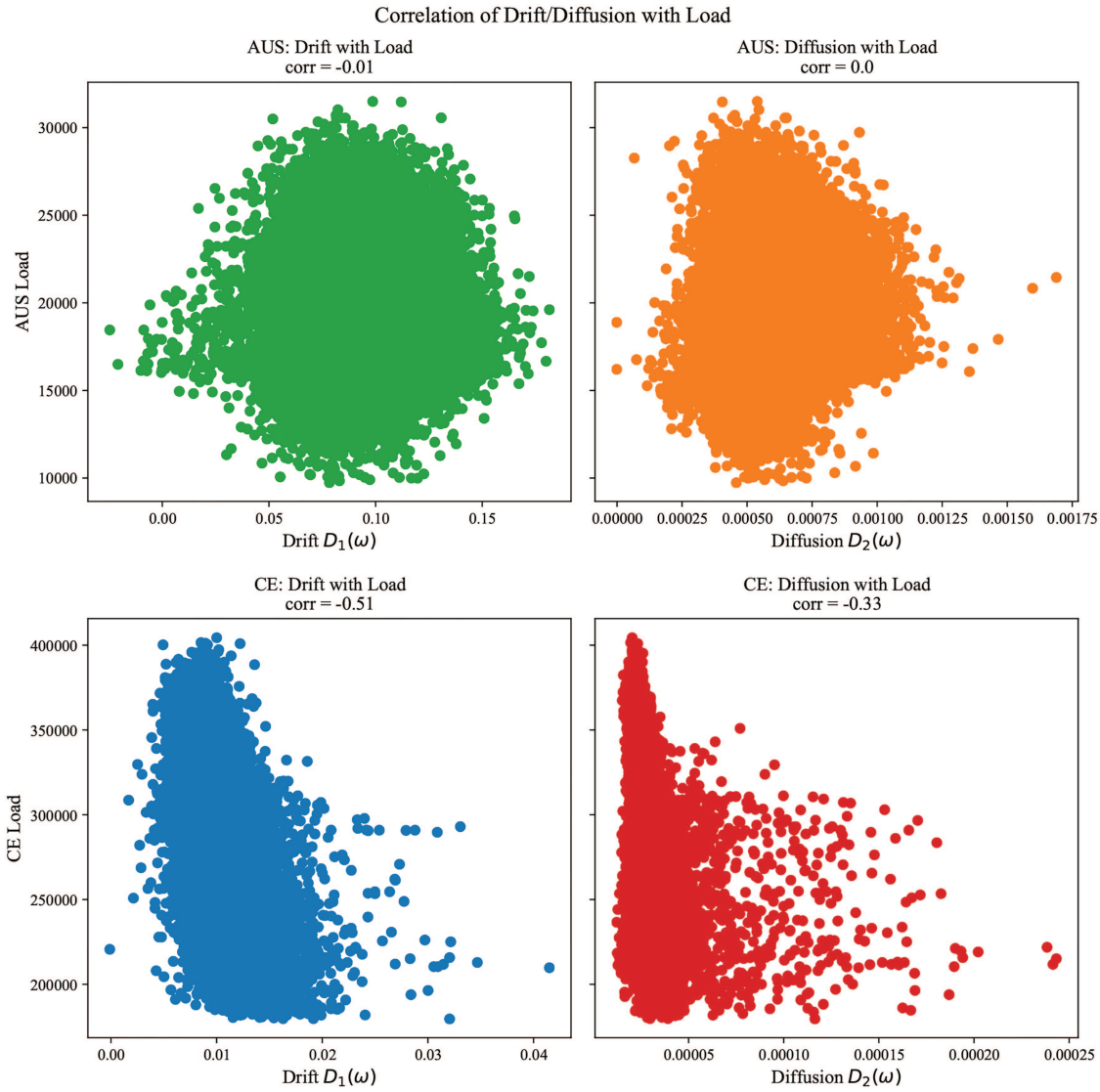

FIG. 16. Correlation of Australian and Continental European drift and diffusion with the respective load.

Figure 16 reveals that the (detrended) Australian drift has virtually no (linear) correlation with the total load. The diffusion is uncorrelated for Australia and has a negative correlation of −0.33 for Continental Europe. Contrary to AUS, the CE drift has the highest correlation with a value of −0.51. These low values of linear correlation further justify the use of non-linear models, such as GBT.

A high load is usually connected to a below-average drift, while a low load is correlated with an above-average drift. This correlation indicates that there is greater control over frequency deviations during periods of low load, which predominantly occur at night. At night, the base power providers, such as nuclear and coal, are mainly responsible for energy generation. They have high inertia and can provide a more stable power generation, unlike, for example, solar power plants, which are only available during the day.

Continental European drift and diffusion both have a negative correlation with the load, which leads to the possibility of them being correlated. This relationship is confirmed by Fig. 17, which reveals a positive correlation of 0.48. In contrast, Australia has no significant correlation between drift and diffusion. This discrepancy is likely due to the variability and lower quality of drift patterns observed in Australian data.

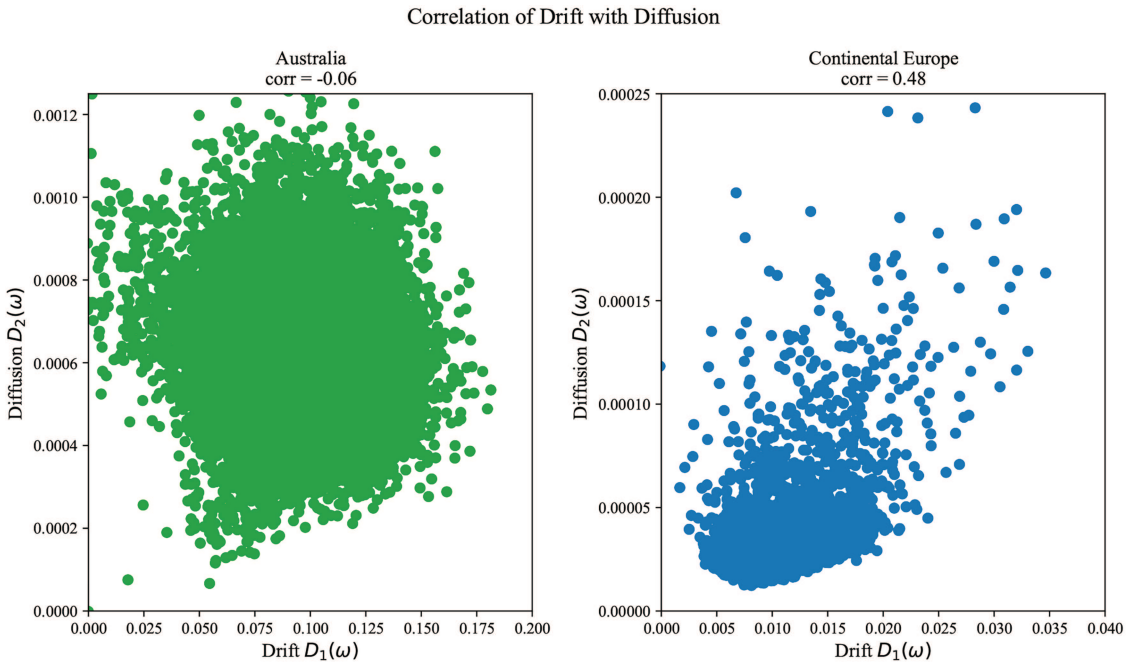

FIG. 17. Correlation of Australian and Continental European drift with the respective diffusion.

REFERENCES

[1] "Renewables 2024 Global status report—Global overview" (2024), https://www.ren21.net/wp-content/uploads/2019/05/GSR2024_GlobalOverview_Full_Report_with_endnotes_web.pdf.

[2] P. Kundur and N. J. Balu, *Power System Stability and Control* (McGraw-Hill, 1994).

[3] A. Fernández-Guillamón, E. Gómez-Lázaro, E. Muljadi, and Á. Molina-García, "Power systems with high renewable energy sources: A review of inertia and frequency control strategies over time," Renewable Sustainable Energy Rev. **115**, 109369 (2019).

[4] L. Löfquist, "Is there a universal human right to electricity? Int. J. Hum. Rights **24**, 711–723 (2020).

[5] T. Drewnick and U. Oberhofer, "Code," (2025), https://github.com/KIT-IAI-DRACOS/Code_ML_for_drift_and_diffusion.

[6] 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**, 43082–43097 (2020).

[7] X. Wen, M. Anvari, L. Rydin Gorjão, G. C. Yalcin, V. Hagenmeyer, and B. Schäfer, "Non-standard power grid frequency statistics in Asia, Australia, and Europe," arXiv:2308.16842[physics] (2023).

[8] J. Machowski, Z. Lubosny, J. W. Bialek, and J. R. Bumby, *Power System Dynamics: Stability and Control* (John Wiley & Sons, 2020).

[9] J. Slingo and T. Palmer, "Uncertainty in weather and climate prediction," Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. **369**, 4751–4767 (2011).

[10] 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).

[11] S. Swain, "Handbook of stochastic methods for physics, chemistry and the natural sciences," Opt. Acta: Int. J. Opt. **31**, 977–978 (1984).

[12] L. Rydin Gorjão and F. Meirinhos, "kramersmoyal: Kramers–Moyal coefficients for stochastic processes," J. Open Source Softw. **4**, 1693 (2019).

[13] "TransnetBW," (2024), https://www.transnetbw.de/en.

[14] "AEMO Australian frequency data," (2024), https://aemo.com.au/energy-systems/electricity/national-electricity-market-nem/system-operations/ancillary-services/frequency-and-time-deviation-monitoring.

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

[16] S. Pütz, J. Kruse, D. Witthaut, V. Hagenmeyer, and B. Schäfer, "Regulatory changes in German and Austrian power systems explored with explainable artificial intelligence," in *Companion Proceedings of the 14th ACM International Conference on Future Energy Systems*, e-Energy '23 Companion (Association for Computing Machinery, New York, 2023), pp. 26–31.

[17] J. Kruse, B. Schafer, and D. Witthaut, "Exploring deterministic frequency deviations with explainable AI," in *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)* (IEEE, Aachen, 2021), pp. 133–139.

[18] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," J. Mach. Learn. Res. **13**, 281–305 (2012).

[19] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2017), Vol. 30.

[20] C. Molnar, *Interpretable Machine Learning*, 2nd ed. (2022).

[21] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," arXiv:1802.03888 [cs, stat] (2019).

[22] T. Weissbach and E. Welfonder, "High frequency deviations within the European power system: Origins and proposals for improvement," in *2009 IEEE/PES Power Systems Conference and Exposition* (IEEE, Seattle, WA, 2009), pp. 1–6.

[23] B. Schafer, M. Timme, and D. Witthaut, "Isolating the impact of trading on grid frequency fluctuations," in *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)* (IEEE, 2018), pp. 1–5.

[24] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," Int. J. Forecasting **32**, 669–679 (2016).

[25] "OpenNEM Australian feature data," (2024), https://opennem.org.au/energy/nem/?range=7d&interval=30m&view=discrete-time.

[26]S. Patel, "Flexible operation of nuclear power plants ramps up," (2019), https://www.powermag.com/flexible-operation-of-nuclear-power-plants-ramps-up/.
[27]"Run-of-the-river hydroelectricity," (2024), Version ID: 1225188783, https://www.mdpi.com/1996-1073/12/20/3987.
[28]A. Ulbig, T. S. Borsche, and G. Andersson, "Impact of low rotational inertia on power system stability and operation," IFAC Proc. Vol. **47**, 7290–7297 (2014).
[29]R. Machlev, L. Heistrene, M. Perl, K. Y. Levy, J. Belikov, S. Mannor, and Y. Levron, "Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities," Energy AI **9**, 100169 (2022).
[30]D. Fryer, I. Strümke, and H. Nguyen, "Shapley values for feature selection: The good, the bad, and the axioms," IEEE Access **9**, 144352–144360 (2021).
[31]"Netztransparenz > regelenergie > daten regelreserve > aktivierte regelleistung," (2024), https://www.netztransparenz.de/de-de/Regelenergie/Daten-Regelreserve/Aktivierte-Regelleistung.
[32]S. Ö. Arik and T. Pfister, "Tabnet: Attentive interpretable tabular learning," in *Proceedings of the AAAI Conference on Artificial Intelligence* (Association for the Advancement of Artificial Intelligence, 2021), Vol. 35, pp. 6679–6687.
[33]J. Wang, J. Wiens, and S. Lundberg, "Shapley flow: A graph-based approach to interpreting model predictions," in *International Conference on Artificial Intelligence and Statistics* (PMLR, 2021), pp. 721–729.
[34]X. Wen, U. Oberhofer, L. R. Gorjão, G. C. Yalcin, V. Hagenmeyer, and B. Schäfer, "Identifying complex dynamics of power grid frequency," in *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems* (ACM, 2024), pp. 408–414.
[35]"OpenNEM: AveragePowerGenerationOver28Days," (2024), https://explore.openelectricity.org.au/.
[36]"ENTSO-E transparency platform," (2024), https://transparency.entsoe.eu/.
[37]T. L. Onsaker, "clean_corrupt_data.py," (2022), https://github.com/thlonsaker/predictability-of-island-power-grids/blob/master/scripts/clean_corrupt_data.py.
[38]T. L. Onsaker, "Pattern-based prediction of islanded power grid frequency," Master's thesis (Norwegian University of Life Sciences, Ås, 2022).
[39]J. Kruse, B. Schafer, and D. Witthaut, "Predictability of power grid frequency," IEEE Access **8**, 149435–149446 (2020).
[40]V. Kosoric, S. Wittkopf, and Y. Huang, "Testing a design methodology for building integration of photovoltaics (PV) using a PV demonstration site in Singapore," Archit. Sci. Rev. **54**, 192–205 (2011).