Machine learning of power grid frequency dynamics and control: prediction, explanation and stochastic modelling

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Johannes Kruse

aus Hamburg

Köln, 2023

Abstract

A reliable supply of electric power is not a matter of course. Power grids enable the transport of power from generators to consumers, but their stable operation constantly requires corrective measures and a careful supervision. In particular, power generation and demand have to be balanced at all times. A large power imbalance threatens the reliability of the power supply and can, in extreme cases, lead to a large-scale blackout. Therefore, the power imbalance is constantly corrected through distinct control schemes.

The power grid frequency measures the balance of power generation and demand. To guarantee frequency stability, and thereby a balance of generation and demand, load-frequency control constantly counteracts large frequency deviations. However, the transition of the energy system to renewable energy sources challenges frequency stability and control. Wind and solar power do not provide intrinsic inertia, which leads to increasingly fast frequency dynamics. Different economic sectors become strongly coupled to the power system, as, for example, the adoption of electric vehicles will interconnect the transport sector and the power system. Finally, wind and solar power are weather-dependent, which increases the variability of power generation. All in all, this gives rise to diverse, interdependent and stochastic impact factors, that drive the balance of power demand and generation, and thus the grid frequency. How can we predict, explain and model frequency dynamics given its strong non-autonomous and stochastic character?

In this thesis, I use machine learning to disentangle the effects of external drivers on grid frequency dynamics and control. First, I propose a prediction model that only uses historic frequency data, but fails in representing external impacts. Therefore, I include time series of techno-economic drivers and model their impact on grid frequency data using explainable machine learning methods. These methods reveal the dependencies between external drivers and frequency deviations, such as the important impact of forecast errors in the Scandinavian grid or the varying effects of different generation types. Finally, I integrate these drivers into a stochastic dynamical model of the grid frequency, which both represents short-term dynamics and long-term trends due to techno-economic impacts. My work complements traditional simulation-based approaches through validation and modelling inspiration. It offers flexible modelling and prediction tools for power system dynamics, which are profitable for systems with diverse impact factors but noisy and insufficient data.

Zusammenfassung

Eine zuverlässige Stromversorgung ist keine Selbstverständlichkeit. Stromnetze ermöglichen die Übertragung elektrischer Energie von den Erzeugern zu den Verbrauchern, aber ihr stabiler Betrieb erfordert ständige Korrekturmaßnahmen und eine sorgfältige Überwachung. Insbesondere müssen Stromerzeugung und -nachfrage jederzeit im Gleichgewicht sein. Ein großes Leistungsungleichgewicht gefährdet die Zuverlässigkeit der Stromversorgung und kann im Extremfall zu einem großflächigen Stromausfall führen. Daher wird das Leistungsungleichgewicht ständig durch verschiedene Regelungssysteme korrigiert.

Die Netzfrequenz ist ein Maß für das Gleichgewicht von Stromerzeugung und Stromnachfrage. Um die Frequenzstabilität und damit ein Gleichgewicht zwischen Erzeugung und Nachfrage zu gewährleisten, wirkt die Regelleistung ständig großen Frequenzabweichungen entgegen. Die Umstellung des Energiesystems auf erneuerbare Energiequellen stellt jedoch eine Herausforderung für die Frequenzstabilität und -regelung dar. Wind- und Solarenergie bieten keine eigene Trägheit, was zu einer beschleunigten Frequenzdynamik führt. Verschiedene Wirtschaftssektoren werden stark an das Stromsystem gekoppelt. Zum Beispiel wird die Einführung von Elektrofahrzeugen den Verkehrssektor und das Stromsystem miteinander verbinden. Schließlich sind Wind- und Sonnenenergie wetterabhängig, was die Variabilität der Stromerzeugung erhöht. Insgesamt führt dies zu vielfältigen, voneinander abhängigen und stochastischen Einflussfaktoren, die das Gleichgewicht von Stromnachfrage und -erzeugung und damit die Netzfrequenz beeinflussen. Wie können wir die Frequenzdynamik vor dem Hintergrund ihrer nicht-autonomen und stochastischen Eigenschaften vorhersagen, erklären und modellieren?

In dieser Arbeit verwende ich maschinelles Lernen, um die Auswirkungen externer Faktoren auf die Netzfrequenzdynamik und -steuerung zu entschlüsseln. Zunächst schlage ich ein Vorhersagemodell vor, das nur historische Frequenzdaten verwendet, aber externe Einflüsse nicht abbilden kann. Um dieses Defizit zu beheben, beziehe ich Zeitreihen technisch-wirtschaftlicher Einflussfaktoren ein und modelliere ihre Auswirkungen auf die Netzfrequenz mithilfe erklärbarer maschineller Lernmethoden. Diese Methoden zeigen die Abhängigkeiten zwischen externen Einflussfaktoren und Frequenzabweichungen auf, wie z. B. den bedeutenden Einfluss von Prognosefehlern im skandinavischen Netz oder die unterschiedlichen Auswirkungen der verschiedenen Erzeugungsarten. Schließlich integriere ich diese Einflussfaktoren in ein stochastisches dynamisches Modell der Netz-

frequenz, das sowohl kurzfristige Dynamiken als auch langfristige Trends aufgrund von technisch-wirtschaftlichen Einflüssen darstellt. Meine Arbeit ergänzt die traditionellen simulationsbasierten Ansätze durch Validierung und Modellinspiration. Sie bietet flexible Modellierungs- und Prognosewerkzeuge für die Dynamik von Stromnetzen, die für Systeme mit verschiedenen Einflussfaktoren, aber verrauschten und unzureichenden Daten von Nutzen sind.

Acknowledgements

First of all, I'd like to gratefully thank my supervisor, Dirk Witthaut, for the support in terms of ideas, discussions and advice, both on the scientific and the personal level. I would also like to thank Benjamin Schäfer for his support, supervision, and personal guidance throughout the dissertation. Furthermore, I'd like to thank Professor Simon Trebst and Professor Pere Colet for reviewing my thesis and for joining my thesis defence committee and Professor Johannes Berg for heading the committee. In addition, I would like to thank the members of the research group for the discussions and helpful inputs, as well as for the IT-management/IT-support that is necessary to keep the research going. Finally, I would like to thank my friends, flatmates and family for their support in completing this work.

Contents

1. Introduction

1.1. The reliability of electricity supply: a matter of course?

Electricity is essential for the daily life of many people around the world. In various countries, central societal functions such as health, security and safety of the people depend on the reliable supply of electricity [6]. For example, a prolonged large-scale blackout will cause the failure of several telecommunication services, thus preventing people to obtain important information during a crisis. Furthermore, all electric rail transport shuts down, road traffic becomes chaotic due to failing traffic lights, water supply stops due to failing electrical pumps, with more failures expected in food supply, the financial and the health care sector. Due to such consequences for the basic functioning of society, power supply is considered a critical infrastructure, which needs special strategical protection [6, 7]. Among the critical infrastructures, such as water supply and transport, power supply is particularly important since almost all other critical infrastructures depend on it [8]. In the future, these interdependencies will grow due to the energy transition. In particular, the rise of electric cars and the shift from gas-based heating to electric heat pumps will potentially increase the electricity demand in the transport and heating sectors [9, 10]. These growing interdependencies with the power system amplify its criticality, such that its reliable operation is of vital importance.

In Europe, the power systems typically exhibit a high level of reliability. The reliability standards require that electricity might be unavailable not more than 3-8 hours per years in the whole grid [11]. In Germany, the average time with unavailable electricity supply per person was around 11 minutes in 2020 (Figure 1.1a), i.e., on average, each person spent 0.002 % of the year without electricity [12]. Therefore, the experience of a reliable electricity supply is a matter of course in daily life.

However, a large technical infrastructure and continuous adaptive measures are necessary to maintain such a high level of system reliability. This becomes clear when decomposing the electricity costs paid by an average private household (Figure 1.1b). Only 26% are paid for actually generating electric power, while 22% account for the grid charges, which are exclusively used to ensure the reliability of electricity supply [12]. In particular, grid charges pay for grid management, i.e., the extension and maintenance of the power grid, and for the supply of adaptive measures called ancillary services. These an-
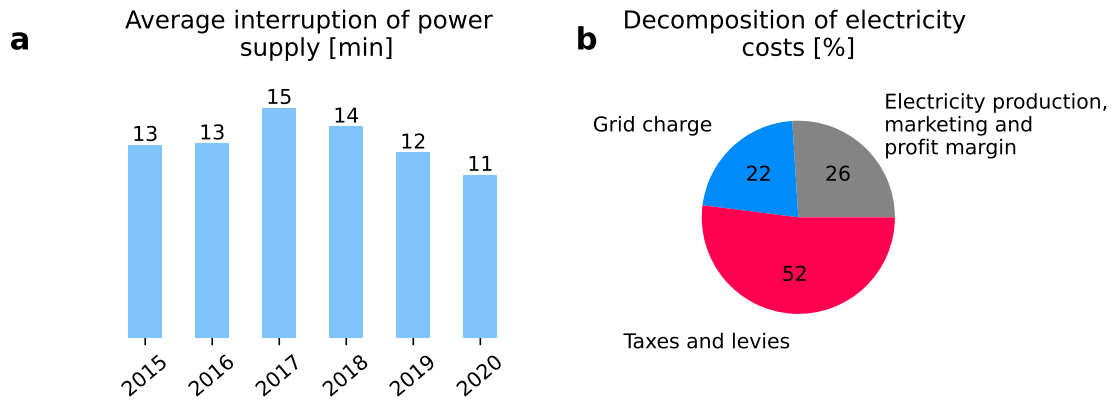
Figure 1.1.: A high reliability of electricity supply requires costly grid management and balancing. **a**: The average duration of power interruptions is constantly low in Germany, thus guaranteeing a reliable electricity supply in daily life [12]. **b**: This reliability requires continuous balancing efforts and infrastructure management, which are reflected in the electricity costs through the grid charge. The pie depicts the average electricity costs paid by private households in the year 2021 [12].

cillary services are continuous measures that control and correct power generation, load and transmission to guarantee a reliable operation of the power system within physical constraints.

Different threats to system reliability require different ancillary services [12]. In Germany, the largest part of grid charges relates to congestion management (69 %) and grid losses (20 %). Congestion occurs in a power grid if the generated power cannot be transported to the consumer because the capacity of the grid is not large enough. Power losses are caused by power transmission and thus have to be paid additionally. The third most costly ancillary service is power balancing and control (8%). It deals with temporary mismatches between power generation and demand and is therefore inherently connected to the power grid frequency.

1.2. Power grid frequency stability and control

The power grid frequency is a central indicator for the reliable operation of power systems [14]. It is measured as the angular velocity of the AC voltages in the power grid. Under normal operating conditions, the voltage angles in large-scale power grids are synchronised at timescales above one second such that I refer to a single "bulk" frequency throughout this thesis. The target value of the grid frequency is $f_{\text{ref}} = 50$ Hz in Europe (both 50 and 60 Hz are used in other regions), but the actual value fluctuates around this reference. These deviations Δf from the reference are caused by imbalances ΔP between power generation and demand. If there is an under-production of electricity ($\Delta P < 0$), the lacking power
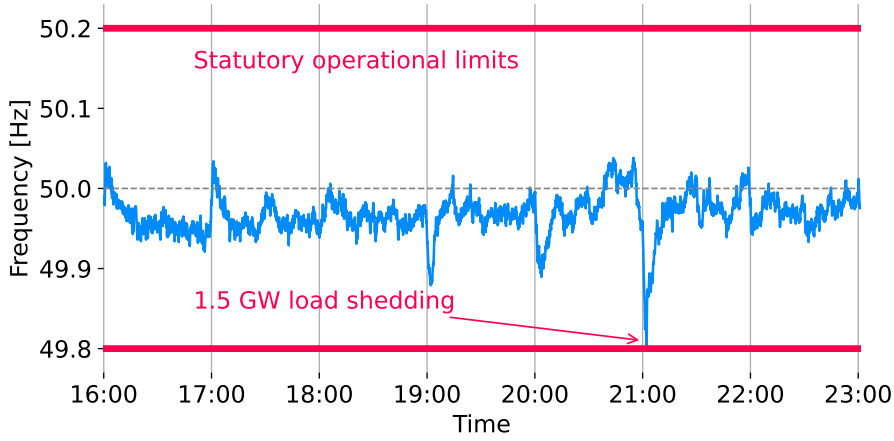
Figure 1.2.: Large grid frequency deviations threaten the reliability of electricity supply. The figure depicts a large frequency deviation in Continental Europe in January 2019, which reached the statutory operational limits of a secure system operation (data from ref. [13]). As an emergency measure 1.5 GW of industrial load were shut down. Within the operational limits, frequency control measures continuously prevent large deviations, thus ensuring a high level of system reliability.

is drawn from the rotational energy E_{rot} of all synchronous generators within the power grid and their rotors slow down thus causing a frequency drop in the AC voltages of the generators (and vice versa for an over-production of electricity). This inertial response of synchronous generators is commonly modelled by the aggregate swing equation [15],

$$\frac{2E_{\text{rot}}}{f_{\text{ref}}} \frac{\mathrm{d}\Delta f}{\mathrm{d}t} = \Delta P. \tag{1.1}$$

The Rate of Change of Frequency (RoCoF) $\mathrm{d}\Delta f/\mathrm{d}t$, increases with the power mismatch, but also with declining rotational energy E_{rot}, which I discuss in Section 1.3.

If grid frequency deviations become too large, they threaten the reliability of the power supply, thus resulting in costly emergency measures. Electric devices in the power grid can only operate safely within a narrow band around the reference frequency. For example, generators can only operate up to a certain minimum frequency (49.5 Hz in Continental Europe) and they shut down below this value to prevent internal damages [16]. This further decreases the frequency and can lead to cascading outages including other types of instabilities such as voltage or angle instability [17]. As an emergency measure, power loads are disconnected from the grid, thus reducing the power imbalance and stabilising the grid frequency [14]. This measure, called load shedding, was for example used to mitigate a severe frequency deviation in January 2019 within the Continental European grid (Figure 1.2). The grid frequency reached the statutory operational limits and 1.5 GW of industrial loads were shed as a countermeasure [18], which is nearly twice the power

consumption of a large German city (such as Cologne). In 2020, load shedding was only needed at nine days in Germany, which however came at a large cost of nearly 32 million Euros to reimburse the affected costumers [12].

To avoid such costly emergency measures, different control schemes constantly act on the frequency and impede deviations beyond statutory operational limits [14]. Primary control is activated on the timescale of seconds. It is governed by a proportional controller $P_{\mathrm{prim}} = K_T \Delta f$ and thus dampens the frequency deviation. The total damping coefficient $K_1 = K_T + K_L$ summarises the effect of primary control and frequency-dependent loads, which also contribute with a small damping coefficient $K_L \ll K_T$. On a timescale of minutes, secondary control activates to restore the grid frequency to its reference value. Secondary control is typically governed by an integral controller $P_{\mathrm{sec}} = K_2 \bar{\theta}$ with $\bar{\theta} = \int^t \Delta f(t') \mathrm{d}t'$ [19], but other implementations are also possible [20].

The resulting dynamics of grid frequency deviations are governed by a non-autonomous differential equation. Combining the inertial response from Eq. (1.1) and the control measures yields

$$\frac{2E_{\mathrm{rot}}}{f_{\mathrm{ref}}} \frac{\mathrm{d}\Delta f}{\mathrm{d}t} = \Delta P - K_1 \Delta f - K_2 \bar{\theta},$$
$$\frac{\mathrm{d}\bar{\theta}}{\mathrm{d}t} = \Delta f. \tag{1.2}$$

Note that I chose a simplified representation of primary and secondary control to make the dynamics analytically accessible and enable their precise understanding, but more detailed and realistic models exist [14].

The dynamical system (1.2) corresponds to a damped harmonic oscillator with external driving ΔP. The complexity arises from ΔP, which varies strongly over time. An important example are deterministic frequency deviations (DFDs), which occur due to the temporary mismatch between the continuously evolving load and the step-like power generation that is traded within fixed market intervals [21]. This temporary power mismatch changes its sign mainly according to the load gradient: If the load rises in the morning, we observe positive frequency peaks, while negative peaks occur during the night where the load decreases. These deterministic peaks are clearly visible in Figure 1.2, where the frequency rises or drops regularly at the beginning of each hour. In addition to these market-based effects, the power mismatch changes due to forecast errors, power outages or other kinds of unscheduled events [22]. This highlights the highly non-autonomous nature of grid frequency dynamics. In the future, these non-autonomous effects will become increasingly complex due to the energy transition.

1.3. Energy transition challenges grid frequency stability

The transition to renewable energy sources introduces new challenges in three main areas [23], which also amplify the complexity of grid frequency dynamics and the threats to system stability.

Increased variability   Most renewable energy sources depend on the variable weather conditions. In particular, the large-scale integration of wind and solar power into the power system increases the variability of power generation on different scales. Most importantly, seasonal and synoptic fluctuations continuously change the power output of renewable sources [24–26]. On larger timescales, the level of available wind power changes significantly from year to year [27] and from decade to decade [28]. Even on smaller timescales of seconds, varying weather conditions shape the fluctuations of renewable power generation [29, 30]. In general, renewable power fluctuations cannot be fully predicted, e.g., day-ahead wind power forecasts typically exhibit forecast errors of up to 20% [31]. The uncertainty and variability of renewables thus introduces power imbalances and thereby affects grid frequency dynamics.

Growing interdependencies   The optimal location of renewable energy sources often lies far away from large centres of load. For example, wind power in Germany is favourably built along the northern coastline, while most of the load concentrates in western and southern Germany [32]. This requires a strong extension of the transmission grid, not only in Germany [32, 33], but also within Europe [34]. Even between different asynchronous power grids that are only connected via DC links, multiple new interconnectors are planned, e.g., to transport hydro power from the Nordic to the Continental European grid [34, 35].

Furthermore, the coupling of electricity markets within Europe will be intensified to balance variable power generation and demand more efficiently. On spot markets, such as the European Power Exchange (EPEX Spot), power is delivered several hours (day-ahead) or several minutes after the market closure (intraday) [36]. The day-ahead and intraday markets in Europe are already strongly coupled, which entails a more effective allocation of import/export capacities, but some countries will still join in the future [37]. Reserves and balancing services, such as secondary control capacities, are also traded on specific markets. These reserve markets will be increasingly coupled as requested by an EU regulation [38], which is currently implemented through projects such as PICASSO [39] and MARI [40].

Finally, different industrial sectors become more coupled, for example due to the elec-

trification of heating (cf. Section 1.1). All in all, these developments increase the interdependencies between the power generation and demand of different regions, power grids, electricity markets and industrial sectors. These developments diversify the impact factors on both power generation and load and thus on the dynamics of the grid frequency.

Low inertia   Wind and solar power are connected to the power grid via inverters [41]. Therefore, they do not provide intrinsic inertia like conventional synchronous generators do [42]. Their contribution to E_{rot} is zero, which directly affects the RoCoF according to Eq. (1.1): More inverter-based generation, i.e., smaller E_{rot}, means larger RoCoFs. This is critical, as different devices in the power system can only cope with limited RoCoFs [43]. Due to the variability of renewable generation, the inertia will also vary over time, thus constantly changing the response characteristics of grid frequency dynamics.

These challenges already affect grid frequency stability and control, as illustrated by current inertia data from Great Britain (Figure 1.3a). Already now, the inertia is nearly always below its critical value, below which acceptable RoCoFs are not guaranteed anymore. Therefore, costly ancillary services have to regulate the largest possible RoCoF (Figure 1.3b) in order to protect electric devices that cannot cope with large RoCoFs.

Increased variability, growing interdependencies and inertia fluctuations enhance the non-autonomous character of grid frequency dynamics (cf. Eq. (1.2)). They intensify the temporal complexity of power imbalances $\Delta P(t)$ and give rise to new inter-dependent impact factors. This setting makes power system dynamics an interesting target for transdisciplinary science [44], and multiple intersections with statistical physics and data analysis exist [45]. In particular, the data-driven analysis of grid frequency dynamics has flourished during the past years.

1.4. Empirical research on grid frequency dynamics and control

In the past years, power system data has become increasingly available on public platforms. Grid frequency measurements from different locations are publicly available with a second or sub-second resolution since 2011 [48]. In Europe, other time series from power system operation are published on a central transparency platform since 2015 [49]. This has enabled a growing number of empirical studies on grid frequency stability and control.

Grid frequency time series have been studied using data and complexity analysis, which yielded complex stochastic characteristics on different timescales. The distribution of frequency deviations exhibits heavy tails in various power grids around the world [50, 51], while its overall shape can be both single-peaked and multi-peaked [52]. The frequency
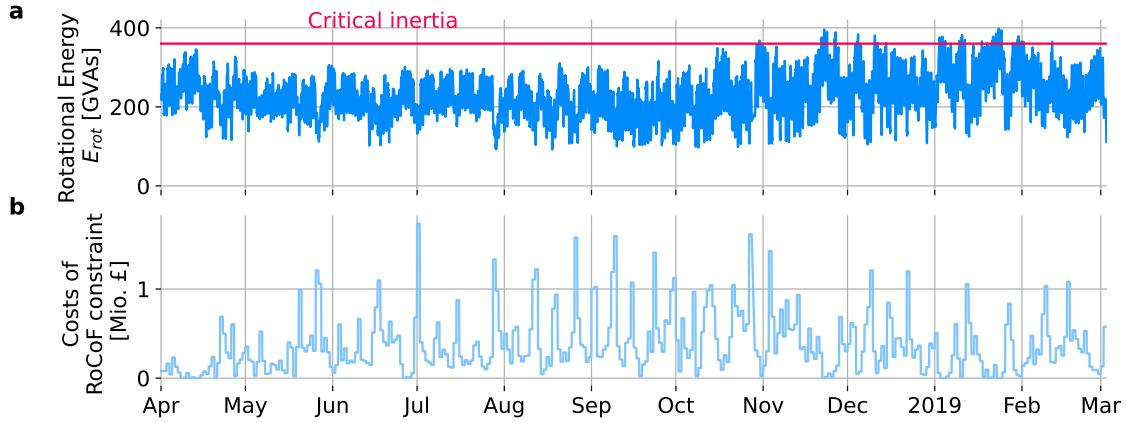
Figure 1.3.: Renewable energy sources challenge grid frequency stability in Great Britain. **a**: Already now, the rotational energy E_{rot}, which is stored in synchronous generators, varies strongly and reaches critically low values due to high shares of renewable generation. Above the critical inertia of 360 GVAs, even the sudden loss of the largest in-feed (1.8 GW [46]) cannot result in RoCoFs larger than 0.125 Hz/s, which is the security constraint for the RoCoF in Great Britain [43]. **b**: Below the critical inertia, ancillary services might be necessary to comply with the RoCoF constraints. The daily costs of these measures are depicted in this panel. The data is taken from ref. [47].

increments often exhibit a non-Gaussian distribution in particular at small timescales [53], but the distribution varies strongly among different power grids [51]. In European grids, the power spectrum reveals strong hourly and daily periodicities in the grid frequency time series [50, 54, 55], which however does not hold for other grids, e.g., on small islands [51]. Detrended fluctuation analysis was used to investigate frequency fluctuations, which yielded multiple characteristic timescales in European frequency data [51, 56, 57]. In particular, on short timescales of seconds, European grid frequency measurements exhibit long-range correlations beyond Brownian motion.

Different stochastic approaches were proposed to model these complex characteristics of the grid frequency. Superstatistics were applied to model the heavy-tailed distribution of frequency deviations in different power grids [50], which demonstrated that statistical parameters of the grid frequency vary at a timescale of 1-5 h. Such non-stationary characteristics were also revealed by superstatistical models of the increments on timescales below 1s [58]. Multiple authors used stochastic differential equations to model grid frequency dynamics [19, 54, 59]. A Langevin equation with Gaussian white noise was able to partly reproduce the heavy tails of grid frequency distributions by incorporating the deterministic effect of electricity markets into the model. Moreover, it reproduced the characteristic periodicities present in the autocorrelation function [19]. The British and Irish frequency was modelled with a stochastic model that reproduced their bi-modal distribution, which suggests deadbands in the primary control as a cause of the bi-modality [60,

61]. Ref. [62] goes beyond Gaussian white noise and modelled the British frequency with fractional non-Gaussian (levy-stable) noise, which reproduced the long-range frequency correlations, as well as the non-Gaussian tails of their distribution. All in all, these studies demonstrate the non-standard statistics of grid frequency dynamics that vary over time at multiple different timescales.

This PhD thesis goes one step further and asks for the drivers of this non-stationary and non-Gaussian process. This is particularly interesting as growing interdependencies diversify the drivers of power system dynamics (cf. Section 1.3). Using grid frequency measurements together with other techno-economic time series from power system operation, I investigated the following questions:

Q.1 Prediction: Can we predict grid frequency dynamics given their non-autonomous stochastic character?

Q.2 Explanation: How do external drivers influence frequency dynamics and control?

Q.3 Stochastic modelling: How can we incorporate these external effects in stochastic models of the grid frequency?

To tackle these questions, I used methods from machine learning and eXplainable artificial intelligence (XAI).

1.5. Scope of the thesis: explainable machine learning for grid frequency stability

Machine learning methods are well suited to model complex, non-linear dependencies and interactions [63]. Modern methods, such as gradient tree boosting [64, 65] and artificial neural networks [66], can process a large number of features by performing feature selection or feature extraction. This makes them perfectly suitable to model the large number of impact factors that drive grid frequency dynamics and control.

However, such complex models are often black-boxes, i.e., the researcher cannot understand how the model actually predicts the target from different features [67]. This is particularly problematic in power system control, which requires trustworthy models to guarantee a reliable electricity supply [68]. Moreover, black-box models (mostly) prevent scientific insights from the modelling process [69]. To learn about drivers and impact factors, I therefore applied methods from XAI. These methods allow us to understand how a black-box model predicts the outputs based on its input factors. Thus, they are promising to provide both an analysis and a prediction tool for power grid frequency deviations based on operational data.

The thesis is structured in three parts. The first part, Section 2, focuses on prediction and the effect of the non-autonomous character of grid frequency dynamics (Topic Q.1). I present a paper that predicts the frequency solely based on its history (univariate prediction), whose limits implicitly reveal the role of external influences that are not present in the model. The second part, Section 3, explicitly studies the effect of external drivers on grid frequency dynamics and control (Topic Q.2). It includes three publications that use explainable machine learning to predict and explain indicators of frequency stability, such as the RoCoF (cf. Section 1.2), as well as frequency control activation. In the last part, Section 4, I turn back to the stochastic models from the literature and incorporate the external drivers studied before (Topic Q.3). In particular, I present a paper with a physics-inspired machine learning model that integrates both a stochastic dynamical model and the effects of external techno-economic features.

In addition to the aforementioned publications, I contributed to three other manuscripts while working on my thesis, which, however, are not part of my thesis: I contributed to ref. [51] as a supporting author, where I analysed spatio-temporal fluctuations of grid frequency time series using principal component analysis. In refs. [70, 71], I contributed with supervision, research design and writing. In particular, we applied my explainable machine learning approach (from Topic Q.2) to reveal interactions between grid frequency stability and power flows between different synchronous power grids [71]. In ref. [70], we used machine learning to predict dynamic stability of power grids under line failures by using static features of the power system.

2. Non-autonomous grid frequency dynamics: strengths and limitations of a univariate prediction

A precise knowledge of grid frequency deviations is central to optimally control power imbalances. The prediction of future deviations can optimise control and balancing, e.g., through model predictive control (cf. Section 5.1). However, this prediction task is challenging due to the non-autonomous nature of grid frequency deviations with its diverse impact factors. In the following section (2.1), I present a univariate prediction model that forecasts the grid frequency solely based on its own history, i.e., without explicitly including external impact factors. By analysing the strengths and limitations of the model, I implicitly reveal the importance of external factors that shape grid frequency dynamics.

The results emphasised the importance of deterministic, daily patterns for predicting the grid frequency, especially in Continental Europe. Using specific hourly patterns as predictor improved the prediction of the frequency trajectory for a horizon of 15-30 minutes. The prediction horizon was strongly reduced at the beginning of a trading interval, where the generation is suddenly adapted due to the interval-based trading on electricity markets (cf. Section 1.2). This performance loss indicates the importance of external impact factors such as electricity trading. Therefore, external techno-economic impact factors have to be included to better describe the non-autonomous grid frequency dynamics.

In this publication, I conducted the research, created the figures and wrote the major part of the manuscript.

2.1. A) Predictability of Power Grid Frequency

[1] Kruse, J., Schäfer, B. & Witthaut, D. Predictability of Power Grid Frequency. *IEEE Access* **8,** 149435–149446. doi:10.1109/ACCESS.2020.3016477 (2020).

Predictability of Power Grid Frequency

JOHANNES KRUSE [ID]1,2, BENJAMIN SCHÄFER [ID]3, AND DIRK WITTHAUT [ID]1,2

[1]Forschungszentrum Jülich GmbH, Institute of Energy and Climate Research - Systems Analysis and
Technology Evaluation (IEKSTE), 52425 Jülich, Germany
[2]Institute for Theoretical Physics, University of Cologne, 50937 Cologne, Germany
[3]School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, U.K.

Corresponding author: Johannes Kruse (jo.kruse@fz-juelich.de)

ABSTRACT The power grid frequency is the central observable in power system control, as it measures the balance of electrical supply and demand. A reliable frequency forecast can facilitate rapid control actions and may thus greatly improve power system stability. Here, we develop a weighted-nearest-neighbour (WNN) predictor to investigate how predictable the frequency trajectories are. Our forecasts for up to one hour are more precise than averaged daily profiles and could increase the efficiency of frequency control actions. Furthermore, we gain an increased understanding of the specific properties of different synchronous areas by interpreting the optimal prediction parameters (number of nearest neighbours, the prediction horizon, etc.) in terms of the physical system. Finally, prediction errors indicate the occurrence of exceptional external perturbations. Overall, we provide a diagnostics tool and an accurate predictor of the power grid frequency time series, allowing better understanding of the underlying dynamics.

INDEX TERMS Power grid frequency, frequency control, power system stability, time series forecasting, k-nearest-neighbours.

I. INTRODUCTION

The electrical power system relies on a constant balance of supply and demand. Abundant energy will speed up generators and lead to an increase of the power grid's (mains) frequency. Similarly, a shortage of generation slows down the same generators and reduces the systems frequency as kinetic energy stored in the generator is transformed into electrical energy. Control systems, ordered from primary to tertiary control, help to ensure the balance of supply and demand by closely monitoring the frequency trajectory and maintaining it close to the desired reference value of $f = 50$ or 60 Hz [1]. Large deviations of the frequency away from the reference require decisive control actions and cause high costs [2].

To optimize the usage of costly control actions, we require a precise understanding of the power grid frequency. This frequency is neither constant nor varying slowly but is instead highly stochastic and subject to multiple external influences [3], [4]. For example, the organization of the energy market leads to deterministic imprints of dispatch activities in the frequency in forms of sudden jumps or drops [5]. Simultaneously, an increasing share of renewable generators decreases

The associate editor coordinating the review of this manuscript and approving it for publication was Jahangir Hossain [ID].

the inertia available in the grid [6] and introduces additional fluctuations [7], [8]. Given this hybrid stochastic and deterministic nature, the question arises to which extend the frequency trajectory is predictable. A precise estimate of the future frequency trajectory would be very beneficial as it would allow an estimate of necessary control power early in time, saving costs [2] and stabilizing the grid [1].

Beyond precise forecasts of the near future trajectories, a fundamental understanding of the power grid frequency dynamics is critical as this one-dimensional time series encodes vast information on the stability and the current state of the power system [9]. Only a solid understanding of how the energy mix, demand patterns and energy market rules impact the power system and its stability will allow us to implement and control highly renewable power systems in the future. As the starting point to develop such an understanding, we study the power grid frequency since frequency data is much more readily available [10] than precise demand or generation values in a given synchronous area.

With the increasing popularity of machine learning techniques [11], there are many tools available to forecast time series, such as the power grid frequency. Recent studies used artificial neural networks (ANN) [12] to predict hourly frequency time series in India based on features such as
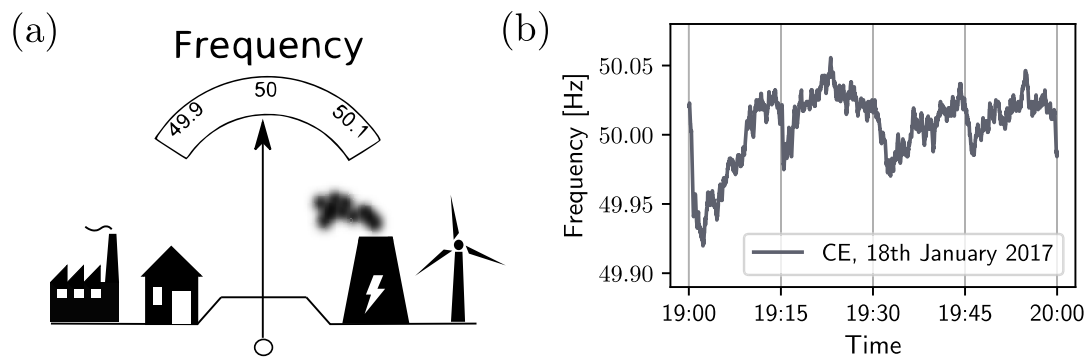
FIGURE 1. The nature of the power grid frequency. (a): The frequency reflects the balance of power demand and generation. Over-production causes a frequency increase and under-production a frequency decline. (b): Example frequency time series from the CE synchronous area [21]. It displays the typical frequency jumps at 15 minute intervals that are caused by the trading on electricity markets and subsequent changes of the power plant dispatch.

wind power generation and power demand. Other authors [9] used a linear state space model and uncertain basis function to predict US frequency time series for up to one second, while a Bayesian network was used to predict the frequency time series for up to 3 minutes [13]. Finally, auto-regressive moving average (ARMA) models have been used in the British grid to achieve prediction horizons of tens to hundreds of seconds [14] and in the US to achieve forecasts of 5 to 30 minutes [15].

We will particularly focus on k-weighted-nearest-neighbour (WNN) methods, which have gained popularity in a variety of fields from biology [16] to financial systems [17], but have also been applied in the energy sector, e.g. to forecast electricity prices [18] or power demand [19]. In contrast to earlier applications of the WNN predictor on the power grid frequency [15], we improve the statistical evaluation of the predictor and introduce a system-specific null model to benchmark its performance. Furthermore, we employ the forecast accuracy not only as a performance measure but as a tool to analyze the dynamics of the power system in general and the interplay of internal and external influences in particular. WNN predictors are particularly well suited for that purpose as they are among the best explainable machine learning algorithms [20].

In this article, we use frequency recordings from several European synchronous areas to motivate the mean frequency (daily profile) as an suitable null model (Section II) and develop a WNN predictor to forecast the time series (Section III). We demonstrate how our predictions outperform the null models in particularly on short prediction horizons and provide in-depth analysis and interpretation of when and how the power grid frequency can be predicted (Section IV).

II. DATASET DESCRIPTION
A. DATA SOURCES AND PRE-PROCESSING
We train and test our frequency predictor on large high-resolution datasets from three different European synchronous areas. In particular, we use publicly available frequency recordings of the years 2015-2018 from the

Continental Europe (CE) [21], the Great Britain (GB) [22] and the Nordic synchronous areas [23], following the naming convention used in [24]. The data from CE and from GB comes with a one-second resolution, while the Nordic data exhibits a resolution of 0.1 s. Moreover, some of the datasets have varying formats and multiple frequency recordings are corrupted or missing. We therefore resample the data to a common one-second resolution and conduct a thorough pre-processing (Supplemental Material). The pre-processed time series are available online [25], thus providing a ready-to-use database to develop new methods for frequency analysis and prediction.

We want to point out that our pre-processing involves the identification and exclusion of corrupted measurements. However, the k-nearest-neighbour method can cope with the resulting holes in the time series. Missing segments are simply ignored during the nearest neighbour search. This is a great advantage of the WNN predictor, as we can harness the full length of the dataset without manipulating it too much.

B. CHARACTERISTICS OF THE FREQUENCY TIME SERIES
The frequency trajectory exhibits deterministic as well as stochastic characteristics, which can be attributed to different dynamics within the power system. Firstly, a frequency deviation generally reflects a mismatch of power generation and demand (Fig. 1(a)). Such a mismatch occurs when the power generation does not match the expected demand curve. The demand itself evolves continuously and shows typical daily, weekly and seasonal patterns [2]. In contrast, the power generation exhibits discontinuous behaviour due to the trading on electricity markets and the resulting changes of the power plant dispatch [5]. In Europe, this trading is operated on various different spot-markets such as the European Energy Exchange Power Spot (EPEX SPOT), which covers countries in Western and Central Europe. The resulting dispatch changes are commonly scheduled for discrete time intervals of one hour, 30 and 15 minutes [26], [27]. The mismatch between the step-like behaviour of the generation and the continuous behaviour of the load leads to regular frequency jumps at the beginning of these trading intervals [3], [5], [10].
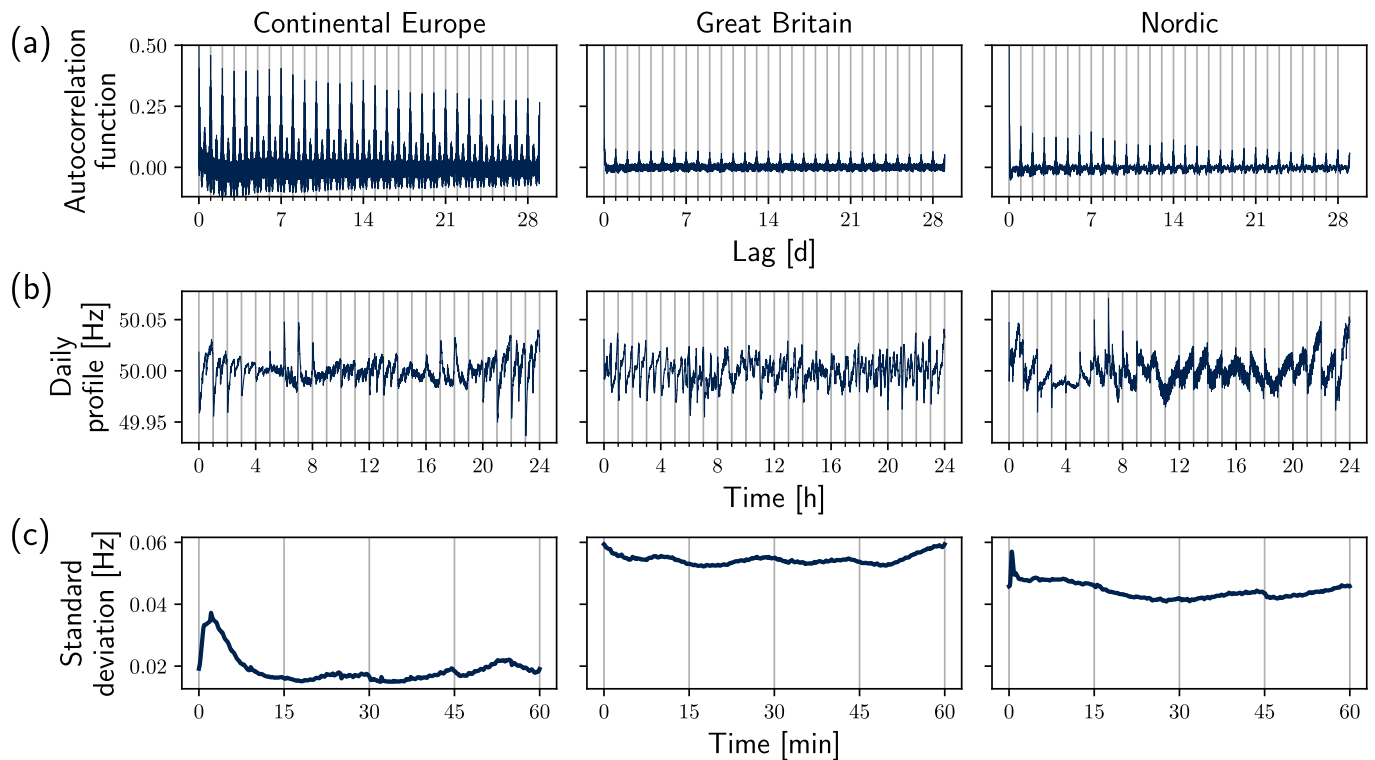
FIGURE 2. The daily profile is an important null model. (a): The autocorrelation functions show significant peaks that repeatedly occur with a period of 24 h. The one-day period thus is the main recurrence period for frequency patterns. Note that the upper limit of the y-axis has been reduced from 1 to 0.5 in order to make even small peaks visible. (b) The daily profile is the average daily pattern that recurs with a one-day (24 h) period. It is most pronounced in CE, where deterministic trading and dispatch actions play and important role. (c): The standard deviation measures the variability among all frequency samples (in the training set) at a fixed time within the hour. The larger CE area displays the lowest variability, with a clear maximum at the beginning of the hour.

Fig. 1(b) shows a frequency sample that displays these typical deterministic jumps after every 15 minute interval.

Secondly, the frequency characteristics are determined by the frequency control schemes. To assure a secure power system operation, these control measures drive back the frequency after a deviation from its reference value of 50 Hz [1]. They thus lead to a characteristic behaviour after a frequency jump or sag, which can for example be observed in Fig. 1(b). On time scales of seconds after a disturbance, the inertia of the rotating generators and the energy supplied by primary control limits the frequency deviation caused by the disturbance. Afterwards, on time scales of several minutes, secondary and tertiary control set in and restore the system to a state of stable operation at the reference frequency [1].

Finally, the frequency characteristics are influenced by other external factors that are of rather stochastic nature. Fluctuations of the demand directly affect the power balance, where demand forecasting errors [28] and large social events [29] can lead to significant unexpected frequency deviations. The variability of renewable energy sources causes additional frequency fluctuations due to its intermittency [30] or due to generation forecasting errors [31]. In summary, the frequency characteristics are thus determined by a complex mix of stochastic and deterministic processes.

C. ANALYSIS OF FREQUENCY PATTERNS
Despite its complex characteristics, the power grid frequency still exhibits regular patterns with a specific recurrence period. We identify this period by searching for regular peaks in the auto-correlation function (ACF) with time lags of up to one month (Fig. 2(a)). In all grid areas, the ACF displays regular peaks with a period of one day. Significant (but less pronounced) peaks with a period of 12 h only show up in the CE data. In CE and GB, the ACF also exhibits regular peaks with shorter periods of 15 min, 30 min and 60 min, but these peaks are much smaller than the daily peak [4]. Frequency patterns are thus most strongly correlated with patterns that occur one or multiple 24h-periods later. We conclude that the one-day period is the main recurrence time for frequency patterns within all three synchronous areas.

The average pattern that belongs to this main recurrence period is the mean daily frequency evolution, which we call the daily profile. A formal definition of the daily profile is given later in (10). The daily profiles of our three datasets exhibit some common feature but also important differences (Fig. 2(b)). All profiles show pronounced frequency jumps at the beginning of the full hour, which reflect the impact of the hourly trading interval. In particular, the CE profile displays sharp peaks of different heights, while the peaks in the GB profile are the least pronounced. The direction and height of the peaks in the CE profile are time-dependent and related to whether the demand curve is rising or falling [5]. These results are consistent with the ACFs in Fig. 2(a). There, we also observe the strongest correlation for the CE data and the lowest correlation for the GB data. The CE frequency is

thus strongly determined by regular daily patterns, while the GB frequency only exhibits weak patterns within this period.

The relevance of regular patterns for the frequency time series is further characterized by the standard deviation (StD) in Fig. 2(c). We calculate the StD for each time within the hour, i.e. the StD at 0 min is computed as the StD of all frequency recordings with time stamps $XX : 00 : 00$ averaging over all hours XX and days. In general, CE exhibits the lowest and GB the highest variability. The StD peaks after the full hour trading event in the Nordic and especially in CE areas, where the StD almost doubles after the full hour trading peak. The exact value of the full hour frequency peak thus exhibits a particularly high uncertainty.

We conclude that CE is a comparatively low-noise system with defining deterministic events that drive the standard deviation. Deterministic patterns are least pronounced in GB, such that random fluctuations are of highest importance compared to the other areas. The Nordic data is mostly in between. The differences between the grid areas can be attributed to different system properties as well as varying regulations for frequency control and market operation. For example, the low variance in the CE area is likely related to its large size [10], which provides much inertia and enables spatial balancing of nodal power mismatches. Moreover, the deadband, i.e. the frequency range without active control, is the largest in GB thus resulting in a high frequency variability [24]. Despite these differences, there is one important common result: In all three cases the main recurrence period of frequency patterns is one day. The same result was found for frequency time series from US grids [15]. This highlights the importance of the daily time scale and the corresponding daily profile for the prediction of future frequency patterns.

III. FORECASTING METHODS

A. WEIGHTED NEAREST NEIGHBOURS

The WNN method predicts future values of a time series by looking for similar patterns in the past. To predict the frequency $f(t)$ for $t \geq t_0$, we cut the historical time series into non-overlapping patterns $\mathbf{F}_n$ with γ data points and a time delay τ:

$$\mathbf{F}_n = \begin{pmatrix} f(t_0 - (n+1)\gamma\tau) \\ f(t_0 - (n+1)\gamma\tau + \tau) \\ f(t_0 - (n+1)\gamma\tau + 2\tau) \\ \dots \\ f(t_0 - n\gamma\tau - \tau) \end{pmatrix}. \quad (1)$$

The vectors $\mathbf{F}_n$ form an embedding of the time series in a space of dimension γ, which is also referred to as delay embedding in the context of time series analysis [32, Chap. 2]. To include the information of all data points, we choose a delay equal to the original time resolution of $\tau = 1$ s.

The WNN predictor searches for patterns $\mathbf{F}_n$ that are similar to the initial pattern $\mathbf{F}_0$, which ends at the prediction start t_0. However, we already know that frequency patterns mainly recur with a period of one day (Section II). Therefore, we only look for similar patterns at the same time of the day, i.e. only

within the set

$$\mathcal{F} = \{\mathbf{F}_n | \exists i \in \mathbb{N} : n\gamma\tau = i \cdot 24\text{h}\}. \quad (2)$$

From this set, we choose those patterns that are closest to the initial pattern in terms of the distance

$$\text{d}(\mathbf{F}_n) = \|\mathbf{F}_n - \mathbf{F}_0\|,$$

with $\| \cdot \|$ denoting the Euclidean distance. Given this metric, we sort the patterns as $\text{d}(\mathbf{F}_{n_1}) \leq \text{d}(\mathbf{F}_{n_2}) \leq \ldots \leq \text{d}(\mathbf{F}_{n_M})$, $M = |\mathcal{F}|$ being the total number of patterns. We then select k patterns with the smallest distance to the initial pattern and obtain the ordered set of nearest neighbours

$$\mathcal{S}_k = \{n_1, n_2, \ldots, n_k | \mathbf{F}_{n_j} \in \mathcal{F}\}. \quad (3)$$

In practice, we use the *scikit-learn* package to search and sort the nearest neighbours [33].

Finally, we assume that trajectories, which were similar in the past, will likely be similar in the future (Fig. 3). Technically, the prediction $f_p(t_0 + \Delta t)$ is therefore given by a weighted average of the trajectories succeeding the k-nearest-neighbours:

$$f_p(t_0 + \Delta t) = \frac{1}{\sum_{j=1}^{k} \alpha_j} \sum_{j=1}^{k} \alpha_j f(t_0 - n_j\gamma\tau + \Delta t). \quad (4)$$

The weights α_n are chosen to decrease with the distance $\text{d}(\mathbf{F}_{n_j})$ which introduces an additional smoothing [32, Chap. 3]. Following [18], we use a linear weighting that has the following form:

$$\alpha_j = \frac{\text{d}(\mathbf{F}_{n_k}) - \text{d}(\mathbf{F}_{n_j})}{\text{d}(\mathbf{F}_{n_k}) - \text{d}(\mathbf{F}_{n_1})}. \quad (5)$$

In practice, we apply the WNN method to predict the time steps $\Delta t \in \{1s, 2s, \ldots, T\}$ with a maximum prediction length of $T = 3600$s. A prediction with maximum length T runs for up to 13 s (on an Intel Core i5-8250U machine with 1.60 GHz processing speed and 23 Gb of RAM). Longer predictions are not relevant, since the superiority of the WNN method over the null models is mostly revealed within the first 30 minutes of the prediction (see Section IV).

B. PERFORMANCE ESTIMATION

During the optimization and evaluation of the WNN predictor, we use the Mean Square Error (MSE) as the central performance measure. In particular, we evaluate the time-dependent MSE of a general predictor $\hat{f}(t_0 + \Delta t)$ for each prediction step Δt by averaging over different starting times t_0^i:

$$\text{MSE}_{\Delta t}(\hat{f}) = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{f}(t_0^i + \Delta t) - f(t_0^i + \Delta t) \right)^2. \quad (6)$$

To select different starting times, we randomly choose $N = 5000$ different start hours h_0^i. The starting time is then given by $t_0^i = h_0^i + \Delta t_0$ where h_0^i counts the hours after the start of 2015 and Δt_0 represents a fixed starting time within the hour. In this way, we account for the frequency dynamics
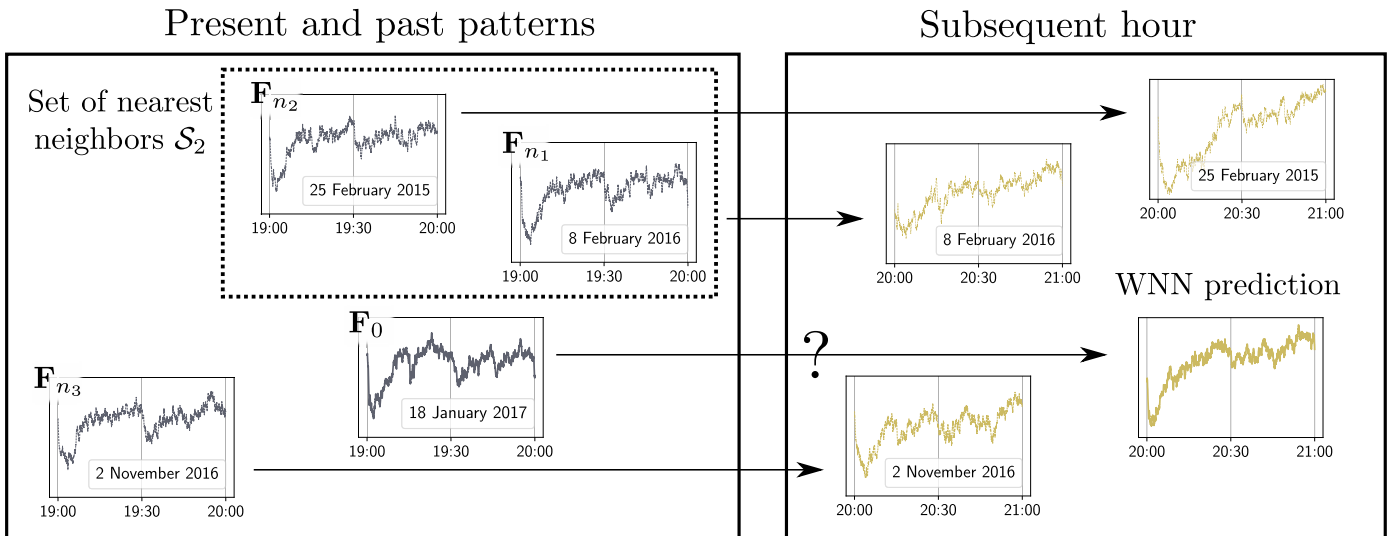
FIGURE 3. The WNN predictor searches for similar patterns in the past. To predict the future of the present (initial) pattern F_0, the WNN method looks for similar patterns F_{n_j} in the past. The patterns that are most similar to the initial pattern form the set of nearest neighbours. Here, we have chosen a set S_2 of two nearest neighbours. The average of their subsequent trajectories generates the WNN prediction.

that crucially depend on the time within the hour as discussed in Section II.

To estimate the out-of-sample performance of our predictor, we split our data into different subsets (equally for all synchronous areas). In general, the years 2015 and 2016 serve as training set, which is searched for nearest neighbours during the WNN prediction. To optimize the hyperparameters of the WNN predictor, we evaluate its MSE on a validation set that comprises the year 2017 (Section III-C). Finally, we define the year 2018 as our test set. On the test set, we compare the performance of our WNN predictor to system-specific null models (Section III-D).

C. HYPERPARAMETER OPTIMIZATION

Our WNN method exhibits two hyperparameters which are the number of nearest neighbours k and the window size (or pattern length) $\gamma\tau$. We use a window size of $\gamma\tau = 60$ min unless stated otherwise, which provides a good prediction at low computational effort. The window size is thus not explicitly optimized, but we investigate its impact on the prediction accuracy in Section IV-E.

In contrast, we strictly optimize the number of nearest neighbours k by using two different approaches. In the *fixed-k* approach, we estimate an optimal number of nearest neighbours by minimizing the time-averaged prediction error $MSE(f_p)$ of the WNN predictor f_p:

$$\text{MSE}(f_p) = \frac{1}{T}\sum_{\Delta t=1s}^{T}\text{MSE}_{\Delta t}(f_p). \qquad (7)$$

In practice, we perform a grid search on the set $\mathcal{G} = \{1, 3, 5, \ldots, 451\}$ to determine a fixed optimal value $k_{opt} \in \mathcal{G}$ for all prediction times $\Delta t \in [1s, T]$. This is how the WNN method is commonly used [18], [19]. We denote this as *fixed-k* WNN prediction.

In the *adaptive-k* approach, we minimize the time-dependent error $MSE_{\Delta t}(f_p)$ (6) for each prediction step

Δt individually. In this way, we account for the very different prediction horizons we investigate in our paper. These range from several seconds to one hour, thus making it highly probable to obtain different optimal k-values for different prediction horizons. In practice, we therefore calculate a time-dependent estimator $k_{opt}(\Delta t)$ for each prediction step Δt by performing a grid search on the set $\mathcal{G}$. To make the estimator more robust against noise, we smooth $k_{opt}(\Delta t)$ using a sliding window with a length of one minute. Finally, the *adaptive-k* WNN prediction is calculated by simply inserting a time-dependent k into (4).

D. NULL MODELS
On our test set, we compare different predictors based on their Root Mean Square Errors (RMSE), which reflects the actual frequency error in Hz:

$$\text{RMSE}(\hat{f}) = \sqrt{\text{MSE}_{\Delta t}(\hat{f})}. \qquad (8)$$

We use two easily interpretable null models to benchmark the performance of the WNN predictor. Our first trivial null model is the reference value of 50Hz, which is also the frequency mean:

$$f_m(t_0 + \Delta t) = 50\,\text{Hz}. \qquad (9)$$

Our second null model is the daily profile. In Section II, we have shown that the daily profile is the most important system-specific pattern that recurs with a period of one day. It should therefore be a benchmark model for every newly proposed frequency predictor. In practice, we calculate the daily profile predictor by averaging over all the patterns in the set $\mathcal{F}$ (from (2)):

$$f_d(t_0 + \Delta t) = \frac{1}{|\mathcal{F}|}\sum_{n\in\mathcal{F}} f(t_0 - n\gamma\tau + \Delta t). \qquad (10)$$

To make its prediction comparable to the WNN predictor, we have restricted the set $\mathcal{F}$ to patterns from the training set.
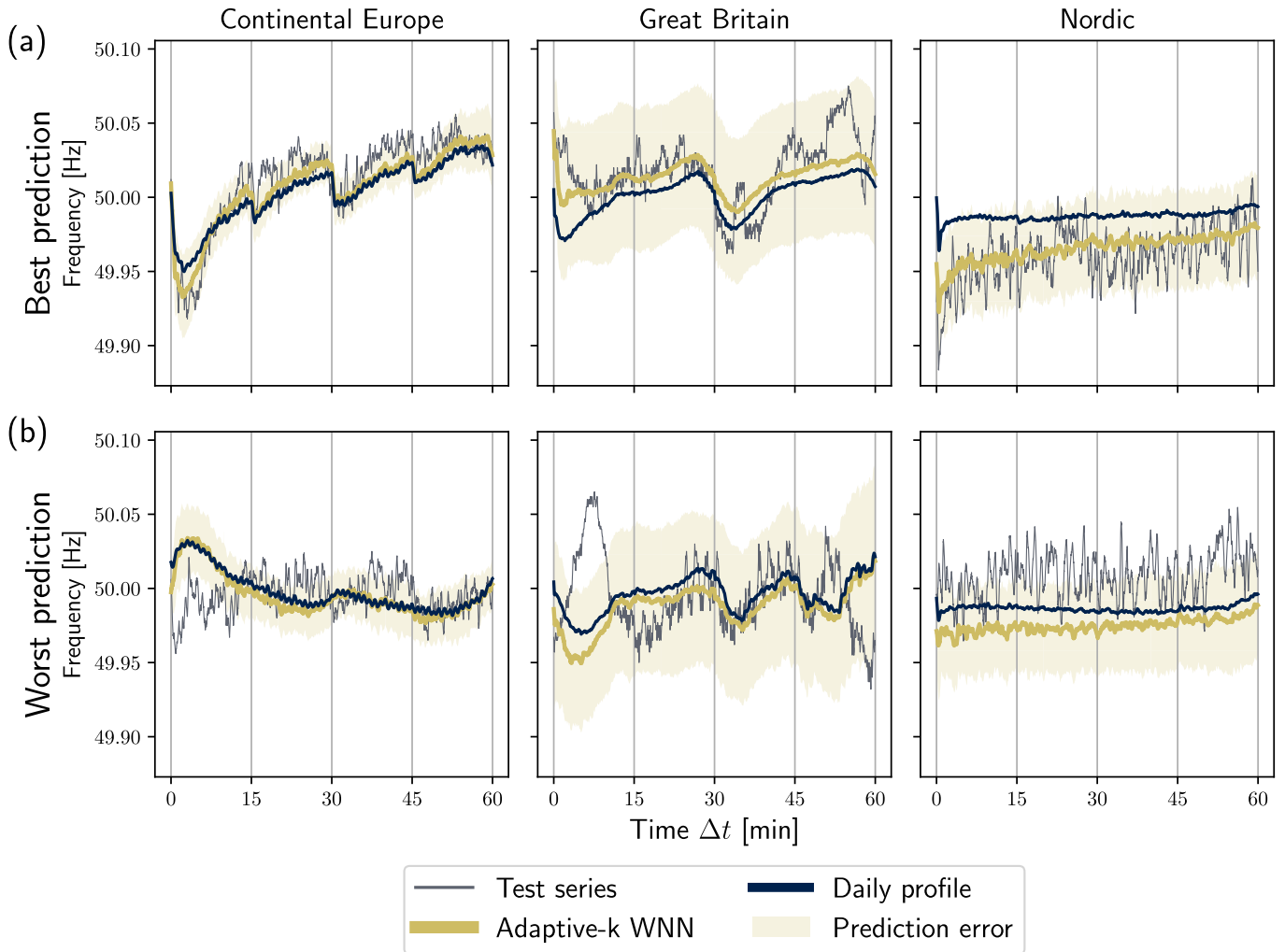
FIGURE 4. The best predictions are a smoothed version of the observed frequency trajectory. Here, we present the best (a) and worst (b) adaptive-*k* predictions from the test set. The selection is based on the relative error RMSE(*f$_p$*)/RMSE(50Hz). With that we account for the difference in variance among the samples, which would automatically result in higher or lower error values. The prediction error $\sigma_{\Delta t}$ (11) equals one standard deviation within the largest set of nearest neighbours used during the prediction. It is thus an upper bound for the standard deviation of the adaptive-*k* WNN prediction.

Note that the WNN predictor (4) converges to the daily profile predictor in the limit $k \rightarrow \infty$ when applying uniform weights.

IV. RESULTS

A. FORECAST EXAMPLES

The best and worst prediction examples give us a first impression about the performance of the WNN predictor (Fig. 4). We complement these examples with an estimate of the prediction uncertainty $\sigma_{\Delta t}$ that is based on the StD of the nearest neighbours:

$$\sigma_{\Delta t}^2 = \langle f(t_0 - n\gamma\tau + \Delta t)^2 \rangle - \langle f(t_0 - n\gamma\tau + \Delta t) \rangle^2. \quad (11)$$

Here, $\langle \cdot \rangle$ denotes the average over all $n \in \mathcal{S}_k$. For the adaptive-*k* WNN, we use $k = \max_{\Delta t} k_{opt}(\Delta t)$, which turns (11) into an upper bound for the uncertainty.

The examples indicate that the best predictions are essentially a smoothed curve of the observed frequency trajectory. The prediction is often very similar to the daily profile, but performs better especially in the first 15 minutes. Even more, the prediction uncertainty provides a good estimate for the short-term variability of the frequency trajectory.

The worst predictions in GB and CE make mistakes at the boundaries but still capture the remaining trajectory (e.g. 30-45 min in GB). In both examples, the daily profile and the WNN forecast predict the same direction for the hourly frequency jump but the observed frequency deviates in the opposite direction. The deviation indicates unforeseen events affecting the grid frequency trajectory, which are also not captured at all by the daily profile. This relation points to a potential application of time series prediction in the posteriori analysis of power system operation. A large forecasting error can serve as a tool to identify external (unforeseen) events.

Meanwhile, the worst prediction in the Nordic area stays nearly constant and the observed frequency randomly oscillates around a shifted value. This exemplifies the weak performance of the WNN predictor for unspecific patterns with strong noise.

B. PERFORMANCE OF FORECASTING METHODS

We evaluate the performance of our forecasting methods by calculating their RMSE on our test set (Fig. 5). The results show that our WNN predictor outperforms both null models in all grid areas. Its RMSE is smallest for CE and largest
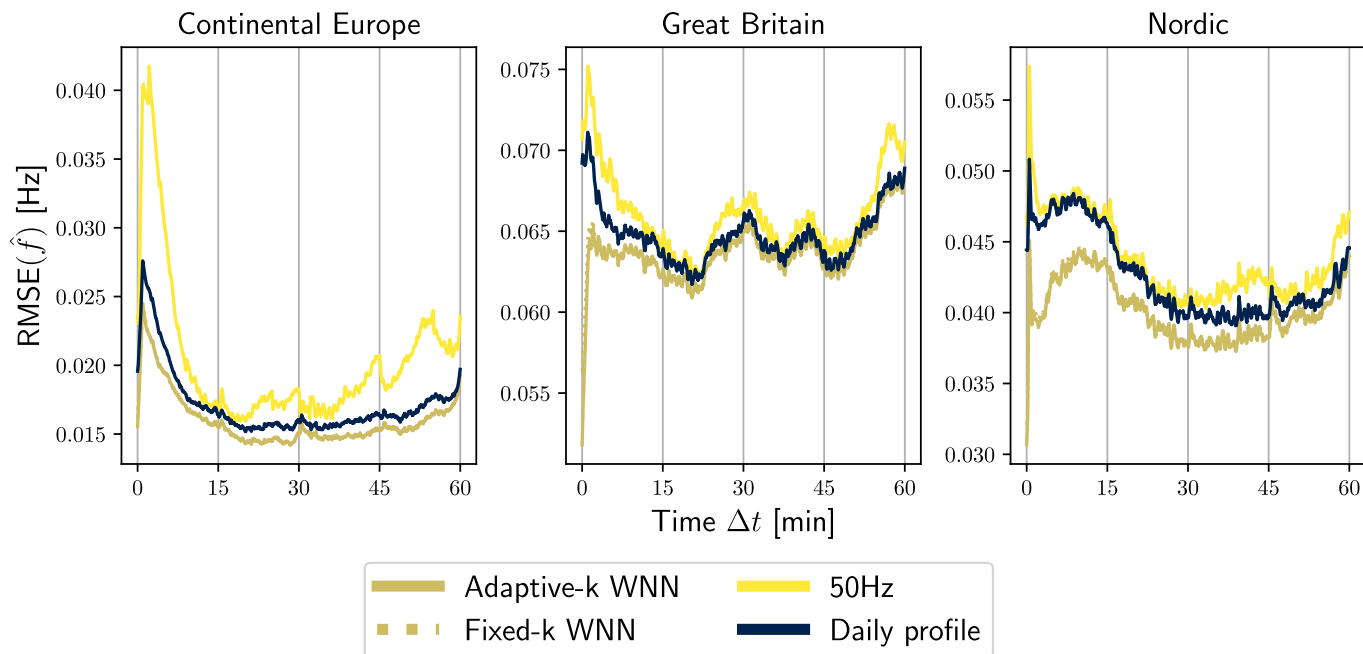
FIGURE 5. The WNN predictors outperform alternatives. The WNN predictor outperforms the null models in all three synchronous areas by returning the smallest RMSE, especially in the first 15-30 minutes. The scale of the y-axis differs between the subplots, since the GB area exhibits much larger errors than the CE area. The RMSE of the WNN predictor is further strongly time-dependent and converges to the daily profile towards the end of the prediction. Note that adaptive-k and fixed-k WNN show a very similar performance and only differ on very small time scales (Fig. 6).

for GB. This relates to Section II where we identified GB as the most stochastic and CE as the most deterministic and thus most predictable grid. The improvement of the WNN predictor relative to the daily profile is largest in Nordic (up to 30%) and smallest in CE (up to 20%). This is due to the fact that the daily profile itself is already a good predictor in CE. Meanwhile, the daily profile performs much worse in the Nordic area, where its RMSE nearly follows the 50Hz prediction error.

Comparing performance over time, we observe that the WNN outperforms the null models especially during the first 15min. As the prediction length increases, the WNN prediction converges to the daily profile. On the other hand, the performance is also clearly affected by the trading events (especially in CE). This time-dependence will be investigated in more detail in Section IV-D and IV-E.

Finally, we note that there is no significant difference between the adaptive-k and the fixed-k WNN predictor for long predictions of up to 60 minutes (Fig. 5). However, we observe a significant difference for very short prediction horizons, which we will discuss in the next section.

C. OPTIMAL NUMBER OF NEAREST NEIGHBOURS
Determining the optimal number of nearest neighbours $k_{opt}(\Delta t)$ can help to better understand the functioning of the WNN predictor. Moreover, it yields valuable information about the grid frequency dynamics in general. We present the optimization results in Fig. 6, which shows the normalized MSE landscape as a function of k and Δt as well as the optimal values $k_{opt}(\Delta t)$. The adaptive number of nearest neighbours tends to increase the more the prediction is in the future. However, the minimum is very flat at most

time steps and both the adaptive-k and the fixed-k predictor lead to very similar errors (in agreement with results from Section IV-B). We only observe a significant difference within the first minute, where the adaptive-k WNN yields up to 5% better results than the fixed-k approach. We conclude that the adaptive approach is slightly better, especially in the first 1 min. We will therefore only apply the adaptive-k WNN method throughout the rest of the paper.

As an application, we can interpret $k_{opt}(\Delta t)$ in terms of the predictability of frequency patterns. A low number of nearest neighbours corresponds to well-defined trajectories that match to some past trajectories accurately. Contrary, a higher number of nearest neighbours $k_{opt}(\Delta t)$ indicates that trajectories are rather unspecific with respect to the history. A large number of trajectories has to be averaged such that the prediction is similar to the daily profile. In particular in the first 15 minutes, the adaptive-k yields very low k values. The frequency trajectory is thus very specific in this time regime. As the prediction time increases, the optimal number $k_{opt}(\Delta t)$ rises. The trajectory thus becomes more unspecific with respect to past patterns and thus less predictable for the WNN predictor. Consistently, the WNN predictor approaches the daily profile at the end of the hour, which we obtain for $k \to \infty$.

D. IMPACT OF THE PREDICTION START
Up to now we have focused on predictions starting at full hours, such that the prediction interval coincides exactly with the main time scale of energy trading and power plant dispatch. We now widen our scope and assess the time-dependence of the WNN performance by initializing the prediction at different starting times Δt_0 (Fig. 7). To still
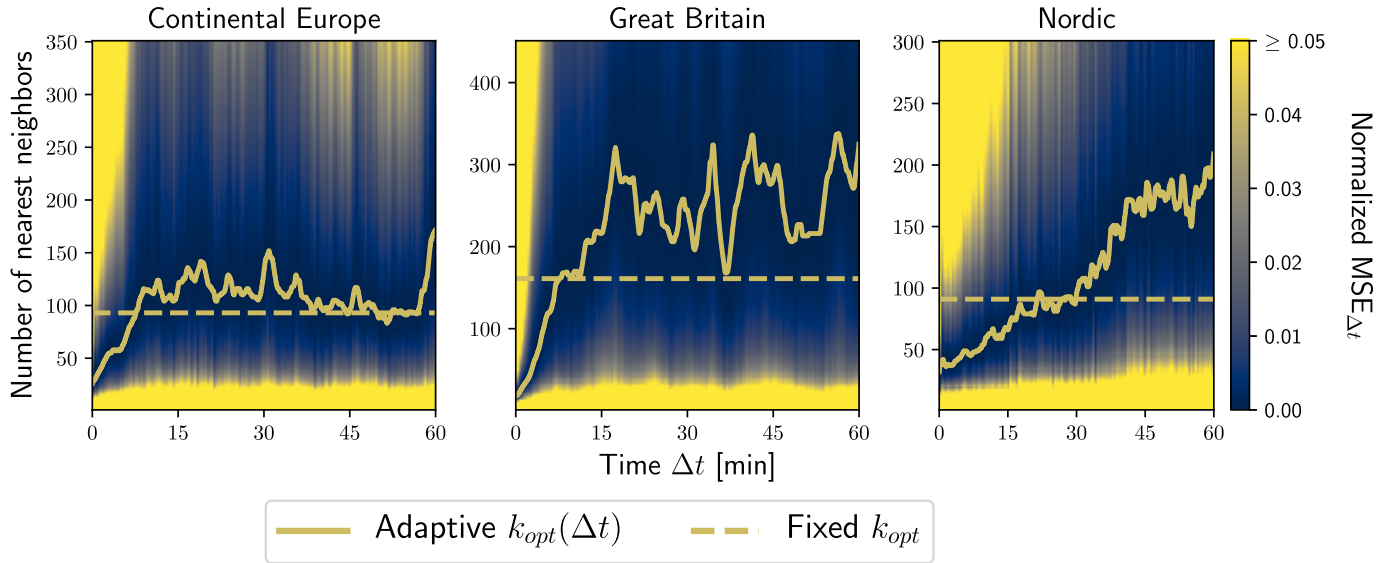
FIGURE 6. The Optimal number of nearest neighbours increases over time. To compare the error landscape for different time steps Δt, we normalize the MSE in this figure. The normalization rescales the MSE to values between zero and one for each time step Δt. The time-dependent minimum of this landscape is the adaptive number of nearest neighbours $k_{opt}(\Delta t)$. The fixed k_{opt} minimizes the aggregated MSE leading to very similar prediction errors in all but the the first minutes.

relate the WNN performance to our null models, we additionally assess its relative error RMSE(f_p)/RMSE(f_d) ("relative RMSE"), which is normalized by the daily profile error RMSE(f_d).

Irrespective of the trading events, we observe two different time regimes depending on the prediction length. During the first 15 minutes, the relative RMSE and the optimal number $k_{opt}(\Delta t)$ are increasing while still being much lower than future values. Here, the frequency dynamics exhibit specific patterns that resemble particular patterns in the past (as described in Section IV-C). This specific memory is lost over time, as the relative RMSE increases continuously during the first 15-30 minutes. In particularly in the CE and Nordic areas, one can identify two clearly distinct time scales of memory loss: Firstly, there is an initial rapid increase of the RMSE and the relative RMSE within approximately one minute. It is followed by a slower, not necessarily monotonous increase of the relative RMSE on timescales up to tens of minutes. This clear separation of time scales is especially visible when energy trading is important, i.e. at full hours being strongest in the CE area. It could be attributed to the grid inertia or to control measures that provide additional memory for a short period of time.

Finally after 15-30 minutes, the relative RMSE reaches a relatively constant level in CE and GB with values close to one. Here, the WNN prediction does not differ much from the daily profile anymore. Meanwhile, the relative RMSE and the optimal number $k_{opt}(\Delta t)$ continue to rise for up to 60 minutes in the Nordic area. Here, the memory of specific historic patterns thus reduces much slower compared to the other areas. We will come back to this effect in Section IV-E.

In addition to the prediction length, the trading events play a crucial role for the prediction. In all grid areas, the RMSE

increases strongly around the one-hour trading event. For CE and Nordic, we observe this also at 15 and 45 minutes. Around these events, the dispatch is changed abruptly, causing large frequency deviations, which are hard to forecast accurately (Fig. 2(c)). The optimal number of nearest neighbours $k_{opt}(\Delta t)$ and the relative RMSE also peak at the trading event. This indicates a lack of specific information about the trading peak and a high uncertainty connected to it. CE is a special case, as its one-hour trading jump is particularly hard to forecast. Interestingly, $k_{opt}(\Delta t)$ decreases again after the peak. The trajectory thus becomes more specific and predictable again, probably due to the control measures reacting to the disturbance in a pre-defined way.

The trading peaks have another important impact on the prediction error. After a trading event, the RMSE loses its dependence on the starting time Δt_0 and joins the error curve of earlier prediction starts. This happens in all grid areas, at latest during the one-hour trading event. In practice, it means that our prediction starting at 55 min performs approximately as well at 60 min as the one that started at 0 min. The information contained in the initial pattern thus looses its significance with the occurrence of a trading event. In other words, the trading events cause a memory loss in the frequency trajectory.

We conclude that the best WNN prediction is always obtained right after the prediction starts. On a time horizon of up to 30 min, the prediction is significantly better than the daily profile. However, this time horizon is considerably shortened if there are trading events, such as the full hour dispatches.

E. IMPACT OF THE WINDOW SIZE
We finalize the discussion of the WNN predictor by shortly investigating the impact of different window sizes. In addition
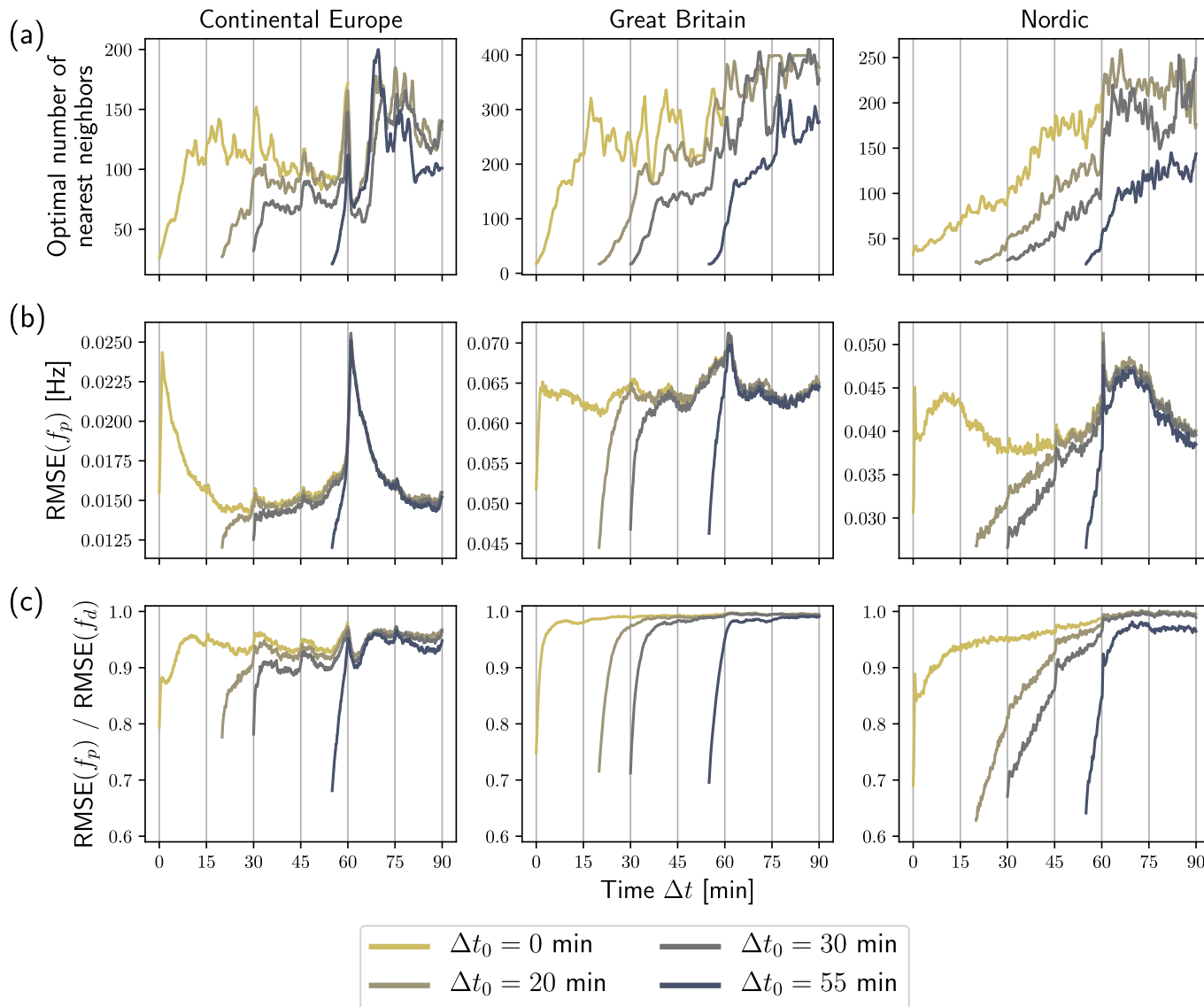
FIGURE 7. Trading events shorten the prediction horizon. Here, we show the optimal number of nearest neighbours $k_{opt}(\Delta t)$ (a), the RMSE (b) and the relative RMSE (c), which is normalized by the daily profile error. Irrespective of the starting time within an hour Δt_0, the predictions perform best within a time horizon of 15min. However, trading events introduce additional uncertainty thus increasing the prediction error and shortening the prediction horizon.

to the window size $\gamma\tau = 60$ min (which we have used throughout this article), we show the prediction errors for $\gamma\tau = 15$ min and 30 min in Fig. 8.

On time scales of several minutes to one hour, there is no significant difference between the predictors in CE and GB. The large window is slightly better than the shorter ones. In contrast, the smallest window performs best in the Nordic area especially in the first 15 minutes. Shorter windows contain more specific information about the near past than longer windows. In the Nordic grid, the significance of very specific historic patterns thus prevail much longer than in the other grids. This is consistent with Section IV-D, where we have seen that the memory of specific historic patterns reduces relatively slow in the Nordic area.

On time scales below one minute the smallest window performs best for all grid areas (inset). Shortly after the

prediction starts, the memory of the last few seconds determines the trajectory. Irrespective of the area, the shorter window thus performs best on this time scale, as it contains more specific information about about the near past of the trajectory.

We conclude that small window sizes are best for prediction horizons below one minute. For several minutes to one hour, large window sizes are slightly better in CE and GB. If computational resources are scarce, smaller window sizes can also be used here, as they are less computationally expensive but only slightly less accurate. In the Nordic area, small window sizes are the best even for several minutes. However, the performance differences are small in all grid areas, which also justifies that we did not systematically determine the optimal value for γ, thus saving computational time during training.
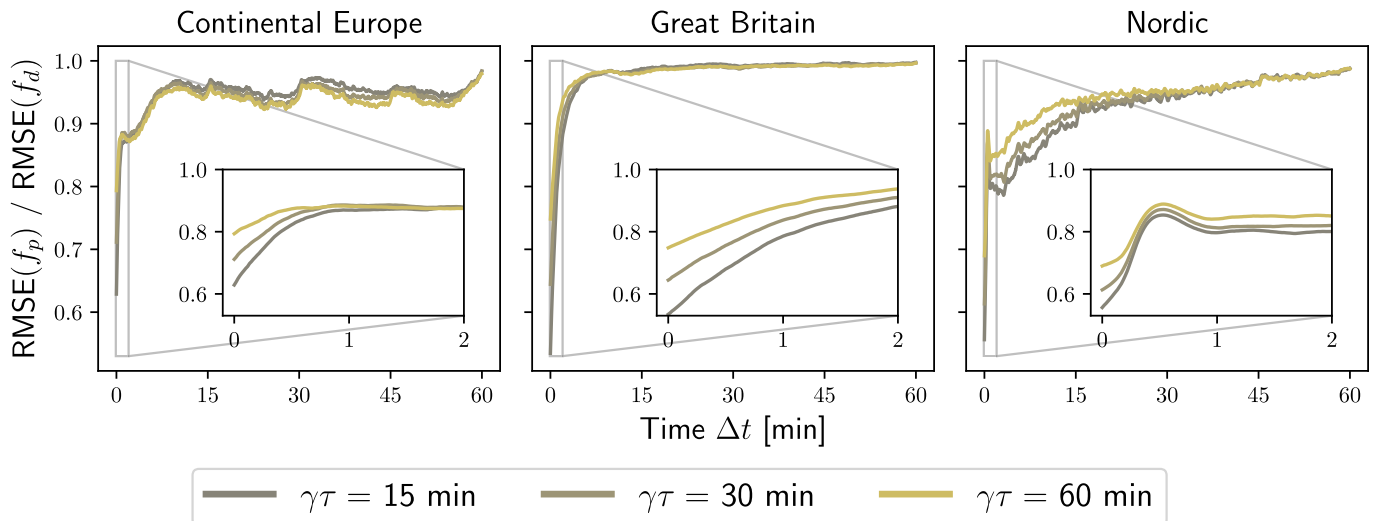
FIGURE 8. Shorter windows predict more accurately at the beginning. The optimal window size (or pattern length) $\gamma\tau$ is different depending on the prediction length. During the first minute, the shortest window performs best in all grid areas, as it contains more specific information about the near past. For several minutes to one hour, the results differ between the areas.

V. DISCUSSION

Summarizing, we have demonstrated how a k-weighted-nearest-neighbour (WNN) approach provides an accurate forecast of the power grid frequency. The predictor performs particularly well when using an adaptive number of nearest neighbours.

Compared to previously existing forecasts of the power grid frequency [9], [12], [14], [15], we make three key contributions: First, we introduce the daily profile as a relevant and system-specific null model. Secondly, we improve the statistical evaluation of the WNN predictor by increasing the amount of training and test data from one month [15] to multiple years. Thirdly, we interpret the time-dependent predictability and optimization results based on the economic and physical dynamics in the different synchronous areas. In that way, we establish machine learning techniques as valuable tools for an a posteriori assessment of power system operation and stability.

Our results can be used to improve power system stability. Since our estimates are more precise than the daily profile, they could be used to estimate necessary control power capacities. This is particularly interesting since we have a solid prediction horizon of about 60 minutes, making slower, typically cheaper forms of control available, instead of purely relying on expensive primary control [1], [2]. Especially during the first 15-30 minutes, our predictor is significantly more accurate than the daily profile and could replace it for planning purposes. Notably, this application is not restricted by computational speed as the WNN predictor only needs a few seconds to generate a forecast. Moreover, our analysis is not limited to any specific grid but can be applied to any power system, given sufficient data to train the algorithm.

We even gained valuable lessons when the predictor performed worst: The largest prediction errors are associated with unforeseen events that are also missed by the daily profile. Therefore, the introduced WNN predictor could also be used as a diagnostics tool to identify external perturbations, where for example renewable generation [34] or singular demand patterns caused by large sports events [29] impact the frequency dynamics. Furthermore, even our worst predictions correctly returned the expected average and standard deviation of the frequency time series for the next hour. Hence, the predictor could be used as a worst-case estimator to determine how much control power will be maximally necessary during the next hour to guarantee stable operation.

Finally, we went beyond pure forecasting of the next sixty minutes of the power grid frequency dynamics but instead achieved a better understanding of the different synchronous areas: Monitoring the number of nearest neighbours allowed us to distinguish deterministic and stochastic behavior of different synchronous areas but also of different time intervals. Our analysis reveals that before the electricity market acts every 15 minutes, the time series becomes less predictable but becomes more predictable after the power has been dispatched. This insight could be used to modify dispatch strategies in order to minimize the unpredictable impact on the frequency, reducing the required control power and thereby saving money.

Our results on the forecast of the power grid frequency can be extended in multiple directions in the future. Firstly, we were restricted by data availability. A similar forecast and interpretation could be developed and applied to power grid frequency time series from other regions in the world, e.g. data from the Eastern Interconnection in the US or from the Irish grid, with its high wind penetration. Secondly, additional features such as wind power generation can be included to better understand the impact of unforeseen pertubations, which are not captured in our univariate forecast. Thirdly, many alternative forecasting methods are available, from artificial neural networks (ANN) [11] and recurrent neural networks (RNN) [35] to classical methods of time series prediction [32]. However, a fully comprehensive review of

all available methods was beyond the scope of this study and will be left for the future. Finally, we are convinced that our approach to forecasting and machine learning as a tool to understand a system's dynamics should also be applied to other time series, such as renewable generation [36], air pollution [37], [38] or the stock market [39].

ACKNOWLEDGMENT

REFERENCES

[1] J. Machowski, J. Bialek, and J. Bumby, *Power System Dynamics: Stability and Control*. Hoboken, NJ, USA: Wiley, 2011.

[2] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power Generation, Operation, and Control*. Hoboken, NJ, USA: Wiley, 2013.

[3] L. R. Gorjao, M. Anvari, H. Kantz, C. Beck, D. Witthaut, M. Timme, and B. Schafer, "Data-driven model of the power-grid frequency dynamics," *IEEE Access*, vol. 8, pp. 43082–43097, 2020.

[4] M. Anvari, L. R. 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.*, vol. 2, no. 1, Mar. 2020, Art. no. 013339.

[5] T. Weißbach and E. Welfonder, "High frequency deviations within the European power system: Origins and proposals for improvement," *VGB Powertech*, vol. 89, no. 6, p. 26, 2009.

[6] F. Milano, F. Dörfler, G. Hug, D. J. Hill, and G. Verbič, "Foundations and challenges of low-inertia systems," in *Proc. Power Syst. Comput. Conf. (PSCC)*, Jun. 2018, pp. 1–25.

[7] P. Milan, M. Wächter, and J. Peinke, "Turbulent character of wind energy," *Phys. Rev. Lett.*, vol. 110, no. 13, Mar. 2013, Art. no. 138701.

[8] M. Anvari, G. Lohmann, M. Wächter, P. Milan, E. Lorenz, D. Heinemann, M. R. R. Tabar, and J. Peinke, "Short term fluctuations of wind and solar power systems," *New J. Phys.*, vol. 18, no. 6, Jun. 2016, Art. no. 063027.

[9] J. Dong, X. Ma, S. M. 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.*, vol. 29, no. 6, pp. 2602–2612, Nov. 2014.

[10] B. Schäfer, C. Beck, K. Aihara, D. Witthaut, and M. Timme, "Non-Gaussian power grid frequency fluctuations characterized by Lévy-stable laws and superstatistics," *Nature Energy*, vol. 3, pp. 119–126, Jan. 2018.

[11] E. Alpaydin, *Introduction to Machine Learning*. Cambridge, MA, USA: MIT Press, 2009.

[12] S. Kaur, S. Agrawal, and Y. P. Verma, "Power grid frequency prediction using ANN considering the stochasticity of wind power," in *Proc. 5th Int. Conf. Comput. Intell. Commun. Netw.*, Sep. 2013, pp. 311–315.

[13] H. Ma and H. Li, "Analysis of frequency dynamics in power grid: A Bayesian structure learning approach," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 457–466, Mar. 2013.

[14] A. Bolzoni, R. Todd, A. Forsyth, and R. Perini, "Real-time auto-regressive modelling of electric power network frequency," in *Proc. IECON - 45th Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2019, pp. 515–520.

[15] W. Bang and J. W. Yoon, "Forecasting the electric network frequency signals on power grid," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2019, pp. 1218–1223.

[16] S. Kollmorgen, R. H. Hahnloser, and V. Mante, "Nearest neighbours reveal fast and slow components of motor learning," *Nature*, vol. 577, pp. 526–530, Jan. 2020.

[17] J. Barkoulas, "Nearest-neighbor forecasts of US interest rates," *Int. J. Banking Finance*, vol. 1, no. 1, pp. 119–140, 2020.

[18] A. T. Lora, J. M. R. Santos, A. G. Exposito, J. L. M. Ramos, and J. C. R. Santos, "Electricity market price forecasting based on weighted nearest neighbors techniques," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1294–1301, Aug. 2007.

[19] G.-F. Fan, Y.-H. Guo, J.-M. Zheng, and W.-C. Hong, "Application of the weighted K-Nearest neighbor algorithm for short-term load forecasting," *Energies*, vol. 12, no. 5, p. 916, Mar. 2019.

[20] G. H. Chen and D. Shah, "Explaining the success of nearest neighbor methods in prediction," *Found. Trends Mach. Learn.*, vol. 10, nos. 5–6, pp. 337–588, 2018.

[21] T. GmbH. *Regelenergie Bedarf+Abruf*. Accessed: Feb. 18, 2020. [Online]. Available: https://www.transnetbw.de/de/strommarkt/systemdienstleistungen/regelenergie-bedarf-und-abruf

[22] N. G. ESO. *Historic Frequency Data*. Accessed: Feb. 18, 2020. [Online]. Available: https://www.nationalgrideso.com/balancing-services/frequency-response-s ervices/historic-frequency-data

[23] F. Oyj. *Frequency–historical Data*. Accessed: Feb. 18, 2020. [Online]. Available: https://data.fingrid.fi/en/dataset/frequency-historical-data

[24] European Commission, Directorate-General for Energy. *Commission Regulation (EU) 2017/1485 of 2 August 2017 Establishing a Guideline on Electricity Transmission System Operation*. Accessed: Feb. 18, 2020. [Online]. Available: http://data.europa.eu/eli/reg/2017/1485/oj/eng

[25] J. Kruse, B. Schäfer, and D. Witthaut. (Apr. 2020). *Pre-processed power grid frequency time series*. [Online]. Available: https://doi.org/10.5281/zenodo.3744121

[26] J. Lin and F. H. Magnago, *Electricity Markets: Theories and Applications*. Hoboken, NJ, USA: Wiley, 2017.

[27] Directorate-General for Energy (European Commission). *METIS Technical Note T4 : Overview of European Electricity Markets*. Accessed: Feb. 18, 2020. [Online]. Available: https://op.europa.eu/s/n8kA

[28] E. Carpaneto and G. Chicco, "Probabilistic characterisation of the aggregated residential load patterns," *IET Gener., Transmiss. Distrib.*, vol. 2, no. 3, pp. 373–382, May 2008.

[29] L. Chen, P. Markham, C.-F. Chen, and Y. Liu, "Analysis of societal event impacts on the power system frequency using FNET measurements," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2011, pp. 1–8.

[30] H. Haehne, J. Schottler, M. Waechter, J. Peinke, and O. Kamps, "The footprint of atmospheric turbulence in power grid frequency measurements," *EPL (Europhys. Lett.)*, vol. 121, no. 3, p. 30001, Feb. 2018.

[31] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renew. Energy*, vol. 37, no. 1, pp. 1–8, Jan. 2012.

[32] B. Schelter, M. Winterhalder, and J. Timmer, Eds., *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, 1st ed. Weinheim, Germany: Wiley-VCH, Oct. 2006.

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[34] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very short-term nonparametric probabilistic forecasting of renewable energy generation—With application to solar energy," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3850–3863, Sep. 2016.

[35] A. Haluszczynski and C. Räth, "Good and bad predictions: Assessing and improving the replication of chaotic attractors by means of reservoir computing," *Chaos: Interdiscip. J. Nonlinear Sci.*, vol. 29, no. 10, Oct. 2019, Art. no. 103143.

[36] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2011, pp. 528–533.

[37] E. Cogliani, "Air pollution forecast in cities by an air pollution index highly correlated with meteorological variables," *Atmos. Environ.*, vol. 35, no. 16, pp. 2871–2877, Jun. 2001.

[38] G. Williams, B. Schäfer, and C. Beck, "Superstatistical approach to air pollution statistics," *Phys. Rev. Res.*, vol. 2, no. 1, Jan. 2020, Art. no. 013019.

[39] Y. Zuo and E. Kita, "Stock price forecast using Bayesian network," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 6729–6737, Jun. 2012.

JOHANNES KRUSE received the B.Sc. and M.Sc. degrees in physics from the Georg-August University of Göttingen, Germany, in 2016 and 2019, respectively. During the final year of his studies, he wrote the master's thesis at the Department of Engineering, Aarhus University, Denmark. He is currently pursuing the Ph.D. degree with the University of Cologne and the Forschungszentrum Jülich, Germany.

BENJAMIN SCHÄFER received the Diploma degree in physics from the University of Magdeburg, Germany, in 2013, and the Ph.D. degree in physics from the University of Göttingen, in 2017. His research for the Ph.D. degree was carried out in Göttingen, Germany; London, U.K.; and Tokyo, Japan. He has worked as a Postdoctoral Researcher with the Max Planck Institute for Dynamics and Self-Organization, Göttingen, and Technical University Dresden, Germany. Since 2019, he has been working as a Marie Skłodowska-Curie Research Fellow with the Queen Mary University of London.

DIRK WITTHAUT received the M.Sc. and Ph.D. degrees in physics from the Technical University of Kaiserslautern, Kaiserslautern, Germany, in 2004 and 2007, respectively. He has worked as a Postdoctoral Researcher with the Niels Bohr Institute, Copenhagen, Denmark, and the Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany. He has been a Guest Lecturer with the Kigali Institute for Science and Technology, Rwanda. Since 2014, he has been leading a Research Group at the Forschungszentrum Jülich, Germany. He is currently an Assistant Professor with the University of Cologne.

● ● ●

Supplemental material for "Predictability of Power Grid Frequency"

Supplementary Material: Data preparation

———————————— ✦ ————————————

The frequency data for this paper has been pre-processed. A thorough pre-processing was necessary, as all datasets contain missing or corrupted data points and exhibit different formats and time resolutions. During the pre-processing we uniformly mark missing and corrupted data points, which offers the opportunity to individually fill, interpolate or exclude these invalid measurements. The whole process consists of two steps (Section 1) and exhibits some free parameters that are fixed based on the data (Section 2). Our code and the ready-to-use pre-processed data are available online [1].

Our time series include frequency measurements from the Continental Europe (CE), the Great Britain (GB) and the Nordic synchronous areas. The raw data is described in Table 1. By the 18th February 2020, the data has been publicly available on the websites of different Transmission System Operators (TSO). We follow the naming convention established in [2] and use "area" synonymous with "synchronous area".

1 PRE-PROCESSING PROCEDURE

1.1 Convert data to common format

The data contains time stamps with varying formats and Daylight Saving Time (DST) changes. We thus process the time stamps and convert them to a uniform format. We point out that we use local time including the DST changes in the processed data. This is relevant for our weighted-nearest-neighbour prediction since we compare frequency patterns based on their time stamp. By using the local time, we account for other important socio-economic patterns (such as the load) that rather follow local time than UTC time. In addition, we convert the time series to a common resolution of one second. We therefore resample the Nordic data by averaging it in non-overlapping windows of 10 data points. Finally, we mark missing time stamps by inserting a NaN-value (Not a Number) into the time series.

1.2 Mark and clean corrupted data

We search for different types of corrupted data. In particular, we identify isolated peaks, too high or low frequency values and too long windows with constant frequency. The identification is based on the frequency $f(t)$ or their increments $\Delta f(t) = f(t) - f(t - 1)$ and follows these definitions:

- A time stamp t contains an isolated peak, if Δf_t and Δf_{t+1} have the opposite sign and are too large, i.e. $|\Delta f_t|, |\Delta f_{t+1}| > \Delta f_c$.

- Following [6], frequencies below 49Hz and above 51Hz are consider as unrealistic and thus as too low (and too high).
- We define a too long constant window as a time intervals of length $T > T_c$ that exhibits increments $|\Delta f| < 10^{-9}$ Hz.

These three types of data points are marked and converted to NaN-values. This offers the possibility to apply custom cleaning methods to the data (for example interpolation or data exclusion). Here, we clean the data by filling (at maximum) N_f NaN-values with the last valid frequency value. During the prediction, we exclude frequency patterns with remaining NaN-values from the simulations.

2 CHOICE OF PRE-PROCESSING PARAMETERS

The above procedure exhibits three free parameters. We select the parameters Δf_c, T_c and N_f in the following way.

- Isolated peaks mainly occur in CE data (Table 2). The increment histogram for CE (Fig. 1(a)) shows minima at $\Delta f = \pm 0.05$ and separated maxima beyond this values. That indicates that the increments beyond this threshold are caused by another process than the regular stochastic frequency movement. The frequency sample in 1(b) also shows that $\Delta f_c = 0.05$ would include most of the isolated peak. The GB and Nordic area do not show as many isolated peaks in their data (Table 2). A manual inspection of the isolated peaks with $\Delta f_c = 0.05$ reveals that no regular data points are accidentally marked as corrupted in the GB and Nordic data. We therefore keep the choice $\Delta f_c = 0.05$.
- We allow intervals of constant values below a length of $T_c = 1$ min. We mainly predict the frequency on time intervals of several minutes, so constant windows with $T < 1$ min are negligible.
- We choose $N_f = 6$ since most of the corrupted CE and GB data can be cleaned in that way (Fig. 2). At the same time, we do not manipulate the time series too much on time scales relevant for our prediction. Note that a large part of the marked data in the Nordic time series will not be filled by $N_f = 6$s. This is mainly due to the large amount of missing data within the Nordic frequency recordings (Table 2).

Using the procedure outlined above, we obtain clean data to be used for training and validation, see also [1], where the clean data can be downloaded.
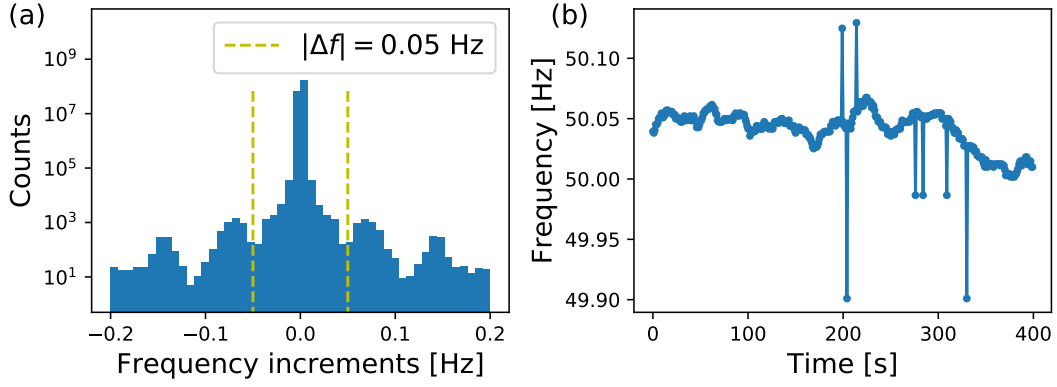
Figure 1. Isolated peaks exhibit increments larger than $\Delta f_c = 0.05$ Hz. The Historgram (a) of the frequency increments in the CE data shows two minima for $\Delta f_c = \pm 0.05$ Hz. The frequency sample (b) from CE confirms that unrealistic isolated peaks exhibit frequency increments larger than 0.05 Hz. Both observations suggest the choice of $\Delta f_c = 0.05$ Hz as a threshold for isolated peaks.
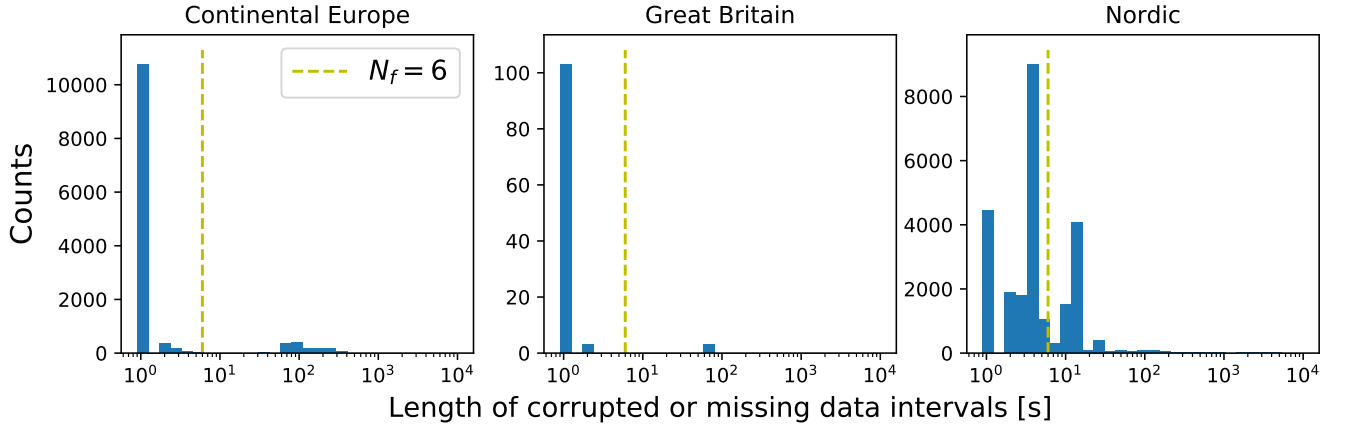

Figure 2. Missing and corrupted data points are marked as NaN-values. As a result, there are many intervals of NaN-values in the processed data. This figure displays the histogram for the lengths of these segments. CE and GB exhibit a particularly large amount of intervals with lengths below 6 time steps. Filling $N_f = 6$ values thus eliminates many corrupted or missing data segments while not changing the data too much.

Table 1
Raw frequency recordings are available from different TSOs. For each synchronous area, we obtained publicly available frequency data from one of the Transmission System Operators (TSO). TransnetBW is a German TSO, while Nationalgrid and Fingrid are British and Finish, respectively. Although we have only used the years 2015-2018 in this paper, we have pre-processed a longer period of time.

TSO (synchronous area)	Time resolution [s]	Pre-processed period
TransnetBW (Continental Europe) [3]	1	2012-02-01 to 2019-08-31
Nationalgrid (Great Britain) [4]	1	2014-01-01 to 2019-07-01
Fingrid (Nordic) [5]	0.1	2015-01-01 to 2019-11-30

Table 2
The number of corrupted data points differs between the synchronous areas. We mark different types of corrupted data points according to the definition in Section 1 and the parameters from Section 2. Note, that we apply this procedure after resampling the data. This table thus shows the number of corrupted and missing data points (or intervals) for the data *after* step (a) in Section 1.

Synchronous area	Isolated peaks	Too high frequency	Too low frequency	Long constant windows	Missing data intervals
CE	12229	10	8630293	1438	21
GB	6	0	0	3	0
Nordic	109	0	0	75	25103

REFERENCES

[1] J. Kruse, B. Schäfer, and D. Witthaut, "Pre-processed power grid frequency time series," https://doi.org/10.5281/zenodo.3744121, Apr. 2020.

[2] European Commission, Directorate-General for Energy, "Commission Regulation (EU) 2017/1485 of 2 August 2017 establishing a guideline on electricity transmission system operation," http://data.europa.eu/eli/reg/2017/1485/oj/eng, accessed 2020-02-18.

[3] TransnetBW GmbH, "Regelenergie Bedarf + Abruf," https://www.transnetbw.de/de/strommarkt/systemdienstleistungen/regelenergie-bedarf-und-abruf, accessed 2020-02-18.

[4] National Grid ESO, "Historic frequency data," https://www.nationalgrideso.com/balancing-services/frequency-response-services/historic-frequency-data, accessed 2020-02-18.

[5] Fingrid Oyj , "Frequency - historical data," https://data.fingrid.fi/en/dataset/frequency-historical-data, accessed 2020-02-18.

[6] B. Schäfer, C. Beck, K. Aihara, D. Witthaut, and M. Timme, "Non-gaussian power grid frequency fluctuations characterized by lévy-stable laws and superstatistics," *Nature Energy*, vol. 3, January 2018.

3. Explainable machine learning of the grid frequency: the effect of techno-economic drivers

Techno-economic impact factors play an important role for grid frequency dynamics. For example, electricity trading affects the grid frequency at regular intervals, thus impeding a univariate forecast (Section 2.1). In the following sections (3.1-3.3), I present machine learning models to infer the impact of techno-economic features on grid frequency stability and control. To explain the black-box models, I used SHapely Additive eXplanation (SHAP) values [72], which allow us to quantify the importance of features and visualise their impact on the model prediction in a consistent way. The results revealed central drivers of frequency stability (two publications in Sections 3.1 and 3.2) and control activation (one publication in Section 3.3).

First, I investigated four important indicators of frequency stability and their relation to techno-economic impact factors (Section 3.1). My SHAP analysis indicated the importance of forecast errors for the Nordic grid frequency, while renewable generation was important in Great Britain and fast generation and load ramps dominated in Continental Europe. Furthermore, dependency analyses revealed opposite impacts of different generation types, which I explained with their different technical characteristics. I contributed to this work by conducting the research, creating the figures, designing the major part of the research and writing the major part of the manuscript.

Second, I took a closer look at one of the stability indicators that quantifies the deterministic frequency deviations (DFDs) due to electricity trading (Section 3.2). As described in Section 1.2, DFDs are strongly connected to the load gradient: if the load rises the frequency rises at the beginning of trading intervals, while it drops if the load decreases. Using SHAP values, I demonstrated that this view is incomplete and revealed the importance of solar ramps and local legislation, thus leading to a better prediction of DFDs in Continental Europe. For this publication, I designed and conducted the research, created the figures and wrote the major part of the manuscript.

Third, I turned to frequency control and studied the impact of techno-economic features on secondary control activation in Germany (Section 3.3). The results again point to the role of different generation types, e.g., the role of slow generation changes of nuclear power stations that can induce temporary power imbalances. Furthermore, I demonstrated

that SHAP values do not always reveal causal relationships, and proposed a smart feature selection to mitigate this problem. Beyond analysis, I also demonstrated how my approach could be used as a transparent forecasting tool for control activation. For this work, I conducted the research, created the figures and wrote half of the paper.

3.1. B) Revealing drivers and risks for power grid frequency stability with explainable AI

[2] Kruse, J., Schäfer, B. & Witthaut, D. Revealing drivers and risks for power grid frequency stability with explainable AI. *Patterns* **2,** 100365. doi:10.1016/j.patter.2021.100365 (2021).

Article

Revealing drivers and risks for power grid frequency stability with explainable AI

Johannes Kruse,[1,2,5,*] Benjamin Schäfer,[3,4] and Dirk Witthaut[1,2]
[1]Forschungszentrum Jülich, Institute of Energy and Climate Research - Systems Analysis and Technology Evaluation (IEK-STE), 52425 Jülich, Germany
[2]Institute for Theoretical Physics, University of Cologne, 50937 Köln, Germany
[3]School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, UK
[4]Faculty of Science and Technology, Norwegian University of Life Sciences, 1432 Ås, Norway
[5]Lead contact
*Correspondence: jo.kruse@fz-juelich.de
https://doi.org/10.1016/j.patter.2021.100365

THE BIGGER PICTURE The transition to a sustainable energy system is challenging for the operation and stability of electric power systems as power generation becomes increasingly uncertain, grid loads increase, and their dynamical properties fundamentally change. At the same time, operational data are available at an unprecedented level of detail, enabling new methods of monitoring and control. To fully harness these data, advanced methods from machine learning must be used.

In this paper, we present explainable artificial intelligence (XAI) as a tool to quantify, predict, and explain essential aspects of power system operation and stability in three major European synchronous areas. We focus on the power grid frequency, which measures the balance of generation and load and thus provides the central observable for control and balancing. Combining XAI with domain knowledge, we identify the main drivers and stability risks, while our model and open dataset may enable further XAI research on power systems.

1 2 3 4 5 **Proof-of-Concept:** Data science output has been formulated, implemented, and tested for one domain/problem

---

SUMMARY

Stable operation of an electric power system requires strict operational limits for the grid frequency. Fluctuations and external impacts can cause large frequency deviations and increased control efforts. Although these complex interdependencies can be modeled using machine learning algorithms, the black box character of many models limits insights and applicability. In this article, we introduce an explainable machine learning model that accurately predicts frequency stability indicators for three European synchronous areas. Using Shapley additive explanations, we identify key features and risk factors for frequency stability. We show how load and generation ramps determine frequency gradients, and we identify three classes of generation technologies with converse impacts. Control efforts vary strongly depending on the grid and time of day and are driven by ramps as well as electricity prices. Notably, renewable power generation is central only in the British grid, while forecasting errors play a major role in the Nordic grid.

INTRODUCTION

The power grid frequency plays a central role for power system control, as it reflects the balance of power generation and demand.[1] An oversupply of power leads to a frequency increase, while a shortage causes a frequency decrease. Large frequency deviations correspond to large power imbalances, which

threaten system stability and may lead to large-scale blackouts.[2] Frequency stability is regarded as a major challenge for the transition to a sustainable energy system because renewable power sources do not provide an intrinsic inertia.[3] Understanding the emergence of large frequency deviations is therefore essential.

Deviations from the reference frequency of 50/60 Hz have distinct causes, which are in turn modified by the complex

interplay of different elements of the energy system. For example, changes in power generation due to electricity trading intervals causes regular frequency jumps,[4] the magnitude of which depends on several technical parameters.[3,5] Fluctuating wind and solar power[6,7] or singular load patterns due to societal events[8] create frequency fluctuations on different scales. To guarantee frequency stability in such a complex and uncertain environment, transmission system operators (TSOs) intensively monitor the system and allocate expensive control reserves as necessary. An improved understanding of the frequency dynamics and external influences could greatly facilitate control efforts and contribute to power system stability. While several studies have investigated univariate correlations to quantify the impact of individual features,[9–11] a comprehensive, data-based analysis is lacking.

Modern machine learning (ML) methods are well suited to this task as they can handle a large number of features and large volumes of data. In recent years, the volume of publicly available energy system data has grown steadily, including frequency recordings[12,13] and data on a variety of external features, such as generation and load time series.[14,15] An optimal basis for analyzing and predicting grid frequency with data-driven models therefore already exists.[16] However, complex ML models do not provide insights on the mapping of input to output.[17,18] This is particularly problematic for critical infrastructures such as power systems, where the black box character poses a security risk.[19,20]

Approaches using explainable artificial intelligence (XAI) could change this. XAI is a quickly growing research field, which covers inherently transparent ML models as well as post-modeling explanations for black box models.[21] Shapley additive explanations (SHAP) values are an example of post-modeling explanations, offering a method of measuring feature effects and avoiding inconsistencies present in other approaches.[22,23] In particular, SHAP values have certain desirable properties, such as additivity, efficiency, and symmetry. SHAP values can be quickly computed for gradient boosted trees,[24] which in turn offer a powerful nonlinear modeling and are particularly suited to tabular data. The combination of tree-based models and SHAP values is already widely used, with applications ranging from medicine[25] to geoscience.[26] In contrast, only a few applications of this methodology have been presented in the field of energy systems analysis to date: for example, to explain solar power forecasts,[27] transient security assessments,[28] or power project failures.[29]

Here, we present an explainable ML model based on gradient boosted trees for selected indicators of frequency stability, and we evaluate its predictive power for three grids in Europe: Continental Europe (CE), the Nordic area, and Great Britain (GB). We demonstrate the benefits of explainability via SHAP values, ranging from coarse-grained global feature importances to detailed dependencies and finally to fine-grained interactions between different external features. In particular, we quantify the impacts of generation and load ramps on the rate of change of frequency (RoCoF) at the beginning of each hour. SHAP values explain the different impacts and roles of different generation technologies. We use aggregated SHAP values to analyze efforts to control generation, which vary strongly depending on the grid and time of the day. We then investigate enduring frequency deviations, which can be attributed to systemic power imbalances, and discuss the role of solar power generation. As data, we utilize the hourly time series of four stability indicators (model outputs or targets) and 66 external features (model inputs) for the years 2015–2019 (see also our Zenodo[30] repository).

Our approach complements established simulation-based methods that predict frequency deviations on the basis of load and generation forecasts. Although simulations can be very accurate, they are reliant on the quality of input data, underlying forecasts, and specific parameters. Data-driven models can reveal additional driving factors, unknown effects, and emerging risks and thus complement and improve existing simulations. For instance, our analysis highlights the role of forecasting errors, which varies depending on the grid.

The next two sections of this paper present the four frequency stability indicators and our ML model. Then, the most important features in each synchronous area are identified before the influence on generator ramps—in particular, on RoCoF predictions—are discussed and nonlinear feature dependencies are revealed. We go on to demonstrate how SHAP analysis reveals feature interactions before concluding with a discussion.

RESULTS

Frequency stability: Indicators and influences

The power grid frequency fluctuates on various timescales, ranging from seconds to weeks.[31] In our model, we aggregated frequency deviations to hourly indicators, which are directly relevant for power system stability (Figure 1; experimental procedures). We analyzed the maximum frequency deviation within the hour (nadir)[32] and the RoCoF,[32] which are of central relevance for grid monitoring and control. Nadirs above a threshold level indicate immediate danger and can be counteracted with measures such as load shedding. High RoCoFs are dangerous because control actions require a few seconds to take effect. In addition, we evaluated two integrated stability measures to account for the duration of frequency deviations. We characterized the variability of hourly time series using the mean square deviation (MSD) from 50 Hz. The MSD also indicates the total (primary) control effort, meaning that a large MSD reflects high operational costs.[33] Finally, we evaluated the integrated frequency deviation (integral), which is proportional to the mean deviation within the hour. Large integrals correspond to a systematic imbalance between the hourly power generation and the demand. Regional differences in the grid frequency within a synchronous area are small during normal operation and are typically damped out after several seconds.[34,35] Although we used local grid frequency measurements, the above indicators characterize frequency stability in an entire synchronous area.

We evaluated these four indicators on an hourly basis, as this timescale is central for power system operation.[37] Electricity is traded predominantly in blocks of one hour, and generation is adapted at the beginning of each hour, leading to deterministic patterns in frequency.[4] When the load decreases continuously during an hour, but the dispatch is set to the hourly mean, then power is scarce at the beginning of the hour and the frequency drops. As a consequence, frequency deviations show a pronounced daily profile, which we use later as a null model to evaluate prediction performance. Another reason for choosing to
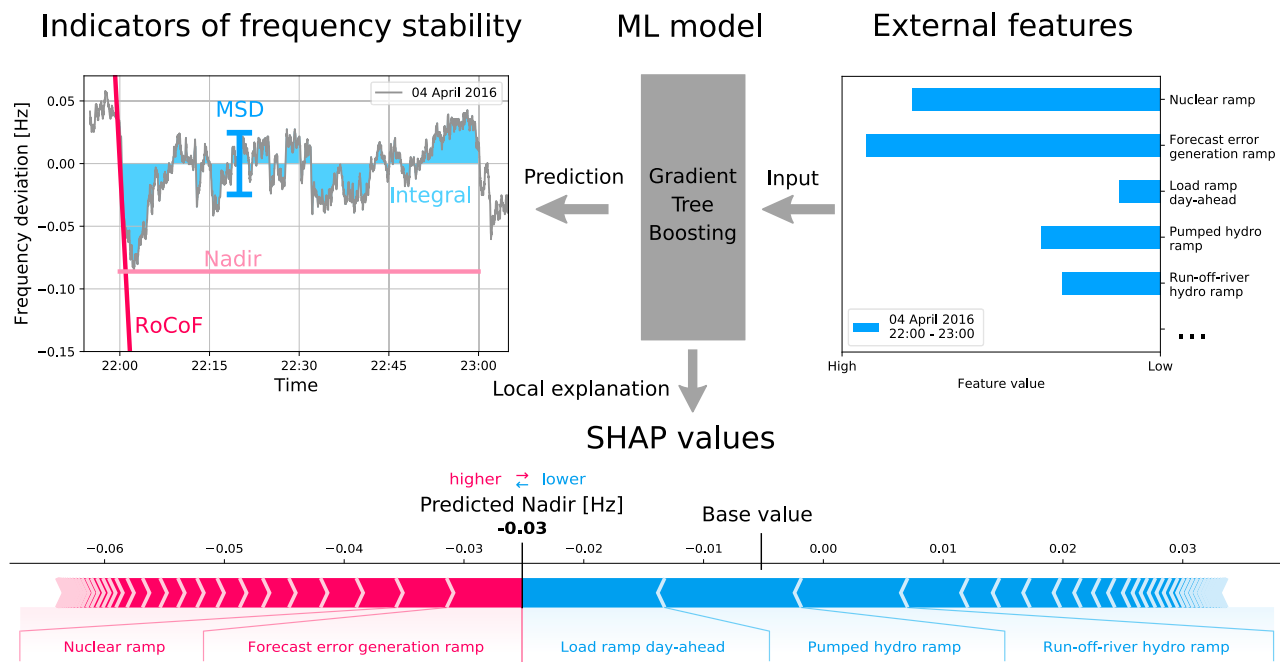
Figure 1. Overview of our explainable ML model
From right to left: using publicly available external features from the ENTSO-E transparency platform,[36] such as load ramps or generation ramps, a gradient tree boosting ML model was constructed to predict indicators of frequency stability. The model was then interpreted using SHAP values, which quantify the effect of the features on the model output in relation to a base value (see experimental procedures). The SHAP value of each feature is shown at the bottom of the figure, separated by white arrows; only the most important features are named. Together with the base value, positive (red) and negative (blue) SHAP values add up to the model prediction. The data represent a sample hour in 2016 from Continental Europe.

evaluate on an hourly basis is that most external features are only publicly available at an hourly resolution.[14]

The assessment of frequency stability indicators conventionally focuses on the transient response after a major disturbance.[1] Many model-based simulation studies have investigated the effects of various parameters on the frequency response, in particular the effect of inertia,[38] as well as effects on the properties of the load-frequency control system.[10] In recent years, ambient and deterministic frequency fluctuations have received more attention in the context of model-based simulations. Studies have highlighted the influence of inertia, control system parameters,[5] and intermittent wind power feed-in[39] on the frequency statistics. Deterministic frequency deviations (DFDs) have been studied using dynamical models[4] and stochastic models[40] revealing the importance of the daily load evolution and generation jumps caused by electricity trading. The main limitation of the simulation approach is that data and parameters are often not publicly available to accurately model all interactions within the power system. For example, load-frequency control systems are operated by individual TSOs, and parameters may have been disclosed to other TSOs.[41] Thus, simplified assumptions are used, which often do not reflect effects present in real-world data.

Over the last few years, comprehensive datasets have become publicly available, enabling an *empirical analysis* of power system frequency stability.[13,35] Most data-driven studies focus on the impact of a single isolated feature and resort to a linear correlation analysis. For instance, studies have quantified

the correlations between different measures of frequency quality and the load value and ramps in the Nordic grid,[11] wind power generation in the Irish grid,[9] load ramps in the British grid,[10] and societal events coinciding with large frequency deviations.[8] A correlation between load and solar ramps as well as trading volumes reflects the role of solar power in power balancing.[42] The relation between wind power generation and large frequency increments in the CE grid has been studied using conditional probabilities by Haehne et al.[7] Although existing studies provide us with essential insights into power system operation and frequency stability, they are limited in two ways. Firstly, linear correlation analyses cannot capture nonlinear dependencies and may thus underestimate or even overlook important effects. Secondly, only one feature/covariate is used in most cases, and the observed effects are not adjusted for other variables. This is problematic when features are correlated, e.g. due to confounding. Modern ML methods can capture multiple dependencies and thus provide more accurate results.[16]

An explainable model for frequency deviations

We developed an explainable ML model to predict indicators of frequency stability from external features (Figure 1; experimental procedures). We used gradient tree boosting (GTB), which produces nonlinear models with state-of-the-art performance for many ML applications[43] while enabling a fast and efficient calculation of SHAP values to explain the predictions.[24] We fed our model with physically meaningful features based on load, generation, and electricity price time series. Our data included both

day-ahead available features, such as the day-ahead predicted load change ("load ramp day-ahead") and ex post available features, such as the error between the day-ahead forecast and the actual total generation change ("forecast error generation ramp"). Finally, we computed SHAP values to quantify how each feature contributes to the model output. For example, in Figure 1 (bottom), the feature "load ramp day-ahead" has a negative contribution (blue), thus causing the predicted nadir to be lower than its average. SHAP values make local predictions more transparent and enable aggregated insights into global feature effects, dependencies, and interactions. However, it should be noted that SHAP values do not guarantee causal relations (see experimental procedures for a more detailed discussion).

Based on its R^2 score, our model outperformed the daily average profile of the stability indicators, which is an important system-specific null model (experimental procedures and supplemental experimental procedures S5). We achieved performances 3.7 (CE), 7.6 (Nordic), and 16.3 (GB) times higher than the daily profile, thus indicating additional important dependencies. Restricting the full model to day-ahead available features resulted in similar performance gains of 2.6 (CE), 3.0 (Nordic), and 8.9 (GB), which opens the possibility of predicting stability indicators a day ahead. The ability to include ex post available features, such as forecast errors, was particularly beneficial in the Nordic area. Here, the full model performed 2.6 times better than the restricted day-ahead model. This indicates the importance of forecast errors for the Nordic frequency dynamics, which we examine in the next section.

Main features affecting frequency deviations

We demonstrated our model explainability on the coarse-grained level of global feature importances, which characterize how much a certain feature affects the hourly frequency stability indicators within the trained model (Figure 2).

In the RoCoF model, only a few features dominated: mainly generation ramps from hydropower and load ramps. The importance of hydropower generation ramps relates to their large ramping speed, which we discuss below. In the Nordic area, the total day-ahead generation ramp is much more important than load ramps for the RoCoF. This suggests that changes in power export and storage may be relevant here, as these are not represented in the load for the area.

The nadir was primarily affected by ramps and their respective forecasting errors. In CE, the day-ahead load ramp was the most important feature. This reflects the importance of DFDs, which are strongly correlated to the direction of the load ramps.[4] In the Nordic grid, the forecast errors of generation and load ramp were by far the most important features, partly explaining the large performance gain when ex post data were included in the model (see above). In contrast, there were many features of almost equal importance in the British nadir model. Here, wind power ramps and solar ramp forecast errors were among the five most important features. This indicates the importance of renewables for frequency deviations in GB.

The MSD behaved similarly to the nadir in CE and in the Nordic grid, with some subtle differences. Load ramps were the most important feature in CE. Forecasting errors again dominated in the Nordic grid, but load and hydropower generation ramps

also played a role. A different situation was found in GB. Day-ahead prices dominated the MSD prediction, with some generation forms (coal and nuclear) coming in at a distant second, while generation ramps did not significantly contribute. These differences point to a more complex behavior of the MSD, which we further discuss below.

Finally, the integral was largely affected by forecasting errors for load and generation ramps, which caused long-lasting power mismatches. This was particularly evident in the Nordic grid, where other features were not as important. In GB, wind power ramps were ranked highly, confirming the importance of renewables. In CE, solar power generation and ramps, as well as nuclear power ramps, were relevant for the prediction. We investigated how the interaction of these two distinct generation types explain model variance.

In summary, CE exhibited strong DFDs that were connected to hourly load and generation ramps. This is consistent with previous results[4,41] (Figure S10). Nordic frequency deviations were strongly connected to forecasting errors, which is in line with Nordic TSOs reporting forecast errors as a driver.[44] In GB, hourly DFDs were less important (supplemental experimental procedures S4) and frequency deviations were mainly affected by renewables, i.e. their fluctuations and forecast errors (cf. Nationalgrid ESO[45]). The total synchronous generation, which approximates the inertia in our model (experimental procedures), affected the British frequency dynamics only in extreme situations where there was very low inertia (supplemental experimental procedures S6). Despite the importance of reduced inertia in renewable energy systems,[3,45] our model showed that the average effect of inertia on the aggregated stability indicators is negligible (Figure 2). This was consistent with other studies on aggregated frequency fluctuations (cf. Vorobev et al.[5]), which found that inertia is important for extreme events but aggregated dynamics are not. It should be noted that we focused on frequent daily fluctuations and stability concerns, which are highly relevant for TSOs and for reducing daily operational costs.[33] This supplements studies focusing on blackouts and cascading failures.[46]

Characterizing the effect of generation ramps

Fast generation ramps significantly affect the hourly RoCoF. For this reason, we went beyond mean feature importances and examined the direction of these dependencies using SHAP values (Figure 3). The effect of ramps is mostly monotonic, meaning that a feature effect either increases or decreases monotonically with the feature value (Figures 3A–3C). Remarkably, the direction of the dependency varies depending on the type of generation and the grid. As expected, hydropower generation ramps were consistently positively correlated (see Figure 3D for CE). The dependency of hard coal ramps for CE was the opposite to the dependency for GB and the Nordic grid.

The observed differences between the generation types can be explained in terms of the *relative* ramping speed of a generation type within a respective area (see experimental procedures on how this speed is estimated). In the Nordic grid, hydropower, a technology capable of fast ramps, is essential and all other generation types must be considered slow in comparison. In GB and CE, non-hydropower dominates the generation mix and technologies with slower ramps than hydropower plants
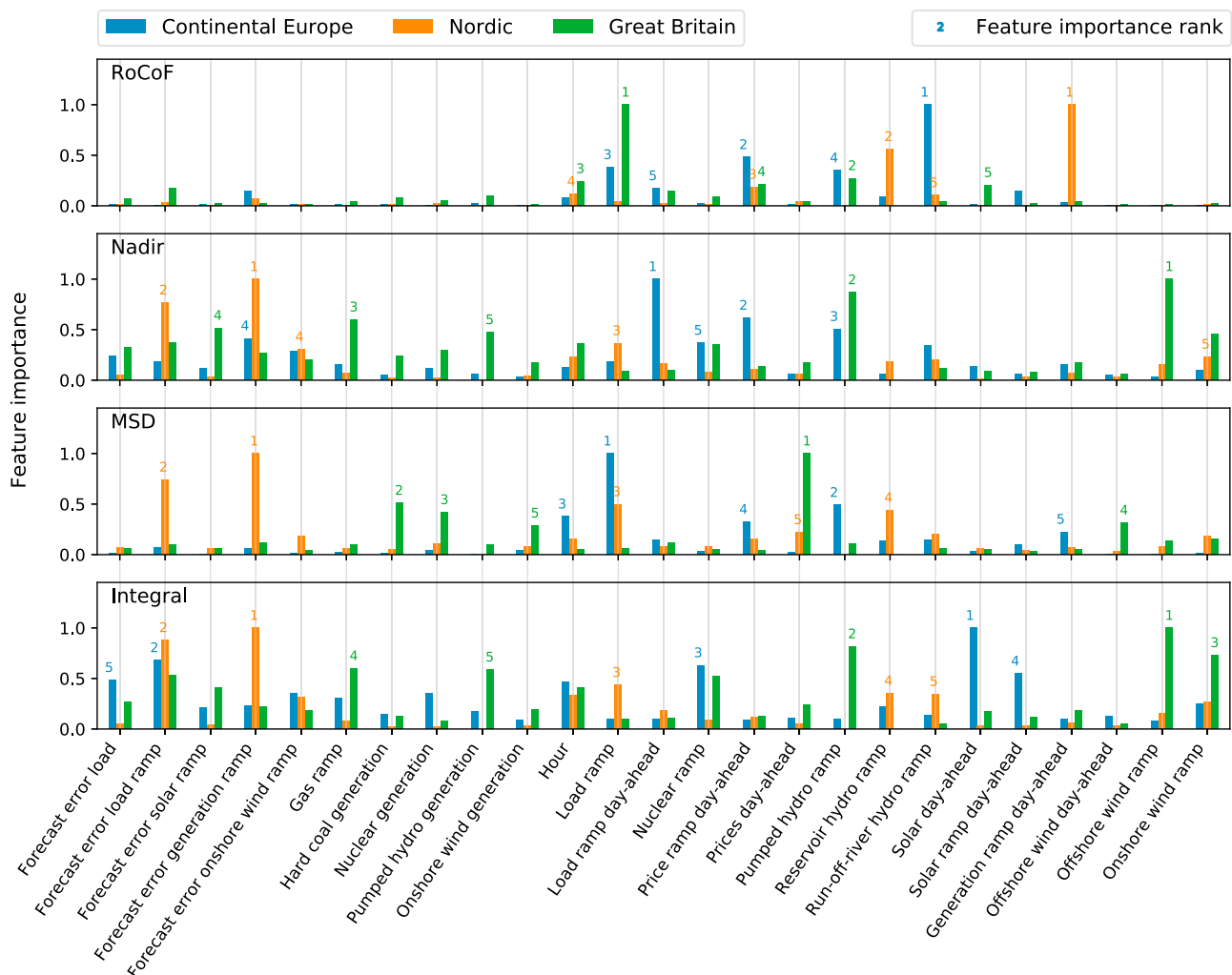
Figure 2. Most important features for predicting frequency stability
The feature importance in our model is measured by the mean absolute SHAP value. The union of the five most important features for each stability indicator and area is shown (see experimental procedures). The importance rank of the five most important features for each area is given above the corresponding bar. While forecast errors and load and generation ramps have a high relative importance, the total synchronous generation is not among the five most important features and its average effect is therefore negligible.

but ramps faster than other generation types play important roles. Notably, hard coal is one of the slow generation types in GB but one of the fast types in CE due to the importance of nuclear and lignite generation in CE, which are even slower than coal. We categorized the generation types using SHAP values for the generation ramps to predict the RoCoF and relative ramping speeds (Figures 3D and 3E). We found that fast generation ramps drove the RoCoF. A positive ramp was associated with more positive frequency jumps. In contrast, ramps of slow generation types offset the RoCoF, leading to a negative correlation. The only exception here was the behavior of gas power plants in GB, which showed a negative correlation despite being fast. This is due to their role as the prime source of balancing reserve in GB.[47] To summarize, the ramps no longer drove the RoCoF, but the RoCoF drove the ramps.

Notably, a model-agnostic data analysis does not produce such consistent results, as our features are strongly correlated

(experimental procedures). For example, the Pearson correlation coefficient between nuclear ramps and RoCoFs in CE is positive (supplemental experimental procedures S3). Instead, the SHAP framework indicates a negative relationship, which we consistently explain with relative ramping speeds.

Relating large control efforts to nonlinear dependencies
Frequency stability indicators often exhibit a complex nonlinear dependency on the features. Using the MSD, an indicator for the (primary) control effort,[33] as an example, we found that the daily profiles of the MSD differed strongly between the three grids (Figure 4). These differences were well reproduced by the ML model and were explained using daily aggregated SHAP values (experimental procedures).

In CE, the control effort peaked around midnight (Figure 4A) due to the nonlinear effects of negative load ramps. Details on this relation are shown in a dependency plot (Figure 4D). Load
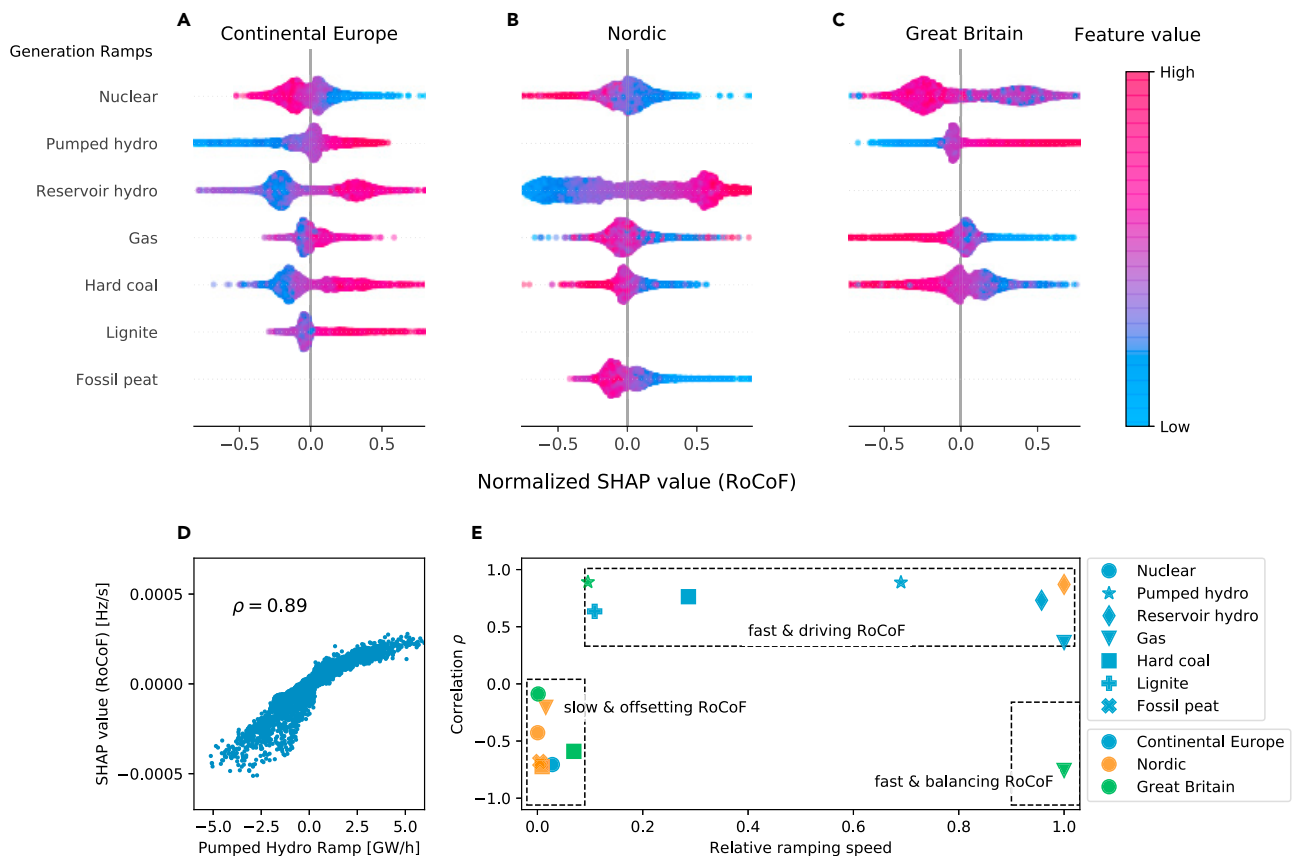
Figure 3. Effect of generation ramps on the RoCoF

(A–C) Examination of the effects of dispatchable, i.e., weather-independent, generation technologies, which generally affect the hourly RoCoF due to their gradual change at the beginning of (hourly) trading intervals.[4] The bee swarm chart depicts the SHAP effects on the RoCoF in the Continental Europe (A), the Nordic (B), and the Great Britain (C) grid areas. For each area and generation type, we normalized the SHAP values by their maximum absolute value to improve visibility. Each colored dot represents one time step in the dataset and the dots pile up vertically to indicate their density on the x axis. The figure only examines generation ramps with a feature importance higher than 0.01 to ensure reliability of results.

(D) Quantification of the direction of the dependencies with the correlation between the feature value and the SHAP effect, shown here for pumped hydro ramps.

(E) Combining the directions with the relative ramping speeds of the generation technologies (see experimental procedures) helps to distinguish RoCoF-driving and RoCoF-offsetting technologies within the three areas.

ramps between −7 and +25 GW/h had a small negative effect on the MSD because such small ramps are easy to control. Outside this range, the effect increased strongly in a nonlinear and asymmetric manner. Negative load ramps had much larger effects than positive load ramps, and they occurred almost exclusively around midnight (see color code). In the Nordic daily profile, load ramps were also the most important feature (Figure 4B), and they showed a very similar nonlinear dependency (Figure 4E). In contrast, the daily MSD profile in GB strongly depended on day-ahead prices (Figure 4C), which had an almost linear dependency (Figure 4F). The control effort peaked during the day in response to high prices in the day-ahead market, while the MSD and the prices were low at night (00:00 to 04:00 h).

Notably, fluctuating renewables did not contribute strongly to the daily MSD profile in our model, although they are an important driver for frequency fluctuations in GB in general (cf. Figure 2). The observed differences between the synchronous areas could be due to different control regulations. For example, in GB,

wind power farms must provide frequency control,[48] and secondary control is allocated manually.[49]

Explaining systematic imbalances with interactions

The SHAP framework explains the role of different features and reveals how predictions depend on the *interaction* of features (see Figure 5 for an application of the integral in the CE grid and experimental procedures for technical details). It should be noted that the ML predictors for the other targets also displayed clear interactions. The most important features were solar and nuclear power ramps, which had a reverse dependency (Figure S13). Without interactions, the SHAP value increased gradually and nonlinearly with the solar ramp (Figure 5B). Strong negative ramps of solar power generation induced an ongoing shortage of power and thus led to negative integrals.

Interactions with nuclear and gas ramps altered the effect of solar ramps by up to 60%, leading to a strong vertical dispersion of the observed SHAP values (Figure 5A). In particular, negative nuclear ramps amplified the effect of negative solar ramps, while
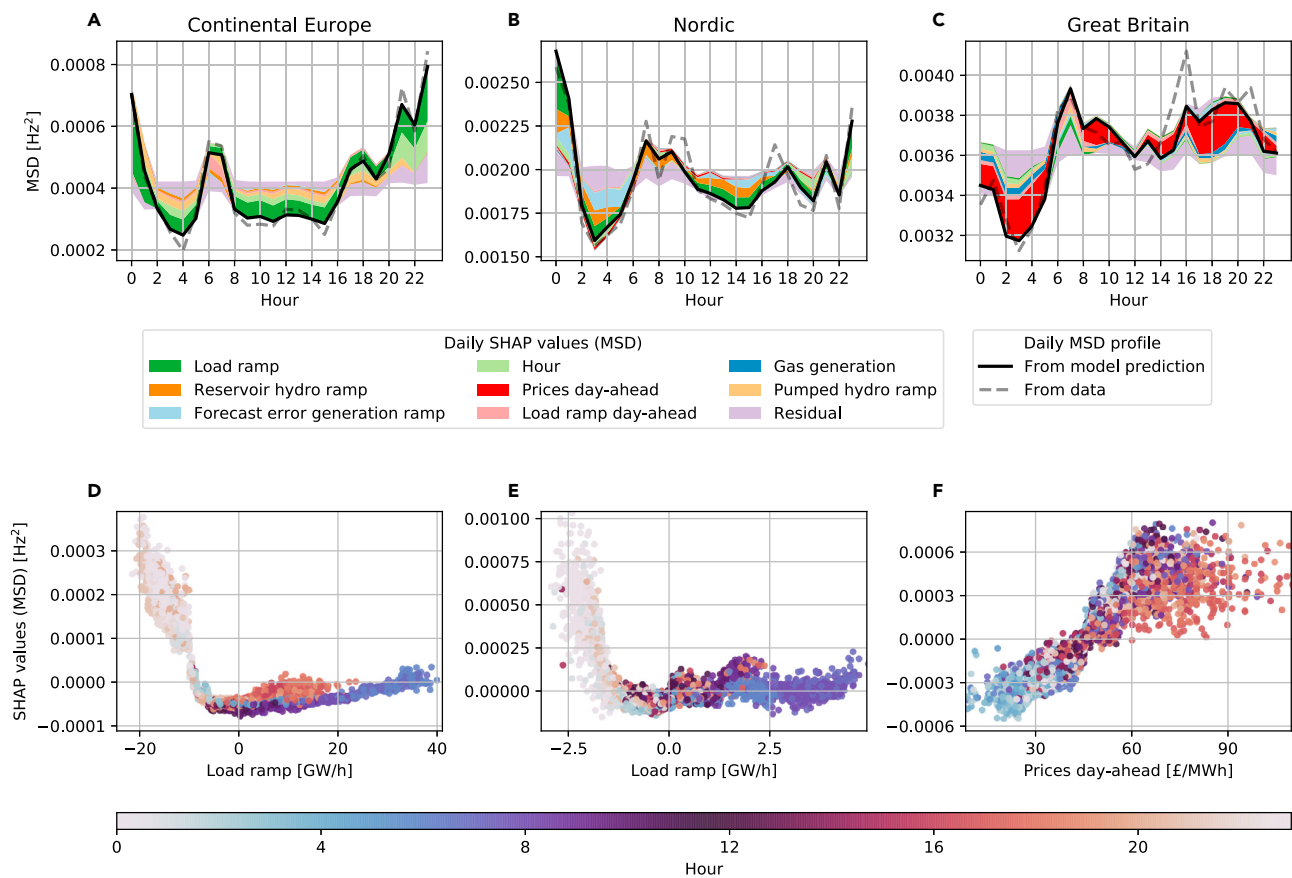
Figure 4. Explaining the daily average control effort with SHAP values

(A–C) The daily average profile of the MSD (dashed line), i.e., the daily average control effort, is very well reproduced by the ML model (solid line), but its shape differs between the areas. We examine these differences with daily SHAP values for the MSD in Continental Europe (A), the Nordic grid (B), and Great Britain (C). Daily SHAP values (see experimental procedures) are sorted such that negative effects are plotted above the prediction line and positive effects below it. The importance of the plotted feature effect decreases the farther away the feature is from the prediction line. Less important features are aggregated in a residual variable.

(D–F) For the Continental Europe (D) and Nordic (E) grids, the dependency plots of the most important daily SHAP effects reveal nonlinear relationships. These relationships explain the large control effort around midnight (color code), while the linear effect of prices explains the low control effort in GB during the night (F).

negative gas ramps dampened the effect on the integral (Figures 5C and 5D). These opposite interactions were again related to different ramping speeds. Nuclear power has the lowest ramping speed in CE, which meant that negative nuclear ramps amplified the continuous ramping behavior in interaction with solar ramps. In contrast, gas power had a fast ramp and therefore often provided balancing power, leading to the opposite effect. In general, these results demonstrate that interactions can influence how strongly a single feature affects power system stability.

DISCUSSION

Our model is based on explainable ML, and it predicts important indicators for power system frequency stability using external features, such as day-ahead electricity prices or total system load. Using real data (ex post analysis), our ML model outperformed the daily profile, a system-specific null model, by a factor of up to 16.3. Using only day-ahead available data, our ML models performed similarly in most cases. When SHAP values were calculated and examined, our model revealed important

features and dependencies, and could thus pave the way for multiple applications.

Our model offers a versatile and substantially improved approach for analyzing risks and drivers of grid frequency stability. Previous data-driven studies analyzed the impact of one external feature on grid frequency fluctuations based on linear correlations[9–11] or conditional probabilities.[7] Such univariate analyses cannot be adjusted for effects of other features, which could be correlated with the feature of interest and may lead to incorrect conclusions (cf. Weißbach et al.[42]). Moreover, univariate, linear dependency analyses underestimate the effects of nonlinearities and ignore feature interactions. Our model includes multiple variables and fits nonlinear dependencies and interactions, which are made transparent by SHAP values. It breaks down the effect of correlated features (as discussed in Figure 3) and reveals otherwise undetectable nonlinear effects (Figure 4) and feature interactions (Figure 5). In addition, our model visualizes feature effects in the daily average evolution of frequency stability (Figure 4), which adds to the many useful visualization tools available
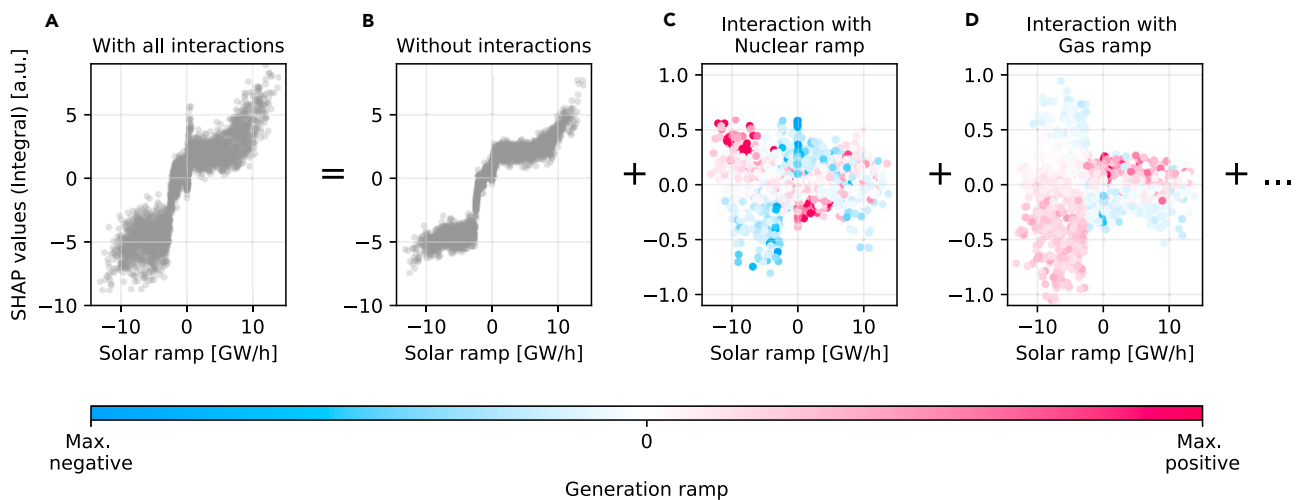
Figure 5. The effect of feature interactions on systematic power imbalances

(A) CE is used to show the SHAP effects of solar ramps on the frequency integral, which are the most important effects in CE (Figure 2). The integral, which represents systematic imbalances, decreases for negative solar ramps, but the SHAP effects vary strongly, as indicated by their vertical dispersion.

(B–D) Using SHAP interaction values (see experimental procedures), this dispersion was broken down into different interaction effects. These effects depend on the generation type, as negative nuclear ramps weaken the effect of negative solar ramps (C), while negative gas ramps lead to an amplified effect (D). Subtracting all interaction effects from the original SHAP values (A) yields the remaining effect of solar ramps (B), which is strongly altered.

in the SHAP framework. Our publicly available model[50] and dataset[30] can be used to predict any frequency stability indicator, thus offering a ready-to-use and flexible tool for analyzing power grid stability.

We applied our model to three different synchronous areas and identified options for improving power grid operation strategies. We discussed four examples for potential applications: (1) first, we showed how generation ramps drove the RoCoF and the nadir and revealed subtle differences between generation types and grids. These insights can help to optimize ramping behavior and mitigate DFDs[51] and improve frequency quality. In particular, hydropower generation ramps should be optimized in response to other ramps. (2) Our results show that forecasting errors play an essential role in the Nordic grid. While TSOs are generally aware of the problem,[44] the SHAP analysis provides a much more detailed view and reveals when and how these errors affect frequency stability. An example is given in Figure 4B, which shows that the features "forecasting errors of the generation ramps" are particularly important during the night. Our model identifies situations where forecasting errors are particularly critical and will thus improve risk awareness in grid operation. (3) Low inertia has been identified as a major threat for the stability of future power grids.[3] Our analysis provides a more finely nuanced view on this topic. In our XAI model, inertia does not generally have a high feature importance. Instead, the impact of inertia on frequency stability is nonlinear and more pronounced for low inertia values (Figure S14). (4) Finally, the predictive power of our XAI model can be harnessed for online grid monitoring and preventive control measures. For instance, a model-predictive frequency restoration reserve has been proposed to mitigate DFDs.[52] The applicability of such predictive control strategies can be extended by data-enabled methods (cf., e.g., Huang et al.[53]).

The main restrictions to our model performance and explainability are due to the quality of available power system data. Firstly, frequency deviations due to renewable fluctuations[7] or load fluctuations[8] occur on timescales that are smaller than the intervals of electricity trading. The limited time resolution of publicly available power system time series restricts both the performance of an ML model and its ability to suggest causal relationships because the time order of events is partly hidden. Secondly, all locations in a synchronous power grid affect the frequency deviations; but in large grid areas, such as CE, many countries provide no or only a limited amount of data.[14] This further emphasizes the need for open data in energy system analysis and design.[54]

Our paper contributes to the applicability of XAI methods in energy systems and engineering sciences in general. Firstly, deriving causal relationships from data is a key challenge in modern ML techniques. With the power grid frequency, we provide a very well-suited test bed; while not all features and interactions are available, there is plenty of domain knowledge to interpret and cross-check XAI results. Secondly, we provide an excellent dataset[30] for applications and the benchmarking of ML methods, such as causal inference or predictive models. While generation data are already publicly available,[36] aggregating these for a whole synchronous area and combining them with frequency stability indicators yields a novel dataset for future usage. Finally, in Figure 4, we explored how daily aggregated SHAP data may be used to explain specific temporal profiles, which could be useful in other ML applications when dealing with strong daily or seasonal trends, e.g., in weather or traffic predictions.[55,56]

In conclusion, we hope that our work will trigger further applications of XAI in energy science, harnessing the strengths of modern ML tools while avoiding the drawbacks of black box approaches, which impede scientific insights[18] and pose security risks.[19] Our model provides insights by explaining feature effects

with SHAP values in the context of the domain science. SHAP dependency and interaction plots visualize the knowledge learned by the model and offer individual explanations for each prediction. The most predictive associations then suggest causal relationships, which can then be validated by domain knowledge or further experiments. For example, we identify RoCoF-driving, RoCoF-offsetting, and RoCoF-balancing generation technologies by connecting our model results to physical ramping rates, thus suggesting different causal relationships. "Suggesting relations" is key here since neither boosted trees nor SHAP guarantee causal relationships but rather indicate associations based on the data. All in all, SHAP values alone do not provide scientific insights, but, when combined with domain knowledge, they can lead to further knowledge.

Future work includes explicitly forecasting the given indicators and classifying whether upcoming events could be problematic for grid operation. Once forecasts or other early warning and control methods have been implemented, our model will need to be retrained using these new, controlled datasets to derive the updated feature-target interactions. Furthermore, while we already outperform the daily profile, the performance of our tree-based predictor could be improved if further features were integrated and our model was compared with other ML prediction models. Finally, such regression models should be complemented by causal inference models to provide clear counterfactual statements and comparisons with XAI approaches.

EXPERIMENTAL PROCEDURES

Resource availability
Lead contact
Further information questions should be directed to the lead author, Johannes Kruse (jo.kruse@fz-juelich.de).
Materials availability
This study did not generate new unique materials.
Data and code availability
The dataset to reproduce our results is available on Zenodo: https://doi.org/10.5281/zenodo.5118352. The Python code used to create our results and the figures is also archived on Zenodo: https://doi.org/10.5281/zenodo.5497609.

Data preparation of frequency stability indicators
In a modern AC power grid, the grid frequency is typically spatially synchronized and its dynamics can be represented by a single bulk time series on time-scales of several seconds and more.[1] In Europe, different synchronous areas exist, which are only inter-connected through DC links and hence display their own frequency dynamics and follow their own specific regulations. We modeled the bulk frequency dynamics for different synchronous areas in Europe, specifically for the CE, Nordic, and GB areas. We used pre-processed frequency time series $\tilde{f}(t)$ with a time resolution of $\tau = 1$ s,[13] which were originally measured by regional TSOs.[57–59]

From the centered frequency time series $f(t) = \tilde{f}(t) - 50$ Hz, we extracted four hourly stability indicators, which are directly relevant for power system operation.[32,33] For the ith hour starting at time t_i, we calculated the (positive or negative) nadir, the integral and the MSD based on the hourly time steps $I_i = \{t_i, t_i + \tau, ..., t_i + \tau\gamma\}$ with $\gamma = 3600$:

$$\text{Nadir}(t_i) = f\left(\arg\max_{t \in I_i}|f(t)|\right),$$

$$\text{Integral}(t_i) = \tau\sum_{t \in I_i}f(t),$$

$$\text{MSD}(t_i) = \frac{1}{\gamma}\sum_{t \in I_i}f^2(t).$$

From the derivative of the frequency time series $\frac{df}{dt}(t)$, we obtained the hourly (positive or negative) RoCoF by looking for the steepest slope within a window $W_i = [t_i - T, t_i + T]$ of length $2T$ around the beginning of the hour t_i:

$$\text{RoCoF}(t_i) = \frac{df}{dt}\left(\arg\max_{t \in W_i}\left|\frac{df}{dt}\right|\right).$$

We estimated the derivative $\frac{df}{dt}(t)$ using a low-pass filter on the frequency increments,[60] i.e., by smoothing the increments $\Delta f(t) = f(t) - f(t-\tau)$ with a rectangular rolling window of length L. We chose the parameters L and T in such a way that they accounted for the different timescales of the RoCoF in the synchronous areas (supplemental experimental procedures S2). This resulted in a choice of $L = T = 60$s for the CE and GB areas, while the Nordic area with its fast hydropower exhibited larger RoCoFs so that we chose $L = T = 30$s instead.

Data preparation of external features
We collected different power system time series as external features to predict frequency deviations in Europe. We retrieved six different sets of publicly available time series from the ENTSO-E transparency platform.[36] These sets comprise the day-ahead load forecast, day-ahead scheduled generation, day-ahead wind and solar power forecast, day-ahead electricity prices, actual load, and actual generation per production type. Most of the time series are available on an hourly basis. Since we predicted stability indicators on an hourly basis, we downsampled a few higher-resolution time series to a common hourly resolution.

We then aggregated the data within the three synchronous areas. Since time series from ENTSO-E are only available for smaller regions within the synchronous areas (i.e., countries), we added up the load and generation data within each area. To aggregate the price data, we calculated area-wide averages weighted by the regional mean load. The time series from the ENTSO-E transparency platform contained multiple missing or corrupted data points,[14] which required a careful aggregation and cleansing procedure (supplemental experimental procedures S1). We deemed area-wide feature aggregation necessary because all locations within the synchronous power grid contribute to large frequency deviations.[1] We additionally prepared selected country-level data for the CE and the Nordic areas. The area-wide aggregated features resulted in a similar or higher model performance than country-level data (supplemental experimental procedures S5). Therefore, we decided to use area-wide aggregated features for this publication. An overview of the available (aggregated) features per area is available in supplemental experimental procedures S1.

Finally, we engineered additional meaningful features based on the hourly ENTSO-E time series $X(t_i)$, which comprise both day-ahead forecast data $X_{D-1}(t_i)$ and actual data $X_D(t_i)$. For each hourly interval $\Delta t = \tau\gamma$, we introduced ramp features (slopes) $(X(t_i) - X(t_i - \Delta t))/\Delta t$, which are inspired by the importance of generation ramps for the CE frequency dynamics.[4] We also added forecast errors $X_{D-1}(t_i) - X_D(t_i)$ and the artificial features of hours (of the day), weekdays, and months. To include the total available inertia as a feature, we calculated the sum of the synchronous generation which approximates to the time-dependent inertia.[38]

In summary, our dataset comprises hourly time series of 4 stability indicators (model outputs or targets) and 66 external features (model inputs) for the years 2015–2019. The dataset is available on Zenodo[30] and our scripts for downloading and preparing the dataset are online at GitHub.[50]

GTB model
To predict indicators of frequency stability from external features, we used GTB. Tree-based ensemble methods, such as GTB, are complex, nonlinear ML models, which makes them suitable for predicting the nonlinear behavior of power grids.[1] They offer a quick method of calculating SHAP values, thus facilitating efficient post-modeling explanation.[24] In addition, they are immune to the effects of feature outliers and perform inherent feature selection, making them robust to the inclusion of correlated or irrelevant features.[16] This is beneficial for our dataset, which exhibits strongly correlated features (supplemental experimental procedures S3) as well as outliers due to bad data quality (supplemental experimental procedures S1).

To fit our GTB model, we used XGBoost, which is a scalable gradient tree boosting system that provides state-of-the-art results for many ML applications.[43] We randomly split our data into a training set (64%), a validation set (16%), and a test set (20%). To optimize the hyperparameters of the XGBoost model, we performed a grid search over selected parameter values and evaluated the performance via 5-fold cross-validation on our training set. To determine the number of trees in the XGBoost models, we performed early stopping on the validation set. Finally, we concatenated the training and validation sets, retrained the model on this data with optimal hyperparameters, and tested its performance on the unseen test set. We also calculated the SHAP values on the test set. The detailed implementation in Python code is available on GitHub[50] and the sets of final hyperparameters are online at Zenodo.[30]

To quantify the model performance, we evaluated the R^2 score, which quantifies the proportion of variability explained by the model. Predicting the true targets results in a score of 1, while always predicting the mean of the target gives a score of 0. To benchmark our predictor, we compared its performance with the daily profile prediction. The daily profile, i.e., the daily average evolution of a target, is the most important recurring pattern of frequency dynamics.[61] Predicting the stability indicators based on their daily profiles thus represents an important null model. Our GTB model consistently outperformed the daily profile for all areas and indicators (see supplemental experimental procedures S5 for a detailed performance evaluation).

Model interpretation with SHAP

SHAP values can explain the output of any ML model.[23] Based on the game-theoretical Shapley values, they attribute a model output to the individual effects of each input feature. In particular, SHAP values quantify the marginal effect of including a feature into the prediction and comparing them with a randomized baseline.[24,62] Within the class of additive feature attributions, they guarantee certain optimal properties, such as local accuracy and consistency.[22] As they are locally accurate, the SHAP values always add up to the total model output. Consistency guarantees that a SHAP value does not decrease if the corresponding feature contributes more to the prediction when the model is altered.

SHAP values represent the feature effects on individual model outputs relative to the base value, which is given by the average prediction (cf. Figure 1). By combining many of these local explanations, SHAP values also offer global insights.[24]

The mean absolute SHAP value measures the global importance of a feature within a model. We identified the five most important features for each stability indicator and area (Figure 2). Figure 2 also displays feature importances for the *union* of these feature sets, i.e. features with an importance rank below five are also displayed. In addition to global feature importances, dependency plots show how the effect of a feature changes with the value of the feature (e.g., Figure 4D). Notably, these dependencies differ from observing relationships in scatterplots or between targets and features in a simple correlation analysis. Such model-agnostic methods cannot distinguish the effect of two correlated features. In contrast, we estimated interventional SHAP values, which quantify the marginal feature effect in the model by breaking down correlations with other features.[63,64]

In addition to first-order attributions, SHAP offers interaction values that attribute the model output to pairs of interacting features.[24] Interaction values decompose the first-order SHAP effects into diagonal effects and pairwise interaction effects (such as in Figure 5). The interaction effects therefore explain the vertical dispersion within the first-order SHAP dependency plots, thus offering scientific insights as well as additional consistency checks for the model applications.

Finally, there is a fundamental difference between predictive models and causal models.[65] Predictive models try to infer the conditional probability of the target given the feature variables by fitting associations. Causal models identify the effect on the target when manipulating or intervening on a feature. ML models, such as the boosted trees used here, are typically predictive models. Using XAI methods to explain how these ML models work reveals only associations learned from the data.[21] In particular, using SHAP values to explain predictive models does not necessarily reflect causal effects.[66] However, causation involves correlation so that predictive and explainable ML models can suggest causal dependencies, which then have to be further validated, e.g. by domain knowledge or causal inference methods.

Aggregated SHAP values

To explain daily average profiles of the model predictions, we visualized the SHAP values in a way that builds on their additivity. Due to their property of "local accuracy," the prediction $f(t)$ at every point in time t can be written as a sum of the respective SHAP values,

$$f(t) = \varphi_0 + \sum_{j=1}^{N} \varphi_j(t), \qquad \text{(Equation 1)}$$

where $\varphi_j(t)$ is the SHAP value of feature j at time t. This property of SHAP values enables a new application in the analysis of daily profiles and other recurrent patterns. The daily profile of the prediction is the average $\langle f(t) \rangle_h$ for the hour h over all days. Based on the SHAP values $\varphi_j(t)$ for feature j ($j = 1, \ldots, N$) and their base value φ_0,[24] we decomposed the daily profile as follows:

$$\langle f(t) \rangle_h = \langle \varphi_0 + \sum_{j=1}^{N} \varphi_j(t_i) \rangle_h = \varphi_0 + \sum_{j=1}^{N} \langle \varphi_j(t) \rangle_h.$$

The daily aggregated SHAP values $\langle \varphi_j \rangle_h$ then explain the daily profile of the prediction. To display the daily SHAP values, such as in Figures 4A–4C, we identified the three most important features according to their average effect $\frac{1}{24}\sum_{h=1}^{24}|\langle \varphi_j \rangle_h|$ on the daily profile in each area. In Figures 4A–4C, we then visualized these features from the union of these sets to display the most important daily SHAP values. The remaining daily SHAP values were aggregated and displayed as a residual variable.

Finally, we add three notes on the interpretation of (daily) aggregated SHAP values. (1) We note that the aggregated SHAP values do *not* coincide with SHAP values of a model trained on the aggregated data. This must be taken into account when interpreting the results. (2) Due to the nonlinearity of an ML model, a large daily SHAP value does not necessarily correspond to a large average for the corresponding feature in that hour. (3) Second-order interactions between features are "fairly" distributed between first-order SHAP values according to the classical Shapley values,[24] i.e., large daily SHAP values can partly relate to strong interaction within this specific hour. To further resolve interactions within the daily SHAP values, the additivity of second-order SHAP values can be used to generate daily profiles of the interactions. This is beyond the scope of this paper.

Relative ramping rates

We used relative ramping rates to validate our SHAP results, particularly for the prediction of the RoCoF. In particular, we quantified the relative ramping speed of each conventional generation technology k within a synchronous area. The ramping speed $\bar{s}_k$ is determined both by the absolute change of generation ΔX_k and the timescale λ_k on which the generator adapts its output to the new set point:

$$\bar{s}_k : = \frac{\Delta X_k}{\lambda_k}.$$

We approximated the typical value of ΔX_k with the median of the absolute generation changes $\Delta X_k \approx \text{Median}_{t_i}|X_k(t_i - \Delta t) - X_k(t_i)|$. The *relative* ramp speed s_k, compared with the fastest technology m within the area, then reads

$$s_k = \frac{\bar{s}_k}{\bar{s}_m} = \frac{\Delta X_k}{\Delta X_m}\frac{\lambda_m}{\lambda_k} \approx \frac{\Delta X_k}{\Delta X_m}\frac{r_k}{r_m}.$$

Finally, we approximated the ratio of timescales λ_k by using the inverse ratio of technology-specific ramping rates r_k.[67] The technology m with the largest absolute ramping speed was determined by the maximum value of $\Delta X_k r_k$.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2021.100365.

AUTHOR CONTRIBUTIONS

J.K., B.S., and D.W. conceived and designed the research. J.K. trained the model and produced the figures. All authors contributed to discussing and interpreting the results and writing the manuscript. B.S. and D.W. contributed equally.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Machowski, J., Bialek, J., Bumby, J., and Bumby, D.J. (2008). Power System Dynamics: Stability and Control (John Wiley & Sons).

2. Pourbeik, P., Kundur, P.S., and Taylor, C.W. (2006). The anatomy of a power grid blackout—root causes and dynamics of recent major blackouts. IEEE Power Energy Mag. *4*, 22–29.

3. Milano, F., Dörfler, F., Hug, G., Hill, D.J., and Verbič, G. (2018). Foundations and challenges of low-inertia systems (invited paper). In 2018 Power Systems Computation Conference (PSCC) (IEEE), pp. 1–25.

4. Weissbach, T., and Welfonder, E. (2009). High frequency deviations within the European Power System: origins and proposals for improvement. In 2009 IEEE/PES Power Systems Conference and Exposition (IEEE), pp. 1–6.

5. Vorobev, P., Greenwood, D.M., Bell, J.H., Bialek, J.W., Taylor, P.C., and Turitsyn, K. (2019). Deadbands, droop, and inertia impact on power system frequency distribution. IEEE Trans. Power Syst. *34*, 3098–3108.

6. Ayodele, T.R., Jimoh, A., Munda, J.L., and Tehile, A.J. (2012). Challenges of grid integration of wind power on power system grid integrity: a review. Int. J. Renew. Energy Res. *2*, 618–626.

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

8. Chen, L., Markham, P., Chen, C.-f., and Liu, Y. (2011). Analysis of societal event impacts on the power system frequency using FNET measurements. In 2011 IEEE Power and Energy Society General Meeting (IEEE), pp. 1–8.

9. Adeen, M., Jónsdóttir, G.M., and Milano, F. (2019). Statistical correlation between wind penetration and grid frequency variations in the Irish network. In 2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), pp. 1–6.

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

11. Persson, M., and Chen, P. (2017). Frequency evaluation of the Nordic power system using PMU measurements. IET Generation, Transmission and Distribution *11*, 2879–2887.

12. Gorjão, L.R., Schäfer, B., and Hassan, G. (2020). Open Access Power-Grid Frequency Database (OSF). https://doi.org/10.17605/OSF.IO/M43TG.

13. Kruse, J., Schäfer, B., and Witthaut, D. (2020a). Pre-Processed Power Grid Frequency Time Series (Zenodo). https://doi.org/10.5281/zenodo.3744121.

14. Hirth, L., Mühlenpfordt, J., and Bulkeley, M. (2018). The ENTSO-E Transparency Platform—a review of Europe's most ambitious electricity data platform. Appl. Energy *225*, 1054–1067.

15. Morrison, R. (2018). Energy system modeling: public transparency, scientific reproducibility, and open development. Energy Strategy Rev. *20*, 49–63.

16. Hastie, T., Tibshirani, R., and Friedman, J. (2016). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition (Springer).

17. Adadi, A., and Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access *6*, 52138–52160.

18. Roscher, R., Bohn, B., Duarte, M.F., and Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. IEEE Access *8*, 42200–42216.

19. Ahmad, T., Zhang, D., Huang, C., Zhang, H., Dai, N., Song, Y., and Chen, H. (2021). Artificial intelligence in sustainable energy industry: status quo, challenges and opportunities. J. Clean. Prod. *289*, 125834.

20. Cremer, J.L., Konstantelos, I., and Strbac, G. (2019). From optimization-based machine learning to interpretable security rules for operation. IEEE Trans. Power Syst. *34*, 3826–3836.

21. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion *58*, 82–115.

22. Lungberg, S.M., Erion, G.G., and Lee, S.-I. (2019). Consistent individualized feature attribution for tree ensembles. arXiv, arXiv:1802.03888 https://arxiv.org/abs/1802.03888.

23. Lundberg, S.M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17 (Curran Associates Inc.), pp. 4768–4777.

24. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nat. Machine Intelligence *2*, 56–67.

25. Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.-W., Newman, S.-F., Kim, J., and Lee, S.-I. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat. Biomed. Eng. *2*, 749–760.

26. Padarian, J., McBratney, A.B., and Minasny, B. (2020). Game theory interpretation of digital soil mapping convolutional neural networks. Soil *6*, 389–397.

27. Kuzlu, M., Cali, U., Sharma, V., and Güler, Ö. (2020). Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools. IEEE Access *8*, 187814–187823.

28. Chen, M., Liu, Q., Chen, S., Liu, Y., Zhang, C., and Liu, R. (2019). XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system. IEEE Access *7*, 13149–13158.

29. Alova, G., Trotter, P.A., and Money, A. (2021). A machine-learning approach to predicting Africa's electricity mix based on planned power plants and their chances of success. Nat. Energy *6*, 158–166.

30. Kruse, J., Schäfer, B., and Witthaut, D. (2021c). Supplementary Data: "Revealing Drivers and Risks for Power Grid Frequency Stability with Explainable AI" (Zenodo). https://doi.org/10.5281/zenodo.5118352.

31. Meyer, P.G., Anvari, M., and Kantz, H. (2020). Identifying characteristic time scales in power grid frequency fluctuations with DFA. Chaos *30*, 013130.

32. Gross, D., Bolognani, S., Poolla, B.K., and Dörfler, F. (2017). Increasing the resilience of low-inertia power systems by virtual inertia and damping. In Proceedings of IREP'2017 Symposium (IREP), p. 64.

33. Tyloo, M., and Jacquod, P. (2021). Primary control effort under fluctuating power generation in realistic high-voltage power networks. IEEE Control. Syst. Lett. *5*, 929–934.

34. Gorjão, L.R., Vanfretti, L., Witthaut, D., Beck, C. and Schäfer, B. (2021). Phase and amplitude synchronisation in power-grid frequency fluctuations in the Nordic grid. arXiv, arXiv:2105.00228, https://arxiv.org/abs/2105.00228.

35. Gorjão, L.R., Jumar, R., Maass, H., Hagenmeyer, V., Yalcin, G.C., Kruse, J., et al. (2020). Open database analysis of scaling and spatio-temporal properties of power grid frequencies. Nat. Commun. 11, 6362.

36. ENTSO-E. (2020). Transparency Platform. https://transparency.entsoe.eu/.

37. Lin, J., and Magnago, F.H. (2017). Electricity Markets: Theories and Applications (John Wiley & Sons).

38. Ulbig, A., Borsche, T.S., and Andersson, G. (2014). Impact of low rotational inertia on power system stability and operation. IFAC Proc. Vol. 47, 7290–7297.

39. Schmietendorf, K., Peinke, J., and Kamps, O. (2017). The impact of turbulent renewable energy production on power grid stability and quality. Eur. Phys. J. B 90, 222.

40. Gorjão, L.R., Anvari, M., Kantz, H., Beck, C., Witthaut, D., Timme, M., and Schäfer, B. (2020). Data-driven model of the power-grid frequency dynamics. IEEE Access 8, 43082–43097.

41. ENTSO-E. (2019). 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.

42. Weißbach, T., Remppis, S., and Lens, H. (2018). Impact of current market developments in Europe on deterministic grid frequency deviations and frequency restoration reserve demand. In 2018 15th International Conference on the European Energy Market (EEM) (IEEE), pp. 1–6.

43. Chen, T., and Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16 (ACM), pp. 785–794.

44. Svenska kraftnät, Statnett, Energinet.dk, and Fingrid. (2016). Challenges and Opportunities for the Nordic Power System. https://www.fingrid.fi/globalassets/dokumentit/fi/yhtio/tki-toiminta/report-challenges-and-opportunities-for-the-nordic-power-system.pdf.

45. Nationalgrid ESO. (2019b). Operability Strategy Report 2019. https://www.nationalgrideso.com/document/159726/download.

46. Schäfer, B., Witthaut, D., Timme, M., and Latora, V. (2018). Dynamically induced cascading failures in power grids. Nat. Commun. 9, 1–13.

47. Nationalgrid ESO. (2019a). Monthly Balancing Services Summary 2019/20. https://www.nationalgrideso.com/document/119361/download.

48. Díaz-González, F., Hau, M., Sumper, A., and Gomis-Bellmunt, O. (2014). Participation of wind power plants in system frequency control: review of grid code requirements and control methods. Renew. Sustain. Energy Rev. 34, 551–564.

49. ENTSO-E Balancing Report 2020 (2020). https://eepublicdownloads.azureedge.net/clean-documents/Publications/Market%20Committee%20publications/ENTSO-E_Balancing_Report_2020.pdf

50. Kruse, J., Schäfer, B., and Witthaut, D. (2021a). Explainable Machine Learning for Power Grid Frequency Stability. https://doi.org/10.5281/zenodo.5497609.

51. Kruse, J., Schäfer, B., and Witthaut, D. (2021b). Exploring deterministic frequency deviations with explainable AI. arXiv, arXiv:2106.09538 https://arxiv.org/abs/2106.09538.

52. Avramiotis-Falireas, I., Troupakis, A., Abbaspourtorbati, F., and Zima, M. (2013). An MPC Strategy for Automatic Generation Control with Consideration of Deterministic Power Imbalances. In 2013 IREP Symposium Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid (IEEE), pp. 1–8.

53. Huang, L., Coulson, J., Lygeros, J., and Dörfler, F. (2021). Decentralized data-enabled predictive control for power system oscillation damping. IEEE Trans. Control. Syst. Technol. https://doi.org/10.1109/TCST.2021.3088638.

54. Pfenninger, S. (2017). Energy scientists must show their workings. Nature 542, 393.

55. Lana, I., Del Ser, J., Velez, M., and Vlahogianni, E.I. (2018). Road traffic forecasting: recent advances and new challenges. IEEE Intell. Transport. Syst. Mag. 10, 93–109.

56. Schultz, M., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L., Mozaffari, A., and Stadtler, S. (2021). Can deep learning beat numerical weather prediction? Philos. Trans. R. Soc. A 379, 20200097.

57. Fingrid Oyj (2020). Frequency—Historical Data. https://data.fingrid.fi/en/dataset/frequency-historical-data.

58. Nationalgrid ESO. (2020). Historic Frequency Data. https://www.nationalgrideso.com/balancing-services/frequency-response-services/historic-frequency-data.

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

60. Frigo, G., Derviškadić, A., Zuo, Y., and Paolone, M. (2019). PMU-based ROCOF measurements: uncertainty limits and metrological significance in power system applications. IEEE Trans. Instrum. Meas. 68, 3810–3822.

61. Kruse, J., Schäfer, B., and Witthaut, D. (2020b). Predictability of power grid frequency. IEEE Access 8, 149435–149446.

62. Sundararajan, M., and Najmi, A. (2020). The Many Shapley Values for Model Explanation (International Conference on Machine Learning, PMLR), pp. 9269–9278.

63. Chen, H., Janizek, J.D., Lundberg, S., and Lee, S.-I. (2020). True to the model or true to the data? arXiv, arXiv:2006.16234 https://arxiv.org/abs/2006.16234.

64. Janzing, D., Minorics, L., and Bloebaum, P. (2020). Feature relevance quantification in explainable AI: a causal problem. In International Conference on Artificial Intelligence and Statistics (PMLR), pp. 2907–2916.

65. Spirtes, P., and Zhang, K. (2016). Causal discovery and inference: concepts and recent methodological advances. Appl. Inform. 3, 3.

66. Ma, S., and Tourani, R. (2020). Predictive and causal implications of using Shapley value for model interpretation. In Proceedings of the 2020 KDD Workshop on Causal Discovery (PMLR), pp. 23–38.

67. Gonzalez-Salazar, M.A., Kirsten, T., and Prchlik, L. (2018). Review of the operational flexibility and emissions of gas- and coal-fired power plants in a future with growing renewables. Renew. Sustain. Energy Rev. 82, 1497–1513.

Supplemental material for "Revealing drivers and risks for power grid frequency stability with explainable AI"

SUPPLEMENTAL EXPERIMENTAL PROCEDURES S1:
EXTERNAL FEATURE AGGREGATION AND DATA CLEANSING

To model frequency stability indicators, we collect publicly available times series of external features from the ENTSO-E Transparency platform.[1] For the synchronous areas investigated here, we aggregate the ENTSO-E time series, which are originally only available for smaller regions within the areas (e.g., countries). However, the time series contain many missing data points, so that we need a careful procedure to aggregate the region contributions within the synchronous areas.

Firstly, we specify region types for which we obtain the best data quality. In all but a few cases, we retrieve country level data. Only in Continental Europe, we retrieve bidding zone data for Italy (North, Center North, Center South, South and Sicilia) and control zone data for Germany (TenneT, TransnetBW, 50Hertz, Amprion), as data quality is better for these regions. For Denmark, we also retrieve bidding zone data, as one zone belongs to the Continental Europe area, while the other belongs to the Nordic area.

Secondly, in the Continental Europe and the Nordic areas, we aggregate the region contributions and propagate missing data points through the data set. For each feature and region, we mark missing time steps as "NaN". Then, during aggregation, we propagate missing values by setting the sum of region contributions to NaN if at least one of them is NaN. For example, if the German load data had a missing value at 10:00 on the 6th of June 2018, the Continental European aggregated load data would have a NaN at this time step. Finally, we clean all aggregated features together with the grid frequency data. In particular, we omit a time step from the *whole* data set if at least one feature is NaN or the corresponding frequency measurements contains missing or corrupted values. This cleansed data set is used in model training and testing.

Notably, we will not have many data points left in the aggregated data set, if the region contributions contain too many NaNs. To avoid this problem, we collect the time series of each feature and region, which we call the *region-variable contribution*, and sort them according to their NaN share. Then, we successively add up the region-variable contributions with increasing NaN share. A contribution is only added if the NaN share S of the aggregated data set *including* the new contribution does not exceed a certain threshold S_T. If the data exceeds the threshold ($S > S_T$), we omit the contribution from the aggregation. We choose $S_T = 37\%$, which we found to result in a good balance between having enough data and including as many contributions as possible. As an example, consider a (hypothetical) area consisting of 2 countries, "A" and "B", with two features "Load" and "Total generation". Initially, we add load data from country A as this (hypothetically) has the lowest NaN share of 30%. Then, we add the total generation from country B with a NaN share of 31%. The contribution is not omitted, since the time steps with NaN values overlap and the final NaN share of the whole data set only yields 33% thus staying below the threshold. However, the last two region-variable contributions might not be added, as they introduce to many new NaN values.

This procedure omits a certain amount of data but allows us to retain a large sample size. In Continental Europe, most of the omitted contributions would increase the feature value by less than 30% on average (Figure S1A) and all omitted feature contributions are smaller than 5% of the total mean load (Figure S1B). In the Nordic area, we only omit the Finish day-ahead solar power forecast, accounting for 0.03% of the area total mean load, and in GB there are no omissions. The aggregated data sets obtained from this procedure contain more than 26800 data points in each area (Table S2). We thus generate large data sets to efficiently learn structures in the data, while still representing most of the load and generation within the areas.

Finally, we obtain 25 different (aggregated) times series of external features. Combining them with additional engineered features, such as forecast errors, we end up with 66 different external features (Table S1), which both contain day-ahead available features (such as the load forecast) and ex-post available features (such as the actual generation per type). None of the synchronous areas exhibits all 66 features and the number of model inputs thus varies between 50 and 64 (Table S2).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES S2:
ROCOF EXTRACTION

The Rate of Change of Frequency (RoCoF) is an indicator of frequency stability, which we use in our study. We extract the RoCoF at the beginning of each hour by smoothing the frequency increments with a rolling window of length L and then looking for the maximum (absolute) RoCoF within a window of $\pm T$ around the full hour.

We choose the values of L and T according to the typical time scale of the RoCoF in the three different synchronous areas. The average hourly evolution of the absolute frequency deviation indicates this time scale (Figure S4). In Continental Europe and Great Britain, the average deviation reaches its maximum 60 s after the full hour, while the Nordic area exhibits its peak already after 30 s. We thus choose $L = T = 60$ s in the Continental Europe and Great Britain areas, but a shorter time scale of $L = T = 30$ s in the Nordic grid area.

Ex-post	Ramps [MW/h]	Load ramp, Total generation ramp, Biomass ramp, Coal gas ramp, Fossil peat 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
	Generation and load [MW]	Load, Total generation, Synchronous generation, Biomass generation, Coal gas generation, Fossil peat 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,
	Forecast errors of generation and load [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
Day-ahead	Generation and load [MW]	Load day-ahead, Scheduled generation, 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 ramp day-ahead [Currency/MWh/h], Prices day-ahead [Currency/MWh], Hour, Weekday, Month

Table S1. All external features in the data set. The units correspond to those used in our publicly available data set.[2]

Area	Number of features	Number of data points
Continental Europe	64	26857
Nordic	58	37154
Great Britain	50	43240

Table S2. Properties of our data sets.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES S3:
BASIC CORRELATION ANALYSIS

A basic correlation analysis between external features and frequency stability indicators can already reveal interesting dependencies. However, this model-agnostic correlation analysis does not account for correlations among the features, which might affect the correlation coefficient between a feature and the stability indicator. Following the main text, we demonstrate this for the effect of nuclear ramps on the RoCoF in Continental Europe.

As depicted in Figure S5, there are various strong correlations between features in all three grid areas. For example, nuclear power generation is positively correlated with the load in Continental Europe. In Figure S6, the features are correlated with our four stability indicators. We observe that nuclear ramps are positively correlated with the RoCoF in Continental Europe, which is not consistent with our SHAP results (see main text). We can explain the positive correlation of nuclear ramps with the hidden relationships to other variables, such as load ramps. Load ramps have a positive effect on the RoCoF in Continental Europe (Figure S10). Due to the strong correlation between load and nuclear power generation (Figure S5) the effect of load ramps can thus "leak" into the correlation coefficients of nuclear ramps. This can explain why we observe a positive correlation between nuclear ramps and the RoCoF in Continental Europe, although nuclear ramps are RoCoF-offsetting in this area, as revealed by SHAP analysis in the main text.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES S4:
DETERMINISTIC FREQUENCY DEVIATIONS

Deterministic frequency deviations (DFDs) occur at the beginning of electricity trading intervals.[4] The generation is adapted in a step-wise manner at the beginning of these intervals, which are mostly hourly time periods. The mismatch between the step-wise generation and the continuously evolving load generates an instantaneous power imbalance, which causes a deterministic frequency jump at the beginning of the hour.

Such DFDs are an important factor for frequency stability in Continental Europe. This is indicated by the time within the hour when the absolute frequency deviation peaks ("Nadir occurrence time"). Figure S12 shows the histograms of these Nadir occurrence times within the hour. In Continental Europe, most of the Nadirs occur in the
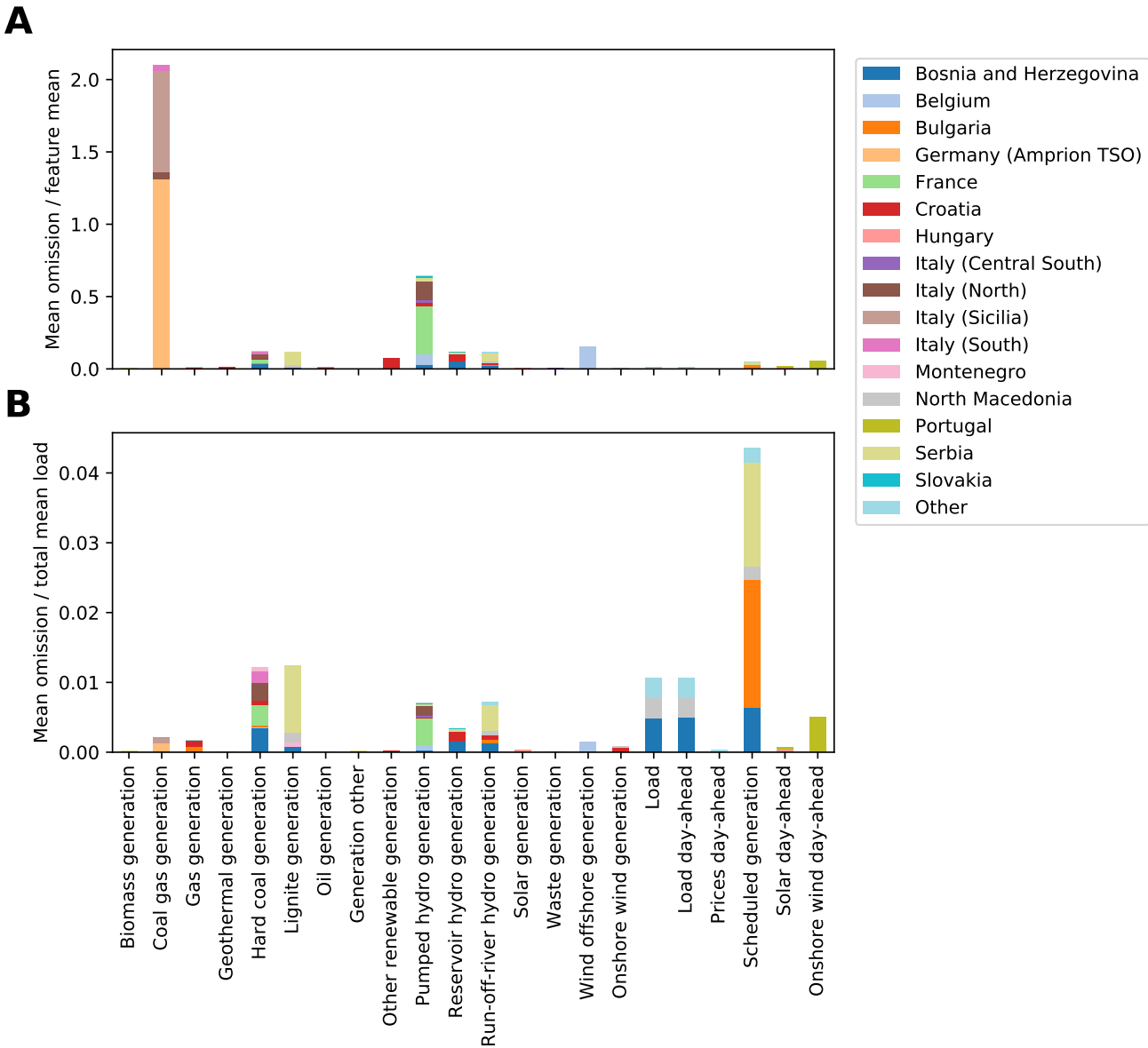
A

B

Figure S1. Omitted data contributions in Continental Europe. (a): We quantify the omitted values with the mean omitted feature value relative to the mean of the included features. The legend indicates the regions (mostly countries) where the omitted contributions come from. Regions with omitted (relative) contributions below 0.8% are aggregated in the "Other" variable. (b) The mean omitted feature value relative to the total mean load of the synchronous area remains below 5%.

first five minutes, which indicates their strong connection to the deterministic electricity trading. In contrast, large deviations in Great Britain occur much more often during the hour and not only at the beginning. This indicates that DFDs play a smaller role in Great Britain than in Continental Europe. The Nordic area is in between, showing strong hourly DFDs as well as Nadirs within the hour.

An extension of the work presented here could consider a threshold for the nadir and turn our regression machine learning task into a classification taks: Will the nadir of the next hour be above or below the security threshold?
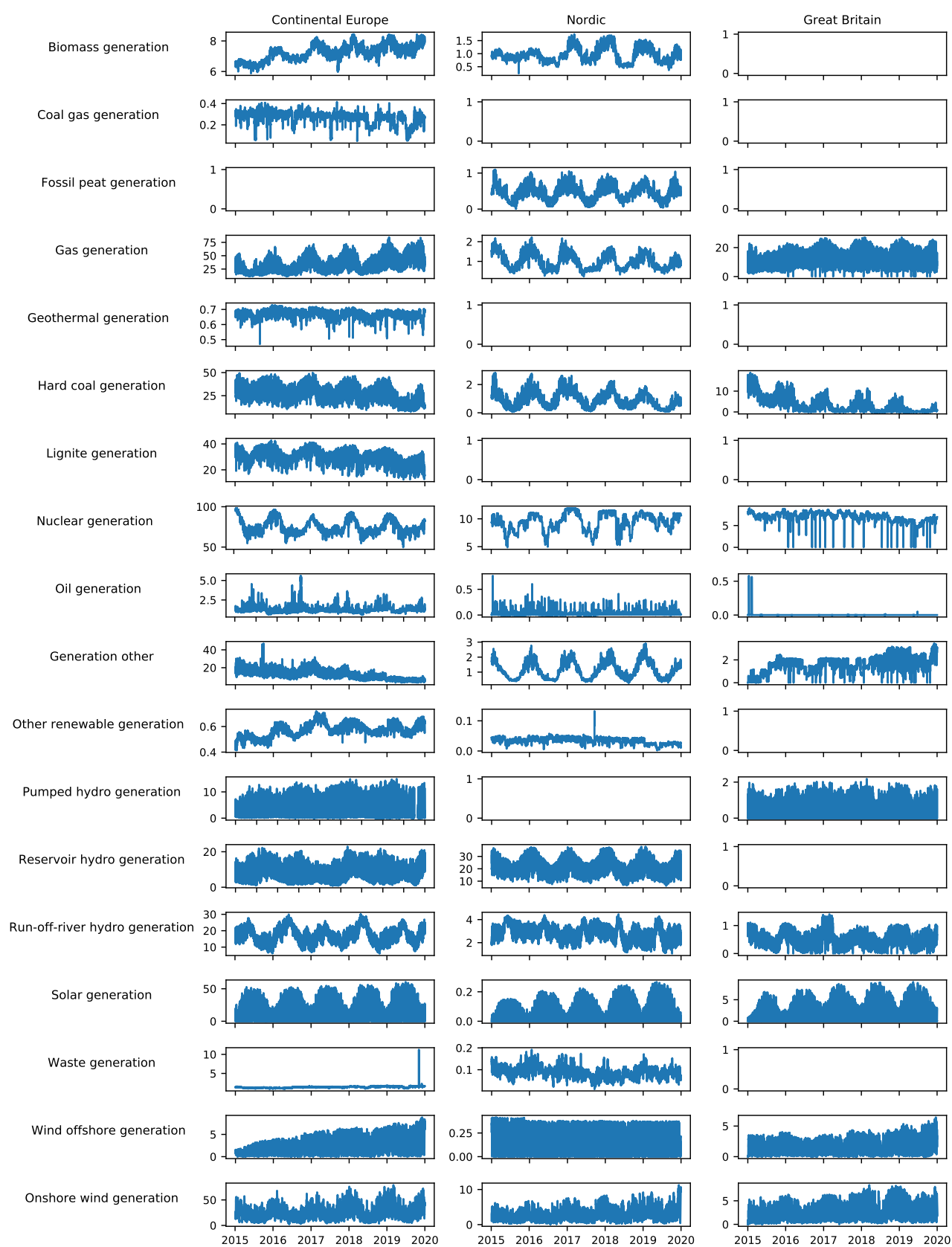
Figure S2. All (aggregated) time series from the ENTSO-E transparency platform:[1] Actual generation per type (in GW).
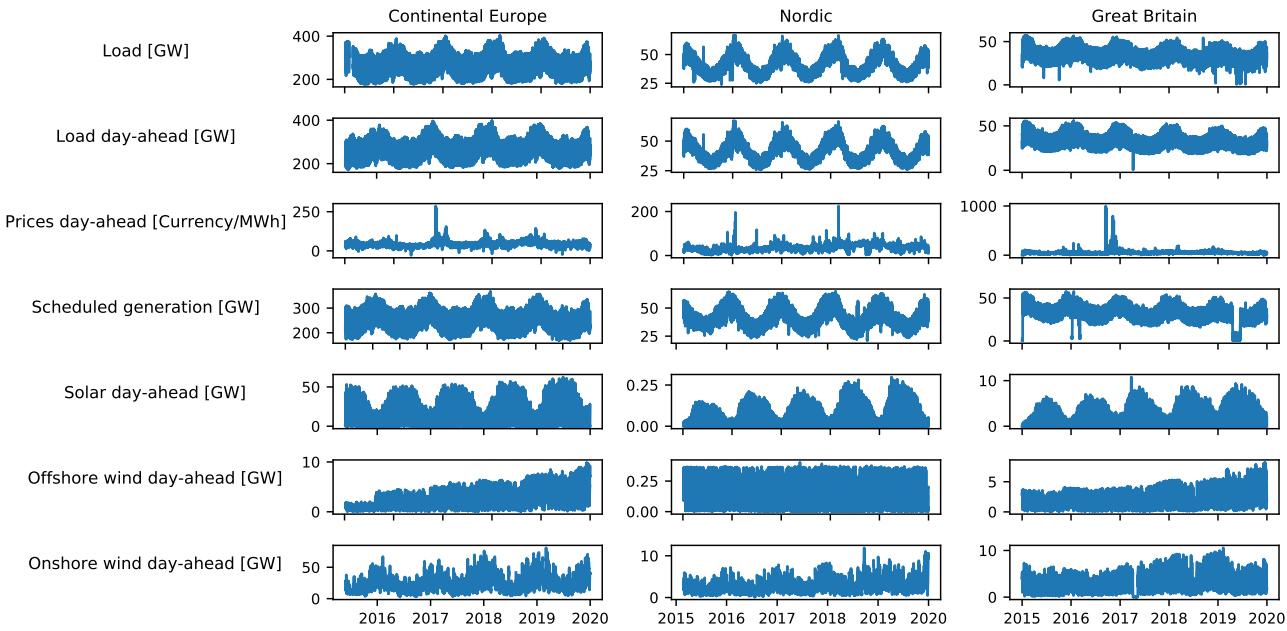
Figure S3. All (aggregated) time series from the ENTSO-E transparency platform:[1] All features except from actual generation per type.
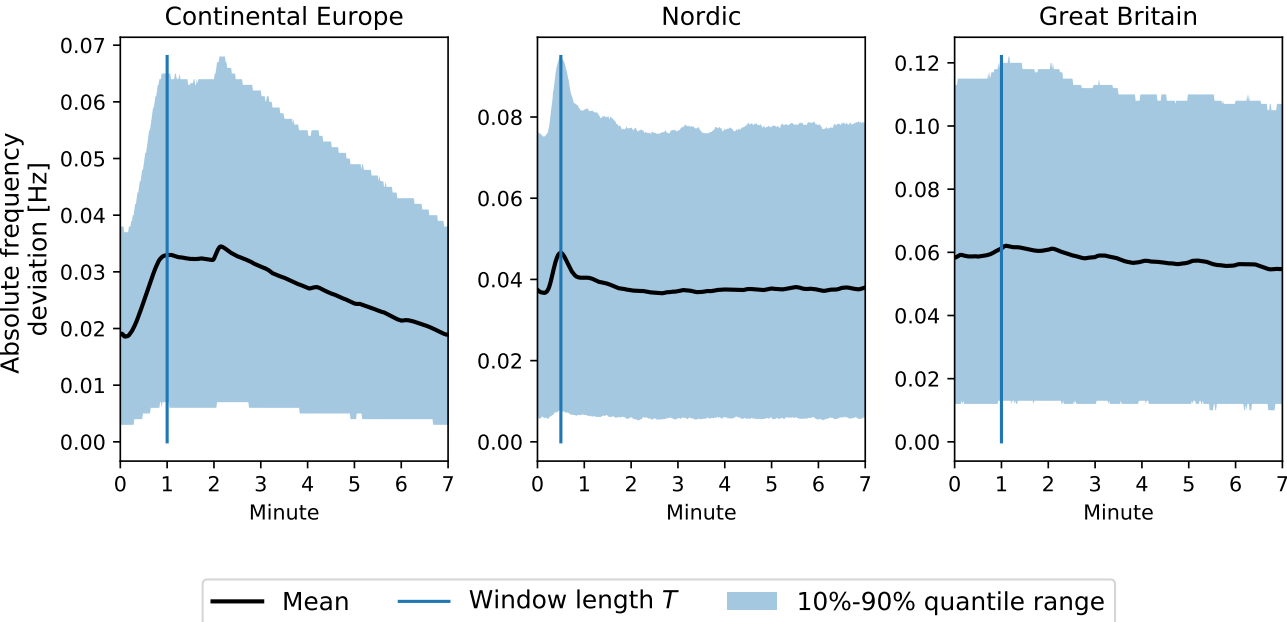

Figure S4. Evolution of hourly absolute frequency deviations. We display the mean evolution of the absolute frequency deviation during the first seven minutes of an hour. The deviation peaks at the beginning of the hour due to the impact of electricity trading.[4] The time scale of this initial increase depends on the grid area. We choose the window length T for the RoCoF extraction according to this time scale.
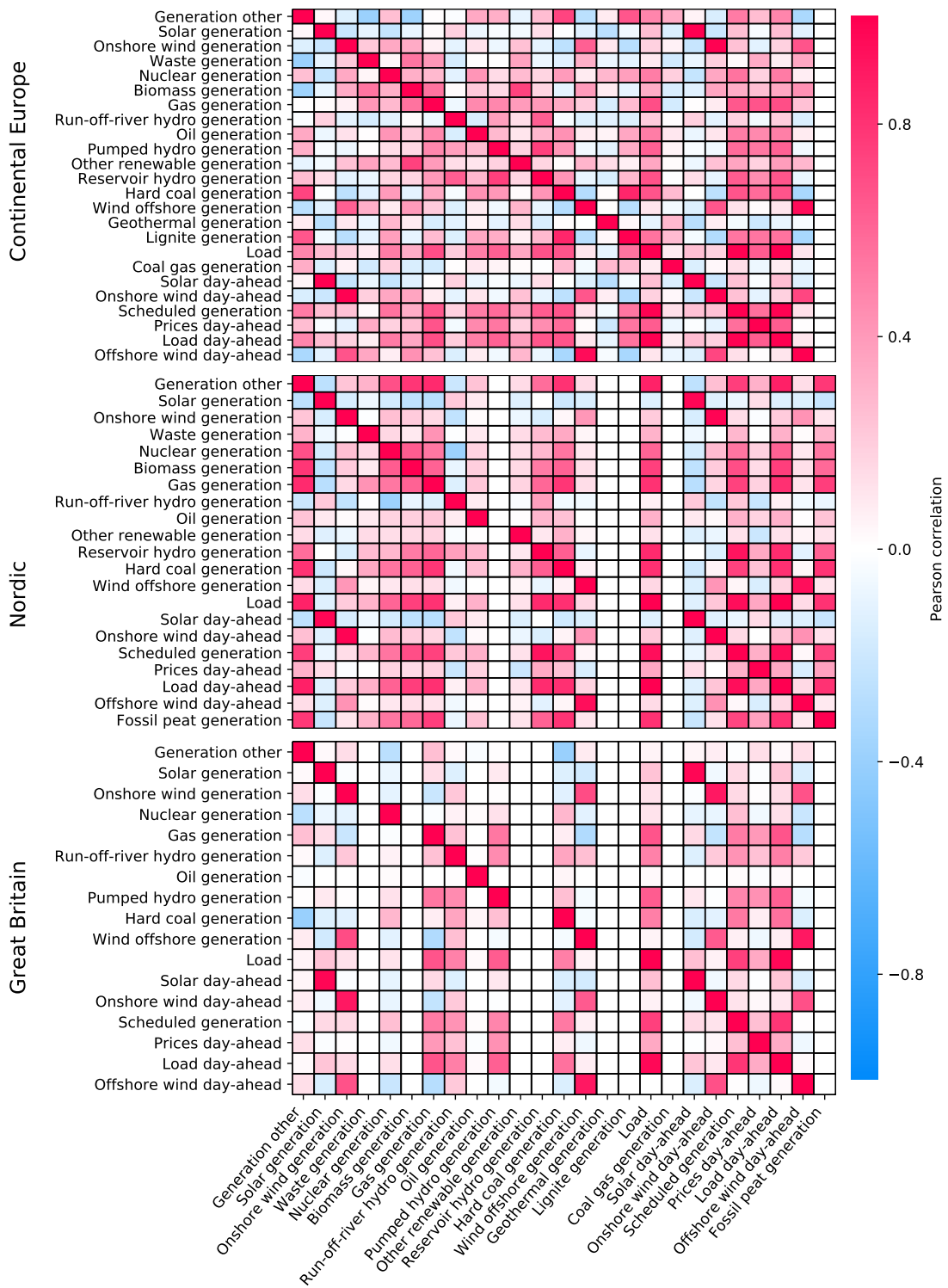
Figure S5. Pearson correlation coefficients between external features. To improve the visibility, we exclude our additional, engineered features from this plot. Features with 0 correlation everywhere have no data in a specific grid (e.g., Fossil peat data is only available in Nordic).
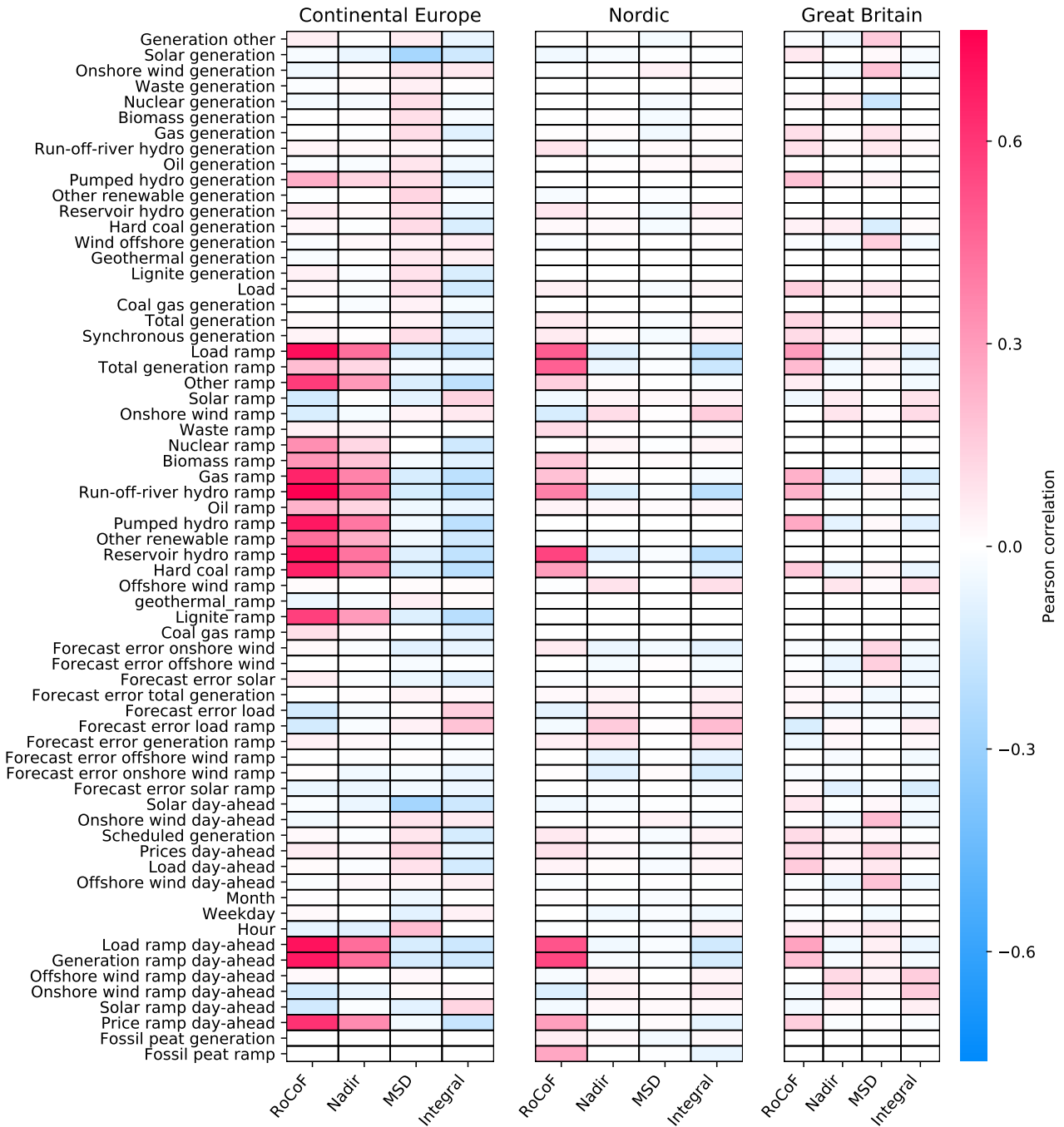
Figure S6. Pearson correlation coefficients between external features and frequency stability indicators. The scale of the colour code is adjusted to the maximum absolute correlation value C_{max} and thus ends at $\pm C_{max}$.

**SUPPLEMENTAL EXPERIMENTAL PROCEDURES S5:
PERFORMANCE EVALUATION OF THE MACHINE LEARNING MODEL**

We evaluate the performance of our Gradient Tree Boosting (GTB) model in terms of the R^2-score, which quantifies the proportion of variability explained by our model. A perfect prediction would result in a score of 1, while predicting the mean of the target results in a score of 0. As a benchmark, we compare the GTB model to the daily profile, which is an important null model for frequency dynamics. In particular, we quantify the gain over the daily profile, which
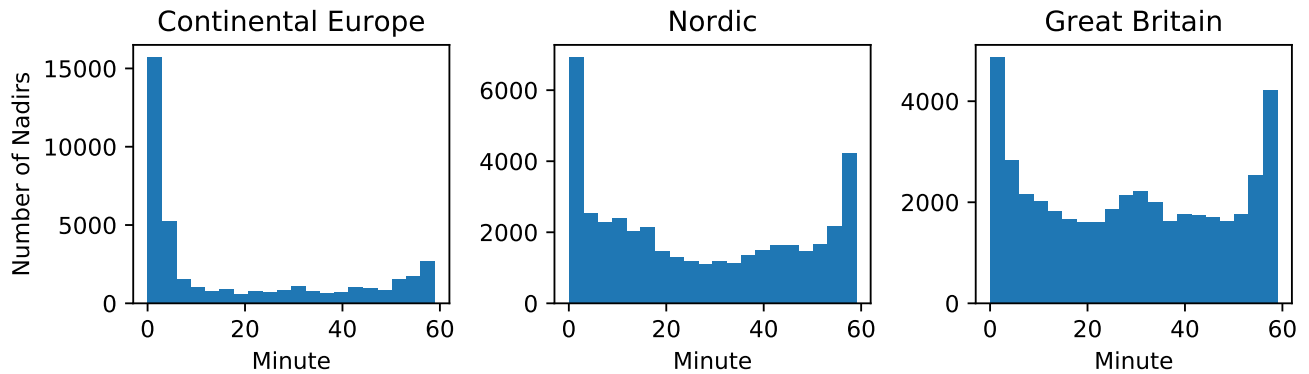
Figure S7. Distribution of Nadir occurrence time. The occurrence time of the nadir is the minute within the hour where the absolute frequency deviation reaches its peak. Its distribution within the hour varies between the grid areas, thus indicating the different importances of DFDs for the grid frequency dynamics.

is the model performance divided by the daily profile performance. Finally, we examine the importance of area-wide feature aggregation and the possibility to predict stability indicators day-ahead. Note that the GTB model used in our main text is referred to as the *full model*. It builds on area-wide aggregated features containing both day-ahead *and* ex-post available data.

The GTB model performs best in Continental Europe, while the performance gain over the daily profile is largest in Great Britain. Figure S8 displays the R^2-score for each stability indicator and each area. We obtain the best predictions in Continental Europe ($R^2 \sim 0.7$) and the lowest scores in Great Britain. In contrast, the performance gain over the daily profile is largest in Great Britain (maximum 16.2) and smallest in Continental Europe (maximum 3.4), while the Nordic area is in between (maximum 7.6). Frequency dynamics in Continental Europe are rather deterministic compared to the stochastic dynamics in Great Britain. Therefore, the prediction is easier and the additional gain through Machine Learning is smaller in Continental Europe. Consistently, the GTB performance is best for the RoCoF as this indicator most strongly reflects the hourly deterministic frequency jumps.

The model performance depends on whether we choose area-wide aggregated features or country-level data (Figure S8). In the Nordic area, we obtain a lower performance when using data from only the largest country (Sweden) instead of aggregating it area-wide. In Continental Europe, the largest country model (using Germany) performs similar to or worse than the aggregated model, and choosing a smaller country (Switzerland) reduces the performance even further. As the grid frequency is affected by all locations within the grid, it is not surprising that data aggregation is important. Overall, the area-wide feature aggregation yields better results than regional data among the areas.

Using only day-ahead available data in our GTB model already outperforms the daily profile for all stability indicators and areas (Figure S9). In Great Britain, the day-ahead model exhibits the strongest performance gain over the daily profile (maximum 8.9), followed by the Nordic area (maximum 3.0) and Continental Europe (maximum 2.6). However, adding ex-post data in the full model can strongly improve the performance, especially in the Nordic area. We quantify this effect in terms of the gain over the day-ahead model, i.e., the full model performance divided by the day-ahead model performance. In the Nordic area, the gain of the full model over the day-ahead model is the largest (maximum 2.6), while it is lowest in Continental Europe (maximum 1.4). The benefits of adding ex-post data in the Nordic area stems from the importance of forecasting errors for the prediction (see main text).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES S6: ADDITIONAL RESULTS WITH SHAP VALUES

We use SHAP values to explain our Machine Learning model for frequency stability indicators. An overview of the most important SHAP dependencies in our model is available in Figures S10, S11, S12 and S13 for each of the four stability indicators. In many cases, we observe non-linear dependencies, which underlines the importance of using a non-linear, complex Machine Learning model such as Gradient Tree Boosting.

Here, we further discuss the effect of synchronous generation on frequency stability indicators. The (total) synchronous generation, which we use as a proxy for the total inertia within the power grid, is not among the eight most important features (Figures S10, S11, S12 and S13). Its feature importance is ranked on places between 19 and 64 in CE, between 38 and 52 in the Nordic area and between 25 and 40 in GB. Overall, the average effect of the (approxi-
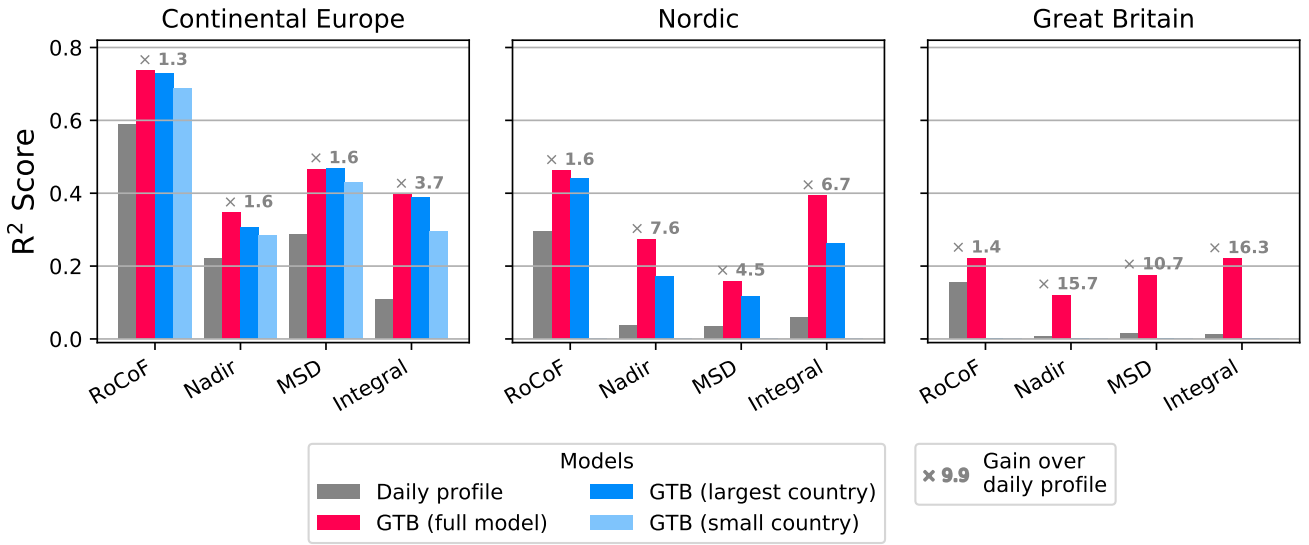
Figure S8. Performance of the full GTB model. The numbers on the bar plots indicate the gain of the GTB model over the daily profile. The full model comprises day-ahead and ex-post available features, which are aggregated area-wide. The country-level models use features from the largest country, i.e., the country with the largest power demand. In Continental Europe, we additionally introduce a model using data from a small country (with smaller average load), which is Switzerland. We retrieve the country-level data from the ENTSO-E transparency platform[1] and construct the same features as in the full model (Table S1).
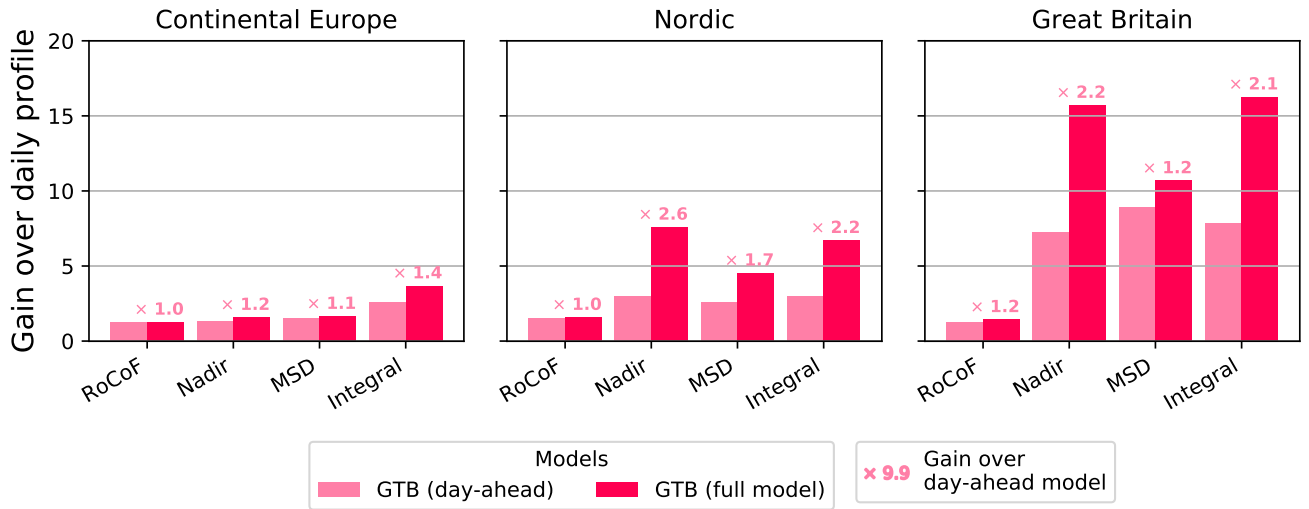

Figure S9. Performance of the day-ahead GTB model. We compare the full model and the day-ahead model in terms of their gain over the daily profile. The numbers on the bar plots indicate the gain of the full model over the day-ahead model.

mated) inertia on the aggregated stability indicators is thus relatively low compared to the most important features. Among the areas, the total synchronous generation is most important in Great Britain, as the feature is constantly among the 40 most important variables. This is consistent with the high share of renewable energy sources in the British power system and the resulting low-inertia situations.[3] The effect of the inertia in Great Britain is depicted in the dependency plots of Figure S14. For all stability indicators, we observe the maximum (absolute) effect of the synchronous generation at low feature values. This is particularly evident for the MSD and the Integral, where the effect of values larger than 20 GW is near to zero. In conclusion, the (approximated) inertia mostly affects frequency stability in Great Britain in extreme situations of low inertia, but the average effect of this feature on our aggregated stability indicators is negligible.

Figure S10. RoCoF dependency plots for the eight most important features.
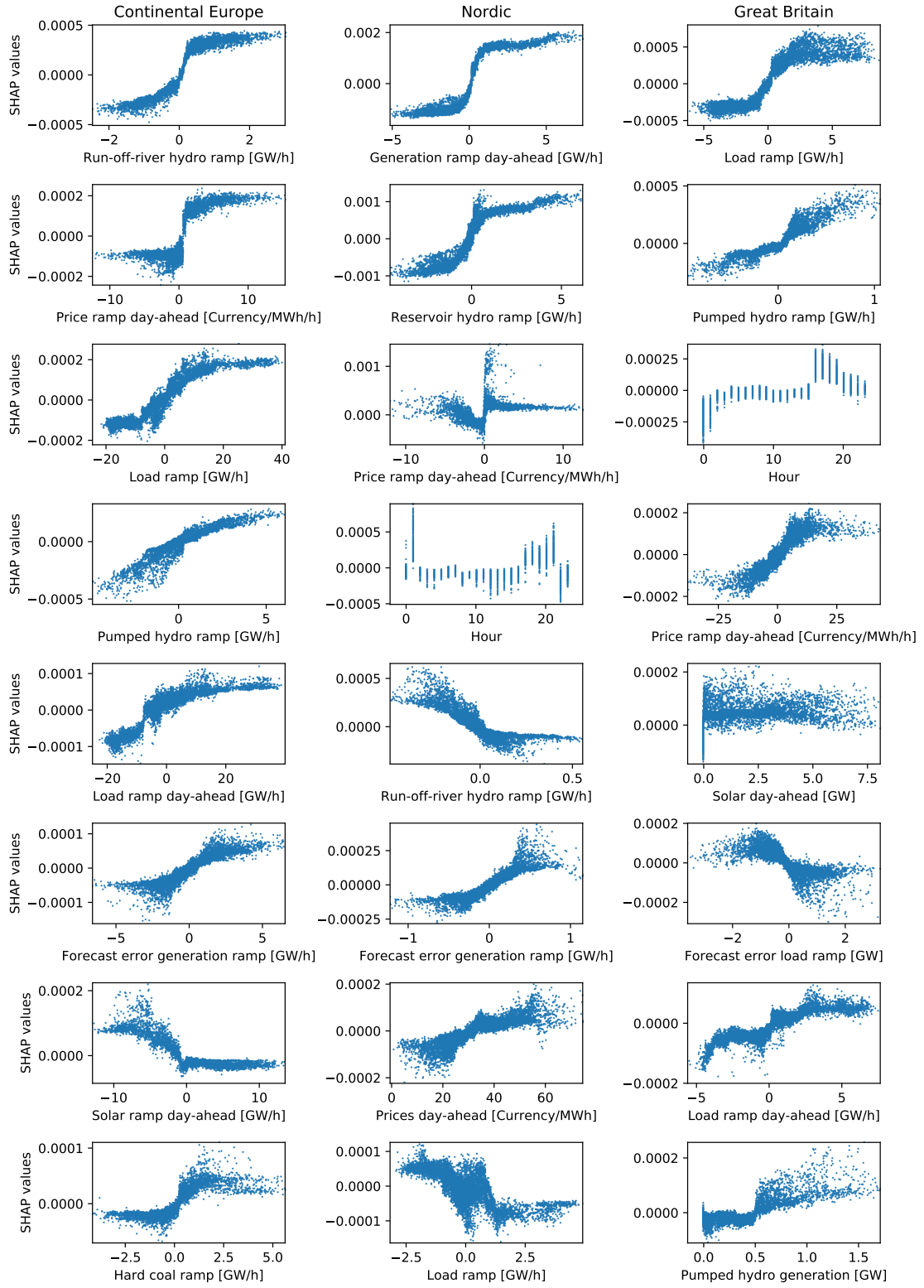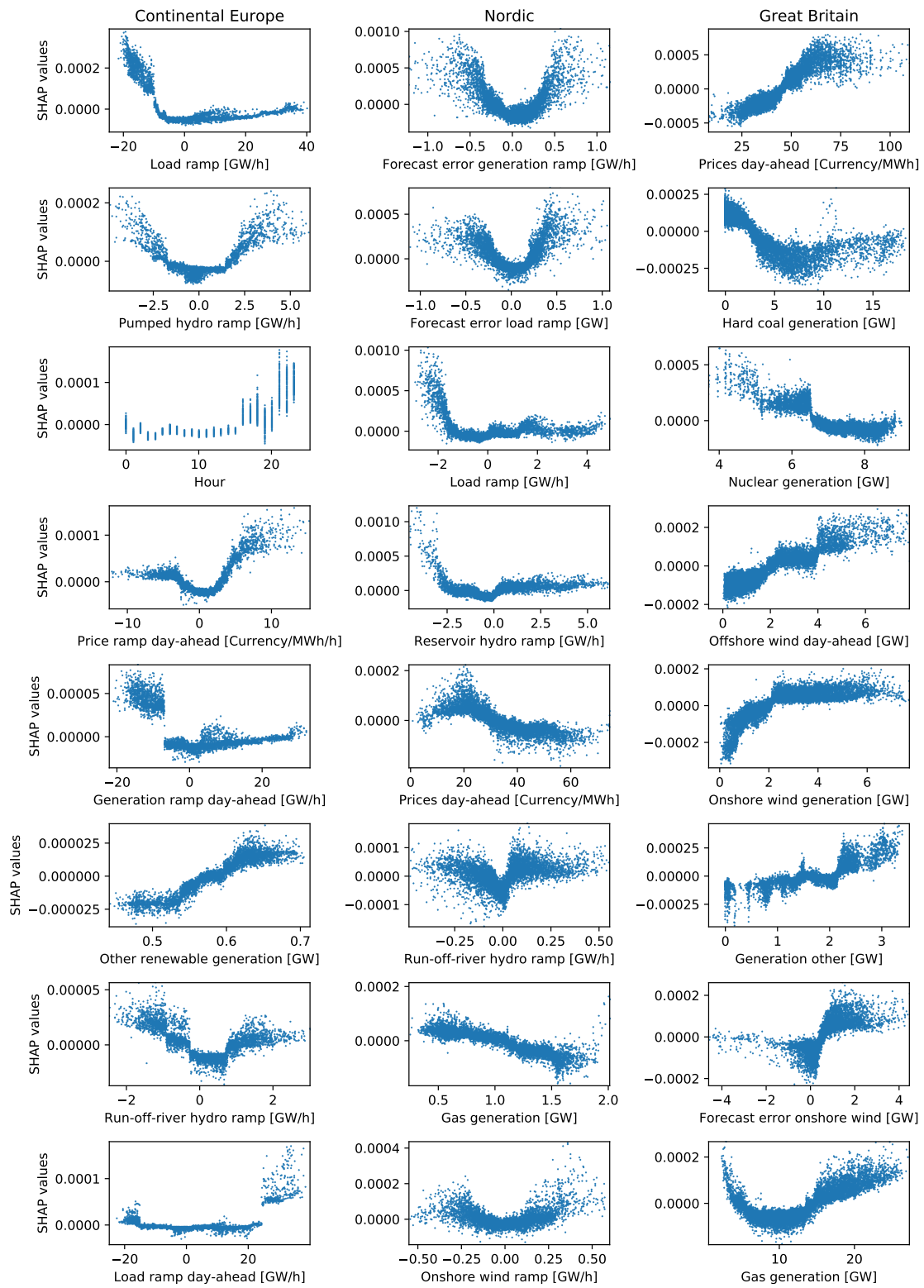
Figure S11. MSD dependency plots for the eight most important features.
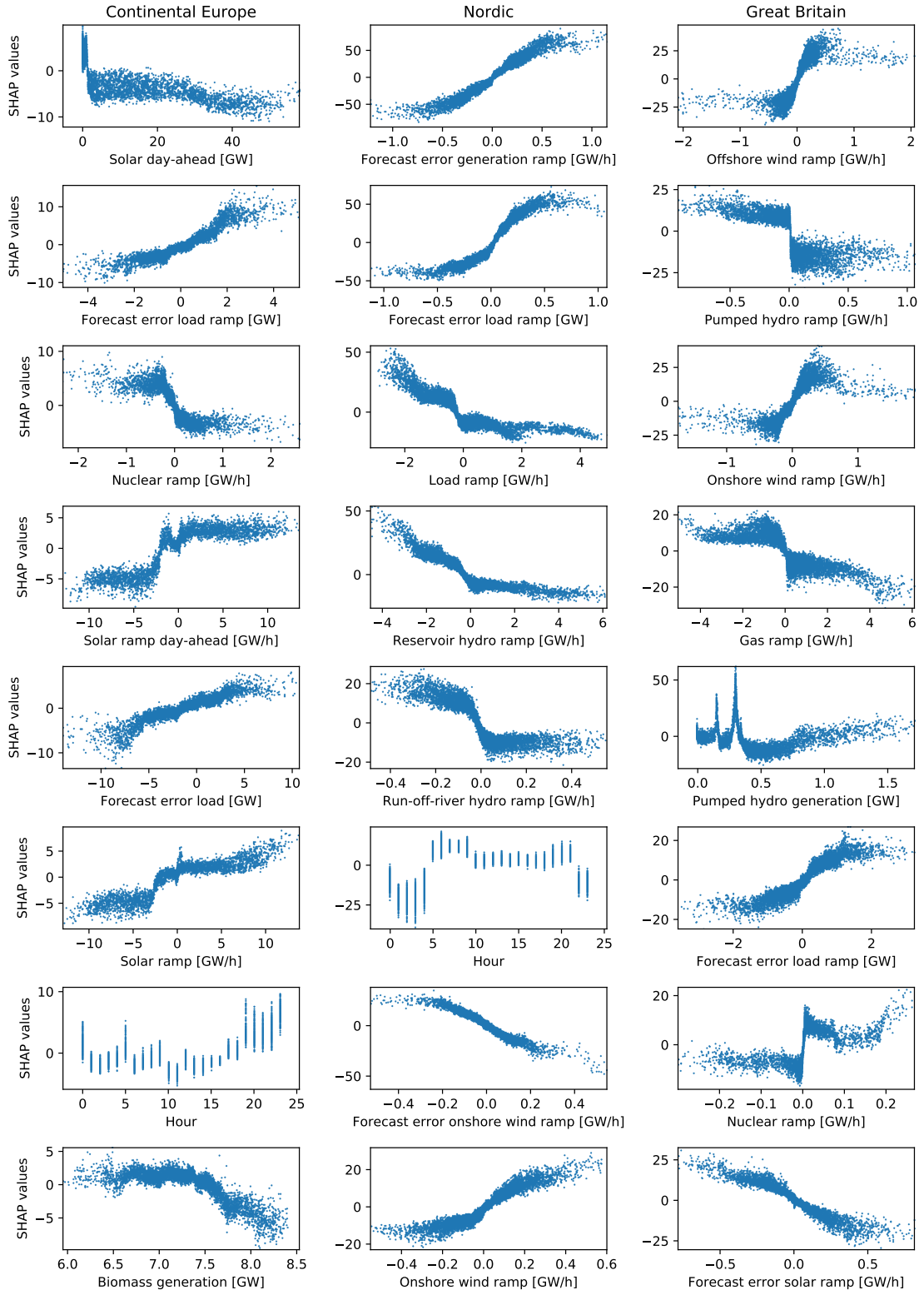
Figure S12. Nadir dependency plots for the eight most important features.
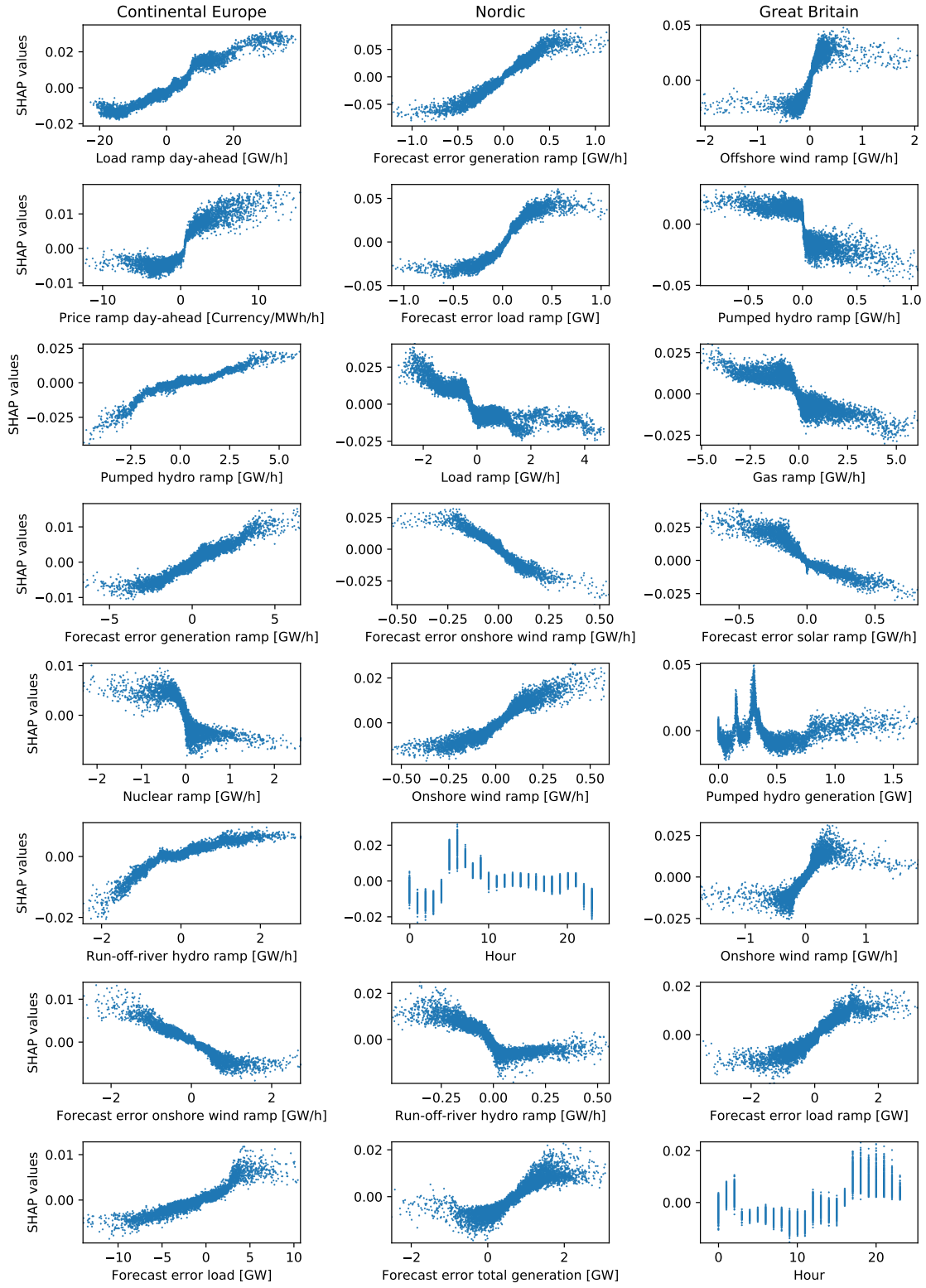
Figure S13. Integral dependency plots for the eight most important features.
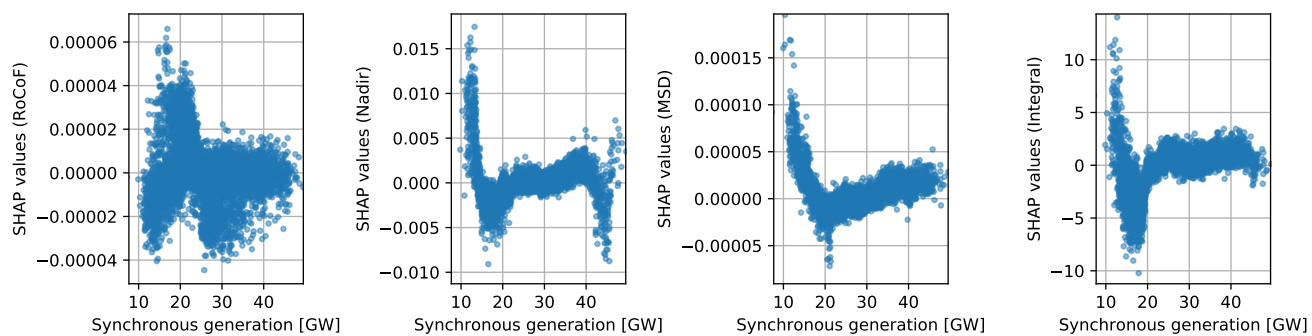
Figure S14. Dependency plots of synchronous generation in Great Britain. Each subplot corresponds to one of our aggregated indicators for frequency stability.

SUPPLEMENTAL REFERENCES

[1] *ENTSO-E Transparency Platform* 2020. `https://transparency.entsoe.eu/`.

[2] Kruse, J., Schäfer, B. and Witthaut, D. 2021. Supplementary data: "Revealing drivers and risks for power grid frequency stability with explainable AI", `https://zenodo.org/record/5118352`.

[3] Milano, F., Dörfler, F., Hug, G., Hill, D. J. and Verbič, G. 2018. Foundations and Challenges of Low-Inertia Systems (Invited Paper), *2018 Power Systems Computation Conference (PSCC)*, IEEE, Dublin, pp. 1–25.

[4] Weissbach, T. and Welfonder, E. 2009. High frequency deviations within the European Power System: Origins and proposals for improvement, *2009 IEEE/PES Power Systems Conference and Exposition*, IEEE, Seattle, pp. 1–6.

3.2. C) Exploring deterministic frequency deviations with explainable AI

[3] Kruse, J., Schäfer, B. & Witthaut, D. *Exploring deterministic frequency deviations with explainable AI* in *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)* (2021), 133–139. doi:10.1109/SmartGridComm51999.2021.9632335.

Exploring deterministic frequency deviations with explainable AI

Johannes Kruse*[†§], Benjamin Schäfer[‡¶], Dirk Witthaut*[†]

*Forschungszentrum Jülich, Institute for Energy and Climate Research - Systems Analysis and Technology Evaluation
(IEK-STE), 52428 Jülich, Germany
[†]Institute for Theoretical Physics, University of Cologne, 50937 Köln, Germany
[‡]School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom
[¶]Faculty of Science and Technology, Norwegian University of Life Sciences, 1432 Ås, Norway
[§] Email: jo.kruse@fz-juelich.de

Abstract—**Deterministic frequency deviations (DFDs) critically affect power grid frequency quality and power system stability. A better understanding of these events is urgently needed as frequency deviations raise the need for substantial control actions and thereby increase cost of operation. DFDs are partially explained by the rapid adjustment of power generation following the intervals of electricity trading, but this intuitive picture fails especially before and around noonday. In this article, we provide a detailed analysis of DFDs and their relation to external features using methods from eXplainable Artificial Intelligence. We establish a machine learning model that well describes the daily cycle of DFDs and elucidate key interdependencies using SHapley Additive exPlanations. Thereby, we identify solar ramps as critical to explain patterns in the Rate of Change of Frequency (RoCoF).**

I. INTRODUCTION

The balance of power generation and demand is central for the stability of our power grids. The grid frequency reflects this balance, since an overproduction of power leads to a rise and an under-supply to a drop of the frequency [1]. The need for power balancing therefore translates into constraints on the grid frequency to limit large deviations from the set point of 50/60 Hz through adequate control measures. This is critical as large frequency deviations, such as deviations of more than 200 mHz in Continental Europe, can trigger the disconnection of loads with severe consequences for the power consumers.

An important threat to frequency stability results from deterministic frequency deviation (DFDs). DFDs occur regularly at the beginning of hourly or sub-hourly intervals and can be observed in various large-scale power grids such as the Continental European [2], the Great Britain [3] and the Nordic synchronous areas [4]. Large DFDs lead to a depletion of frequency control reserves at the beginning of hourly or sub-hourly intervals, thus making the system vulnerable to addi-

tional, unforeseen disturbances or failures [5]. For example, the combination of a DFD and a measurement failure led to an extreme deviation of nearly 200 mHz on the 10th of January 2019, in Continental Europe. In this area, the number of large frequency deviations has increased constantly thus requiring an intensified control of large DFDs [5].

A common model explains DFDs through the effect of scheduled changes in the power generation due to block-wise electricity trading [6]. In this view, the direction of DFDs during the day is often related to the load ramp, i.e., to the slope of the load curve [4], [6], [7] and a connection between large frequency events and the load ramp is suggested [3], [8]. However, not only load ramps but many other external features such as electricity prices or forecasting errors affect power grid frequency deviations [9].

Methods from Machine Learning (ML) are excellent candidates to model the complex effects of multiple features on power grid frequency fluctuations. Due to control measures, frequency dynamics exhibit complex non-linear dependencies [1] and measurement errors further modify many publicly available grid frequency measurements [10]. Moreover, drivers of frequency deviations such as load or generation ramps are strongly correlated [9]. Modern ML methods, such as Gradient Tree Boosting, are able to extract non-linear dependencies even from noisy data with correlated features [11], which is not possible in a simple correlation analysis. They can further harness the growing amount of power system data, which has been made publicly available in the past years [12], [13]. However, complex ML models are often black-boxes, which impede a scientific understanding of the model structure [14].

Tools from eXplainable Artificial Intelligence (XAI) enable us to understand and visualise the dependencies captured by the ML model [15]. In particular, the recently introduced SHapely Additive exPlanations (SHAP) values offer a numerically efficient way to quantify the impact of different features on the model output [16]. Within the set of methods to measure feature effects, SHAP values guarantee certain optimal properties and avoid inconsistencies within other approaches [17], [18]. Combining non-linear ML models and SHAP values thus enables us to examine the effect of multiple external features
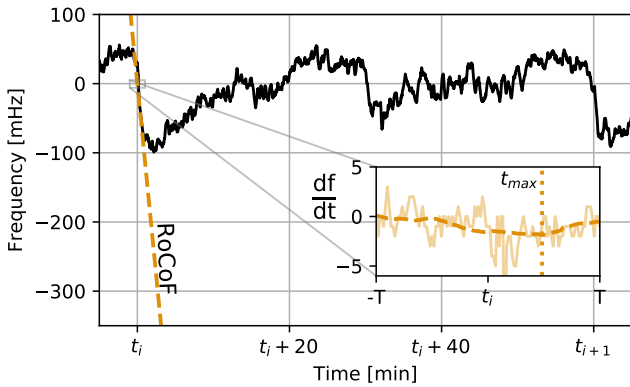
Fig. 1. Quantifying DFDs with the hourly RoCoF. The grid frequency exhibits regular jumps, particularly at the beginning of the hour t_i, which are deterministic frequency deviations (DFDs). To measure the DFD at the full hour t_i, we calculate the corresponding Rate of Change of Frequency (RoCoF) from the grid frequency time series $f(t)$ (see inset). We first estimate the derivative $\mathrm{d}f/\mathrm{d}t$ by extracting and smoothing the increments $f(t+1\mathrm{s})-f(t)$ (dashed orange line in the inset). We then evaluate the maximum slope in an interval around the full hour (see text for details). The figure depicts the grid frequency from 10 to 11 p.m. on April 14, 2017, in Continental Europe [19].

on DFDs in a consistent way.

Here, we explore hourly DFDs in the Continental European (CE) synchronous area with an explainable ML model introduced in ref. [9]. The ML model predicts hourly DFDs from 54 different features such as the actual load and day-ahead electricity prices. In combination with SHAP values, it offers a versatile tool for the ex-post analysis of DFDs and their drivers. In Section II, we first introduce a common load-based model of DFDs, which is based on the effect of scheduled-based generation, and then continue with our ML model for DFDs. In Section III, we first evaluate how well the load-based DFD model reproduces the daily pattern of DFDs in CE. Then, we explore the deficits of the load-based model with SHAP values and finally discuss a refined physical view of DFDs, which integrates the insights from our explainable ML.

II. METHODS

A. The problem of deterministic frequency deviations (DFDs)

DFDs are regular deviations of the grid frequency that occur at the beginning of hourly or sub-hourly intervals. For instance, the grid frequency sample from an evening hour in CE (Fig. 1) exhibits a steep slope at the full hour and after 30 minutes. These large slopes even persist after averaging over multiple days, which demonstrates their deterministic nature [6]. In particular, the direction of the slope shows a regular daily profile. In the morning, the deviations typically point upward and in the evening we typically observe negative slopes. We quantify the DFDs via the Rate of Change of Frequency (RoCoF) at the beginning of the hour (Fig. 1).

Within this study, we focus on the CE synchronous area, which is the largest synchronous area in Europe, covering 26 countries, hundreds of millions of customers and spanning

from Portugal to Turkey and from Italy to Denmark. Numerous Transmission System Operators (TSOs) operate the CE area and they coordinate via the European Network of Transmission System Operators for Electricity (ENTSO-E), where for example grid codes are defined and data are made available via a transparency platform [20].

DFDs are extensively observed in the CE grid and have been explained as a mismatch in supply and demand: As electricity is traded in blocks of fixed time, the smooth demand curve is typically approximated by a step-wise generation curve, thereby causing a particularly large mismatch of scheduled generation and load at the beginning of the trading blocks [6]. Such deterministic deviations are very problematic as they bind Frequency Containment Reserve (FCR), which then is not available in case further unexpected fluctuations, accidents or attacks occur, as highlighted by a special ENTSO-E report [5]. To address this problem, several countries within the CE area have increased the number of dispatch intervals, leading to updates of the scheduled generation every 15 minutes instead of every hour. Still, electricity is mostly traded on an hourly basis and therefore the hourly interval remains the most important time scale for power system operation in CE. Consistently, hourly DFDs continue to dominate the daily profile of the grid frequency [2]. Therefore, we focus on the hourly RoCoF to analyse DFDs in this study.

B. Modelling DFDs

A common model for (hourly) DFDs is based on different temporal behaviour of the load and the generation, which is crucially determined by the electricity market [6]. In Continental Europe, electricity is still mostly traded in hourly blocks, such that power generators adapt their output to the new set point at the beginning of each hour. This leads to a step $\Delta P_L(t_i)$ in the power generation at the beginning of the hour t_i, as shown in Fig. 2a. Since the load varies only slowly, the step-like generation introduces power imbalances around the full hour. For example, at 06:00 the imbalance changes from under- to oversupply due to the positive step ΔP_L. The slope of the resulting frequency jump is then proportional to the generation step [7], which implies a linear model $y_L(t_i)$ for the RoCoF:

$$y_L(t_i) = a \cdot \Delta P_L(t_i). \tag{1}$$

Here, the value of the generation step $\Delta P_L(t)$ is estimated by the load ramp, i.e., the hourly change of the total load $\Delta L(t)$ in CE. This estimate is based on the assumption that most of the load is covered by schedule-based generation. The schedule coincides with the hourly average of the load, such that the generation steps approximate the load ramp. In that way, the direction of the DFDs is inherently correlated to the load ramp, so that we call the above approach the *load-based* model.

To visualise the assumptions of the load-based model, we construct load and generation curves on sub-hourly resolution (Fig. 2a). We design a continuous load curve $L(t)$ with a resolution of one minute by applying a cubic spline interpolation
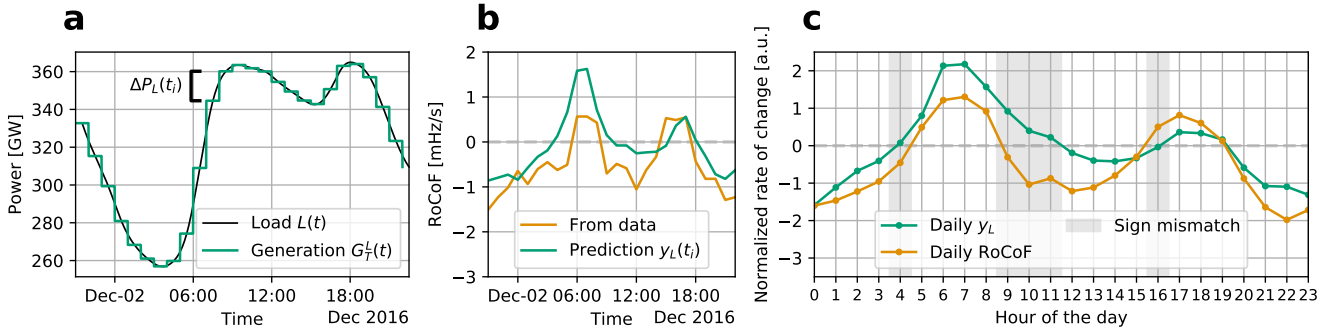
Fig. 2. A common load-based model cannot reproduce the daily pattern of DFDs. a: A common model explains hourly DFDs with the power jump $\Delta P_L(t_i)$ that results from scheduled changes in the generation due to hourly trading on the electricity markets. b: Based on the generation steps ΔP_L, we a estimate a linear model $y_L(t)$ for the hourly RoCoF, which we call the load-based model (1). The prediction reproduces the data only partially and yields an R^2-score of 0.37 on the test set. The sample depicted in panel a and b is included in the test set. c: The load-based model cannot fully reproduce the daily pattern of DFDs, which is indicated by a sign mismatch between the daily average prediction (green) and the data-based RoCoF (orange).

to the load values, which only have an hourly resolution (see Section II-C). The generation values for each minute $G_T^L(t)$ follow a step function with hourly steps ΔP_L.

C. Data sources and preparation

To evaluate the load-based DFD model, we use a set of pre-processed, publicly available time series from the CE power system, which covers the years 2015 to 2019 [21]. The data set contains hourly RoCoFs of the CE grid frequency, as well as multiple hourly-resolved *external features*, such as the load and the power generation. The frequency data was originally recorded by the TSO TransnetBW [19] and the feature data derived from the ENTSO-E Transparency Platform [20].

The hourly RoCof is extracted from grid frequency measurements $f(t)$ using the procedure sketched in Fig. 1. Based on frequency data with 1 s resolution [19], we estimate the derivative df/dt by first extracting the increments $f(t+1s) - f(t)$. Then, we smooth the increments with a rectangular rolling window of length $L = 30$ s. Finally, we identify the time t_{max} in an interval around the full hour $[t_i - T, t_i + T]$, where the absolute derivative $|df/dt|$ reaches its maximum. We then set $\text{RoCoF}(t_i) = df/dt(t_{max})$ using $T = 30s$ (cf. [9]).

The external features serve as input to the load-based model and to our ML model. From the data set of ref. [21], we only include a subset of 54 features into our model. The subset includes the actual load and generation per type, as well as their hourly ramps (i.e., gradients) and forecast errors. Moreover, we include day-ahead prices together with their ramps, and three variables indicating the hour, the weekday and the month. As our focus is on the ex-post explanation of DFDs and not on prediction, we here use the correct post-hoc values of all features and omit day-ahead forecasts which are in principle available in the data set. As the grid frequency is affected by power imbalances in all locations of the synchronous area, the features represent the area-wide aggregated values of the variable. For example, the "load ramp" represents the slope of the aggregated load within the whole CE synchronous area. We refer to ref. [9] for a detailed

description of the data set, including the processing and the aggregation of the external features.

D. Machine learning model for hourly RoCoF

We apply the explainable ML model from ref. [9] to explore the impact of external features on the DFDs. The model uses a Gradient Tree Boosting model [22] for the prediction of hourly RoCoFs from external features, which is then explained through SHAP values.

For the model training and evaluation, we randomly split the data set into a test set (20%) and a training set (80%). In contrast to ref. [9], we additionally include the data of a continuous 24h interval (from December 2016) in our test set to allow for a visualisation of the predicted time series (see Fig. 2b). Moreover, we only use a subset of 54 features as inputs to our ML model (see Section II-C). To optimise the model hyper-parameters, we use 5-fold cross-validation and grid search on the training set. Then, we retrain the optimal model on the whole training set and calculate the ML prediction $y_{ML}(t_i)$ of the RoCoF for every hour t_i in the unseen test set. We evaluate the performance on the test set using the R^2-score, which represents the share of variability in the hourly RoCoF explained by the model. In the same way, we also estimate the linear model (1) on the training set (via least-squares) and evaluate it on the test set. In addition to the R^2-score, we evaluate the model predictions in terms of their daily average profile. A good model should reproduce the daily profile of RoCoFs estimated from the frequency data.

To explain the model output, we finally calculate the SHAP values on the test set [17]. SHAP values quantify the (positive or negative) effect of each feature on the model prediction for each hour t_i relative to the average prediction. SHAP values attribute the feature effects based on Shapely values from game theory and thus exhibit certain optimal properties such as consistency and local accuracy. For example, local accuracy guarantees that SHAP values sum up to the model prediction.

By combining SHAP values of individual hours t_i, we can also understand global relations in the model. For example, dependency plots depict the relation between features and targets, e.g., between the load and the hourly DFDs [17]. In addition, we explore daily aggregated SHAP values that explain the daily average profile of the model prediction [9]. The daily aggregated SHAP values build on the additivity (local accuracy) of SHAP. They reflect the average impact of a feature on the RoCoF at a certain hour of the day and are therefore perfectly suited to explain the daily DFD pattern. For details on the ML model training, performance evaluation and model explanation via SHAP values we refer to ref. [9].

E. Data and code availability

The input and output data for our model is publicly available on Zenodo [21] and the python code for this study can be obtained from GitHub [23].

III. RESULTS

A. Evaluation of load-based DFD model

The load-based predictor $y_L(t_i)$ (1) partially explains the direction of DFDs, but its overall performance is limited. Fig. 2b depicts the model prediction for a continuous time interval within the test set. The prediction reproduces the general trend of the RoCoF with upward jumps in the morning and in the afternoon, and downward jumps around noonday and during the night. However, the R^2-score only yields a value of 0.38, which reflects the mismatch of data and prediction in Fig. 2b.

Evaluating the daily profile of the prediction confirms this observation (Fig. 2c). The overall shape of the daily RoCoF profile is reproduced, but the direction of the jumps does not align with the sign of the predicted RoCoF in 5 of 24 hours. In particular, the load is increasing on average between 09:00 and 11:00, but the average RoCoF is negative within these hours. We observe such a sign mismatch also at 04:00 and 16:00. The direction of the load ramp thus cannot fully explain the direction of the DFDs within the day.

Note, that adding a constant bias to the linear model (1) increases the performance to $R^2 = 0.52$. However, the prediction is merely shifted and its daily profile still exhibits the wrong sign in 5 hours within the day (not shown in the figure). Consequently, an additional bias does not improve the explanation of the daily DFD pattern through the load-based model.

B. Exploring DFDs with XAI

We explain the deficits of the load-based model by exploring the daily DFDs with our explainable ML model (Fig. 3). As shown in Fig. 3b, the ML prediction $y_{ML}(t_i)$ approximates the RoCoF in nearly every hour of the sample, which is consistent with the high R^2-score of 0.73 on the test set. We obtain an even higher correspondence for the daily profile (Fig. 3a). The ML prediction (solid line) reproduces the daily RoCoF profile (dashed line) nearly perfectly. The model thus captures dependencies that are important to predict and explain the hourly RoCoF, and in particular the daily pattern of DFDs.

Using our ML model explanation, daily aggregated SHAP values reveal the impact of external features on the daily DFD pattern (Fig. 3a). The SHAP values describe the contribution of each feature on the model prediction both in magnitude and sign. In the figure, the daily SHAP values are represented by coloured areas. Areas below the prediction line (solid) indicate a positive effect of the feature on the RoCoF in that hour, while areas below the line reflect a negative impact. We observe a positive impact of the load ramp on the RoCoF in the morning between 04:00 and 11:00, where the load is still rising (see Fig. 2c). This is in line with the load-based model, which predicts a positive RoCoF within this period due to the rising load. However, the actual RoCoF becomes negative between 09:00 and 11:00. Here, the contribution of the load ramp is still positive, but other features now have a strong impact on the model outcome. In fact, the ML model mainly attributes this decrease to the negative impact of hydro power ramps, which leads to a correct prediction of the downwards DFD direction.

To further explain the contradiction between the ML prediction and the load-based model in the morning, we connect the ML prediction to the daily evolution of solar and hydro power (Fig. 3c). Solar power ramps up slowly between 05:00 and 08:00. As the load already increases strongly during this time (see Fig. 2c), fast conventional generators, such as pumped hydro, ramp up to cover the load. Between 09:00 and 11:00 the load is still rising, but solar power starts to ramp up strongly and pumped hydro thus has to ramp down. As solar power ramps are still slow compared to hydro ramps, the generation step at the beginning of the hour is dominated by fast negative hydro ramps. Their direction is opposite to the load ramp, which is not captured in the load-based model. However, the ML model integrates this effect as depicted in the dependency plot Fig. 3d, which displays the SHAP values together with the feature value. Negative hydro ramps have a strong negative impact on the predicted RoCoF. The ML model integrates this effect (Fig. 3a) and thus reproduces the direction of the DFD correctly.

In the afternoon, fast generation ramps and a smoothing effect of solar power explain the deficits of the load-based model. At 16:00 the load does not change strongly and the load-based model predicts a slightly negative RoCoF (see Fig. 2c). However, solar power ramps down strongly in that time, so that flexible generators, such as pumped hydro, ramps up to cover the load (Fig. 3c). The ML model captures this effect and the daily SHAP values in Fig. 3a clearly show the positive effect of hydro power at 16:00, which leads to an upward DFD at this time. Interestingly, we also observe large positive SHAP values of solar ramps between 15:00 and 17:00. According to the dependency in Fig. 3e, negative solar power ramps strongly increase the RoCoF, while positive ramps have a small negative effect. The impact of negative ramps most probably relates to a smoothing effect solar power. Solar power
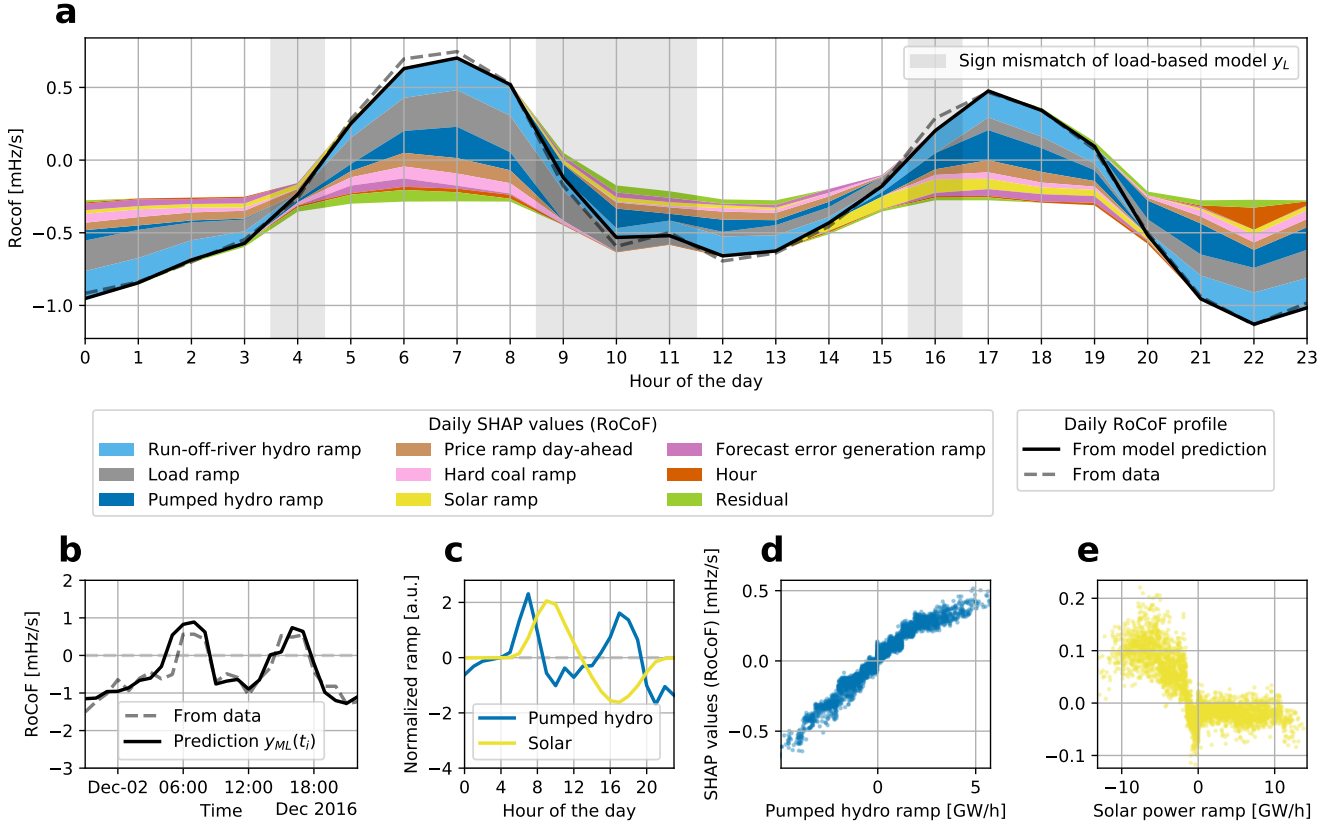
Fig. 3. Revealing the effect of external features on daily DFDs with machine learning. a: We model the hourly RoCoF with an ML model based on external features, such as load ramps or forecast errors. By interpreting the model with SHAP, we can calculate daily (aggregated) SHAP values that explain the average impact of a feature on the model prediction for each hour of the day (coloured areas). For example, load ramps (grey) have a positive impact on the predicted RoCoF (solid line) at 07:00 and a negative impact at 22:00. b: The ML prediction $y_{ML}(t_i)$ predicts the hourly RoCoF with high accuracy in this sample, yielding an overall performance of $R^2 = 0.73$ on the test set. c: Driven by the SHAP results, we explain the sign mismatch of the load-based model with the daily ramps of solar power and fast generation, such as pumped hydro. d-e: Pumped hydro ramps have a positive relation to the RoCoF and solar ramps a non-linear dependency. The strong effect of negative pumped hydro ramps thus explains the deficits of the load-based model between 09:00 and 11:00 in the morning.

ramps down slowly thus following the load more smoothly than scheduled generation, which offsets the RoCoF.

The sign mismatch of the load-based model at 04:00 is not explained by the ML model and probably relates to an overall bias in the hourly RoCoF. The ML prediction at 04:00 exhibits the average RoCoF value of -0.29 mHz/s (Fig. 3a), which cannot be related to feature effects as SHAP values only explain the prediction *relative* to the average. The average RoCoF reflects a bias of the hourly DFD towards negative values. For example, this bias can stem from different ramping speeds for upwards and downwards regulation of generators, or from a systematic bias in the power imbalance due to market mechanisms. While our ML model reproduces the bias of the average RoCoF, a physical explanation thereof is still open.

Finally, we point to the impact of local legislation that is revealed by the "hour" feature in Fig. 3a. The hour exhibits its only (strong) effect at 22:00. At this time wind power farms in Germany are shut down due to environmental protection in favour of bats and noise prevention [5]. This leads to a strong

negative effect on the RoCoF, which is well captured in our ML model, but not included in the load-based model at all.

C. A refined view of DFDs

The ML results lead us to a refined view of the generation behaviour that cause hourly DFDs (Fig. 4a). To include the continuous effect of solar power into the load-based model, we calculate modified generation steps by subtracting the hourly solar power ramp $\Delta G_S(t_i)$:

$$\Delta P_S(t_i) = \Delta P_L(t_i) - \Delta G_S(t_i), \qquad (2)$$

In this way, we define a refined linear model that predicts the RoCoF based on the power jump $\Delta P_S(t_i)$:

$$y_S(t_i) = a \cdot \Delta P_S(t_i) + b. \qquad (3)$$

Here, we introduced the intercept b to model the negative bias of the RoCoF as discussed in Section III-B.

To visualise the refined model, we calculate the solar generation $G_S(t)$ for every minute with a cubic spline interpolation
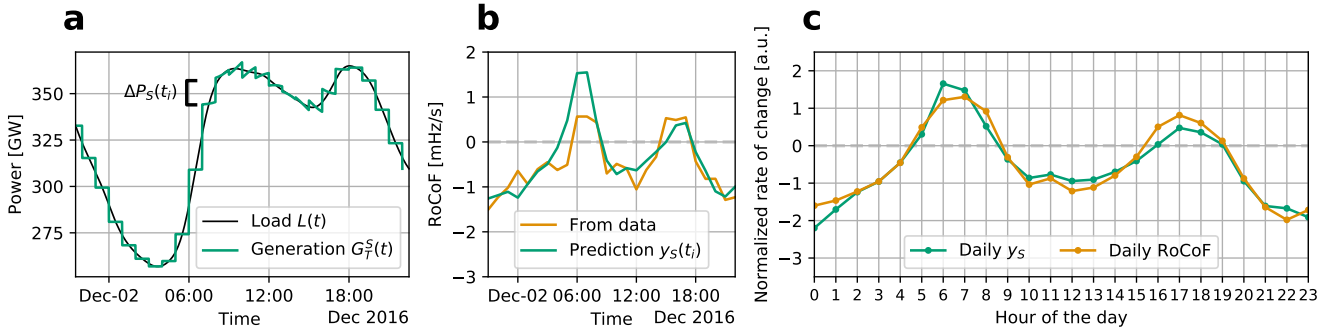
Fig. 4. A refined physical model better reproduces daily DFDs. a: Based on our ML results, we calculate modified generation steps $\Delta P_S(t_i)$ that do not include continuous solar ramps. Instead, we interpolate the hourly solar power data to obtain a minute resolved solar generation $G_S(t)$. The panel sketches the new (total) generation $G_T^S(t)$ which combines the smooth solar generation and generation steps of size $\Delta P_S(t_i)$. b: By using a refined linear model $y_S(t_i)$ of the modified steps ΔP_S (3), we improve the prediction of the load-based model. The refined model better reproduces the sample (orange line), with an overall performance of $R^2 = 0.63$ on the test set. c: The refined model also reproduces the direction of DFDs throughout the day.

of the hourly solar power time series in our data set. The refined generation curve is then given by $G_T^S(t) = G_L(t) + G_S(t)$, where G_L has hourly steps ΔP_S. The resulting sketch in Fig. 4a depicts the same sample as in Fig. 2a, but the generation steps have another direction particularly at 09:00. We clearly see that the continuous solar power ramps up faster than the load between 09:00 and 10:00. As explained through our ML model, this leads to fast downwards ramps of flexible generation, as observed in Fig. 4a at 09:00. Notably, we can also observe the smoothing effect of solar power between 13:00 and 15:00. Here, the solar generation smoothly follows the load thus leading to a less severe generation step than predicted through the load-based model.

The refined model of hourly power steps outperforms the load-based model and explains the daily DFD pattern successfully. The prediction of the continuous sample in Fig. 4b better aligns with the data, as compared to the load-based model, and the overall R^2-score yields 0.64 on the test set. Furthermore, the refined model correctly reproduces the direction of the daily RoCoFs in each hour of the day (Fig. 4c). Together with a model bias, the approximation of hourly generation ramp *without* the contribution of solar power already leads to a good explanation of the daily DFD pattern.

IV. CONCLUSION

In summary, we have modelled and explored the relation between DFDs and external features such as load and generation ramps using explainable ML. Our ML model explains 73% of the variation within the slope of hourly DFDs. The evaluation of a common, load-based DFD model yields a much lower performance of 38% and does not entirely explain the daily pattern of DFDs in Continental Europe. Using daily aggregated SHAP values, we offer another way to explain the daily DFD pattern, thus revealing multiple important external features that go beyond a pure load-based view. Including this information into the load-based predictor improves its performance to 64% and gives a refined view on daily DFD patterns.

Our explainable ML model reveals that solar ramps and scheduled (fast) generation ramps are important for the daily DFD pattern. Solar power ramps continuously during the day. Meanwhile, fast scheduled generation ramps, from e.g. hydro power, have to balance this behaviour thus leading to DFDs that can be opposite to the slope of the load curve. This is consistent with other studies that point to the connection between solar power and generation ramps on 15 min basis due to intra-day electricity trading [24], [25].

Our explainable ML approach provides an alternative to physical models of the grid frequency, but also complements them. Physical models such as the DFD model in ref. [6] have limits when modelling the impact of multiple external features on frequency stability for each hour. Control parameters and highly resolved generation data are mostly not publicly available, such that simulation have to use simplifying assumptions. Our ML model therefore offers an alternative by using publicly available data and estimating a model for DFDs without physical simplifications. However, the ML approach can also complement physical models by pointing to unknown dependencies and important features that have to be included to improve physical modelling. For example, we have presented a refined model of hourly generation steps using our ML insights, which improves the common load-based view.

As frequency deviations have been increasing in Continental Europe, a limitation of large DFDs is urgently needed [5]. Data-based predictions can be combined with control actions to prevent large DFDs [26], thus saving costs for control actions and making the grid less vulnerable. Explainable ML models can offer predictions [9], but they also contribute by better understanding drivers and risks of DFDs in an increasingly complex power system.

REFERENCES

[1] J. Machowski, J. Bialek, J. Bumby, and D. J. Bumby, *Power System Dynamics: Stability and Control*. New York: John Wiley & Sons, Ltd., 2008.

[2] B. Schäfer, C. Beck, K. Aihara, D. Witthaut, and M. Timme, "Non-Gaussian power grid frequency fluctuations characterized by Lévy-stable laws and superstatistics," *Nature Energy*, vol. 3, no. 2, pp. 119–126, Feb. 2018.

[3] S. Homan, N. Mac Dowell, and S. Brown, "Grid frequency volatility in future low inertia scenarios: Challenges and mitigation options," *Applied Energy*, vol. 290, p. 116723, May 2021.

[4] Z. W. Li, O. Samuelsson, and R. Garcia-Valle, "Frequency deviations and generation scheduling in the nordic system," in *2011 IEEE Trondheim PowerTech*, Jun. 2011, pp. 1–6.

[5] 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.

[6] 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*. Seattle: IEEE, Mar. 2009, pp. 1–6.

[7] L. R. 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*, vol. 8, pp. 43 082–43 097, 2020.

[8] M. Persson and P. Chen, "Frequency evaluation of the Nordic power system using PMU measurements," *Transmission Distribution IET Generation*, vol. 11, no. 11, pp. 2879–2887, 2017.

[9] J. Kruse, B. Schäfer, and D. Witthaut, "Revealing drivers and risks for power grid frequency stability with explainable AI," Preprint at https://arxiv.org/abs/2106.04341, 2021.

[10] J. Kruse, B. Schäfer, and D. Witthaut, "Predictability of Power Grid Frequency," *IEEE Access*, vol. 8, pp. 149 435–149 446, 2020.

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

[12] L. Hirth, J. Mühlenpfordt, and M. Bulkeley, "The ENTSO-E Transparency Platform – A review of Europe's most ambitious electricity data platform," *Applied Energy*, vol. 225, pp. 1054–1067, Sep. 2018.

[13] 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," *Nature Communications*, vol. 11, no. 1, p. 6362, Dec. 2020.

[14] R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke, "Explainable Machine Learning for Scientific Insights and Discoveries," *IEEE Access*, vol. 8, pp. 42 200–42 216, 2020.

[15] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[16] 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," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56–67, Jan. 2020.

[17] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. New York: Curran Associates Inc., Dec. 2017, pp. 4768–4777.

[18] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent Individualized Feature Attribution for Tree Ensembles," Preprint at https://arxiv.org/abs/1802.03888, 2019.

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

[20] "ENTSO-E Transparency Platform," https://transparency.entsoe.eu/, 2020.

[21] J. Kruse, B. Schäfer, and D. Witthaut, "Supplementary data: "Revealing drivers and risks for power grid frequency stability with explainable AI"," https://zenodo.org/record/5118352, 2021.

[22] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York: ACM, Aug. 2016, pp. 785–794.

[23] J. Kruse, "johkruse/XAI-for-DFDs: v0.1.0," https://github.com/johkruse/XAI-for-DFDs, doi: 10.5281/zenodo.5497955, 2021.

[24] C. Koch and L. Hirth, "Short-term electricity trading for system balancing: An empirical analysis of the role of intraday trading in balancing Germany's electricity system," *Renewable and Sustainable Energy Reviews*, vol. 113, p. 109275, Oct. 2019.

[25] T. Weißbach, S. Remppis, and H. Lens, "Impact of Current Market Developments in Europe on Deterministic Grid Frequency Deviations and Frequency Restauration Reserve Demand," in *2018 15th International Conference on the European Energy Market (EEM)*, Jun. 2018, pp. 1–6.

[26] I. Avramiotis-Falireas, A. Troupakis, F. Abbaspourtorbati, and M. Zima, "An MPC strategy for automatic generation control with consideration of deterministic power imbalances," in *2013 IREP Symposium Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid*, Aug. 2013, pp. 1–8.

3.3. D) Secondary control activation analysed and predicted with explainable AI

[4] Kruse, J., Schäfer, B. & Witthaut, D. Secondary control activation analysed and predicted with explainable AI. *Electric Power Systems Research* **212,** 108489. doi:`10.1016/j.epsr.2022.108489` (2022).

Secondary control activation analysed and predicted with explainable AI[☆]

Johannes Kruse [a,b,*], Benjamin Schäfer [c,d,e], Dirk Witthaut [a,b]

[a] *Institute for Energy and Climate Research - Systems Analysis and Technology Evaluation (IEK-STE), Forschungszentrum Jülich, 52428 Jülich, Germany*
[b] *Institute for Theoretical Physics, University of Cologne, 50937 Köln, Germany*
[c] *School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom*
[d] *Faculty of Science and Technology, Norwegian University of Life Sciences, 1432 Ås, Norway*
[e] *Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany*

ARTICLE INFO

ABSTRACT

The transition to a renewable energy system challenges power grid operation and stability. Secondary control is key in restoring the power system to its reference following a disturbance. Underestimating the necessary control capacity may require emergency measures, such that a solid understanding of its predictability and driving factors is needed. Here, we establish an explainable machine learning model for the analysis of secondary control power in Germany. Training gradient boosted trees, we obtain an accurate ex-post description of control activation. Our explainable model demonstrates the strong impact of external drivers such as forecasting errors and the generation mix, while daily patterns in the reserve activation play a minor role. Training a prototypical forecasting model, we identify forecast error estimates as crucial to improve predictability. Generally, input data and model training have to be carefully adapted to serve the different purposes of either ex-post analysis or forecasting and reserve sizing.

1. Introduction

Balancing and control is central for the stable operation of power systems. Secondary control is one of three measures that are typically installed to enforce the balance between power supply and demand [1]. While primary control acts within a few seconds after a disturbance and stabilises the frequency, secondary control activates fully after a few minutes and restores the frequency back to its reference value. Secondary control, also referred to as automatic Frequency Restoration Reserve (aFRR) in Europe, activates automatically according to the local power mismatch of the control area. Meanwhile, a lack of control reserves requires costly emergency measures such as load shedding. For an appropriate reserve sizing and optimal control design we thus need a precise modelling and a good understanding of the required aFRR volumes. Furthermore, predicting future aFRR volumes can be helpful for trading and bidding strategies.

Data-driven models have already proven to be excellent candidates for modelling and predicting aFRR. In the past years, power system data has become increasingly publicly available, thus enabling transparent data-driven analysis and prediction [2,3]. Koch et al. have used multiple regression and data analysis tools to disentangle the German paradox of increasing renewable penetration and decreasing imbalance volumes [4]. Similarly, Ref. [5] applies a data-driven analysis to examine the impact of 15 min intra-day trading on imbalances and control volumes in Germany. Apart from ex-post analysis, data-driven forecasting methods have been developed for the aFRR market to allow for optimal bidding strategies [6]. From the perspective of Transmission System Operators (TSOs), data-driven forecasting methods have been used to optimise the dimensioning of aFRR capacities based on historic data [7]. While simple probabilistic methods use parametric models to estimate the probabilities of power imbalances, more advanced methods apply machine learning to predict system imbalances from external features [8]. In this context, Artificial Neural Networks [9], (non-parametric) kernel density estimation and k-means clustering [10] and LASSO [11] have been applied to predict aFRR volumes. However, complex machine learning models are often black-boxes [12]. In particular, we cannot easily understand what the artificial intelligence has learnt, which impedes scientific insights and discoveries [13].
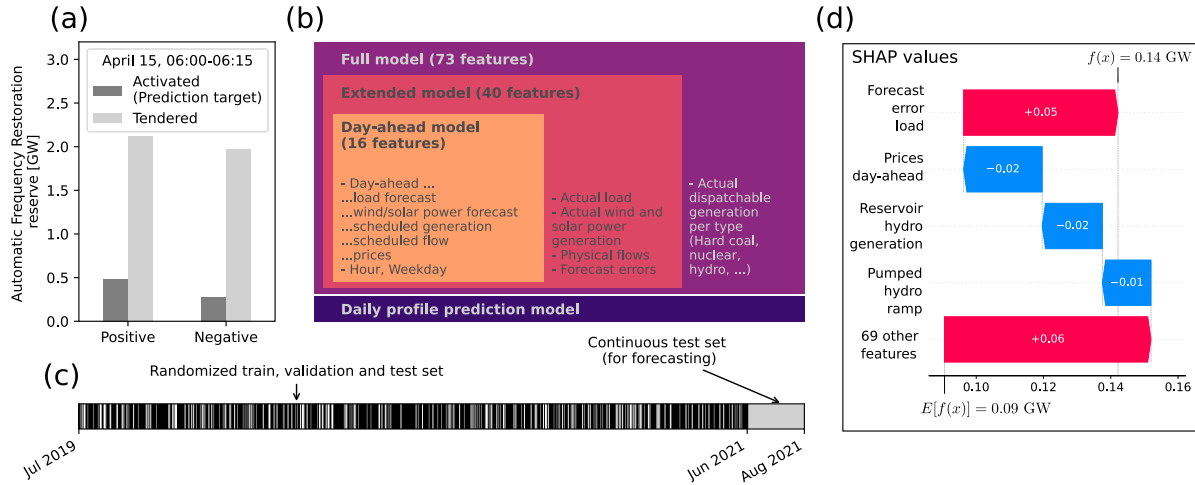
Fig. 1. An explainable machine learning model to predict the activation of automatic Frequency Restoration Reserve (aFRR). (a): Our model predicts the activated amount of positive and negative aFRR in Germany with a 15 min resolution. We downloaded activated power and tendered power demands from the German TSOs [15], which is depicted here for one time step in 2021. (b) We considered four different models. A daily profile, only dependent on the historic aFRR, and three models utilising data from the ENTSO-E transparency platform [14]: a prediction-oriented model using only day-ahead values, an extended model including actual load, wind/solar generation, physical flows and thereby forecast errors and a full model that also comprises actual dispatchable generation. (c): We trained and tested these models on randomised train, validation and test sets with data from July 2019 to May 2021. The last two month of the data set served as a continuous test set for time series forecasting. (d): To explain our model predictions, we used SHAP values [16], which quantify the impact of each feature on the model prediction $f(x)$ relative to the expected prediction $E[f(x)]$ (the base value). The figure refers to the prediction of positive aFRR for the time step used in panel (a).

Here, we leverage tools from eXplainable Artificial Intelligence (XAI) to gain insights about the drivers of aFRR activation. Using publicly available data [14,15], we build a machine learning model for the ex-post analysis of historic data as well as a day-ahead predictor. Within this study, aFRR predictors are first and foremost used as analysis tools to explore the predictability of aFRR, and not necessarily as precise forecasting tools directly applicable to reserve sizing. We interpret the models with SHapely Additive exPlanation (SHAP) values [16], which enable rich explanations of ex-post models as well as transparent day-ahead predictors. As our case study, we focus on the German aFRR and demonstrate how loss functions and data sets have to be adapted for either use case. Our publicly available data set [17] comprises two years of 15 min resolved data with 85 features from the German and European power system as well as the corresponding aFRR volumes.

We start in Section 2 by describing our data collection and pre-processing procedure as well as the machine learning model. In Section 3, we use our model for an ex-post analysis of dependencies and important features in the German aFRR system. In Section 4, we switch to the day-ahead prediction of aFRR volumes and demonstrate the impact of different loss functions and feature sets, before closing with a discussion in Section 5.

2. Methods

2.1. Frequency restoration reserve: Markets and data

The design of balancing markets determines the procurement and activation of control power. In Germany, the TSOs procure their aFRR demand through an anonymous auction [6,15]. First, suppliers of aFRR have to fulfil pre-qualification criteria such as a minimum activation speed to participate in the auction. As of September 2020, hydro- and gas-driven power plants represent the largest part (75%) of the prequalified aFRR capacity [18]. Second, the TSOs tender a demand for the required reserve capacity and energy, which is done for all four German TSOs together (grid control cooperation). Then, prequalified parties can sell reserve capacity (in MW) on the capacity market, which they must hold available. In addition, a supplier can omit this step and directly bid at the energy market (in MWh). The energy market determines the suppliers that actually deliver the balancing energy.

The activation of aFRR depends on the local power imbalance. The TSOs activate aFRR according to the imbalance of the control area. The activation is typically proportional to the integrated imbalance, which is the case for a standard PI-controller [1]. Suppliers with bids on the energy market are activated successively starting with the lowest price until the demand is met. However, the International Grid Control Cooperation (IGCC) further changes the activated control volume [15]. The IGCC avoids the activation of counter-acting aFRR in different countries through an imbalance netting among the 17 operational IGCC member states in Europe (as of July 2021) [19].

Both the procurement and activation of aFRR exhibit fixed time scales and deadlines. The TSOs tender capacity demands one week ahead and the capacity market closes one day ahead of delivery, while the energy market closes one hour ahead. On both markets, aFRR is sold for 4 h blocks separately for negative (downward) and positive (upward) regulation. These time scales also determine the resolution of the available aFRR data.

For our prediction study, we used publicly available aFRR data from July 2019 to July 2021 [15]. We downloaded tendered capacity demands, which come with a 4 h resolution, and activated aFRR with a resolution of 15 min (Fig. 1a). We used the Germany-wide activated aFRR (in GW) as the target of our prediction, while the tendered demands only serve as a benchmark. We note that the market design during our period of investigation changed: The energy market was only introduced in November 2020. Before, both the capacity price and the energy price were submitted to the day-ahead capacity market and energy bids with shorter lead times were not possible.

2.2. Input features and prediction models

As inputs for our prediction model, we used publicly available power system features from the ENTSO-E Transparency platform [14]. The feature preparation included data collection, aggregation, upsampling and feature engineering.

Following [20], we collected six feature types for Germany from the ENTSO-E Transparency platform: Day-ahead forecast data for wind and solar power, load forecasts, day-ahead scheduled generation, day-ahead prices, actual generation per type and actual load. In addition, we included pumped hydro consumption and cross-border power flows

in this study. The flows comprise day-ahead and total commercial exchanges and physical flows. For each flow feature, the in and out flows between Germany and its neighbours were directly aggregated into one import–export balance reflecting the total (positive or negative) flow into Germany.

To model the impact of imbalance netting (IGCC), we aggregated the features across the other IGCC member states (excluding Germany). We only included day-ahead forecast data, since the actual imbalance netting between the countries should be reflected in the actual cross-border flows. The aggregation follows the procedure in Ref. [20].

In line with the prediction target, we used features with 15 min resolution. The German data mostly exhibits the required 15 min resolution, except from day-ahead prices, scheduled generation and flow variables. These have an hourly resolution, so that 4 hourly steps were padded with the same value. The same procedure is applied to IGCC features that often come with only 1 h resolution. Day-ahead load and renewable forecasts are treated differently; here we used a linear interpolation for upsampling for the continuous nature of the variable.

To enhance interpretability, we finally added engineered features to our input data. For each feature, we constructed ramps (gradients from time $t - 1$ to t) and day-ahead forecast errors. Positive forecast errors indicate an overestimation of the actual values by the day-ahead forecast. The flows yield two forecast errors, the day-ahead error between commercial exchanges and physical flows, as well as the unscheduled flows (the difference between total exchanges and physical flows).

From this data, we constructed four models containing different feature sets (Fig. 1b): The day-ahead model contains only day-ahead available data, the extended model also includes actual load, renewable generation, physical flows and thereby forecast errors and the full model also comprises actual conventional generation. Conventional generators such as hydro or nuclear power participate in frequency control [18], which opens the question whether we can predict the activated control *without* including the actual output of participating generation types. The fourth model, the daily profile, predicts the daily mean evolution of the activated aFRR based on historic aFRR data, without using additional features. We use a 5th model in Section 4, which is a variation of the day-ahead model. It additionally contains the day-ahead features (day-ahead load, wind and solar forecasts, . . .) aggregated over the other IGCC member states.

2.3. Model training, evaluation and interpretation

We used Gradient Boosted Trees (GBTs) to predict the activated aFRR from power system features. GBTs offer complex non-linear models and perform inherent feature selection [21], which is beneficial for the case of strongly correlated time series features in the power system [20]. Moreover, tree-based methods are highly interpretable and offer efficient ways to compute model explanations [16]. We used the LightGBM implementation of GBTs, which enables a particularly fast way of model training [22].

To train our model, we split the data set into train, test and validation set (Fig. 1c). First, we set aside the last two months of the data set as a continuous test set for time series forecasting. Then, we randomly split the remaining part into a train set (64%), a validation set (16%) and a test set (20%). We optimised the hyper-parameters of our LightGBM model via grid search and 5-fold cross validation on the train set, while performing early stopping of the boosting rounds on a validation set. Finally, we retrained the model with optimal hyper-parameters on the union of the train and validation set. This model was used to evaluate the performance on the randomised or the continuous test set. Note that we used two different loss functions for training and two different evaluation metrics, which we will specify depending on the use case.

We interpreted the trained model with SHapely Additive exPlanation (SHAP) values [16]. SHAP values quantify the (positive or
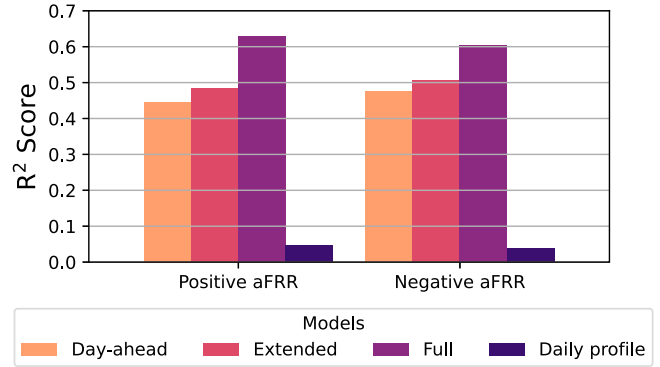

Fig. 2. Prediction performance varies strongly with the model type and feature set. For our ex-post analysis, we trained our model with an L^2 loss and used the randomised train and test sets to predict positive (left) and negative (right) activated aFRR. In terms of the R^2-score, the full model including actual hydro power generation performed best which is likely a manifestation of reversed causality as hydro power supplies most of the aFRR in Germany. Daily patterns in the aFRR activation are very weak such that the daily profile performed badly.

negative) impact of each feature on an individual model prediction relative to a base value. They avoid inconsistencies present in other feature attribution methods [23] and fulfil certain optimal properties [16]. For example, their *local accuracy* guarantees that SHAP values sum up to the model prediction. Fig. 1d depicts a sample prediction, where "Forecast error load" has a positive and "Prices day-ahead" has a negative impact on the prediction. Adding up all feature contributions and the base value (the expected prediction) yields the model output $f(x)$.

2.4. Data and code availability

Our data set, including the prediction targets and features, as well as the results of our hyper-parameter optimisation is available on Zenodo [17]. The code, including the details on our feature pre-processing and the hyper-parameter optimisation, is also archived on Zenodo [24].

3. Ex-post analysis of aFRR operation

We first consider the application of our machine learning model for the ex-post analysis of the aFRR system. Model explainability is essential for any analysis, and we resorted to SHAP values for this task (cf. Section 2.3). For analysis purposes, we aimed at reproducing the system trajectory as well as possible and thus chose a standard L^2 loss function (cf. Fig. 4a) and evaluated the performance of the model by the R^2 score. For model training, evaluation and interpretation we used the randomised train, validation and test set (cf. Fig. 1c).

We found that the tree-based model reproduces the true system trajectory with an R^2 score in the range 0.45–0.63 depending on the feature used in training set and analysis (Fig. 2). The R^2 score reflects the share of variability in the data that is explained by our model [25]. Hence, we conclude that the machine learning model accounts for roughly half of the variability.

Taking into account the full set of features provides the best R^2 score as expected, but it does not necessarily provide the deepest insight into the drivers of the aFRR system. This was revealed by the SHAP framework, which quantifies both the relation of individual predictions and feature values as well as the global feature importance. Inspecting the results for the full feature set in Fig. 3, we found that the most important features (for both positive and negative aFRR) are given by actual generation or generation ramps of hydro power plants and the ramp of nuclear power generation in the case of negative aFRR. The hydro power plants can be switched and controlled rather rapidly and thus provide a major share of aFRR power. In fact, they make up the
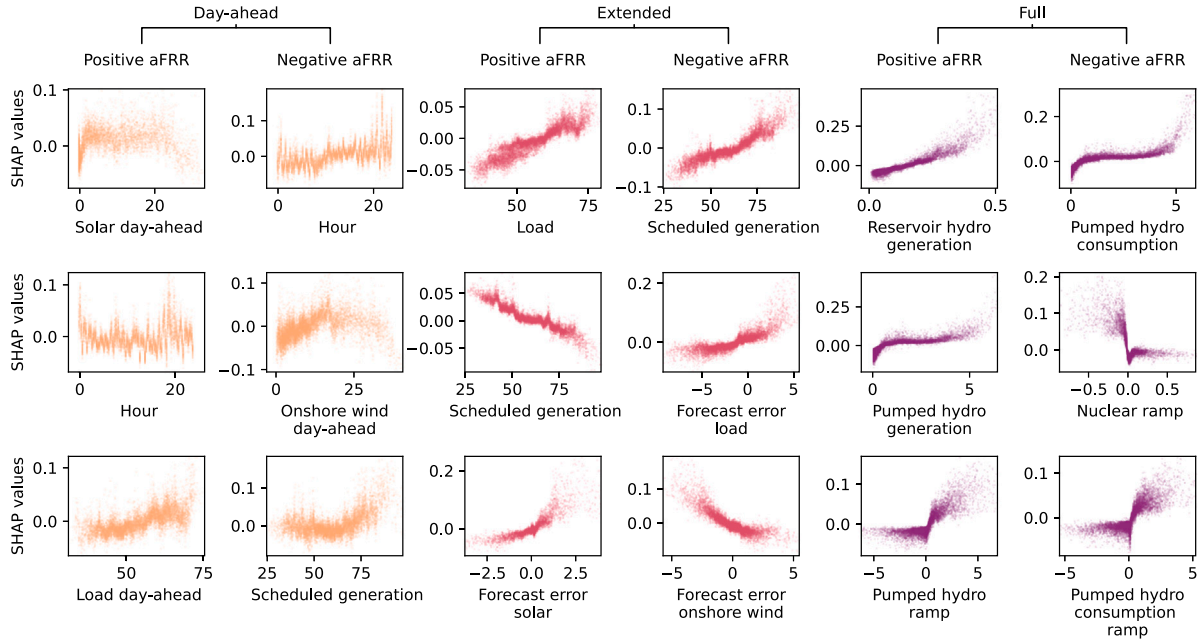
Fig. 3. Partial dependency plots can reveal important drivers of activated aFRR depending on the model type. Dependency plots depict the relation between SHAP values and the feature value, thus characterising the dependency of the model output on the feature. We show the three most important features for the day-ahead (left), extended (centre) and full model (right) with decreasing mean absolute SHAP value (from top to bottom). The day-ahead and extended models did not include actual generation that supplies aFRR. Thus, the most important dependency plots for these models likely reflect causal drivers such as forecast errors and mismatches between load and scheduled generation. In contrast, the full model fitted relations between hydro power and aFRR activation, which probably reflect a reverse causation and must be interpreted differently in the ex-post analysis. In this figure, all variables have the unit GW, except from ramp features (GW/h) and the "hour" feature (h).

largest amount of prequalified aFRR capacity in Germany [18]. Hence, it is likely that we here observed a case of reverse causation: The application of aFRR caused a strong activity of the respective hydro power plants and *not* vice versa. Hence, the model with the full feature set rather explained how aFRR is provided and not why. For instance, the model suggested that positive aFRR is predominantly provided by increasing hydro generation while negative aFRR predominantly provided by increasing pumped hydro consumption. One might extend the analysis beyond hydro power features to further infer which generation types actually contribute to aFRR provision.

In the case of negative aFRR, nuclear power ramps were among the most important features. Nuclear power plants can in principle provide aFRR, but they account for less than 2% of the total prequalified capacity in Germany [18]. Hence, we conclude that the dependency of aFRR activation and nuclear ramps likely reflects a causal relation. Nuclear power plants typically ramp slowly and continuously, which can lead to imbalances of power generation and load. In fact, a previous study has shown a strong dependency between nuclear ramps and long-lasting deviations of the grid frequency [20], which then results in the activation of aFRR.

Restricting the model to the extended feature set provided a different picture. The R^2 score decreases to 0.50 (negative aFRR) and 0.48 (positive aFRR), respectively (Fig. 2). In this case, the most important features were given by the scheduled generation, the load and forecasting error of load and renewable generation (Fig. 3). The high importance of forecasting errors is absolutely consistent with our expectation on the causal interrelation of the aFRR systems. Forecasting errors generally lead to an imbalance of generation and load, causing a deviation of the grid frequency from its reference value. The control system will thus demand the activation of aFRR power.

The high feature importance of the scheduled generation and the total load reflects the general requirements for secondary control power: If the scheduled generation is already high, it is more likely that the total generation has to be reduced via negative aFRR than it has to be increased via positive aFRR. Similarly, if the scheduled generation is already low, it is more likely to be increased via positive aFRR than

it has to be reduced via negative aFRR. Hence, the SHAP values for positive aFRR decreased with the scheduled generation, while the SHAP values for negative aFRR increased with the scheduled generation. The reverse relation was found for the load. The large impact of actual load and scheduled generation (with reverse signs) suggests that the difference between actual load and scheduled generation per 15 min interval is very relevant for the aFRR. This intuitively makes sense as the aFRR activation depends on the area control error, see also [5].

Remarkably, we did not observe a high feature importance of the ramps of load and the dispatchable generation, which are central for the understanding of deterministic frequency deviations (DFDs), in particular the Rate of Change of Frequency (RoCoF) [20]. These DFDs are caused by short-term imbalances due to different adaption/ramping behaviour of generation and load [26]. The low importance of ramps in the aFRR model suggests that DFDs are mostly compensated by primary control. The low importance of DFDs for the aFRR activation is consistent with other studies that show a strong decrease in deterministic aFRR peaks between 2012 and 2018 due to the introduction of 15 min intra-day trading in December 2011 [4,5].

In the day-ahead model, the most important features were the total generation (or similarly load), as well as wind and solar power forecasts and the hour of the day (Fig. 3). Dependency plots showed a strong vertical dispersion, such that the dependencies were less clear than in the other models. This observation, as well as the high importance of the hour, showed that more specific information is missing in this model, which is consistent with the additional performance decrease in the day-ahead model compared to the extended and full model (Fig. 2). Notably, load and generation were both used heavily by the day-ahead and the extended models with similar but slightly different dependencies. In the extended model, low generation led to smaller predictions in negative aFRR, while a low load led to smaller predictions in positive aFRR.

Finally, we remark that a good analysis model is not necessarily a good prediction model. In the current framework this becomes apparent if we do not choose a randomised test set, but a continuous test set in the end of the available training interval, i.e., replace an interpolation
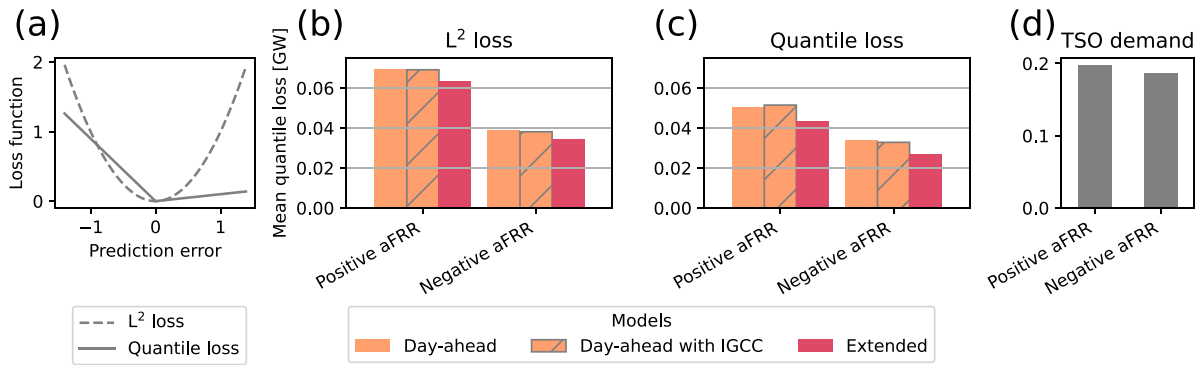
Fig. 4. Loss functions have to fit the use case of the model. (a): For the day-ahead prediction of activated aFRR, we trained our model both with the L^2 loss (squared error) and the quantile loss (90% quantile) and evaluated its performance on the continuous test set (cf. Fig. 1c). In this figure, a positive prediction error indicates an overestimation of the true value by the prediction. (b) and (c): The mean quantile loss (90% quantile) represents the weighted costs of over- and underestimating the aFRR volume, where an underestimation is much more costly than procuring too much aFRR. The quantile loss model performed best and further minimised these costs compared to the L^2 loss model. While the L^2 loss suits best for explaining the aFRR trajectory ex-post, the quantile loss better fits the task to predict required aFRR volumes day-ahead. (d): Tendered aFRR demands from the Transmission System Operators (TSOs) strongly overestimated the activated aFRR in most cases thus yielding a high mean quantile loss.

by an extrapolation task. In this case the R^2 score dropped dramatically to values below 0.38 (full model), 0.12 (extended model) and 0 (day-ahead model), respectively (plots not shown, see code for details [24]). R^2 scores below 0 indicate a performance that is worse than simply predicting the average aFRR volume at each time step [25]. The overall low performance might be related to the fact that the entire aFRR system evolves quite strongly during time, for instance regulatory framework has been adapted repeatedly during the analysis period (cf. Section 2.1).

4. Day-ahead predictability of activated aFRR

So far, we utilised explainable machine learning to understand aFRR activation ex-post. Within this section, we explore the predictability of aFRR, i.e., we propose a prototypical forecasting model based solely on day-ahead inputs. Our goal is not to obtain a ready-to-use forecasting tool but instead to understand the possibilities and limitations of forecasting aFRR activation in general.

The L^2 loss used in the analysis part was well-suited to predict the actual aFRR on average but often underestimated the necessary control. Indeed, the L^2 loss treats predictions underestimating control identical to those overestimating it, while in reality a shortage of control is much more costly: If the necessary control exceeds the estimated control, i.e., the control available in back-up generators, the frequency cannot be restored back to its original set point, making it vulnerable in case of further disturbances and thereby increasing the risks of generator disconnections and load shedding [1].

Hence, moving from a pure ex-post analysis towards a predictive model, we introduced a new loss function, namely the quantile loss, see Fig. 4. Similar to the L^2 loss it penalises predictions more the further they are away from the actual value. In contrast to the L^2 loss, the quantile loss penalises asymmetrically: Overestimating costs is punished much less than underestimating costs (Fig. 4a). In particular, we employed the 90% quantile, i.e., we expect to overestimate the target value in 90% of the cases by penalising underestimation nine times more than overestimation. Consistently, we did not compute the R^2 score when evaluating model performance but the mean quantile loss, which can be interpreted as the weighted average cost of false prediction and procurement of aFRR capacity. Any underestimation of control needs causes larger costs than any overestimation. Here, we consider the 90% quantile corresponding to weight factors of 10 : 1, but the model can be readily adjusted for the actual costs of a TSO (which are not available to us). In contrast to the previously used R^2 score, a model performs better when its mean quantile loss is low.

Indeed, when comparing models trained on L^2 loss and quantile loss, we noted a substantial improvement by introducing the adequate loss function (Figs. 4b and 4c). Within each loss function, moving from the day-ahead to the extended model again yielded a further improvement, consistent with the results from Section 3. Furthermore, we considered an expansion of the day-ahead model where we included the International Grid Control Cooperation (IGCC) as a new feature, see Section 2.2 for details. Including the IGCC altered the prediction quality only slightly, regardless of the chosen loss function, and hence we did not investigate it in detail in the following. Finally, the TSO (tendered) demand systematically overestimated the necessary control by far (see also Fig. 5) and hence led to the highest costs according to the loss considered here (Fig. 4d).

The differences between day-ahead, extended and tendered demand became even more clear when visualising aFRR time series (Figs. 5a and 5b): The tendered capacity demand was almost constant and always overestimated the actual demand substantially. The true time series showed pronounced peaks, which were not fully reproduced by the day-ahead model. Meanwhile, the extended model more closely resembled the true time series. Using SHAP, we were able to investigate why the predictions of the day-ahead and the extended model differed. For individual days, we identified forecast errors in solar generation and ramps as the main reasons (Figs. 5c and 5d). More systematically, we revealed forecast errors in both solar and wind power to drive the mismatch of the predicted power (Figs. 5e and 5f), with solar power playing a more important role for positive and wind power for negative aFRR.

To obtain a comprehensive comparison between day-ahead, extended and full model using the quantile loss, we compared their most important features in Fig. 6. Based on the design of the models, only the full model could use specific actual generation information, e.g., on hydro or nuclear power, while forecast errors were available to both extended and full model and the day-ahead model was restricted to estimates. Intriguingly, each model used some of its unique features extensively and hence they typically had a high rank, yielding results consistent with the L^2 loss results discussed in Fig. 3. Although different features were available, we still noticed some overlap: For positive aFRR both the extended and the day-ahead model utilised load (or load-day-ahead) as a top feature, while negative aFRR was predicted also based on forecast errors of load and scheduled generation respectively. This indicates that the absolute value of the load (or the necessary generation) is important for predicting aFRR usage. Furthermore, the extended model heavily used forecast errors of volatile renewable generation, while the day-ahead model relied on the day-ahead renewable forecasts. This suggests that our day-ahead model tried to estimate
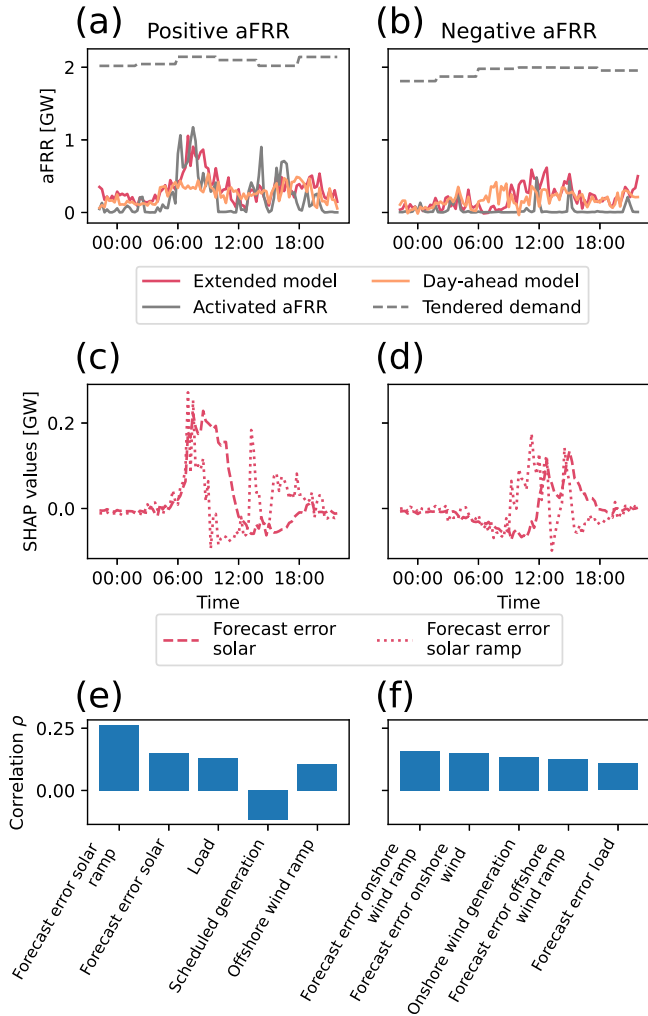
Fig. 5. SHAP values identify volatile renewable forecast errors as driving the loss in the day-ahead aFRR prediction. (a) and (b): The figures depict the prediction of the quantile loss models for July 14, 2021, as part of the continuous test set (cf. Fig. 1c). While the day-ahead model already covers activated aFRR volumes for many time steps, it underestimates the volumes in certain periods, where the extended model performed better. (c) and (d): The SHAP values represent the impact of solar forecast errors on the prediction of the extended model (trained with quantile loss). For example between 9:00 and 10:00, the extended model better estimated the positive aFRR mostly due to the large impact of solar forecast errors in the model, thus pinpointing the source of prediction errors in the day-ahead model. (e) and (f): The correlation ρ between the underestimated power in the day-ahead model and SHAP values of the extended model generalises this observation. We display the 5 features with largest absolute correlations for positive and negative aFRR individually, showing the highest correlations for solar and wind forecast errors.

potential forecast errors from the day-ahead generation forecasts since forecast errors appeared as the actual drivers of control activation, indicated by the increased performance of the extended model. Ideally, forecast errors would be as small as possible, thereby reducing the need for aFRR. If this is not possible, including uncertainties in the forecasts might give better estimates of necessary aFRR.

5. Discussion

Concluding, we have demonstrated how boosted trees and SHapley Additive exPlanations (SHAP) offer versatile tools for investigating secondary frequency control both for ex-post interpretation and also when forecasting trajectories.

Including all available features in the ex-post analysis yielded the most accurate description, both for negative and positive aFRR. Interestingly, the aFRR behaved very differently from deterministic frequency deviations (DFDs). DFDs are mostly driven by ramps [27] and are already well-described by daily profiles [5,26]. Contrary, the aFRR depends much more on the mismatch between actual load and scheduled generation due to forecasting errors and not as much on ramps, thus not showing a pronounced daily profile.

The day-ahead forecasts of the necessary aFRR mostly relied on day-ahead estimations for the volatile renewable generation and also on the total load and generation. Using SHAP, we did not only obtain a most open and interpretable model but also had the opportunity to identify the cause for mispredictions, e.g., pinpointing them to large solar and wind power generation forecast errors (Fig. 5). This analysis allowed us to understand the predictability of aFRR in general, while a ready-to-use forecasting tool would require further optimisation, for example a fine-tuning of the loss function or adding a security margin to prevent the underestimation of aFRR.

Several important lessons are to be learnt when applying machine learning to power system analysis: Firstly, good performance and interpretability of machine learning models can be achieved by combining complex models with ex-post interpretations, such as boosted trees combined with SHAP used here, or right-off starting with white-box models, i.e., choosing techniques that are inherently interpretable [28]. Hence, they allow reducing the usage of black-box machine learning models, which pose severe security concerns [29]. Secondly, the selection of input features is critical when answering research questions via machine learning: Using ex-post analysis of all available generation data, we observed hydro power as critical. But this is likely a reverse causality: Generation in hydro power plants does not raise the need for secondary control but the necessary secondary control is provided by hydro power plants. Furthermore, when excluding features, in our case moving from the full towards the day-ahead model, different features will be used for similar predictions. Therefore, before deeming a feature essential based on a single feature set, regularisation methods [21] should be considered and different feature sets used to train a model. Thirdly, loss functions are critical to tune the model towards desired performance. In the case of control power, underestimating the power is much more costly than overestimating it and this has to be reflected in the loss function. Ideally, loss functions are directly related to the actual costs of false predictions in the system under investigation. While a square loss (L^2 loss) is adequate as an ex-post analysis tool, a quantile loss is more appropriate for predictions.

Frequency restoration capacities have previously been estimated using machine learning [9–11], including the application of quantile loss functions. Compared to these earlier studies, we make a clear comparison between forecasting and ex-post analysis and use the same methods for both tasks, namely boosted trees analysed via SHAP values. Thereby, we emphasise the interpretability of a machine learning approach instead of only optimising performance. By interpreting analysis results, we obtain insights into the system itself and thereby generate value beyond individual algorithms.

Our study of day-ahead aFRR forecasting using different loss functions is a starting point to develop appropriate predictors for aFRR demand. Such a prediction could then be used to optimise the capacity procurement for aFRR day-ahead thus saving costs and freeing flexible capacity for other usage. To move towards such as general predictor, the presented analysis can be further extended, e.g., by incorporating load and generation data from other countries within the same synchronous area or investigating the aFRR in other synchronous areas. Before employing our forecast to estimate aFRR needs, it is necessary to test the effect of time delays in aFRR activation, which can lead to an underestimation of control within the prediction interval. From the interpretation side, SHAPs offer further analysis tools, such as interaction analysis [20] not used here in detail, which could prove useful to disentangle the influence of individual features on aFRR
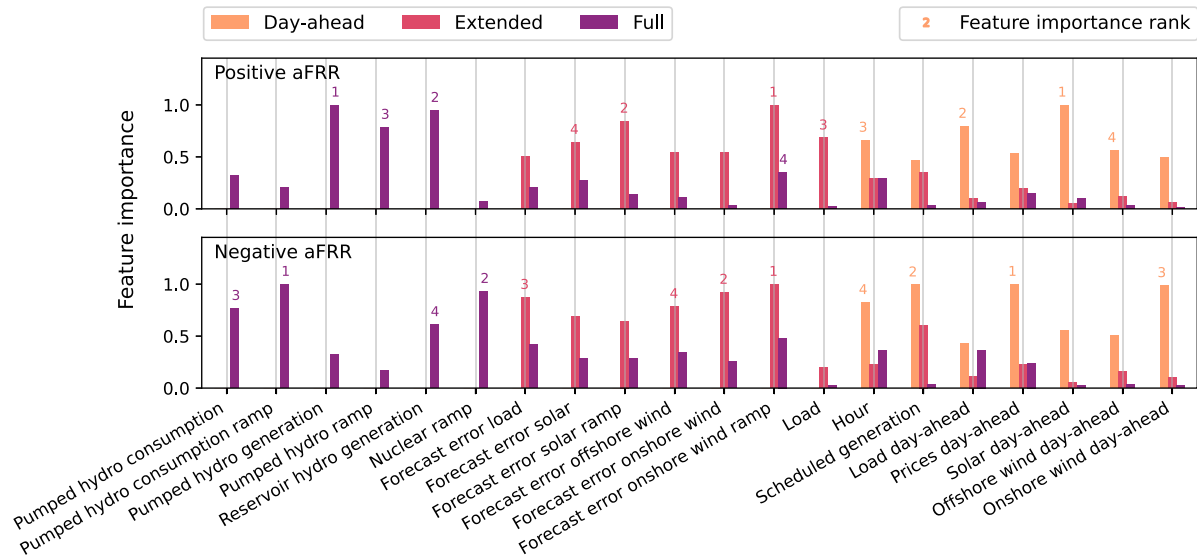
Fig. 6. An overview of feature importances reflects consistent overlaps as well as differences between the model types. We display the mean absolute SHAP value as a measure for the feature importance for all three model types trained with the quantile loss (90% quantile). To be able to compare with Fig. 3, we evaluated the SHAP values on the randomised test set (cf. Fig. 1c). The figure shows the union of the four most important features across all model types and targets and the numbers on top of the bars indicate the rank of the four most important features. From left to right, the models become more restrictive in the feature set thus resulting in differences between the most important features. However, (day-ahead) load as well as features related to wind and solar power were consistently important in all models.

needs. Finally, aFRR is part of a regulatory framework and interacts heavily with current market rules. At the moment, it remains an open question how the predictability and the impact of specific features changed over time.

CRediT authorship contribution statement

Johannes Kruse: Investigation, Formal analysis, Visualization, Conceptualization, Writing. **Benjamin Schäfer:** Supervision, Writing, Conceptualization, Funding acquisition. **Dirk Witthaut:** Project administration, Supervision, Writing, Conceptualization, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Machowski, J. Bialek, J. Bumby, Power System Dynamics: Stability and Control, John Wiley & Sons, Ltd., New York, 2008.

[2] L. Hirth, J. Mühlenpfordt, M. Bulkeley, The ENTSO-E transparency platform – A review of Europe's most ambitious electricity data platform, Appl. Energy 225 (2018) 1054–1067.

[3] R. Morrison, Energy system modeling: Public transparency, scientific reproducibility, and open development, Energy Strategy Rev. 20 (2018) 49–63.

[4] C. Koch, L. Hirth, Short-term electricity trading for system balancing: An empirical analysis of the role of intraday trading in balancing Germany's electricity system, Renew. Sustain. Energy Rev. 113 (2019) 109275.

[5] T. Weißbach, S. Remppis, H. Lens, Impact of current market developments in Europe on deterministic grid frequency deviations and frequency restoration reserve demand, in: 2018 15th International Conference on the European Energy Market, EEM, 2018, pp. 1–6.

[6] M. Merten, F. Rücker, I. Schoeneberger, D.U. Sauer, Automatic frequency restoration reserve market prediction: Methodology and comparison of various approaches, Appl. Energy 268 (2020) 114978.

[7] D. Jost, A. Braun, R. Fritz, Dynamic dimensioning of frequency restoration reserve capacity based on quantile regression, in: 2015 12th International Conference on the European Energy Market, EEM, IEEE, 2015, pp. 1–5.

[8] K. De Vos, N. Stevens, O. Devolder, A. Papavasiliou, B. Hebb, J. Matthys-Donnadieu, Dynamic dimensioning approach for operating reserves: Proof of concept in Belgium, Energy Policy 124 (2019) 272–285.

[9] D. Jost, A. Braun, R. Fritz, S. Otterson, Dynamic sizing of automatic and manual frequency restoration reserves for different product lengths, in: 2016 13th International Conference on the European Energy Market, EEM, 2016, pp. 1–5.

[10] M. Bucksteeg, L. Niesen, C. Weber, Impacts of dynamic probabilistic reserve sizing techniques on reserve requirements and system costs, IEEE Trans. Sustain. Energy 7 (4) (2016) 1408–1420.

[11] A. Essl, A. Ortner, R. Haas, P. Hettegger, Machine learning analysis for a flexibility energy approach towards renewable energy integration with dynamic forecasting of electricity balancing power, in: 2017 14th International Conference on the European Energy Market, EEM, 2017, pp. 1–6.

[12] A. Adadi, M. Berrada, Peeking inside the Black-Box: A survey on explainable artificial intelligence (XAI), IEEE Access 6 (2018) 52138–52160.

[13] R. Roscher, B. Bohn, M.F. Duarte, J. Garcke, Explainable machine learning for scientific insights and discoveries, IEEE Access 8 (2020) 42200–42216.

[14] ENTSO-E transparency platform, 2020, https://transparency.entsoe.eu/.

[15] 50Hertz Transmission GmbH, Amprion GmbH, TransnetBW GmbH and TenneT TSO GmbH, Regelleistung.net, 2021, https://www.regelleistung.net/ext/.

[16] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (1) (2020) 56–67.

[17] J. Kruse, B. Schäfer, D. Witthaut, Supplementary data: Secondary control activation analysed and predicted with explainable AI, 2021, http://dx.doi.org/10.5281/zenodo.5497500.

[18] 50Hertz Transmission GmbH, Amprion GmbH, TransnetBW GmbH and TenneT TSO GmbH, Capacity of prequalified providers in Germany, 2021, https://www.regelleistung.net/ext/download/pq_capacity.

[19] ENTSO-E balancing report 2020, 2020, https://eepublicdownloads.entsoe.eu/clean-documents/Publications/Market%20Committee%20publications/ENTSO-E_Balancing_Report_2020.pdf.

[20] J. Kruse, B. Schäfer, D. Witthaut, Revealing drivers and risks for power grid frequency stability with explainable AI, Patterns 2 (11) (2021).

[21] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed., Springer, New York, 2016.

[22] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: A highly efficient gradient boosting decision tree, in: Advances in Neural Information Processing Systems, 30, Curran Associates, Inc., 2017.

[23] S.M. Lundberg, G.G. Erion, S.-I. Lee, Consistent individualized feature attribution for tree ensembles, 2019, Preprint at arXiv:1802.03888.

[24] J. Kruse, B. Schäfer, D. Witthaut, Explainable machine learning for secondary control activation in Germany, 2021, http://dx.doi.org/10.5281/zenodo.6417324.

[25] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Comput. Sci. 7 (2021) e623.

[26] T. Weissbach, 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, 2009, pp. 1–6.

[27] J. Kruse, B. Schäfer, D. Witthaut, Exploring deterministic frequency deviations with explainable AI, in: 2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, SmartGridComm, 2021, pp. 133–139.

[28] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115.

[29] T. Ahmad, D. Zhang, C. Huang, H. Zhang, N. Dai, Y. Song, H. Chen, Artificial intelligence in sustainable energy industry: Status Quo, challenges and opportunities, J. Cleaner Prod. 289 (2021) 125834.

4. Stochastic modelling of grid frequency dynamics with techno-economic drivers

Machine learning models are well suited to explain risks and drivers of grid frequency dynamics. Using tools from explainable AI, I successfully predicted stability indicators and control activation from techno-economic features, which revealed important impact factors and dependencies (Section 3.1-3.3). These models focused on large-scale frequency dynamics represented by temporally aggregated targets, as stability indicators and control activation are defined for quarter-hourly or hourly time intervals. In this section (4.1), I extend the prediction target to short-term dynamics on a second timescale. However, techno-economic features are only available at hourly timescales, which represent the large-scale trends of power system dynamics [73]. This challenges their integration into short-term prediction models. As a solution, I propose a physics-inspired machine learning (PIML) model, which integrates a stochastic model of short-term dynamics and a neural network to assimilate techno-economic features.

As a use case, I demonstrated the PIML approach for grid frequency dynamics in Continental Europe. The PIML prediction of 15 minute frequency dynamics outperformed important benchmarks such as the daily average profile. In particular, the prediction model can forecast the next 15 minutes of the time series by using only day-ahead available features. Furthermore, the model learned time-dependent dynamical parameters, which showed a strong variability, thus emphasising the importance of non-autonomous models for grid frequency dynamics. Again, I used SHAP values to explain the model, which revealed important drivers such as the wind feed-in and fast generation ramps. Finally, I showed that synthetic time series generated by the model reproduce central characteristics of grid frequency time series, such as the heavy tails of their distribution. I contributed to this work by conducting the research, creating the figures, designing the research and writing the major part of the paper.

4.1. E) Physics-inspired machine learning for power grid frequency modelling

[5] Kruse, J., Cramer, E., Schäfer, B. & Witthaut, D. *Physics-inspired machine learning for power grid frequency modelling*. Preprint at `https://doi.org/10.48550/arXiv.2211.01481`. 2022.

Physics-inspired machine learning for power grid frequency modelling

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

[1]*Forschungszentrum Jülich, Institute for Energy and Climate Research (IEK-STE), 52428 Jülich, Germany*
[2]*Institute for Theoretical Physics, University of Cologne, Köln, 50937, Germany*
[3]*Forschungszentrum Jülich GmbH, Institute of Energy and Climate Research,*
Energy Systems Engineering (IEK-10), Jülich 52428, Germany
[4]*Institute for Automation and Applied Informatics,*
Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany
(Dated: November 4, 2022)

The operation of power systems is affected by diverse technical, economic and social factors. Social behaviour determines load patterns, electricity markets regulate the generation and weather-dependent renewables introduce power fluctuations. Thus, power system dynamics must be regarded as a non-autonomous 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-inspired 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. Using the integrated model, we identify and explain the parameters of the dynamical system from the data, which reveals 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 emphasises the importance of modelling power system dynamics as a stochastic non-autonomous 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 utilise 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 modelling of frequency dynamics in large-scale power systems is profitable but complex due to its non-autonomous character. Stochastic dynamical models have successfully reproduced central characteristics of frequency measurements such as their non-standard distributions [13–15]. Such models can be used to generate synthetic frequency time series, which are, for example, employed to optimise 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 non-autonomous dynamical system, which makes grid frequency modelling 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

* d.witthaut@fz-juelich.de

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 is 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].

In this work, we propose physics-inspired machine learning (PIML) to approach these challenges. Compared to numerical simulations, PIML models can perform better in solving ill-posed problems with noisy insufficient data and imperfect physical models [22, 23]. Moreover, they can better generalise from small amounts of data than common machine learning methods, and they efficiently solve inverse problems of differential equations in situations with insufficient data or incomplete models [22]. Notably, inverse problems are particularly important for power system control to estimate hidden states or dynamical parameters from measurements [24]. The inverse problem of inferring system parameters from input/output data is known as system identification [25] and PIML models offer a promising tool for such applications [26].

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 an analytic 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 a feed-forward artificial 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 SHapely Additive eXplanation (SHAP) values [27, 28] to extract the dependency between dynamical parameters and techno-economic features. All in all, the model bridges the gap between the large-scale behaviour of interdependent energy systems and markets and the short-term dynamics of the power system.

The article is organised as follows. In Sec. II we introduce the physics-inspired 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-inspired model of the power grid frequency including a stochastic description of the frequency dynamics on coarse scales and the interaction with techno-economic features. Furthermore, we interpret the model in terms of power system operation and discuss the implementation as an artificial neural network. The detailed implementation of our data preparation and model pipeline, as well as all input data and the results are available on Zenodo [29, 30].

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

Our starting point is a stochastic model for the dynamics of the grid frequency as illustrated in Fig. 1a. The rate of change of the frequency at a 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 $\omega(t) = 2\pi(f(t) - f_{\mathrm{ref}})$, we have the equation of motion

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

where M is the aggregated inertia constant. The power imbalance on the right-side has been decomposed into three contributions. The term $P_{\mathrm{im}}(t)$ denotes sustained power imbalances, for instance due to a mismatch of the load and the scheduled generation of dispatchable power pants (cf. Fig. 1b). For time intervals of a quarter-hour, we can approximate the time dependence of these imbalances by an affine linear function

$$P_{\mathrm{im}}(t) = M \cdot (q + rt), \qquad (2)$$

where q models the power step of scheduled generation and r represents the continuous drift of the load (cf. Fig. 1c). The term $P_{\mathrm{noise}}(t)$ describes short-term fluctuations of the power balance, which we modelled as

$$P_{\mathrm{noise}}(t) = M \cdot D \cdot \xi(t), \qquad (3)$$

where $\xi(t)$ is white noise with a standard normal distribution and D quantifies the strength of short-term power fluctuations. The term $P_{\mathrm{control}}(t)$ denotes the balance of primary and secondary load-frequency control, which can be modelled by a proportional-integral law as

$$P_{\mathrm{control}}(t) = -\frac{M}{\tau}\omega(t) - \frac{M}{\kappa^2}\underbrace{\int_{t_i}^{t}\omega(t')\,\mathrm{d}t'}_{=:\theta(t)} \qquad (4)$$

with time constants τ and κ. We note that some simplifications are necessary to keep the model tractable. For

instance, Eq. (4) neglects the existence of a small deadband in the proportional control law.

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

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

The parameter τ quantifies the effective primary control time scale. Equation (5) resembles a driven harmonic oscillator with an eigenfrequency κ, which can be interpreted as the intrinsic time scale of secondary control. In contrast, the effective time scale of secondary control is approximated by τ/κ^2 [13], as the frequency decays with this time constant in the overdamped case [14]. Note that these are only effective parameters which were rescaled by the inertia M. For example, the actual primary control strength is M/τ (Eq. (4)). However, the whole model is invariant under a scaling of M (cf. appendix A 2) such that it is only possible to estimate the ratio of parameters and the inertia.

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

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

As we show in appendix A 3, the FPE is solved by a multivariate Gaussian distribution

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

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

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

if the parameters satisfy the ordinary differential equations

$$\frac{d}{dt}\mu_\theta = \mu_\omega$$
$$\frac{d}{dt}\mu_\omega = q + rt - \tau^{-1}\mu_\omega - \kappa^{-2}\mu_\theta$$
$$\frac{d}{dt}\sigma_\theta^2 = 2\sigma_{\theta\omega}$$
$$\frac{d}{dt}\sigma_\omega^2 = \sigma^2\omega - \tau^{-1}\sigma_{\theta\omega} - \kappa^{-2}\sigma_\theta^2$$
$$\frac{d}{dt}\sigma_{\theta\omega} = -2\tau^{-1}\sigma_\omega^2 - 2\kappa^{-2}\sigma_{\theta\omega}. \tag{8}$$

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 solve the moment equations (8) analytically, thus solving the entire stochastic dynamic model, which vastly simplifies the analysis.

B. Power system operation and interdependecies

The SDE for the frequency dynamics in Eq. (5) contains several parameters, describing the load-frequency control system (τ, κ), or the power imbalances on different time scales (r, q, D). The parameters are not constant, but change during the day. For instance, the market based scheduling of conventional power plants causes characteristic imbalances of generation and load [31, 32]. Electricity is traded on the spot markets in blocks of 15, 30 or 60 minutes, leading to characteristic patterns of the power imbalance $P_{im}(t)$ illustrated in Fig. 1b,c. The shape of these patterns, as well as other properties of the power system, change in time due to the influence of a variety of techno-economic features.

We thus propose a model that integrates the internal dynamics of the frequency-control system, the stochastic noise, and the impact of various techno-economic features. In every 15 minute interval, the frequency is modelled by the SDE (5). The system parameters τ, κ, D, r, q change from interval to interval, depending on the influence of external techno-economic features as detailed below. This dependence is modelled by a FFNN, that is trained such that the stochastic dynamics best fits the recorded time series. More precisely, the FFNN constitutes a parameter model $\mathcal{F}_{NN} : \boldsymbol{X} \mapsto \boldsymbol{\vartheta}$, where $\boldsymbol{X}$ summarises the values of techno-economic features (Fig. 1d). The vector $\boldsymbol{\vartheta}$ includes the system parameters τ, κ, D, r, q 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.

We thus establish a model that links different temporal and technological scales, from the slow evolution of electricity markets to the fast dynamics of the power grid frequency. The integrated model predicts a probability distribution $\mathcal{P}(\omega, \theta; t)$ for an entire interval of 15 minutes 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 the ENTSO-E Transparency platform [18]. Following ref. [28], we downloaded load forecasts, day-ahead scheduled generation, day-ahead forecast data for wind and solar power, day-ahead electricity prices, actual generation per type and actual load. In addition, we included pumped hydro consumption (cf. [33]), as well as net scheduled, i.e., market-based flows and net physical flows between Continental Europe and other synchronous areas (cf. [34]).
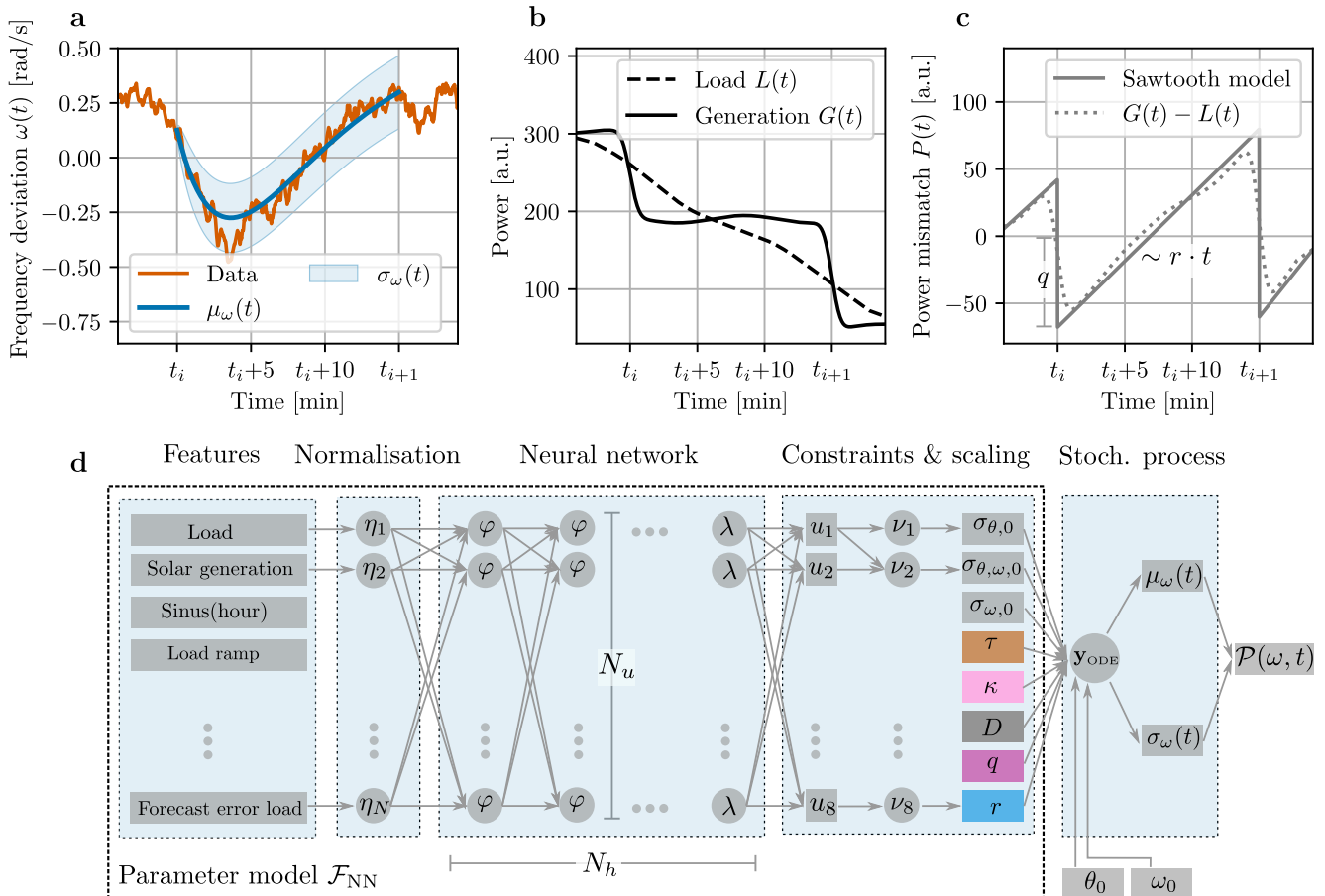
FIG. 1. A physics-inspired machine learning model for the power grid frequency. **a** In each interval $[t_i, t_{i+1}]$, we modelled the grid frequency with a stochastic process that provides a normal distribution with time-dependent mean $\mu_\omega(t)$ and standard deviation $\sigma_\omega(t)$. The panel depicts our model prediction of the Continental European grid frequency at $t_i = 23{:}00$ on the 28th of October 2018 in comparison to the recorded data. **b,** Deviations from the reference frequency are partly driven by deterministic power imbalances $P_{\mathrm{im}}(t) = M \cdot P(t)$. 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 $P_{\mathrm{im}}(t)$, additional stochastic imbalance fluctuations $P_{\mathrm{noise}}(t)$ and the load-frequency control $P_{\mathrm{control}}(t)$. The parameter model $\mathcal{F}_{\mathrm{NN}}$ predicts the model parameters, i.e., the imbalance and control parameters (colour highlight), as well as the initial covariances, from $N = 77$ techno-economic features by using a feed-forward neural network (FFNN).

We prepared the features for each 15 minute interval. This opens the possibility to capture the effects of intraday electricity markets that operate in 15 minute intervals, while the day-ahead electricity market typically acts every 60 minutes [8]. To this end, we determined the time resolution of each feature in each country, which can vary due to different market designs. Then, we upsampled 60 minute data using linear interpolation for load and renewable generation data, and forward padding for all other feature types.

To gain additional interpretable input data, we engineered physically meaningful features and aggregated the data area-wise (cf. [28, 33]). We included forecast errors (day-ahead minus actual), ramps (time derivative of a feature) and unscheduled flows (scheduled minus physical flows). Since all features are only available on country-

level, the data was finally aggregated within the whole Continental European area to represent the aggregated impact on the grid frequency (the detailed implementation is available on Zenodo [29]).

The grid frequency recordings used in this work were taken from ref. [35], which provides pre-processed frequency data from the German transmission system operator TransnetBW [36].

D. Artificial neural network model

The integrated model provides a probabilistic prediction of the power grid frequency for every 15-minute interval $[t_i, t_{i+1}]$. The architecture of the model is depicted in Fig. 1d.

As input, we used the techno-economic features $\boldsymbol{X}(i) = (X_1(i), ..., X_N(i))^T$ for each time interval $[t_i, t_{i+1}]$. In the first step, each feature X_k was normalised 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 normalised features were fed into a 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 initialisation, 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 axis leading to inhomogeneous loss landscapes which can make optimisation inefficient and more difficult [37]. 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 behaviour of the solution. Rescaling and constraints were implemented with parameter-specific functions $\nu_j(u_j)$, which are described in appendix C.

After rescaling, the output $\boldsymbol{\vartheta}(i)$ of the parameter model $\mathcal{F}_{\text{NN}}$ 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). The 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 $\mu_{\theta,0}(i) = \int_{t_i-60s}^{t_i} \omega(t')\,\mathrm{d}t'$. 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 107650 data points from 2015 to 2019. In particular, it includes features $\boldsymbol{X}(i)$ and frequency time series $\boldsymbol{\omega}(i) = (\omega(t_i), ..., \omega(t_i + t_{\max}))^T$ for each interval i. To assess the time-dependence of the performance, we modelled and predicted subsets of the 15 minute interval with $t_{\max} < 15$ min (cf. Sec. III A), but the full interval ($t_{\max} = 15$ min) was used in all other cases.

We quantified the ability of the model to predict the stochastic frequency dynamics by the negative log-likelihood. For a given time interval $\mathcal{I} = [t_i, t_i + t_{\max}]$, the negative log-likelihood is defined as

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

TABLE I. Parameter choices during hyperparameter optimisation. Sig(x) is the sigmoid function and tanh(x) the hyperbolic tangent.

	Possible values
Learning rate	10^{-4}, 10^{-3}, 10^{-2}
Dropout rate	0, 0.1, 0.2, 0.3
N_u	64, 128
N_h	3,5,7
Activation $\varphi(x)$	Sig(x), tanh(x)

where $P(\omega; t | \boldsymbol{\vartheta})$ is the marginal of the PDF (7) evaluated at the measured data $\omega(t)$. The log-likelihood is a negatively oriented metric, i.e., smaller values represent a better performance.

To train the FFNN, we initialised the weights using the Glorot uniform initialiser [38]. Using data from 2015 to 2017, we trained the weights with stochastic gradient descent using the ADAM optimiser with a fixed learning rate [37]. As a loss function we chose the negative log-likelihood (9), summed over all quarter-hour intervals in the training set. The model hyperparameters were optimised using random search on data from 2018 (as a validation set) and with parameter choices defined in Tab. I. 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 log-likelihood 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 the 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 the global mean and variance of the whole frequency time series.

Finally, we interpreted our parameter model $\mathcal{F}_{NN} : \boldsymbol{X} \mapsto \boldsymbol{\vartheta}$ with SHapely Additive eXplanation (SHAP) values [27], 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 importances and dependencies extracted by the FFNN. In particular, we used KernelSHAP [39], which approximates SHAP values for any machine learning model.

Our FFNN model is implemented with tensorflow [40] and tensorflow probability [41] and we used keras tuner for hyperparameter optimisation [42].

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 analyse 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 scenarios 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-inspired 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).

The machine-learning model outperforms elementary benchmarks irrespective of the prediction horizon (Fig. 2a-d). Using all available techno-economic features, the full model yielded lower median loss values than the daily profile and the constant model for each prediction length $t_{\max}$. Restricting the feature set to day-ahead available data yields a similar performance, which enables us to forecast future frequency deviations better than the daily profile.

However, the performance slightly deteriorates with increasing prediction length (Fig. 2e). The absolute log-likelihood of different data sets, i.e., of different $t_{\max}$ values, cannot be compared. We therefore employed the relative loss increase between the machine learning model and the daily profile as a measure, which mainly exhibits negative values as our model outperforms the daily profile. Predicting only the first $t_{\max} = 360$ seconds yielded a better performance then predicting the full interval ($t_{\max} = 900$ s). This points to a potential limitation of our model at the end of the prediction interval, which is likely due to the approximate treatment of the power imbalance $P_{\mathrm{im}}(t)$ (cf. Fig. 1c). Our model assumes a discrete step at the start of an interval, while real power plants start ramping up or down continuously earlier at the end of the previous interval [20]. Hence, the frequency $\omega(t)$ at the end of an interval is already driven by the dispatch in the future interval. This aspect is not included in the physics-inspired model, while it is present in the daily profile.

The prediction examples in Fig. 2f-i illustrate the strengths and limitations of our model. The intervals with the best model performance at 00:00 and 06:00 demonstrate how our model outperforms the daily profile by far. A remarkable aspect is observed when inspecting the intervals with the worst performance: As the physics-inspired model fails to capture the dynamics, so does the daily profile, albeit at a different magnitude (Fig. 2g,i). The limitations of our model due to continuous generation ramps turned up in Fig. 2i: The frequency increased and then ramped down expectedly due to the rising load in the morning, which causes upwards deterministic frequency deviations. However, the frequency ramped up again towards the end of the interval, which is not covered in our model (see above), but slightly visible in the daily profile.

In the following sections, we explain dynamical parameters based on techno-economic features, among others. The full model better suits for explanation as it also includes actually measured features, such as forecast errors. Therefore, we only focus on the full model and predict the full interval (with $t_{\max} = 15$ min) 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, 43], our parameter model $\mathcal{F}_{NN}$ extracts time-dependent system parameters $\boldsymbol{\vartheta}(i)$ for each time interval i, which mirror the local dynamical properties of load-frequency control (cf. Fig. 1). Note that we only estimate effective parameters that also contain the impact of the inertia (cf. Sec. II), which we discuss later in Sec. IV.

The inferred parameters strongly change during the day, which illustrates the importance of time-dependent dynamical modelling (Fig. 3a-e). The daily profile of the primary control time scale shows variations of 14 %, while the intrinsic secondary control time scale varies by 16 %. The power imbalance parameters show even stronger variations, with the short-term fluctuation strength D varying by 61 % and the deterministic parameters q and r changing by 550 % and 442 %.

The imbalance parameters show distinct patterns that reveal physically meaningful impact factors on the grid frequency. We inferred upwards power steps q and negative drifts r in the morning around 06:00 and in the evening around 18:00, while the opposite behaviour 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. Fig. 1b,c). Moreover, the inferred power steps q peak at the beginning of the hour, while being smaller during the rest of the hour. This indicates the dominance of generation
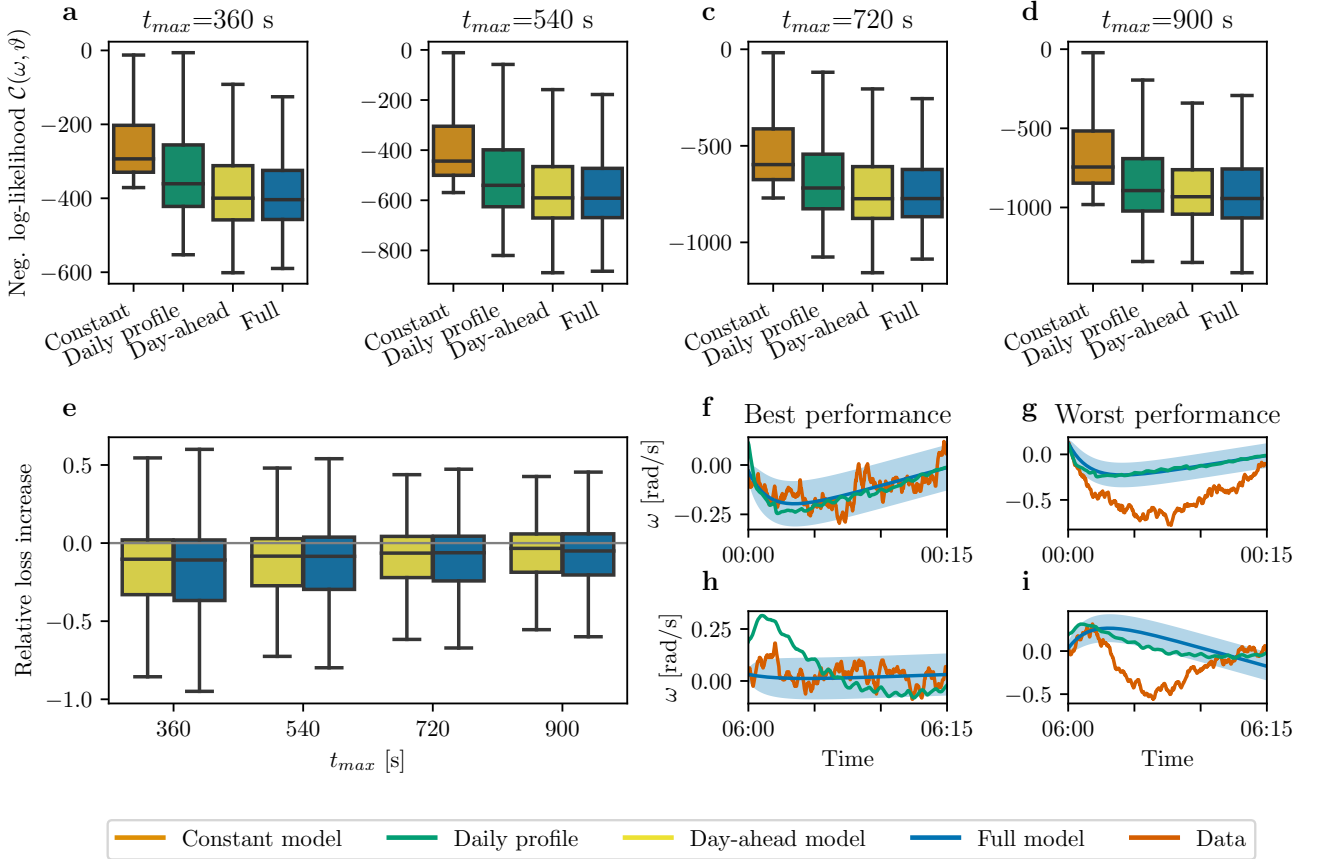
FIG. 2. Probabilistic machine learning model outperforms elementary benchmarks. **a,b,c,d,** We quantify the performance of our probabilistic frequency prediction with the negative log-likelihood loss, which implicates better performance if values are lower. Irrespective of the prediction length $t_{\max}$, the day-ahead and full models outperform the the benchmark models, namely the constant model and the daily profile predictor. **e,** The relative loss increase between our model and the daily profile provides a comparison between different prediction lengths, yielding a slight decrease of the performance with longer predictions. **f,g,h,i,** Prediction examples with best and worst performance at 00:00 and 06:00 illustrate the strengths and limitations of our model.

changes within 1 h intervals over changes in 15 min periods, which is consistent with the characteristics of the European electricity markets: In fact, much more generation volume is traded on the hourly day-ahead market then on the quarter-hourly intraday market in Europe [44].

The inferred parameters generally agree with estimates from the literature [14], with small differences due to redundancies between secondary control and deterministic deviations (Fig. 3f). The panel depicts the ratio between the time average of our absolute parameter estimates and the reference value from the literature $\vartheta_j^{\mathrm{ref}}$ (cf. appendix B), which was extracted by a time-independent model. Our model (with standard scaling defined in appendix C), inferred values for τ and D that are very similar to the reference, i.e., the ratio to the reference is near one. The secondary control was weaker in our model (larger time scale) and the deterministic power mismatch was also weaker, i.e., our model inferred smaller absolute

generation steps q and drifts r. As discussed in ref. [14], there is a redundancy between secondary control and deterministic power drifts, because both can drive the frequency back to its reference. This makes it generally difficult to obtain unique estimates, which might explain the difference to the literature values for κ.

The adequate inference of dynamical parameters is greatly facilitated by our implementation of appropriate scaling steps. Fig. 3f depicts the parameter estimates if no scaling is applied (triangle markers), which resulted in much stronger deviations from the literature values. This is probably due to the large difference in scale between the parameters, which renders the FFNN training inefficient and unstable due to the very heterogeneous loss landscape [37]. Note that small variations of the scaling coefficients (defined in appendix C) did not strongly change the parameter estimates (square markers) such that the results seem to be independent of the exact choice of the scaling.
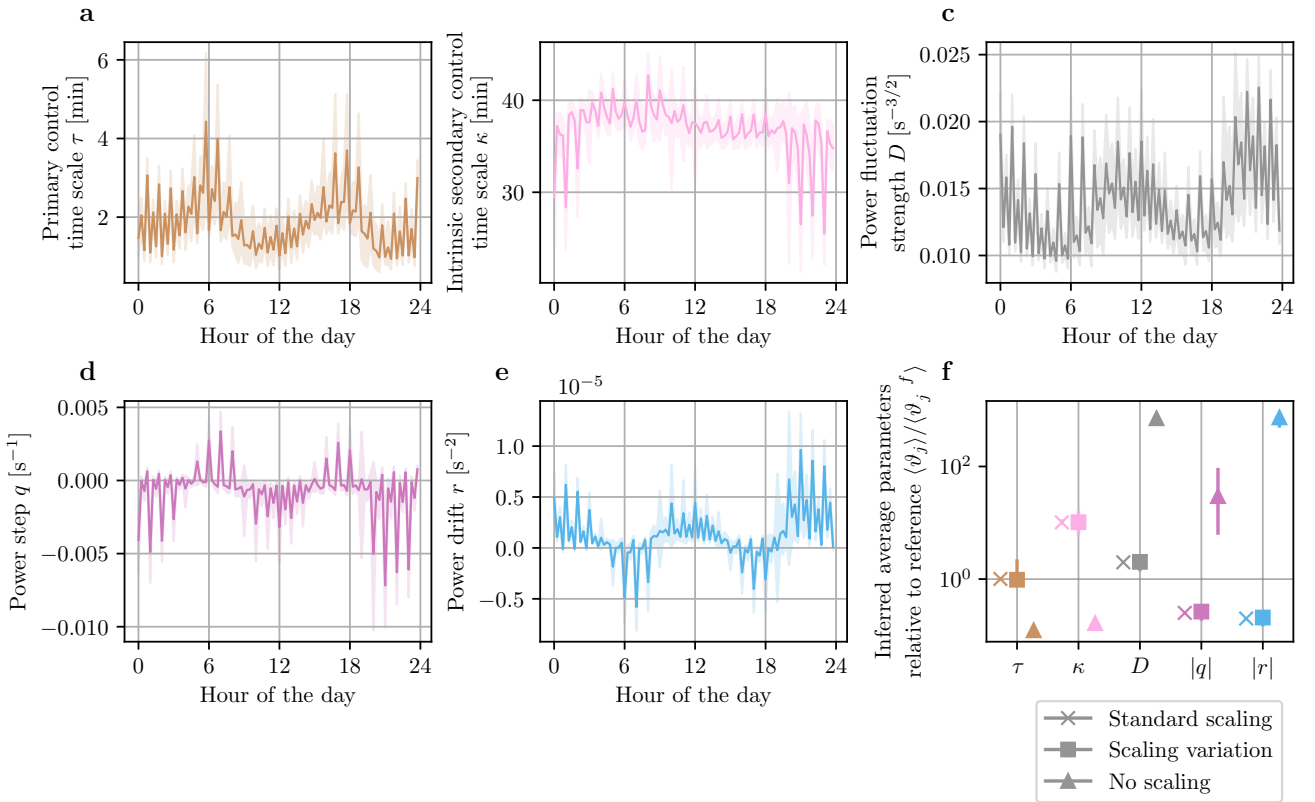
FIG. 3. Physics-inspired machine learning model infers dynamical system parameters. **a,b,c,d,e** The inferred parameters show strong daily patterns, which underlines the importance of time-dependent dynamical properties in our model. The panels depict the daily means (solid lines) and the range between the 25 % and 75 % quantile (area). **f,** The ratio between the average inferred parameters and the reference value $\vartheta_j^{\mathrm{ref}}$ obtained from time-independent stochastic inference (cf. appendix B). 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 Tab. II), with 10 random weight initialisations for each combination of scaling parameters. The ensemble means (markers) and the data range (errorbar) indicate the relation between the inferred parameters and the reference, which shows particularly large deviations if no scaling is applied.

2. Techno-economic drivers of dynamical system properties

Using SHAP values [27], 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 and r and the primary control time scale τ, we analysed feature importances quantified by the mean absolute SHAP value (Fig. 4a), as well as dependencies (Fig. 4b-d) that display SHAP values for different feature values.

The power step q was mostly determined by generation ramps and forecast errors (Fig. 4a,b). The most important feature, forecast error generation ramp, represents the difference between day-ahead scheduled generation ramps (on a 1h basis) and actual values of the total generation ramps that also include intraday trading within 15 minute intervals. Therefore, they mirror the additional 15 minute ramps that are not included in the day-ahead generation ramps thus making the feature

essential for the model to estimate the power steps q. In addition, 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 [28]. Accordingly, the PIML model yielded a high importance of hydro power ramps, which are among the fastest in the European power system. Interestingly, positive ramps exhibited a smaller effect then negative ramps. This probably relates to the limitation of our model in representing upwards deterministic deviations. Upwards deviations typically start before the start of the interval (cf. Sec. III C), while downward ramps rather follow the approximation of a discrete power step at $t = t_i$. Therefore, the remaining upward power step at t_i represents only a part of the total step. Hence, the model sees a step which is smaller than for downwards deviations and thus assumes a weaker effect of positive hydro ramps on the step q.

The drift r of the deterministic mismatch mirrors continuous changes of the load (Fig. 1b,c). Consistently,
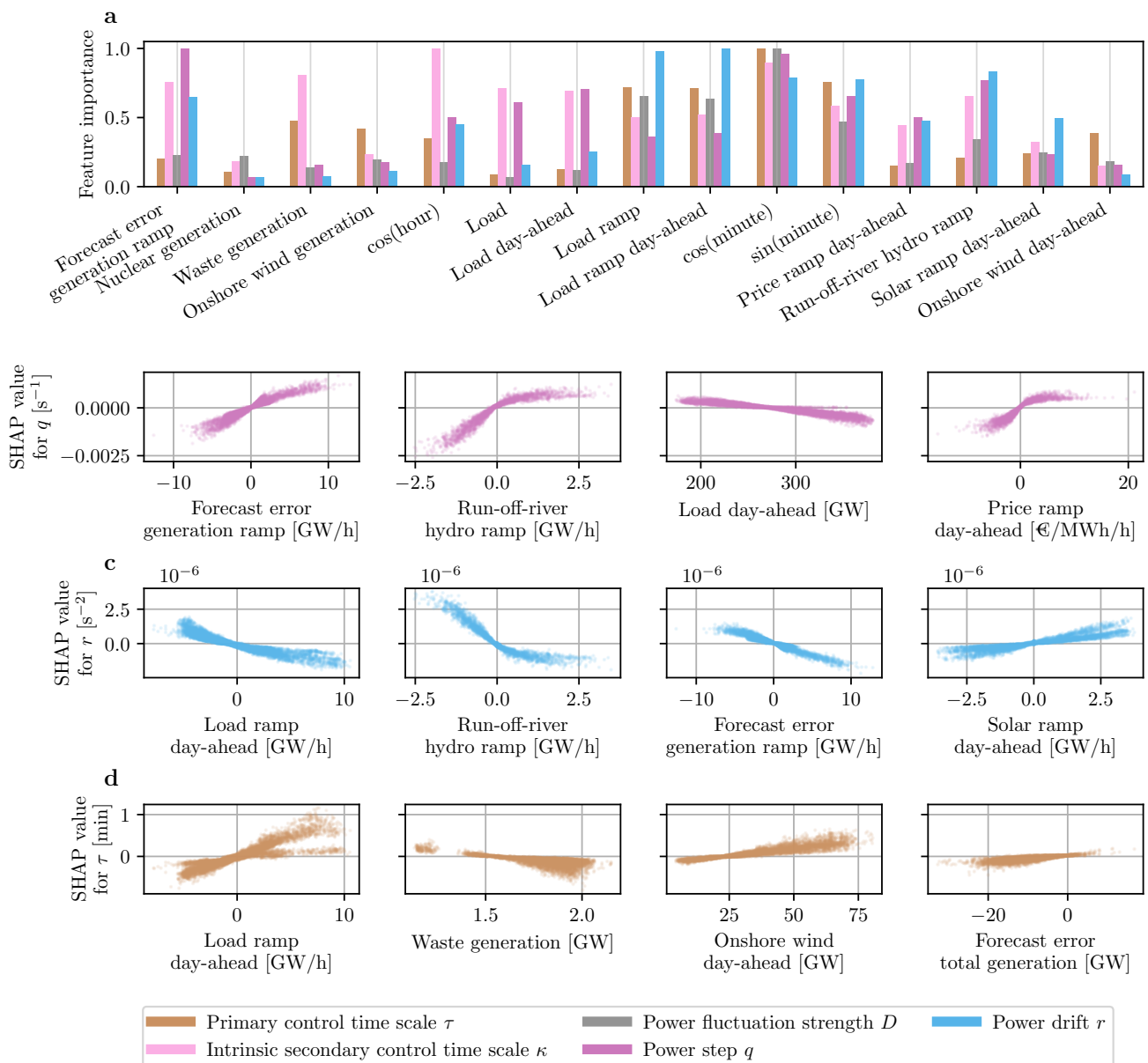
FIG. 4. SHAP values reveal dependencies between techno-economic features and dynamical system parameters. **a,** The mean absolute SHAP values quantify the overall importance of a feature, which varies strongly among the different dynamical system parameters. **b,c,d,** The relation between feature values 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 main text). In this panel, we only display the four most important techno-economic features, i.e., we excluded the time encoding features such as sinus of the minute "sin(minute)". Before the selection, we also excluded the actual features, e.g., actual load ramps, if a day-ahead version exists, e.g., load-ramp day-ahead, because the dependencies of these actual and day-ahead available features were largely the same.

load ramps obtained the highest feature importance for r with positive load ramps leading to negative slopes r (Fig. 4a,c). Solar ramps were also ranked highly, but their dependency showed the opposite behaviour. This mirrors the fact that in addition to the load, solar power also shapes the slow evolution of the deterministic mis-

match $P_{im}(t) \sim r \cdot t$ [32]: 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 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. 4c).

The effective time scale of primary control τ was determined by load ramps, waste power generation and wind power (Fig. 4a,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 [45]. 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. 4d), thus pointing to other causes such as an increased variability.

Finally, note that the hour and minute features were very important, in particular for τ, κ and D (Fig. 4a). This points to missing information in the feature set, so that the model relies on an average daily behaviour using the hour and minute features.

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 optimisation models [46, 47]. Given a data set of external features, synthetic time series are obtained as follows. For every interval i, we applied the FFNN to predict the system parameters $\boldsymbol{\vartheta}(i)$. 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 August 25 to September 05 in our test set, for which the first hours are shown in Fig. 5a.

Power grid frequency trajectories exhibit several highly characteristic stochastic properties [48]: The distribution of both frequency $\omega(t)$ and its increments $\Delta\omega_T(t) = \omega(t+T) - \omega(t)$ is heavy tailed (Fig. 5b,c). Large deviations and large jumps 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. 5d). All these characteristic patterns were well reproduced by our PIML model.

Moreover, the daily pattern strongly determines the grid frequency in Continental Europe [49]. Our synthetic time series adequately reproduced this important pattern, especially during the evening and night (Fig. 5e,f). In particular, the samples captured the specific dynamics of August and September 2019 in contrast to the daily profile predictor, which predicts a strong downward ramp at 20:00. Upward frequency deviations, for example at 05:00 and 06:00 in the morning, are reproduced less accurately than negative deviations at 23:00 (Fig. 5g). Our model does not capture continuous (conventional) generation ramps that start before the beginning of the interval (cf. Sec. III A). These effects were particularly strong for upward ramps (e.g., at 05:50 in Fig. 5g) and less dominant for downward ramps (e.g., at 23:50 in Fig. 5e), which explains the model deviations in the morning interval.

Furthermore, 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 non-autonomous character of the power system. The changing system parameters induce heavy tails in the frequency distribution, without the need for heavy-tailed power fluctuations, cf. the discussion in [15, 48].

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-inspired 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-inspired model. First, we provided a probabilistic prediction of the grid frequency in intervals of 15 minutes. Our model outperformed the daily average profile of the grid frequency, which already is a good predictor in Continental Europe [28]. This was also possible when using only day-ahead available techno-economic features, thus providing the possibility to forecast grid frequency dynamics 15 minutes ahead. Previous grid frequency predictors only used historic frequency data as inputs [49–51]. Approaches that integrated external features previously focused on aggregated frequency deviations [28, 52], which is also a cause of the data quality. Techno-economic features are typically available only on aggregated time scales of 15 min or 1h, while the frequency fluctuates on 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-inspired model. It connects the time-aggregated features with dynamical parameters of a stochastic process that well describes the short term grid frequency
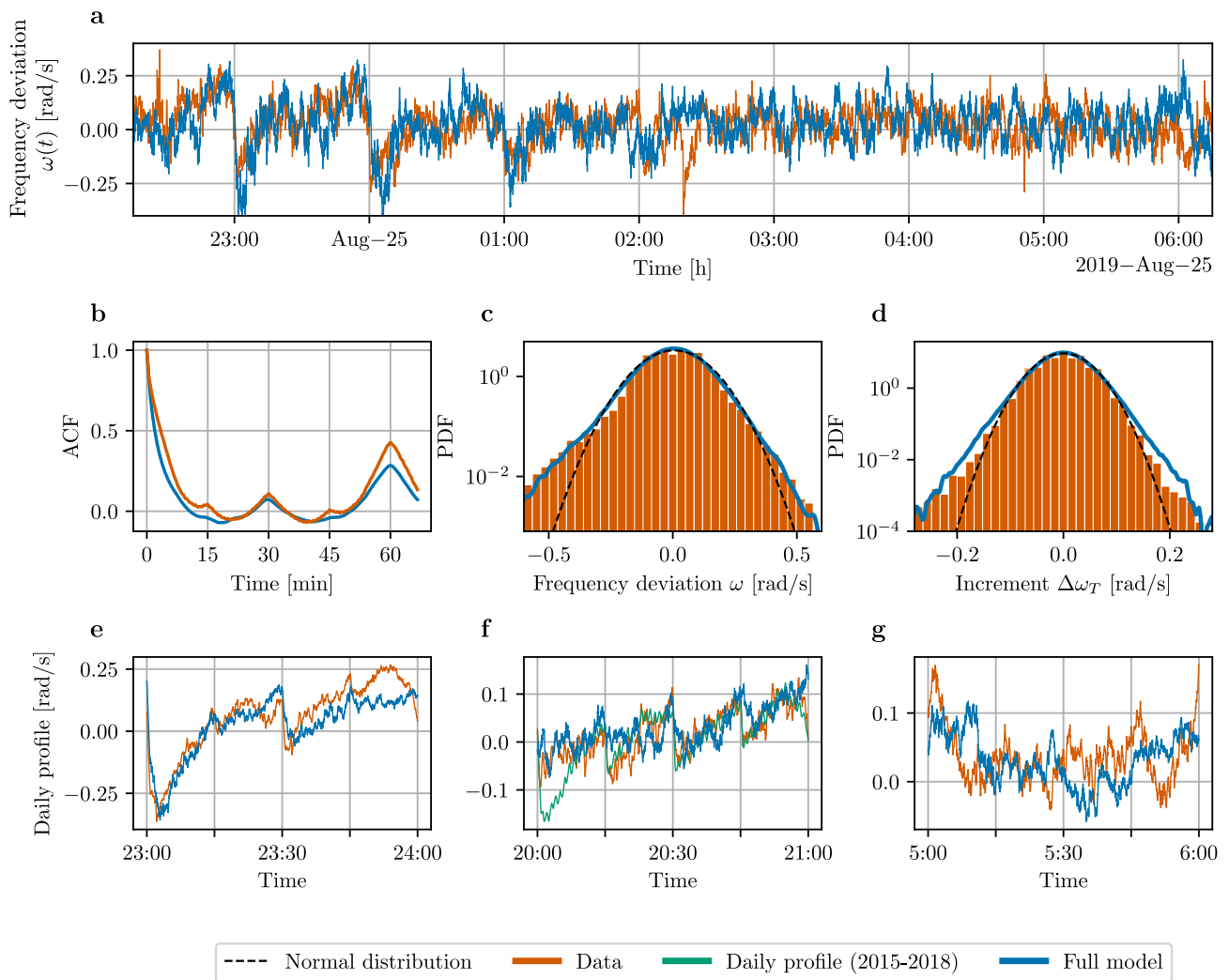
FIG. 5. Physics-inspired machine learning model generates synthetic grid frequency time series with characteristic properties. **a,** We sampled a synthetic frequency trajectory from our probabilistic model for the test period August 25 to September 05 in our test set, which was the longest period without missing or corrupted data points. The panel depicts the start of this period showing a good agreement between real and synthetic time series. **b,c,d,** The autocorrelation function (ACF) and the probability density function (PDF) of the frequency deviation $\omega(t)$ and its increments $\Delta\omega_T(t) = \omega(t+T) - \omega(t)$ ($T = 10$ s) have heavy tails, which is well reproduced by our synthetic time series. **e,f,g.** The daily average profiles of the real and synthetic frequency time series within the test period. For 20:00, we additionally display the daily profile computed from the training set (2015-2018). Our model better fits the frequency in this hour and therefore grasps local time-dependent properties of its daily evolution.

fluctuations.

Second, our model provides a tool for system identification and explanation. The model inferred the effective system parameters for every 15 minute interval 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, 53]. As an example, we extracted the time-varying nature of deterministic power imbalances, which

arises due to the step-wise evolution of scheduled conventional generation. The inferred power steps were particularly large for the first 15 minutes within the hour, which is consistent with the large share of generation traded at hourly day-ahead markets [44]. The strong time variation of the inferred parameters indicates the importance of modelling the grid frequency as a non-autonomous 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 are harder to control and thus effectively lead 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 scenario 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. Most interestingly, we only applied Gaussian white noise, but well reproduced the heavy-tailed distribution due to the time-dependent parameters of our stochastic model of power imbalance fluctuations and the control system. 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 easily be transferred to other grids.

In the context of power system dynamics and control, physics-inspired machine learning methods have become popular during the past years [54, 55]. Classical physics-informed neural networks (PINNs) are commonly applied to the differential equations directly [56], which we circumvented by solving our system analytically. However,

an analytical solution is not possible anymore when including non-linearities such as deadbands. In the future, our model can be modified to leverage classical PINNs to also treat non-linearities and more generic power system dynamics. Previous applications of PINNs to power system dynamics have successfully addressed autonomous dynamics [55, 57]. We contribute to these developments by proposing a model that explicitly models non-autonomous dynamics, which may greatly advance the application of physics-inspired machine learning in the energy sector.

[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, Nature Climate Change **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 and Energy Magazine **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, Renewable 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* (John Wiley & Sons, Inc., 2017).

[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, Nature Com-

munications **13**, 4593 (2022).

[10] C. Han, H. Hilger, E. Mix, P. C. Böttcher, M. Reyers, C. Beck, D. Witthaut, and L. R. 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* (John Wiley & Sons, Ltd., 2008).

[12] F. Milano, F. Dörfler, G. Hug, D. J. Hill, and G. Verbic, Foundations and Challenges of Low-Inertia Systems (Invited Paper), in *2018 Power Systems Computation Conference (PSCC)* (IEEE, 2018) pp. 1–25.

[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 Transactions on Power Systems **34**, 3098 (2019).

[14] L. R. 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 Transactions on Power Systems (in press) (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, Preprint at `https://ifisc.uib-csic.es/en/publications/dynamical-model-for-power-grid-frequency-fluctuati/` (2022).

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

[23] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Reviews of Modern Physics **91**, 045002 (2019).

[24] 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 Transactions on Power Systems **34**, 3188 (2019).

[25] L. Ljung, Perspectives on system identification, Annual Reviews in Control **34**, 1 (2010).

[26] J. Stiasny, G. S. Misyris, and S. Chatzivasileiadis, Physics-Informed Neural Networks for Non-linear System Identification for Power System Dynamics, in *2021 IEEE Madrid PowerTech* (2021) pp. 1–6.

[27] 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, Nature Machine Intelligence **2**, 56 (2020).

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

[29] J. Kruse, E. Cramer, B. Schäfer, and D. Witthaut, Physics-inspired machine learning for power grid frequency modelling (v0.1.0), `https://doi.org/10.5281/zenodo.7274829` (2022).

[30] J. Kruse, E. Cramer, B. Schäfer, and D. Witthaut, Supplementary data: "Physics-inspired machine learning for power grid frequency modelling", `https://doi.org/10.5281/zenodo.7273665` (2022).

[31] 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, 2009) pp. 1–6.

[32] J. Kruse, B. Schäfer, 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, 2021) pp. 133–139.

[33] J. Kruse, B. Schäfer, and D. Witthaut, Secondary control activation analysed and predicted with explainable AI, Electric Power Systems Research **212**, 108489 (2022).

[34] S. Pütz, B. Schäfer, D. Witthaut, and J. Kruse, Revealing interactions between HVDC cross-area flows and frequency stability with explainable AI, Preprint at `https://arxiv.org/abs/2204.10727` (2022).

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

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

[37] 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, Physics Reports A High-Bias, Low-Variance Introduction to Machine Learning for Physicists, **810**, 1 (2019).

[38] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* (JMLR Workshop and Conference Proceedings, 2010) pp. 249–256.

[39] S. M. Lundberg and S.-I. Lee, A unified approach to interpreting model predictions, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17 (2017) pp. 4768–4777.

[40] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, TensorFlow: a system for Large-Scale machine learning, in *12th USENIX symposium on operating systems design and implementation (OSDI 16)* (2016) pp. 265–283.

[41] 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 at `https://arxiv.org/abs/1711.10604` (2017).

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

[43] M. Anvari, L. R. Gorjão, M. Timme, D. Witthaut, B. Schäfer, and H. Kantz, Stochastic properties of the frequency dynamics in real and synthetic power grids, Physical review research **2**, 013339 (2020).

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

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

[46] E. Cramer, L. R. 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).

[47] 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, Computers & Chemical Engineering **166**, 107923 (2022).

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

[49] J. Kruse, B. Schäfer, and D. Witthaut, Predictability of Power Grid Frequency, IEEE Access **8**, 149435 (2020).

[50] W. Bang and J. W. Yoon, Forecasting the Electric Network Frequency Signals on Power Grid, in *2019 International Conference on Information and Communication Technology Convergence (ICTC)* (2019) pp. 1218–1223.

[51] 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 Transactions on Power Systems **29**, 2602 (2014).

[52] S. Kaur, S. Agrawal, and Y. P. Verma, Power Grid Frequency Prediction Using ANN Considering the Stochasticity of Wind Power, in *2013 5th International Conference and Computational Intelligence and Communication Networks* (2013) pp. 311–315.

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

[54] S. Chatzivasileiadis, A. Venzke, J. Stiasny, and G. Misyris, Machine Learning in Power Systems: Is It Time to Trust It?, IEEE Power and Energy Magazine **20**, 32 (2022).

[55] G. S. Misyris, A. Venzke, and S. Chatzivasileiadis, Physics-Informed Neural Networks for Power Systems, in *2020 IEEE Power Energy Society General Meeting (PESGM)* (2020) pp. 1–5.

[56] 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, Journal of Computational Physics **378**, 686 (2019).

[57] B. Huang and J. Wang, Applications of Physics-Informed Neural Networks in Power Systems - A Review, IEEE Transactions on Power Systems (in press) (2022).

[58] A. Ulbig, T. S. Borsche, and G. Andersson, Impact of Low Rotational Inertia on Power System Stability and Operation, IFAC Proceedings Volumes 19th IFAC World Congress, **47**, 7290 (2014).

[59] E. Weitenberg, Y. Jiang, C. Zhao, E. Mallada, C. De Persis, and F. Dörfler, Robust Decentralized Secondary Frequency Control in Power Systems: Merits and Trade-offs, IEEE Transactions on Automatic Control **64**, 3967 (2019).

[60] P. C. Böttcher, A. Otto, S. Kettemann, and C. Agert, Time delay effects in the control of synchronous electricity grids, Chaos: An Interdisciplinary Journal of Nonlinear Science **30**, 013122 (2020).

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

[62] C. W. Gardiner, *Handbook of stochastic methods*, Vol. 3 (Springer, 2009).

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

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

Appendix A: A stochastic model of grid frequency dynamics and control

1. The aggregated swing equation

Grid frequency deviations $\Delta f(t) = (f(t) - f_{\text{ref}})$ from the reference $f_{\text{ref}} = 50$ Hz 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 [58]:

$$P_{\text{rot}} = \frac{2HS_B}{f_{\text{ref}}} \frac{\mathrm{d}\Delta f}{\mathrm{d}t} = 2\pi M \frac{\mathrm{d}\Delta f}{\mathrm{d}t}. \tag{A1}$$

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. The parameter M then denotes the aggregated inertia of the grid.

Secondly, damping power is provided via primary control, also refered to as frequency containment reserve (FCR), and frequency-sensitive loads [58],

$$P_{\text{prim}} = K_1 \Delta f, \tag{A2}$$

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 one or two orders of magnitude larger than K_L [11]. Their values are often provided in the *per unit* (pu) system with $K_1^{pu} = f_{\text{ref}} K_1/P_0$, as the 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 (FRR), restores the frequency back to its reference f_{ref}. Secondary control is typically implemented as an integral controller (but other implementations exist)[59]:

$$P_{\text{sec}} = K_2 \bar{\theta}, \tag{A3}$$

with the integrated frequency deviation

$$\bar{\theta}(t) = \int_{t_i}^{t} \Delta f(t') \, \mathrm{d}t'. \tag{A4}$$

The parameter K_2 is the secondary control gain ($[K_2] = $ W), which reads $K_2^{pu} = f_{\text{ref}} K_2/P_0$ in the pu system. For example, in Great Britain a typical value of $K_2^{pu} = 0.05/$s is reported [13]. In interconnected power grids, secondary control may also be used to reduce unscheduled flows between different control areas [60]. These control actions are applied reciprocally in two areas and thus have only minor effects for the overall frequency dynamics.

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

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

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

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

2. Stochastic differential equations

Following reference [14], we modelled the power imbalances $\Delta P(t) = P_{\text{im}}(t) + \bar{D}\xi(t)$ as a sum of "deterministic" power imbalances $P_{\text{im}}(t)$ 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 [61]. Stochastic deviations are modelled as Gaussian noise $\xi(t)$ defined as the derivative of a Wiener process W_t that has independent, normally distributed increments $\mathrm{d}W_t$ with zero mean and variance $\langle \mathrm{d}W_t^2 \rangle = \mathrm{d}t$.

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

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

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

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

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

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

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

$$\mathrm{d}\theta = \omega\,\mathrm{d}t, \tag{A12}$$

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

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

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

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

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 [62]. Using Itô calculus, one obtains

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

3. Solution of the Fokker-Planck equation

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

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

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

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

if the parameters satisfy the ordinary differential equations

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

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

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

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

$$\frac{d}{dt}\sigma_{\theta\omega} = -2\tau^{-1}\sigma_\omega^2 - 2\kappa^{-2}\sigma_{\theta\omega}. \tag{A17}$$

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

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

In terms of the characteristic function, the FPE reads

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

$$= is\left[\left(P(t) + i\tau^{-1}\frac{\partial}{\partial s} + i\kappa^{-2}\frac{\partial}{\partial r}\right) + r\frac{\partial}{\partial s} - \frac{D^2}{2}s^2\right]\phi(\boldsymbol{u};t), \tag{A19}$$

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

$$\phi(\boldsymbol{u};t) = \exp\left(i\boldsymbol{u}^\top\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{u}^\top\Sigma\boldsymbol{u}\right), \tag{A20}$$

$$\phi(r,s;t) = \exp\left(i(r\mu_\theta + s\mu_\omega) - \frac{1}{2}(r^2\sigma_\theta^2 + s^2\sigma_\omega^2 + 2rs\sigma_{\theta\omega})\right).$$

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

$$\mathcal{L}\phi(\boldsymbol{u};t) = is\left[\left(P(t) + i\tau^{-1}\frac{\partial}{\partial s} + i\kappa^{-2}\frac{\partial}{\partial r}\right) + r\frac{\partial}{\partial s} - \frac{D^2}{2}s^2\right]\exp\left(i(r\mu_\theta + s\mu_\omega) - \frac{1}{2}(r^2\sigma_\theta^2 + s^2\sigma_\omega^2 + 2rs\sigma_{\theta\omega})\right),$$

$$= \left[isP(t) - \frac{D^2}{2}s^2 - (\tau^{-1}s - r)(i\mu_\omega - s\sigma_\omega^2 - r\sigma_{\theta\omega}) - \kappa^2 s(i\mu_\theta - r\sigma_\theta^2 - s\sigma_{\theta\omega})\right]\phi(\boldsymbol{u};t). \tag{A21}$$

Now we proceed with the left-hand side,

$$\frac{\partial}{\partial t}\phi(\boldsymbol{u};t) = \left[i\left(s\frac{d\mu_\omega}{dt} + r\frac{s\mu_\theta}{dt}\right) - \frac{1}{2}\left(r^2\frac{d\sigma_\theta^2}{dt} + s^2\frac{d\sigma_\theta^2}{dt} + 2rs\frac{d\sigma_{\theta\omega}}{dt}\right)\right]\phi(\boldsymbol{u};t). \tag{A22}$$

Inserting the equations (A17) then yields

$$\frac{\partial}{\partial t}\phi(\boldsymbol{u};t) = \left[isP(t) - \frac{D^2}{2}s^2 - (\tau^{-1}s - r)(i\mu_\omega - s\sigma_\omega^2 - r\sigma_{\theta\omega}) - \kappa^2 s(i\mu_\theta - r\sigma_\theta^2 - s\sigma_{\theta\omega})\right]\phi(\boldsymbol{u};t), \tag{A23}$$

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

4. Moment equations

The ordinary differential equations for the parameters (A17) can also be obtained in a more direct way, once we know that the PDF remains Gaussian for all times. In fact, we can exploit 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 $g(\boldsymbol{X})$ of the random variable $\boldsymbol{X}$ in Eq. (A14), Itô's lemma reads [63],

$$dg = \left[(\boldsymbol{\nabla}g)^T\boldsymbol{\mu} + \frac{1}{2}\text{Tr}\left(\boldsymbol{D}^\text{T}\boldsymbol{H}_\text{g}\boldsymbol{D}\right)\right]dt + (\boldsymbol{\nabla}g)^T\boldsymbol{D}d\boldsymbol{W}, \tag{A24}$$

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

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

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

$$\frac{d\langle\theta\rangle}{dt} = \langle\omega\rangle, \tag{A26}$$

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

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

$$\frac{\mathrm{d}\langle \theta^2 \rangle}{\mathrm{d}t} = 2\langle \theta\omega \rangle, \tag{A28}$$

$$\frac{\mathrm{d}\langle \omega^2 \rangle}{\mathrm{d}t} = 2P(t)\langle \omega \rangle - 2\tau^{-1}\langle \omega^2 \rangle - 2\kappa^{-2}\langle \omega\theta \rangle + D^2, \tag{A29}$$

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

In this derivation, we used $\langle \omega \mathrm{d}W \rangle = 0$ and $\langle \theta \mathrm{d}W \rangle = 0$. 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 Eq. (A17).

5. Solution of the moment equations

We now provide a semi-analytic solutions for the ordinary differential equations (A17) describing the evolution of the paramaters 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 $\boldsymbol{y}_d = (\mu_\theta, \mu_\omega)^T$,

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

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

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

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

$$\frac{\mathrm{d}\boldsymbol{y}_s}{\mathrm{d}t} = \boldsymbol{A}_s \boldsymbol{y}_s + \boldsymbol{b}_s, \tag{A34}$$

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

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

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 $\dot{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{y}$ with time-independent coefficients $\boldsymbol{A}$ is given by

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$+ \sigma_{\omega,\theta,0}\left(-2\tau e^{\lambda_{s,1}t} + \frac{8\tau^2}{\kappa^2}\frac{\lambda_{s,2}\kappa^2(4\tau)^{-1}+1}{\lambda_{s,2} - \lambda_{s,3}}e^{\lambda_{s,2}t} + \frac{8\tau^2}{\kappa^2}\frac{\lambda_{s,3}\kappa^2(4\tau)^{-1}+1}{\lambda_{s,3} - \lambda_{s,2}}e^{\lambda_{s,3}t}\right)$$

$$+ \sigma_{\omega,0}^2\left(-2\tau e^{\lambda_{s,1}t} + \frac{2\lambda_{s,2}\tau^2 - \lambda_{s,2}\kappa^2 - 2\tau}{\lambda_{s,3} - \lambda_{s,2}}e^{\lambda_{s,2}t} + \frac{2\lambda_{s,3}\tau^2 - \lambda_{s,3}\kappa^2 - 2\tau}{\lambda_{s,2} - \lambda_{s,3}}e^{\lambda_{s,3}t}\right)\right]\frac{1}{\kappa^2 - 4\tau^2}, \tag{A50}$$

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. (Semi-)Analytical solution of inhomogeneous equations

The general solution of the inhomogeneous ODE $\dot{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}(t)$ with time-independent coefficient $\boldsymbol{A}$ is given by the sum of the homogeneous solution and an inhomogeneous contribution $\boldsymbol{y}_{in}(t)$

$$\boldsymbol{y}_{\mathrm{ODE}}(t) = \boldsymbol{y}_h(t) + \boldsymbol{y}_{in}(t),$$

$$= \boldsymbol{y}_h(t) + \boldsymbol{U}(t)\int_0^t \boldsymbol{U}^{-1}(t')\boldsymbol{b}(t')\mathrm{d}t'. \tag{A51}$$

We first provide a semi-analytical solution, which leaves the integration of the inhomogeneity $b_d(t)$ to a numerical routine. This enables us to flexibly insert different power function $P(t)$, as we will see below.

The ω-components of the inhomogeneous contributions yield

$$\mu_{\omega,in}(t) = \frac{1}{\lambda_{d,1} - \lambda_{d,2}}\left[\lambda_{d,1}e^{\lambda_{d,1}t}\int_0^t P(t')e^{-\lambda_{d,1}t'}\mathrm{d}t' - \lambda_{d,2}e^{\lambda_{d,2}t}\int_0^t P(t')e^{-\lambda_{d,2}t'}\mathrm{d}t'\right], \tag{A52}$$

$$\sigma_{\omega,in}^2(t) = \frac{D^2}{\kappa^2 - 4\tau^2}\left[\frac{2\tau^2}{\lambda_{s,1}}(1 - e^{\lambda_{s,1}t}) + \frac{2\tau/\lambda_{s,2} - 2\tau^2 + \kappa^2}{\lambda_{s,3} - \lambda_{s,2}}(1 - e^{\lambda_{s,2}t}) + \frac{2\tau/\lambda_{s,3} - 2\tau^2 + \kappa^2}{\lambda_{s,2} - \lambda_{s,3}}(1 - e^{\lambda_{s,3}t})\right]. \tag{A53}$$

Note that the deterministic part 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. (A52) can become very large as $\lambda < 0$ for stable systems thus causing numerical problems. To use this semi-analytical solution during neural network training, one has to mitigate these numerical problems, e.g., by strongly restricting the parameter space.

We used a fully analytical solution for the case $P(t) = q + rt$, thus avoiding these numerical problems. In this case, the inhomogeneous solution of the deterministic part yielded

$$\mu_{\omega,in}(t) = \frac{1}{\lambda_{d,1} - \lambda_{d,2}}\left[e^{\lambda_{d,1}t}\left(q + \frac{r}{\lambda_{d,1}}\right) - e^{\lambda_{d,2}t}\left(q + \frac{r}{\lambda_{d,2}}\right) + r\left(\frac{1}{\lambda_{d,2}} - \frac{1}{\lambda_{d,1}}\right)\right]. \tag{A54}$$

TABLE II. Properties of dynamical system parameters that are predicted by the FFNN (cf. Fig. 1d). The parameters are subject to several physical constraints which are summarised 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 the 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	ϑ_7	ϑ_8
Name	$\sigma_{\theta,0}$	$\sigma_{\theta,\omega,0}$	$\sigma_{\omega,0}$	τ	κ	D	q	r
Physical constraints	$\sigma_{\theta,0} > 0$	$\lvert\sigma_{\theta,\omega,0}\rvert \leq \sigma_{\theta,0}\sigma_{\omega,0}$	$\sigma_{\omega,0} > 0$	$0 < \tau \leq \kappa/2$	$\kappa > 0$	$D > 0$	$r \in \mathbb{R}$	$q \in \mathbb{R}$
Scaling s_j	0.01	-	0.1	-	100	0.01	10^{-3}	10^{-6}
Scaling variation	$\{1\}$	-	$\{0.1\}$	-	$\{1000, 100\}$	$\{0.01, 0.1\}$	$\{10^{-3}, 10^{-2}\}$	$\{10^{-5}, 10^{-6}\}$
Minimum v_j	10^{-3}	-	10^{-3}	10	30	10^{-4}	-	-
Range of $\nu_j(u_j)$	(v_1, ∞)	$(-\sigma_{\theta,0}\sigma_{\omega,0}, \sigma_{\theta,0}\sigma_{\omega,0})$	(v_3, ∞)	$(v_4, \kappa/2)$	(v_5, ∞)	(v_6, ∞)	$(-\infty, \infty)$	$(-\infty, \infty)$

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

Appendix B: Parameter values from the literature

Based on the definition in Eqs. (A12)-(A13), Gorjao et al. inferred time-independent parameter values using the Kramers-Moyal expansion [14]. The authors employed the actual grid frequency instead of the angular velocity. Thus, we rescaled the results according to $D \to 2\pi D$, $q \to 2\pi q$, $r \to 2\pi r$, $\theta \to 2\pi\theta$, while τ and κ stayed the same.

$$\tau = 120 \ s\,, \tag{B1}$$

$$\kappa = 183 \ s\,, \tag{B2}$$

$$D = 0.007 \ s^{-3/2}\,, \tag{B3}$$

$$q = 0.0042 \ s^{-2}\,, \tag{B4}$$

$$r = 0.000009 \ s^{-3}. \tag{B5}$$

In the main text, we use these parameter values as reference values and therefore call them $\vartheta_j^{\mathrm{ref}}$. Note that the authors of ref. [14] 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 $r = 2q/t_{max}$. Moreover, the parameter q was specified separately for the full hour and for every (other) quarter of the hour, so we took the average value here.

Appendix C: Parameter Scaling and Constraints

The developed PIML model includes a layer that rescales the parameters and ensures some physical constrains (Fig. 1d). The outputs u_j of the FFNN do not necessarily fulfil the physical constraints of the parameters ϑ_j (cf. Tab. II), 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. appendix B), but the outputs u_j of the initialised FFNN typically have the same scale due uniform random initialisation of the weights [64]. This will yield large initial errors along certain parameter axis thus leading to inhomogeneous loss landscapes which can make optimisation inefficient and more difficult [37].

Therefore, we added a constraint and scaling layer that applies functions ν_j to the FFNN outputs. The results then represent the parameter estimates $\vartheta_j = \nu_j(u_j)$. First, the functions ν_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 $\mathrm{Sig}(u) = (1 + \exp(-u))^{-1} \in (0, 1)$ was used to ensure that $\tau \leq \kappa/2$ holds. Numerical imprecision can lead to a violation of these constraints so that we added a safety factor $\delta = 0.999$ in some cases. Second, the factors s_j, which mirror the typical scale of parameters ϑ_j, are applied to make the optimisation more efficient. Third, minimum values are added in certain cases to ensure numerical stability during optimisation. 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 Tab. II:

$$\nu_1(u_1) = \mathcal{S}_+(u_1)s_1 + v_1,$$
$$\nu_2(u_2) = \delta \tanh(u_1)\nu_1(u_1)\nu_3(u_3),$$
$$\nu_3(u_3) = \mathcal{S}_+(u_3)s_3 + v_3,$$
$$\nu_4(u_4) = \left(\frac{2v_4}{\nu_5(u_5)} + \delta \mathrm{Sig}(u_4)\left(1 - \frac{2v_4}{\nu_5(u_5)}\right)\right)\frac{\nu_5(u_5)}{2},$$
$$\nu_5(u_5) = \mathcal{S}_+(u_5)s_5 + v_5,$$
$$\nu_6(u_6) = \mathcal{S}_+(u_6)s_6 + v_6,$$
$$\nu_7(u_7) = u_7 s_7,$$
$$\nu_8(u_8) = u_8 s_8.$$

To test the impact of the scaling with s_j, we varied the scaling parameters according to Tab. II. 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 10 different random initialisation of the FFNN weights. Finally, we trained 10 initialisations using the standard scaling defined in Tab. II and no scaling with $s_j = 1$ (cf. Sec. III B).

5. Discussion and Outlook

In this thesis, I have developed machine learning models for the prediction, explanation and stochastic modelling of the power grid frequency dynamics. The grid frequency is a central indicator for balance and control in power system operation and thereby a relevant non-autonomous, stochastic dynamical system to investigate. I presented a univariate prediction method that approximates future grid frequency deviations better than important benchmarks such as the daily frequency pattern. However, the impact of external drivers such as electricity trading limits its performance, which points to the importance of external impact factors beyond the intrinsic dynamics of load-frequency control. Using explainable machine learning, I inferred the impact of such external drivers on indicators of frequency stability and control activation. The dependency on techno-economic drivers varied between different power grids, and I revealed the specific roles of different generation types, forecast errors and renewable generation. Turning back to grid frequency dynamics, I incorporated the effects of external drivers into stochastic dynamical models to describe large-scale trends as well as small-scale frequency fluctuations. This yielded a physics-inspired machine learning model that successfully reproduced important stochastic characteristics of grid frequency dynamics. The model heavily built on the time-dependent impact of the external features, which emphasises the importance of non-autonomous modelling for power system dynamics.

In the following sections 5.1-5.3, I discuss the impact and the limitations of my work with respect to the three main topics Q.1-Q.3, and give an outlook on possible extensions and methodological developments. Then, in Section 5.4, I discuss my thesis in the context of power system engineering with a focus on the relation between simulation-based and data-driven methods for power system stability and control. Finally, in Section 5.5, I close with an outlook on other types of non-autonomous dynamical systems and their data-driven prediction.

5.1. Grid frequency prediction and its applications

The frequency predictors developed in this thesis outperformed important benchmarks. They predicted frequency deviations 15-30 minutes ahead (Section 2) or even day-ahead (Section 4) by using day-ahead available techno-economic features as input data.

Such frequency predictors enable multiple applications in power system operation. Pre-

dicting typical trajectories facilitates the detection of outliers and measurement errors [74, 75]. The prediction can then be used to impute missing or corrupted data points. Furthermore, grid frequency predictions are useful to operate battery energy storage systems (BESS), which are increasingly used for primary and secondary frequency control [76, 77]. To optimise the state-of-charge of the BESS, model predictive control can be used in combination with a grid frequency prediction [78]. Model predictive control was also proposed to anticipate and control deterministic frequency deviations, which can enhance frequency stability through preventive actions [79].

However, the operation of BESS often requires prediction horizons around one hour [80]. Such a prediction horizon enables the BESS owner to optimise the bidding strategy on intraday electricity markets that operate in intervals of 15 or 60 minutes. The univariate frequency predictor in Section 2.1, has a prediction horizon of 15-30 minutes. Other univariate predictors from the literature have a similar performance [55] or even shorter prediction horizons [74, 81]. This can impede their application for BESS operation in certain markets. To reach longer prediction horizons, one has to include external drivers, which I demonstrated in Sections 3 and 4. My prediction models that include external features offer a day-ahead prediction and thus a possibility to optimise bidding strategies on intraday markets.

Notably, the application of grid frequency predictions for control systems has certain limits, because frequency data already includes the effect of control measures. Assuming the frequency trajectory to be fixed is only appropriate if the system under investigation has a negligible feedback effect on the grid frequency, e.g., in the case of a single wind turbine [82]. Frequency response through batteries, however, impacts the frequency trajectory. Therefore, model predictive control typically includes a dynamical system model that is able to predict the effect of control interventions on the system [83] or such a model is estimated based on observational data [84].

In the context of power system control, the prediction of power imbalances instead of frequency deviations can be of great importance. In the past years, power imbalance data was scarcely available and time series were only uploaded with 15 minute or 60 minute resolution [73, 85]. Recently, data with higher resolution became available [73, 85], thus opening the possibility to forecast even short-term power imbalances. Imbalance prediction can be used for preventive control actions of system operators [86] or for estimating the necessary amount of control reserves for the upcoming day or week [87, 88]. In Section 3.3, I presented a model that predicts control activation on a 15-minute basis, which outperformed important benchmarks. In the future, similar models based on techno-economic features can be trained to predict short-term power imbalances, which can yield great benefits for frequency control and reserve sizing.

5.2. Explainable machine learning: causality vs. correlation

In this thesis, I used explainable machine learning to infer important risks and drivers of frequency stability from operational data. In particular, I used a multivariate approach to model grid frequency stability and control based on multiple techno-economic features from electricity markets and power system operation (Section 3).

This multivariate approach has clear advantages compared to previous empirical analyses. Previous empirical studies on grid frequency data employed simple correlation analysis [89–91] and conditional probability assessment [53] to quantify the impact of external features. However, as already pointed out in ref. [92], correlations between external features can lead to wrong conclusions concerning their impact on the target variable. I have demonstrated this effect in Section 3.1, where I analysed the impact of nuclear ramps, i.e., the change of generation of nuclear power plants. The correlation between nuclear ramps and the RoCoF was positive, but the SHAP analysis of my multivariate model yielded a negative dependency. The negative dependency was consistent with domain knowledge (namely, that nuclear ramps are too slow to drive the RoCoF). This revealed the inadequacy of simple correlation analysis, which most probably was caused by a correlation among the features. The multivariate model adjusted for the effect of multiple features and therefore better revealed the remaining actual effect of nuclear ramps.

Nevertheless, such machine learning methods have certain limits related to the inference of causal relations. To be precise, my goal is to estimate causal effects of techno-economic drivers on grid frequency dynamics. However, as stated in refs. [93, 94], SHAP values do not reveal causal relationships between variables in reality, but only the causal impact of a feature on the model prediction. This impact does not necessarily correspond to a causal impact in reality, because machine learning models commonly learn associations and not causalities. I have addressed this limitation by interpreting my SHAP results with domain knowledge, thus identifying physically meaningful dependencies (Section 3.1). For example, the model revealed a strong importance of forecast errors for the Nordic grid frequency dynamics, which was consistent with reports from Nordic system operators [95]. In particular, the SHAP analysis provided a comprehensive quantitative analysis of how and when forecast errors are important, going far beyond the previous state-of-the-art. As another example, my model extracted a strong impact of hydro power on the activation of secondary control power in Section 3.3. Using domain knowledge, I exposed this dependency as a possible reverse causality, i.e., control activation caused an increase in hydro power generation and not vice versa. This enabled the design of restricted feature sets that better allow for a causal interpretation.

Recently, different studies have addressed the problem of causality and SHAP values by

incorporating causal knowledge into the model explanations. Causal shapely values [96] and asymmetric shapely values [97] attribute the model prediction to individual features by using causal domain knowledge. Shapley flow [98] generalises these approaches in an axiomatic way by reformulating the problem as credit attribution to edges on a directed acyclic graph. This method decomposes the model prediction into the effect of causal relations (edges) in a causal graph, which has to be defined beforehand. In the future, such tools can improve the interpretability of my explainable machine learning models by explicitly incorporating engineering knowledge about the relation of power system features.

In general, this discussion targets the problem of identifying and modelling causal relations from observational data [99]. Learning causation requires us to quantify the impact of interventions instead of just learning conditional probabilities. The field of causal inference and causal discovery comprises a large amount of literature. For example, a major goal is the estimation of interventional effects, such as the average treatment effect, or the reconstruction of causal relationships as causal graphs, for which different estimation methods exist. Causal inference is straightforward for experimental data, where the treatment is applied while controlling all other variables. However, such experiments are typically impossible in large techno-economic systems such as energy markets and power systems. We just have observational data with variables that are mutually dependent. Therefore, advanced causal inference and discovery methods are necessary, which can facilitate empirical analyses of power systems in the future. Multiple python packages implement such methods, e.g., an end-to-end library for causal inference "DoWhy" [100],"EconML"[101] to estimate heterogeneous treatment effects from observational data via machine learning, and the "Causal Discovery Toolbox" [102] for recovering causal graphs from observational data. Such tools can be helpful for power system analysis, but they have to be carefully adapted to deal with the specific data sources and the causal domain knowledge in energy system research.

5.3. Stochastic modelling with insufficient data

In Section 4, I investigated how to incorporate external drivers into stochastic dynamical models of the grid frequency. To this end, I proposed a physics-inspired machine learning model, which successfully predicted frequency trajectories and reproduced central stochastic characteristics of frequency time series.

External drivers are essential for grid frequency dynamics, and multiple time series of these features are publicly available. As I explained in Section 1.2 and 1.3, the grid frequency is affected by diverse impact factors such as power imbalances due to forecast errors, renewable power fluctuations or short-term load fluctuations. Even the parameters of the dynamical system such as the inertia can vary over time [103] due to the varying

operational conditions. Time series of renewable power generation, load and other important features are publicly available [73] and contain information about grid frequency dynamics or their dynamical parameters.

One of the main challenges is that such techno-economic time series are typically available only at an hourly timescale, while frequency dynamics are much faster. For example, the ramping behaviour of conventional generation on timescales of minutes, i.e., their change of power generation, strongly affects the deterministic frequency deviations that occur due to the market-based dispatch of the generators (cf. Section 1.2). Publicly available generation time series typically have a resolution of 15 to 60 minutes [73], which does not capture these important dynamics.

Using such techno-economic time series in dynamical system models therefore requires a careful data assimilation, but it can greatly facilitate realistic modelling. For example, ref. [104] assimilated load and generation time series to successfully reproduce grid frequency measurements from the Gran Canaria island. The authors interpolated load and renewable generation that come with a 10-minute resolution, while dynamical parameters such as the inertia were calculated directly from properties of the individual power plants.

In large-scale power systems, such as Continental Europe, the integration of techno-economic times series into dynamical models is more challenging. Load and generation time series contain various missing or unrealistic values, and time series for individual power plants are scarce [49]. Parameters of load-frequency control change over time and differ among local control regions [105], but dynamical models mostly use a time-independent uniform control factor [19, 21, 105], such that information on detailed time-dependent parameters is not available (to the best of my knowledge). Even the inertia, which is essential for frequency dynamics (Section 1.2), is typically only approximated from generation time series instead of calculating exact values [91, 103] due to the unavailability of sufficient data.

In this context, I contribute with my physics-inspired machine learning model by systematically integrating various operational time series of large-scale power systems into dynamical models. Instead of assimilating the data via self-engineered functions, I inferred the relationships with an artificial neural network and a maximum-likelihood approach. This generic method offers the flexibility to be applied to any power system, even if the data is incomplete and insufficient.

Notably, this inference task has certain limits. Suitable solutions are non-unique and there are redundancies between different parameters (cf. Section 4.1). In the future, these problems could be explored by increasing training times to better approach a global optimum. Another option is to further constrain the parameter space and force dynamical parameters into certain intervals based on domain knowledge. Finally, the physical model can also be inaccurate so that the correct parameters are not an optimum in the training

process anymore. Possible model extensions and variations include more detailed representations of deterministic imbalances, e.g., concerning the ramping behaviour of conventional generators, the integration of non-linearities, e.g., the deadbands of primary control, or better models of the control loops, which is particularly important for inverter-based renewable generation [42].

5.4. Relation to simulation-based methods from engineering

The power grid is a human-made system, and physical models of its devices are therefore well established in engineering science [14]. However, when modelling measured data, simulation-based approaches often use simplifications and not all model parameters are exactly known [105]. For instance, precise operation strategies are rarely disclosed, and market outcomes depend on unknown human decisions. Data-driven methods can complement simulation-based methods by validating results and inspiring new model-based studies. As a first example, simulation-based studies suggest that inertia is not important for ambient, i.e., day-to-day frequency fluctuations in Great Britain [60]. Notably, inertia is one of the most critical topics for frequency stability in renewable power systems because it determines the RoCoF after a disturbance such as an exceptional generator outage (cf. Section 1.3 and ref. [42]). In Section 3.1, I found that inertia has a negligible feature importance in my model of day-to-day frequency deviations, thus validating the previous finding of ref. [60]. Second, I found an unexpected relation between electricity prices and frequency variability in Great Britain, which could inspire model-based explanations (Section 3.1).

Due to the energy transition, machine learning techniques will become even more beneficial. The integration of great numbers of decentralised renewable energy sources and the coupling of diverse energy sectors and devices can render the optimisation of power system security intractable [106]. Machine learning models can be faster, more efficient and more accurate, thus allowing for secure system operation under increasingly complex conditions [107].

However, an important challenge is to create trustworthy machine learning tools for the application in safety-critical domains such as power system operation and control [68]. My approach of using explainable models based on SHAP values opens the possibility to make transparent predictions, thus increasing the trust in such data-driven applications [67]. Recently, SHAP values were shown to reproduce Power Transfer Distribution Factors (PTDFs), a well established sensitivity index for power flows, in a machine learning model of a test grid, thus emphasising the potential of SHAP for power system applications [68]. Furthermore, hybrid approaches that incorporate physics into machine learning tools have

been proposed for power systems [108]. They add trustworthiness by producing models that closely approximate certain physical laws. I contribute to these developments with my physics-inspired model in Section 4. Here again, machine learning and simulation-based methods complement each other, e.g., by using feedback loops between simulation and machine learning to enhance training sets and prediction performance [106].

5.5. Data-driven modelling of non-autonomous dynamical systems

In this thesis, I investigated grid frequency dynamics and control, which can be described by an elementary system of dynamical equations (1.2). However, the grid frequency is driven by external power imbalances $\Delta P(t)$, thus turning it into a non-autonomous dynamical system. The complex nature of power imbalances in large interconnected power grids challenges the prediction and modelling of these dynamics.

These challenges primarily arise due to a lack of sufficient data that quantifies the external driving (cf. Section 5.3). (1) The data is only available with low temporal and spatial resolution, it is noisy and includes measurement errors and missing data points. (2) Some physical variables, such as the inertia, are not directly observable in large-scale power systems. (3) The data is observational and not experimental, i.e., it was not extracted under controlled experimental conditions, and the resulting interdependencies between variables might not relate to causal effects. Modelling the impact of external drivers on frequency dynamics with such data is not straightforward.

In this thesis, I demonstrated that machine learning can be used to still benefit from such data. In particular, explainability was crucial, as it enables us to check the plausibility of results with domain knowledge or physical models. I showed how transparency and explainability can be achieved through explanation tools such as SHAP or by integrating physics and machine learning in physics-inspired models.

The above challenges also arise in other areas of power system stability, e.g., rotor-angle stability [14]. The synchronisation of multiple generators in a power grid is strongly affected by external effects such as line outages under different operating conditions. Here, detailed dynamical data is typically not available, because it is too expensive to gather for every realisation of power system operation. Instead, static features are readily available, which however do not directly provide the dynamical information. We have studied how machine learning can contribute in leveraging static features for dynamic stability assessment in such situations with reduced availability of dynamical data [70].

Beyond power systems, we encounter similar challenges when modelling dynamics of epidemics, climate or financial markets. Epidemics exhibit strongly non-autonomous dynamics as the spreading of diseases is affected by multiple factors such as seasonality, polit-

ical decisions and societal behaviour [109, 110]. Explainable machine learning tools, such as physics-inspired neural networks, have been used to model such dynamics [111]. The climate is driven by external forcing, e.g., due to human-induced climate change. Typically, not all relevant system states can be measured and data is noisy, including measurement errors. Meteorological data therefore has to be assimilated into physical models, for which machine learning can be used, too [112]. Finally, financial markets exhibit diverse external impact factors and multiple machine learning applications have been developed to forecast its non-autonomous dynamics [113]. In my thesis, I discussed generic challenges and opportunities of using machine learning to model power system dynamics. This contributes to a better understanding of how to model non-autonomous and stochastic dynamical system in general.

Bibliography

1. Kruse, J., Schäfer, B. & Witthaut, D. Predictability of Power Grid Frequency. *IEEE Access* **8,** 149435–149446 (2020).

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

3. Kruse, J., Schäfer, B. & Witthaut, D. *Exploring deterministic frequency deviations with explainable AI* in *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)* (2021), 133–139.

4. Kruse, J., Schäfer, B. & Witthaut, D. Secondary control activation analysed and predicted with explainable AI. *Electric Power Systems Research* **212,** 108489 (2022).

5. Kruse, J., Cramer, E., Schäfer, B. & Witthaut, D. *Physics-inspired machine learning for power grid frequency modelling.* Preprint at `https://doi.org/10.48550/arXiv.2211.01481`. 2022.

6. Petermann, T., Bradke, H., Lüllmann, A., Poetzsch, M. & Riehm, U. *What happens during a blackout: Consequences of a prolonged and wide-ranging power outage* (Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB), 2011).

7. Laugé, A., Hernantes, J. & Sarriegi, J. M. Critical Infrastructure Dependencies: A Holistic, Dynamic and Quantitative Approach. *International Journal of Critical Infrastructure Protection* **8,** 16–23 (2015).

8. Lukasik, S. Vulnerabilities and Failures of Complex Systems. *International Journal of Engineering Education* **19,** 206–212 (2003).

9. European Commission *et al. EU Reference Scenario 2016: Energy, transport and GHG emissions - Trends to 2050* . `http://doi.org/10.2833/554470`. 2016.

10. Orths, A. *et al.* Flexibility From Energy Systems Integration: Supporting Synergies Among Sectors. *IEEE Power and Energy Magazine* **17,** 67–78 (2019).

11. ENTSO-E. *European Resource Adequacy Assessment - 2021 Edition.* `https://eepublicdownloads.azureedge.net/clean-documents/sdc-documents/ERAA/ERAA_2021_Executive%20Report.pdf` (3rd Nov. 2022).

12. Bundesnetzagentur. *Monitoring report 2021.* `https://www.bundesnetzagentu` `r.de/EN/Areas/Energy/Companies/DataCollection_Monitoring/start.` `html` (3rd Nov. 2022).

13. TransnetBW GmbH. *Regelenergie Bedarf + Abruf.* `https://www.transnetbw.d` `e/de/strommarkt/systemdienstleistungen/regelenergie-bedarf-und-` `abruf` (3rd Nov. 2022).

14. Machowski, J., Bialek, J. & Bumby, J. *Power system dynamics, stability and control* (John Wiley & Sons, New York, 2008).

15. Ulbig, A., Borsche, T. S. & Andersson, G. Impact of Low Rotational Inertia on Power System Stability and Operation. *IFAC Proceedings Volumes. 19th IFAC World Congress* **47,** 7290–7297 (2014).

16. Union for the Coordination of Transmission of Electricity. *Final report on the system disturbance on 4 November 2006* provided by the successor organization of the UCTE, the European Network of Transmission System Operators for Electricity (ENTSO-E). `https://www.entsoe.eu/fileadmin/user_upload/_library/` `publications/ce/otherreports/Final-Report-20070130.pdf` (14th June 2021).

17. Pourbeik, P., Kundur, P. S. & Taylor, C. W. The Anatomy of a Power Grid Blackout - Root Causes and Dynamics of Recent Major Blackouts. *IEEE Power and Energy Magazine* **4,** 22–29 (2006).

18. ENTSO-E. *Continental Europe Significant Frequency Deviations - January 2019.* `https://eepublicdownloads.entsoe.eu/clean-documents/news/2019/` `190522_SOC_TOP_11.6_Task%20Force%20Significant%20Frequency%` `20Deviations_External%20Report.pdf` (3rd Nov. 2022).

19. Gorjão, L. R. *et al.* Data-driven model of the power-grid frequency dynamics. *IEEE Access* **8,** 43082–43097 (2020).

20. Weitenberg, E. *et al.* Robust Decentralized Secondary Frequency Control in Power Systems: Merits and Tradeoffs. *IEEE Transactions on Automatic Control* **64,** 3967–3982 (2019).

21. Weissbach, T. & Welfonder, E. *High Frequency Deviations within the European Power System: Origins and Proposals for Improvement* in *2009 IEEE/PES Power Systems Conference and Exposition* (IEEE, Seattle, 2009), 1–6.

22. Hirth, L. & Ziegenhagen, I. Balancing Power and Variable Renewables: Three Links. *Renewable and Sustainable Energy Reviews* **50,** 1035–1051 (2015).

23. Witthaut, D. *et al.* Collective nonlinear dynamics and self-organization in decentralized power grids. *Reviews of Modern Physics* (2021).

24. Bloomfield, H. C., Brayshaw, D. J., Shaffrey, L. C., Coker, P. J. & Thornton, H. E. The Changing Sensitivity of Power Systems to Meteorological Drivers: A Case Study of Great Britain. *Environmental Research Letters* **13,** 054028 (2018).

25. Heide, D. *et al.* Seasonal optimal mix of wind and solar power in a future, highly renewable Europe. *Renewable Energy* **35,** 2483 (2010).

26. Staffell, I. & Pfenninger, S. The Increasing Impact of Weather on Electricity Supply and Demand. *Energy* **145,** 65–78 (2018).

27. Collins, S., Deane, P., Gallachóir, B. Ó., Pfenninger, S. & Staffell, I. Impacts of Inter-Annual Wind and Solar Variations on the European Power System. *Joule* **2,** 2076–2090 (2018).

28. Wohland, J., Omrani, N. E., Keenlyside, N. & Witthaut, D. Significant Multi-Decadal Variability of German Wind Energy Generation. *Wind Energy Science Discussions,* 1–21 (2019).

29. Anvari, M. *et al.* Short Term Fluctuations of Wind and Solar Power Systems. *New Journal of Physics* **18,** 063027 (2016).

30. Anvari, M., Wächter, M. & Peinke, J. Phase Locking of Wind Turbines Leads to Intermittent Power Production. *EPL (Europhysics Letters)* **116,** 60009 (2016).

31. Foley, A. M., Leahy, P. G., Marvuglia, A. & McKeogh, E. J. Current Methods and Advances in Forecasting of Wind Power Generation. *Renewable Energy* **37,** 1–8 (2012).

32. Pesch, T., Allelein, H.-J. & Hake, J.-F. Impacts of the transformation of the German energy system on the transmission grid. *Eur. Phys. J. Special Topics* **223,** 2561 (2014).

33. 50Hertz Transmission GmbH, Amprion, TenneT TSO & TransnetBW. *Netzentwicklungplan Strom.* `http://www.netzentwicklungsplan.de` (3rd Nov. 2022).

34. ENTSO-E. *Ten-year network development plan: TYNDP 2020 Main Report.* `https://tyndp.entsoe.eu/documents` (3rd Nov. 2022).

35. Tellefsen, T., van Putten, J. & Gjerde, O. Norwegian Hydropower: Connecting to Continental Europe. *IEEE Power and Energy Magazine* **18,** 27–35 (2020).

36. Lin, J. & Magnago, F. H. *Electricity Markets: Theories and Applications* (John Wiley & Sons, Inc., Hoboken, 2017).

37. Gomez, T. *et al.* European Union Electricity Markets: Current Practice and Future View. *IEEE Power and Energy Magazine* **17,** 20–31 (2019).

38. European Commission. *Commission Regulation (EU) 2017/2195 of 23 November 2017: Establishing a Guideline on Electricity Balancing (Text with EEA Relevance).* `http://data.europa.eu/eli/reg/2017/2195/2021-03-15` (3rd Nov. 2022).

39. ENTSO-E. *PICASSO Network Code.* `https://www.entsoe.eu/network_codes/eb/picasso/` (3rd Nov. 2022).

40. ENTSO-E. *MARI Network Code.* `https://www.entsoe.eu/network_codes/eb/mari/` (3rd Nov. 2022).

41. Carrasco, J. *et al.* Power-Electronic Systems for the Grid Integration of Renewable Energy Sources: A Survey. *IEEE Transactions on Industrial Electronics* **53,** 1002–1016 (2006).

42. Milano, F., Dörfler, F., Hug, G., Hill, D. J. & Verbič, G. *Foundations and Challenges of Low-Inertia Systems (Invited Paper)* in *2018 Power Systems Computation Conference (PSCC)* (2018), 1–25.

43. Nationalgrid ESO. *Operability Strategy Report 2019.* `https://www.nationalgrideso.com/document/159726/download` (3rd Nov. 2022).

44. Brummitt, C. D., Hines, P. D., Dobson, I., Moore, C. & D'Souza, R. M. Transdisciplinary electric power grid science. *Proceedings of the National Academy of Sciences* **110,** 12159–12159 (2013).

45. Clemson, P. T. & Stefanovska, A. Discerning Non-Autonomous Dynamics. *Physics Reports. Discerning Non-Autonomous Dynamics* **542,** 297–368 (2014).

46. ENTSO-E. *Rate of Change of Frequency (RoCoF) Withstand Capability.* `https://docstore.entsoe.eu/Documents/Network%20codes%20documents/NC%20RfG/IGD_RoCoF_withstand_capability_final.pdf` (3rd Nov. 2022).

47. National Grid ESO. *Data Portal.* `https://data.nationalgrideso.com/` (3rd Nov. 2022).

48. Gorjão, L. R., Schäfer, B. & Hassan, G. *Open Access Power-Grid Frequency Database.* `https://doi.org/10.17605/OSF.IO/M43TG`. 2020.

49. Hirth, L., Mühlenpfordt, J. & Bulkeley, M. The ENTSO-E Transparency Platform – A Review of Europe's Most Ambitious Electricity Data Platform. *Applied Energy* **225,** 1054–1067 (2018).

50. Schäfer, B., Beck, C., Aihara, K., Witthaut, D. & Timme, M. Non-Gaussian Power Grid Frequency Fluctuations Characterized by Lévy-Stable Laws and Superstatistics. *Nature Energy* **3,** 119–126 (2018).

51. Rydin Gorjão, L. *et al.* Open Database Analysis of Scaling and Spatio-Temporal Properties of Power Grid Frequencies. *Nature Communications* **11,** 6362 (2020).

52. Deng, X., Li, H., Yu, W., Weikang, W. & Liu, Y. *Frequency Observations and Statistic Analysis of Worldwide Main Power Grids Using FNET/GridEye* in *2019 IEEE Power & Energy Society General Meeting (PESGM)* (2019), 1–5.

53. Haehne, H., Schottler, J., Waechter, M., Peinke, J. & Kamps, O. The Footprint of Atmospheric Turbulence in Power Grid Frequency Measurements. *EPL (Europhysics Letters)* **121,** 30001 (2018).

54. Anvari, M. *et al.* Stochastic properties of the frequency dynamics in real and synthetic power grids. *Physical review research* **2,** 013339 (2020).

55. Bang, W. & Yoon, J. W. *Forecasting the Electric Network Frequency Signals on Power Grid* in *2019 International Conference on Information and Communication Technology Convergence (ICTC)* (2019), 1218–1223.

56. Meyer, P. G., Anvari, M. & Kantz, H. Identifying Characteristic Time Scales in Power Grid Frequency Fluctuations with DFA. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **30,** 013130 (2020).

57. Rydin Gorjão, L., Vanfretti, L., Witthaut, D., Beck, C. & Schäfer, B. Phase and Amplitude Synchronization in Power-Grid Frequency Fluctuations in the Nordic Grid. *IEEE Access* **10,** 18065–18073 (2022).

58. Gorjão, L. R., Schäfer, B., Witthaut, D. & Beck, C. Spatio-Temporal Complexity of Power-Grid Frequency Fluctuations. *New Journal of Physics* **23,** 073016 (2021).

59. Gorbunov, A., Peng, J. C.-H., Bialek, J. W. & Vorobev, P. Can Center-of-Inertia Model Be Identified From Ambient Frequency Measurements? *IEEE Transactions on Power Systems* **37,** 2459–2462 (2022).

60. Vorobev, P. *et al.* Deadbands, Droop, and Inertia Impact on Power System Frequency Distribution. *IEEE Transactions on Power Systems* **34,** 3098–3108 (2019).

61. Mele, F. M., Ortega, A., Zarate-Minano, R. & Milano, F. *Impact of Variability, Uncertainty and Frequency Regulation on Power System Frequency Distribution* in *2016 Power Systems Computation Conference (PSCC)* (IEEE, Genoa, Italy, 2016), 1–8.

62. Kraljic, D. Towards Realistic Statistical Models of the Grid Frequency. *IEEE Transactions on Power Systems (in press)* (2022).

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

64. Chen, T. & Guestrin, C. *XGBoost: A Scalable Tree Boosting System* in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, 2016), 785–794.

65. Ke, G. *et al.* Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* **30** (2017).

66. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).

67. Barredo Arrieta, A. *et al.* Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion* **58,** 82–115 (2020).

68. Chatzivasileiadis, S., Venzke, A., Stiasny, J. & Misyris, G. Machine Learning in Power Systems: Is It Time to Trust It? *IEEE Power and Energy Magazine* **20,** 32–41 (2022).

69. Roscher, R., Bohn, B., Duarte, M. F. & Garcke, J. Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access* **8,** 42200–42216 (2020).

70. Titz, M., Kaiser, F., Kruse, J. & Witthaut, D. *Predicting Dynamic Stability from Static Features in Power Grid Models Using Machine Learning* . Preprint at `https://doi.org/10.48550/arXiv.2210.09266`. 2022.

71. Pütz, S., Schäfer, B., Witthaut, D. & Kruse, J. *Revealing Interactions between HVDC Cross-Area Flows and Frequency Stability with Explainable AI* . Preprint at `https://doi.org/10.48550/arXiv.2204.10727`. 2022.

72. Lundberg, S. M. *et al.* From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature Machine Intelligence* **2,** 56–67 (2020).

73. ENTSO-E. *ENTSO-E Transparency Platform.* `https://transparency.entsoe.eu/` (3rd Nov. 2022).

74. Dong, J., Ma, X., Djouadi, S. M., Li, H. & Liu, Y. Frequency Prediction of Power Systems in FNET Based on State-Space Approach and Uncertain Basis Functions. *IEEE Transactions on Power Systems* **29,** 2602–2612 (2014).

75. Dong, J., Ma, X., Djouadi, S. M., Li, H. & Kuruganti, T. *Real-Time Prediction of Power System Frequency in FNET: A State Space Approach* in *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)* (2013), 109–114.

76. Fleer, J. & Stenzel, P. Impact Analysis of Different Operation Strategies for Battery Energy Storage Systems Providing Primary Control Reserve. *Journal of Energy Storage* **8,** 320–338 (2016).

77. Merten, M., Olk, C., Schoeneberger, I. & Sauer, D. U. Bidding Strategy for Battery Storage Systems in the Secondary Control Reserve Market. *Applied Energy* **268,** 114951 (2020).

78. Khalid, M. & Savkin, A. V. *Model Predictive Control Based Efficient Operation of Battery Energy Storage System for Primary Frequency Control* in *2010 11th International Conference on Control Automation Robotics Vision* (2010), 2248–2252.

79. Avramiotis-Falireas, I., Troupakis, A., Abbaspourtorbati, F. & Zima, M. *An MPC Strategy for Automatic Generation Control with Consideration of Deterministic Power Imbalances* in *2013 IREP Symposium Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid* (2013), 1–8.

80. Xu, B., Oudalov, A., Poland, J., Ulbig, A. & Andersson, G. BESS Control Strategies for Participating in Grid Frequency Regulation. *IFAC Proceedings Volumes. 19th IFAC World Congress* **47,** 4024–4029 (2014).

81. Bolzoni, A., Todd, R., Forsyth, A. & Perini, R. *Real-Time Auto-Regressive Modelling of Electric Power Network Frequency* in *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society* **1** (2019), 515–520.

82. Guo, F. & Schlipf, D. A Spectral Model of Grid Frequency for Assessing the Impact of Inertia Response on Wind Turbine Dynamics. *Energies* **14,** 2492 (2021).

83. Ersdal, A. M., Imsland, L. & Uhlen, K. Model Predictive Load-Frequency Control. *IEEE Transactions on Power Systems* **31,** 777–785 (2016).

84. Hewing, L., Wabersich, K. P., Menner, M. & Zeilinger, M. N. Learning-Based Model Predictive Control: Toward Safe Learning in Control. *Annual Review of Control, Robotics, and Autonomous Systems* **3,** 269–296 (2020).

85. 50Hertz Transmission GmbH, Amprion GmbH, TenneT TSO GmbH & TransnetBW GmbH. *Regelleistung.net: Internetplattform zur Vergabe von Regelleistung.* `https://www.regelleistung.net/ext/` (3rd Nov. 2022).

86. Salem, T. S., Kathuria, K., Ramampiaro, H. & Langseth, H. Forecasting Intra-Hour Imbalances in Electric Power Systems. *Proceedings of the AAAI Conference on Artificial Intelligence* **33,** 9595–9600 (2019).

87. De Vos, K. *et al.* Dynamic Dimensioning Approach for Operating Reserves: Proof of Concept in Belgium. *Energy Policy* **124,** 272–285 (2019).

88. Jost, D., Braun, A. & Fritz, R. *Dynamic Dimensioning of Frequency Restoration Reserve Capacity Based on Quantile Regression* in *2015 12th International Conference on the European Energy Market (EEM)* (2015), 1–5.

89. Persson, M. & Chen, P. Frequency Evaluation of the Nordic Power System Using PMU Measurements. *Transmission Distribution IET Generation* **11,** 2879–2887 (2017).

90. Adeen, M., Jónsdóttir, G. M. & Milano, F. *Statistical Correlation between Wind Penetration and Grid Frequency Variations in the Irish Network* in *2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)* (2019), 1–6.

91. Homan, S., Mac Dowell, N. & Brown, S. Grid Frequency Volatility in Future Low Inertia Scenarios: Challenges and Mitigation Options. *Applied Energy* **290,** 116723 (2021).

92. Weißbach, T., Remppis, S. & Lens, H. *Impact of Current Market Developments in Europe on Deterministic Grid Frequency Deviations and Frequency Restauration Reserve Demand* in *2018 15th International Conference on the European Energy Market (EEM)* (2018), 1–6.

93. Ma, S. & Tourani, R. *Predictive and Causal Implications of Using Shapley Value for Model Interpretation* in *Proceedings of the 2020 KDD Workshop on Causal Discovery* (PMLR, 2020), 23–38.

94. Janzing, D., Minorics, L. & Bloebaum, P. *Feature Relevance Quantification in Explainable AI: A Causal Problem* in *International Conference on Artificial Intelligence and Statistics* (PMLR, 2020), 2907–2916.

95. Svenska kraftnät, Statnett, Fingrid & Energinet.dk. *Challenges and Opportunities for the Nordic Power System.* https://www.fingrid.fi/globalassets/dokum entit/fi/yhtio/tki-toiminta/report-challenges-and-opportunities -for-the-nordic-power-system.pdf (3rd Nov. 2022).

96. Heskes, T., Sijben, E., Bucur, I. G. & Claassen, T. *Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models* in *Advances in Neural Information Processing Systems* **33** (Curran Associates, Inc., 2020), 4778–4789.

97. Frye, C., Rowat, C. & Feige, I. *Asymmetric Shapley Values: Incorporating Causal Knowledge into Model-Agnostic Explainability* in *Advances in Neural Information Processing Systems* **33** (Curran Associates, Inc., 2020), 1229–1239.

98. Wang, J., Wiens, J. & Lundberg, S. *Shapley Flow: A Graph-based Approach to Interpreting Model Predictions* in *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics* (PMLR, 2021), 721–729.

99. Schölkopf, B. *et al.* Toward causal representation learning. *Proceedings of the IEEE* **109,** 612–634 (2021).

100. Sharma, A. & Kiciman, E. *DoWhy: An End-to-End Library for Causal Inference* . Preprint at `https://doi.org/10.48550/arXiv.2011.04216`. 2020.

101. Battocchi, K. *et al. EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation.* `https://github.com/microsoft/EconML` (3rd Nov. 2022).

102. Kalainathan, D. & Goudet, O. *Causal discovery toolbox: Uncover causal relationships in python* . Preprint at `https://doi.org/10.48550/arXiv.1903.02278`. 2019.

103. ENTSO-E. *Inertia and Rate of Change of Frequency (RoCoF).* `https://eepu blicdownloads.azureedge.net/clean-documents/SOC%20documents/ Inertia%20and%20RoCoF_v17_clean.pdf` (3rd Nov. 2022).

104. Martınez-Barbeito, M., Gomila, D. & Colet, P. *Dynamical Model for Power Grid Frequency Fluctuations: Application to Islands with High Penetration of Wind Generation* . Preprint at `https://ifisc.uib-csic.es/en/publications/ dynamical-model-for-power-grid-frequency-fluctuati/`. 2022.

105. ENTSO-E. *Report on Deterministic Frequency Deviations.* `https://consult ations.entsoe.eu/system-development/deterministic_frequency_ deviations_report/user_uploads/report_deterministic_frequency_ deviations_final-draft-for-consultation.pdf` (3rd Nov. 2022).

106. Stiasny, J., Chevalier, S., Nellikkath, R., Sævarsson, B. & Chatzivasileiadis, S. *Closing the Loop: A Framework for Trustworthy Machine Learning in Power Systems* . Preprint at `https://doi.org/10.48550/arXiv.2203.07505`. 2022.

107. Wang, J. *et al.* On Machine Learning-Based Techniques for Future Sustainable and Resilient Energy Systems. *IEEE Transactions on Sustainable Energy,* 1–15 (2022).

108. Misyris, G. S., Venzke, A. & Chatzivasileiadis, S. *Physics-Informed Neural Networks for Power Systems* in *2020 IEEE Power & Energy Society General Meeting (PESGM)* (2020), 1–5.

109. Coutinhoa, F. A. B., Burattinia, M. N., Lopeza, L. F. & Massada, E. Threshold Conditions for a Non-Autonomous Epidemic System Describing the Population Dynamics of Dengue. *Bulletin of Mathematical Biology* **68,** 2263–2282 (2006).

110. Pei, L. & Zhang, M. Long-Term Predictions of Current Confirmed and Dead Cases of COVID-19 in China by the Non-Autonomous Delayed Epidemic Models. *Cognitive Neurodynamics* **16,** 229–238 (2022).

111. Grimm, V., Heinlein, A., Klawonn, A., Lanser, M. & Weber, J. Estimating the Time-Dependent Contact Rate of SIR and SEIR Models in Mathematical Epidemiology Using Physics-Informed Neural Networks. *ETNA - Electronic Transactions on Numerical Analysis* **56,** 1–27 (2021).

112. Schultz, M. G. *et al.* Can Deep Learning Beat Numerical Weather Prediction? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **379,** 20200097 (2021).

113. Thakkar, A. & Chaudhari, K. Fusion in Stock Market Prediction: A Decade Survey on the Necessity, Recent Developments, and Potential Future Directions. *An International Journal on Information Fusion* **65,** 95–107 (2021).

A. Appendix

A.1. Erklärung zur Dissertation

gemäß der Promotionsordnung vom 12. März 2020

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht.
Ich versichere an Eides statt,

- dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat;

- dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie,

- dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde.

Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.
Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Teilpublikationen

A) Kruse, J., Schäfer, B. & Witthaut, D. Predictability of Power Grid Frequency. *IEEE Access* **8,** 149435–149446. doi:10.1109/ACCESS.2020.3016477 (2020). (Section 2.1)

B) Kruse, J., Schäfer, B. & Witthaut, D. Revealing drivers and risks for power grid frequency stability with explainable AI. *Patterns* **2,** 100365. doi:`10.1016/j.patter.2021.100365` (2021). (Section 3.1)

C) Kruse, J., Schäfer, B. & Witthaut, D. *Exploring deterministic frequency deviations with explainable AI* in *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)* (2021), 133–139. doi:`10.1109/SmartGridComm51999.2021.9632335`. (Section 3.2)

D) Kruse, J., Schäfer, B. & Witthaut, D. Secondary control activation analysed and predicted with explainable AI. *Electric Power Systems Research* **212,** 108489. doi:`10.1016/j.epsr.2022.108489` (2022). (Section 3.3)

E) Kruse, J., Cramer, E., Schäfer, B. & Witthaut, D. *Physics-inspired machine learning for power grid frequency modelling.* Preprint at `https://doi.org/10.48550/arXiv.2211.01481`. 2022. (Section 4.1)

---

4. November 2022, Johannes Kruse