Predictability of Power Grid Frequency

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

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 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

---

The associate editor coordinating the review of this manuscript and approving it for publication was Jahangir Hossain[iD].
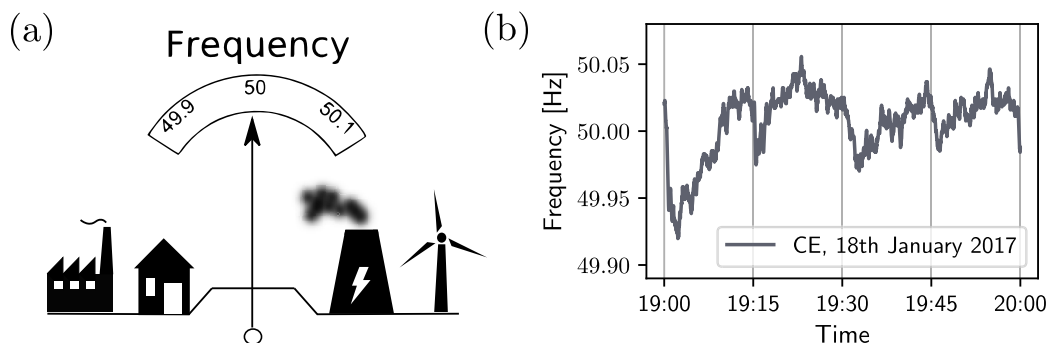
FIGURE 1. The nature of the power grid frequency. (a): The frequency reflects the balance of power demand and generation. Over-production causes a frequency increase and under-production a frequency decline. (b): Example frequency time series from the CE synchronous area [21]. It displays the typical 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].
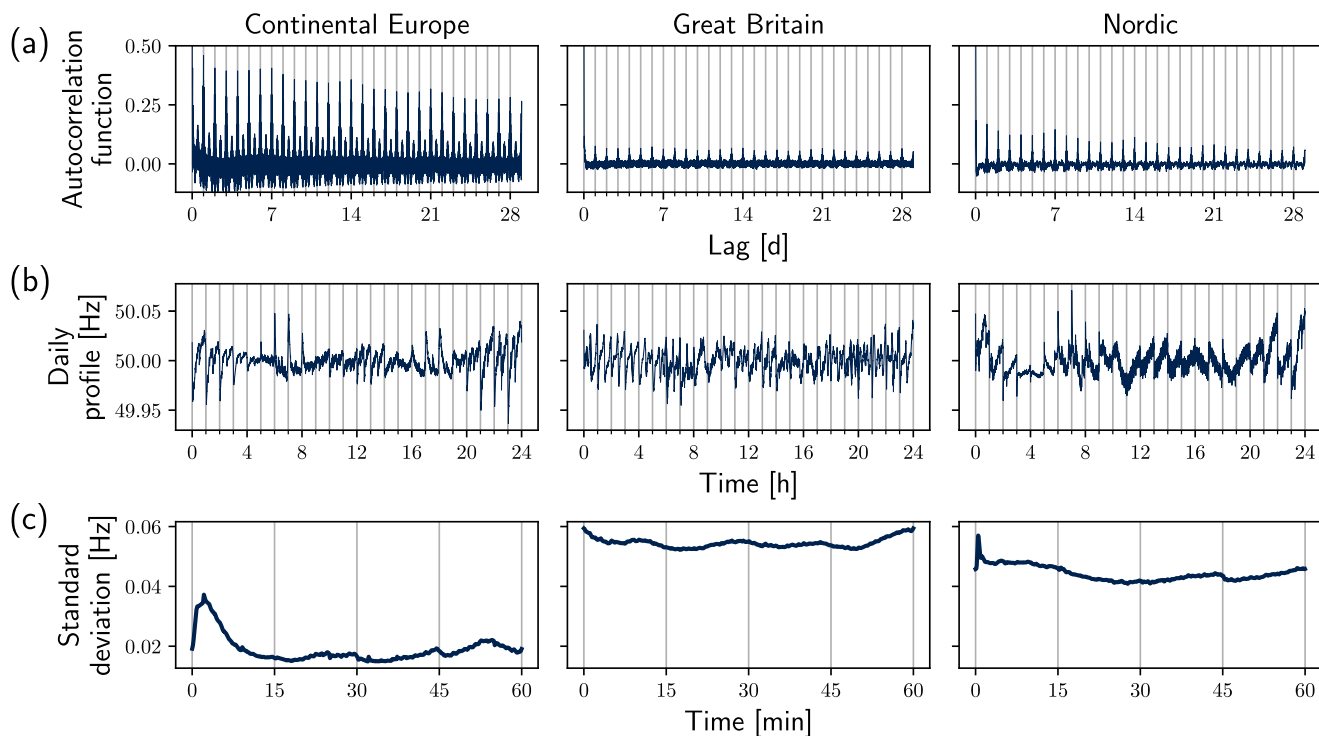
FIGURE 2. The daily profile is an important null model. (a): The autocorrelation functions show significant peaks that repeatedly occur with a period of 24 h. The one-day period thus is the main recurrence period for frequency patterns. Note that the upper limit of the y-axis has been reduced from 1 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:

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

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

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

with $\| \cdot \|$ denoting the Euclidean distance. Given this metric, we sort the patterns as $d(\mathbf{F}_{n_1}) \leq d(\mathbf{F}_{n_2}) \leq \ldots \leq d(\mathbf{F}_{n_M})$, $M = |\mathcal{F}|$ 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 $d(\mathbf{F}_{n_j})$ which introduces an additional smoothing [32, Chap. 3]. Following [18], we use a linear weighting that has the following form:

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

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
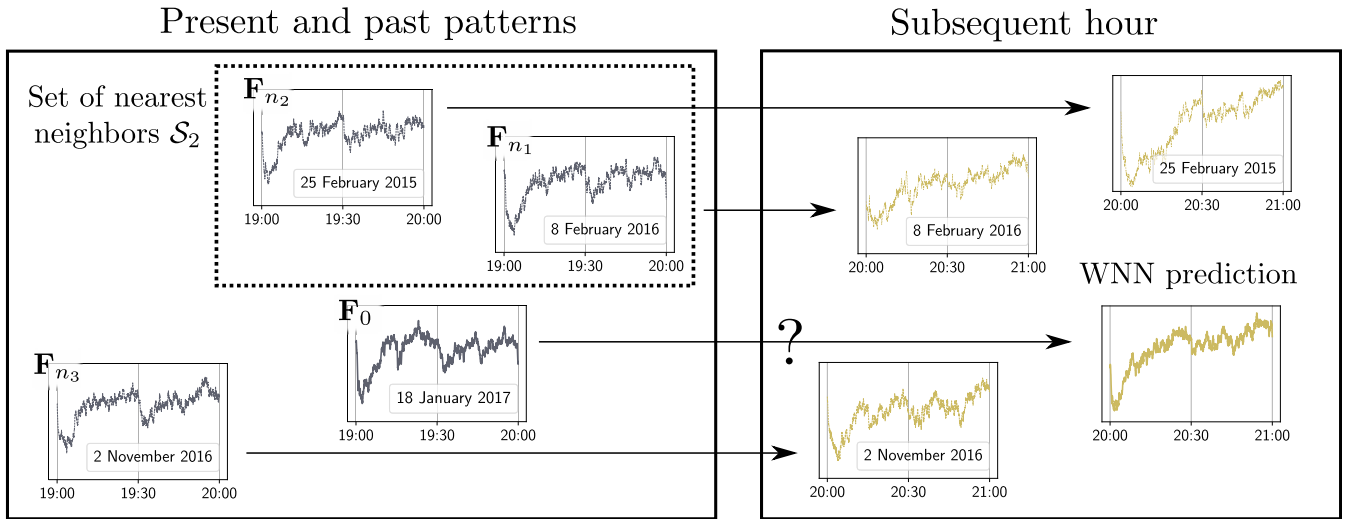
FIGURE 3. The WNN predictor searches for similar patterns in the past. To predict the future of the present (initial) pattern $\mathbf{F_0}$, the WNN method looks for similar patterns $\mathbf{F}_{n_j}$ in the past. The patterns that are most similar to the initial pattern form the set of nearest neighbours. Here, we have chosen a set $\mathcal{S}_2$ 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 $\mathrm{MSE}(f_p)$ of the WNN predictor f_p:

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

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:

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

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:

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

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.
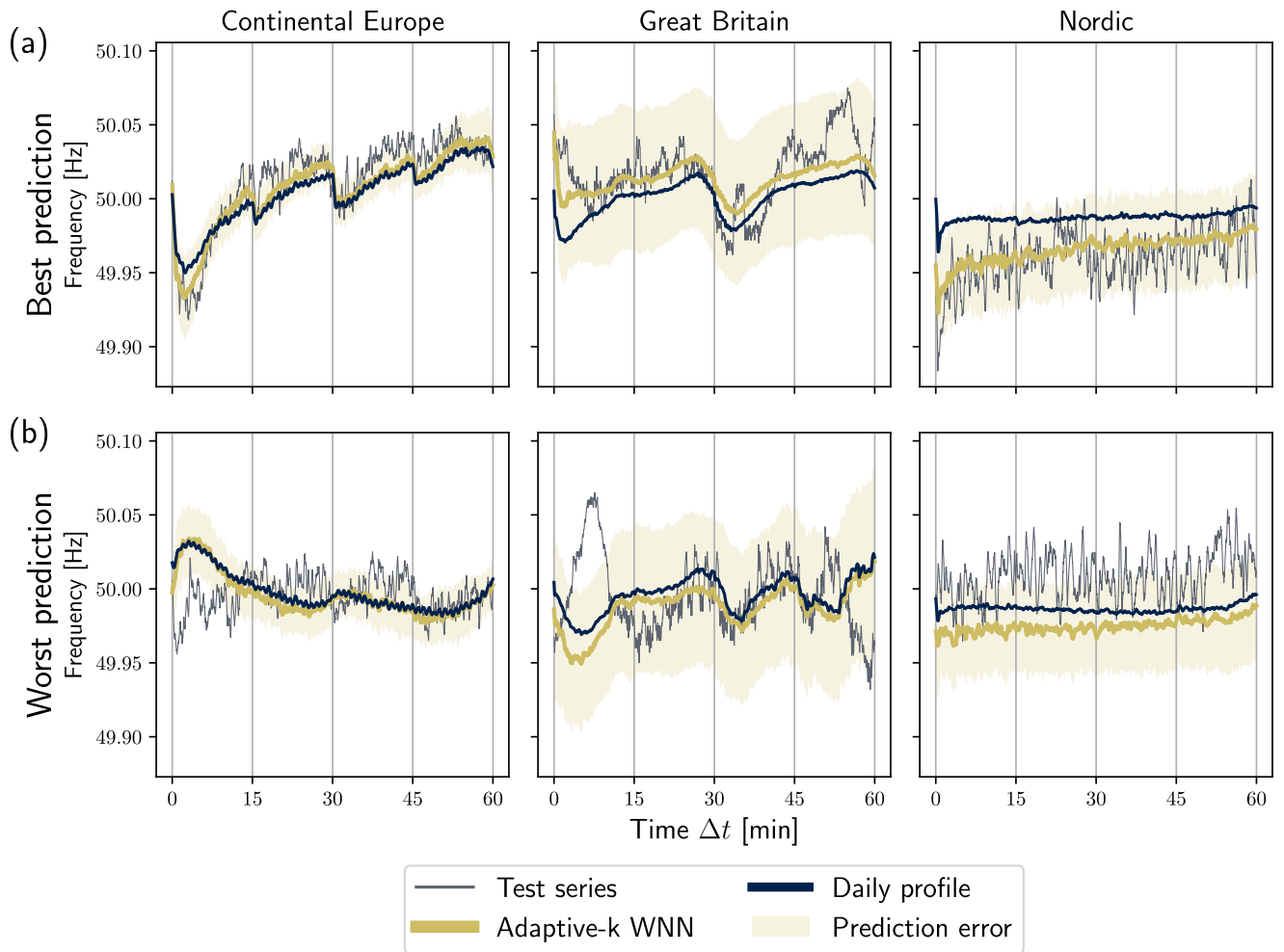
FIGURE 4. The best predictions are a smoothed version of the observed frequency trajectory. Here, we present the best (a) and worst (b) adaptive-*k* predictions from the test set. The selection is based on the relative error RMSE(f_p)/RMSE(50Hz). With that we account for the difference in variance among 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 $k \to \infty$ 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
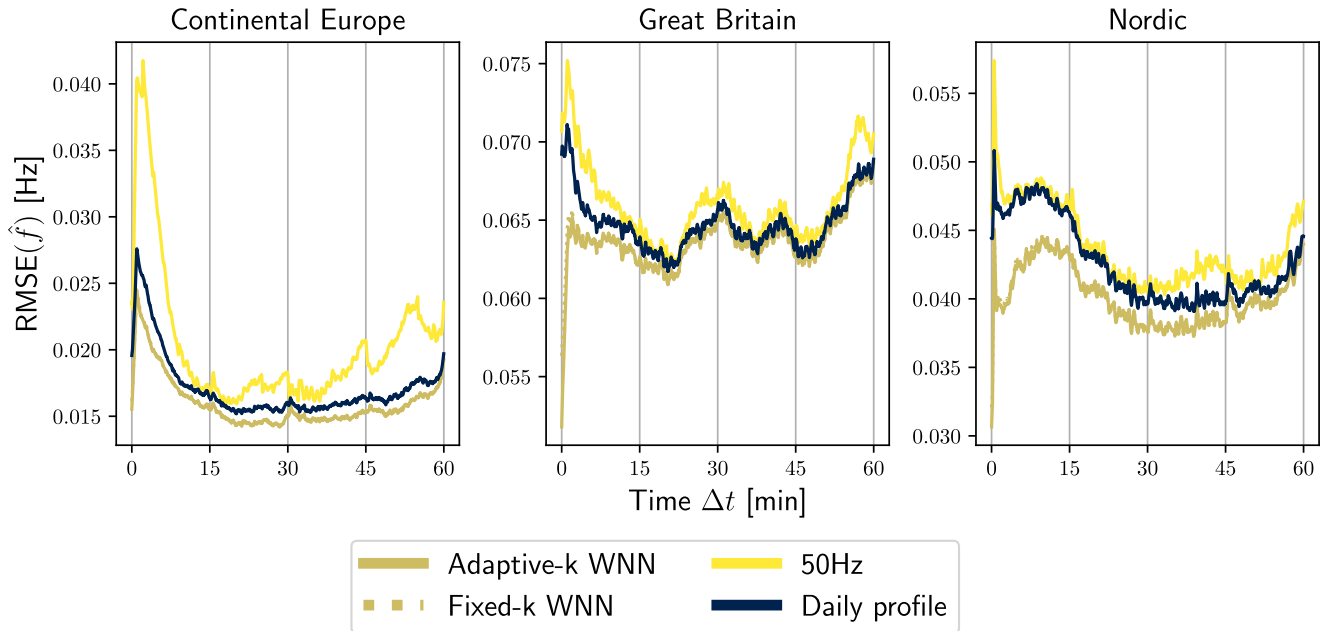
FIGURE 5. The WNN predictors outperform alternatives. The WNN predictor outperforms the null models in all three synchronous areas by returning the smallest RMSE, especially in the first 15-30 minutes. The scale of the y-axis differs between the subplots, since the GB area exhibits much larger errors 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.

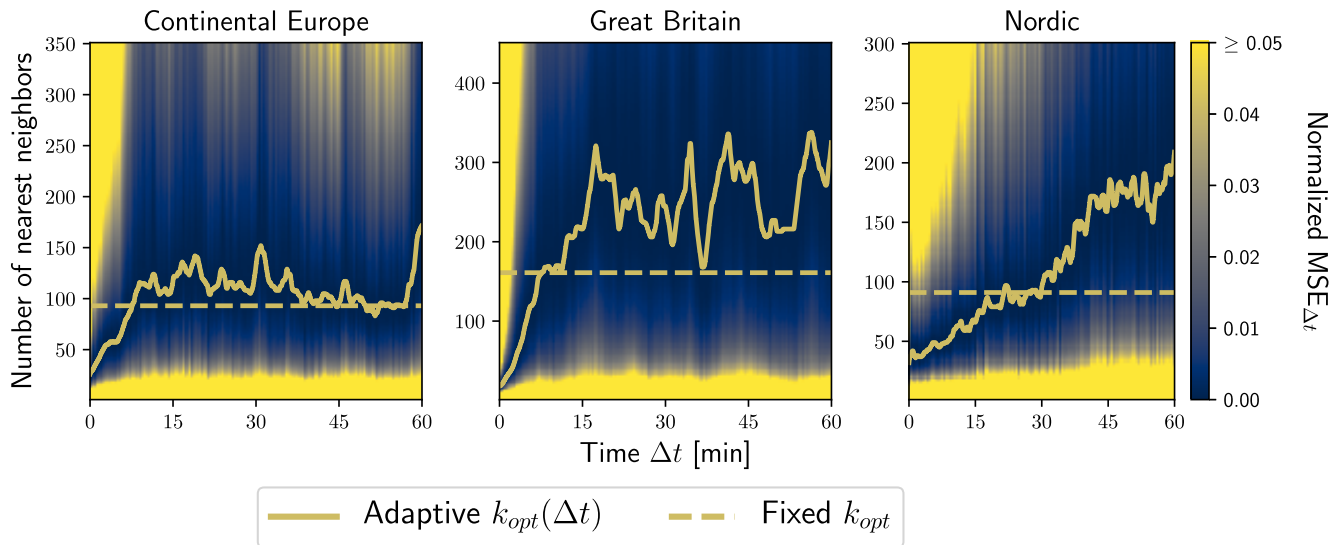### C. OPTIMAL NUMBER OF NEAREST NEIGHBOURS
Determining the optimal number of nearest neighbours $k_{opt}(\Delta t)$ can help to better understand the functioning of the WNN predictor. Moreover, it yields valuable information about the grid frequency dynamics in general. We present the optimization results 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
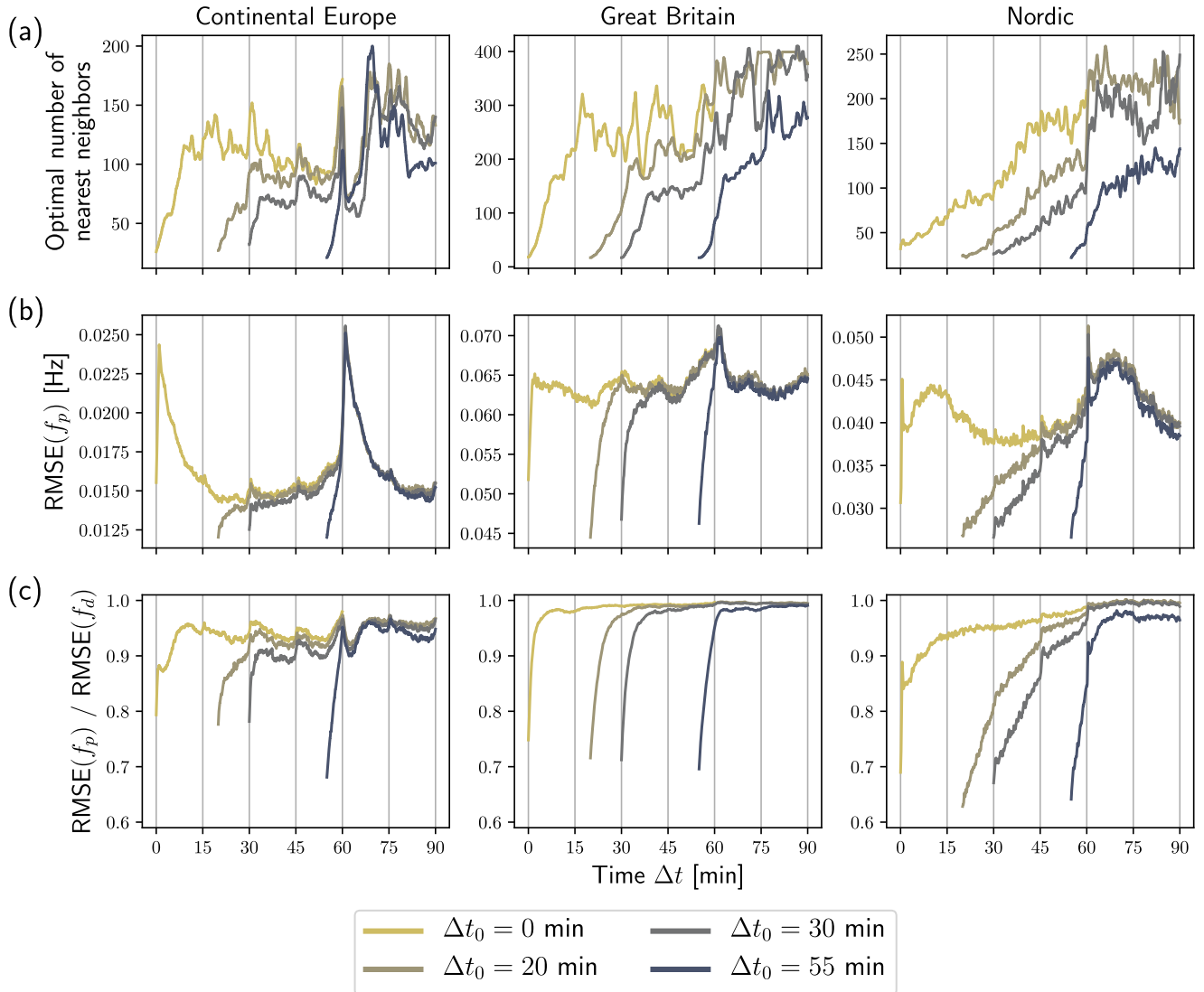
FIGURE 7. Trading events shorten the prediction horizon. Here, we show the optimal number of nearest neighbours $k_{opt}(\Delta t)$ (a), the RMSE (b) and the relative RMSE (c), which is normalized by the daily profile error. Irrespective of the starting time within an hour Δt_0, the predictions perform 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.
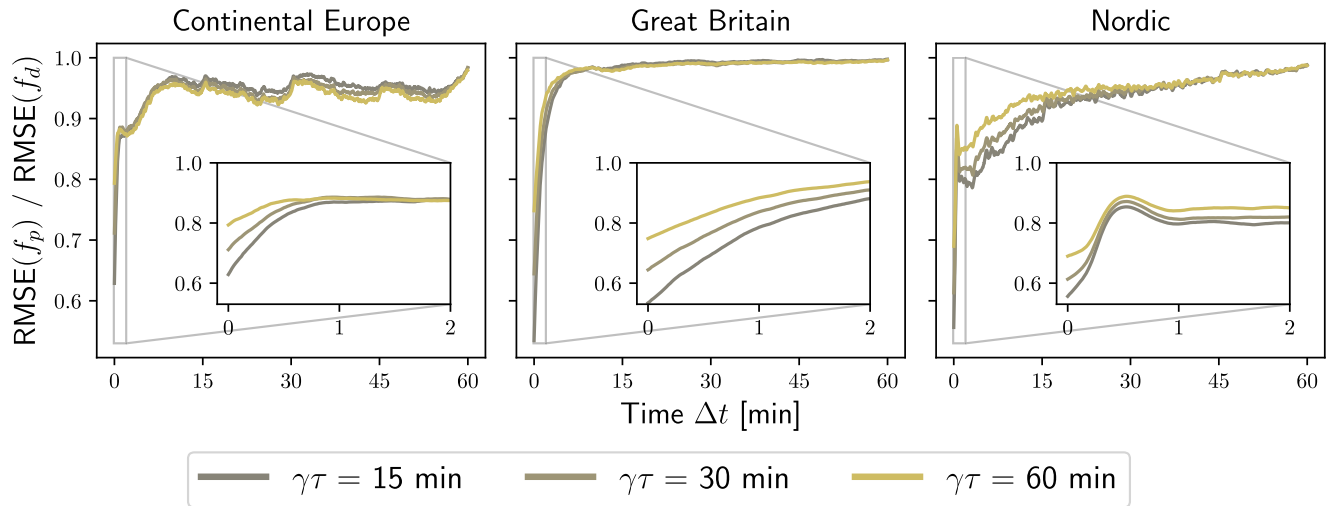
FIGURE 8. Shorter windows predict more accurately at the beginning. The optimal window size (or pattern length) $\gamma\tau$ is different depending on the prediction length. During the first minute, the shortest window performs best in all grid areas, as it contains more specific information about the 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
The authors thank Mark Thiele for fruitful discussions.

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: Interdiscipl. 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.

● ● ●