

PAPER • OPEN ACCESS

Predicting the power grid frequency of European islands

To cite this article: Thorbjørn Lund Onsaker *et al* 2023 *J. Phys. Complex.* **4** 015012

View the [article online](#) for updates and enhancements.

You may also like

- [Infrastructures for the development on Bawean Island, Gresik, East Java-Indonesia](#)
Y Milla Ardiani and Noegroho
- [Study on the Solid Waste Properties and Disposal Methods of Island at Home and Abroad](#)
Ruicheng Dong, Xiao Chen, Zhibin Ding et al.
- [The Potential of Renewable Energy Generations at Barrang Caddi Island](#)
I Kitta, S Manjang, I Rachmaniar et al.



PAPER

OPEN ACCESS

RECEIVED
27 September 2022

REVISED
6 February 2023

ACCEPTED FOR PUBLICATION
20 February 2023

PUBLISHED
6 March 2023

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Predicting the power grid frequency of European islands

Thorbjørn Lund Onsaker¹, Heidi S Nygård¹ , Damiá Gomila² , Pere Colet² , Ralf Mikut³ ,
Richard Jumar³ , Heiko Maass³ , Uwe Kühnapfel³ , Veit Hagenmeyer³ and Benjamin Schäfer^{3,*}

¹ Faculty of Science and Technology, Norwegian University of Life Sciences, 1432 Ås, Norway

² Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), Campus Universitat Illes Balears, E-07122 Palma de Mallorca, Spain

³ Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany

* Author to whom any correspondence should be addressed.

E-mail: benjamin.schaefer@kit.edu

Keywords: Machine learning, forecasting, power grid, time series analysis, Europe, frequency synchronisation

Abstract

Modelling, forecasting and overall understanding of the dynamics of the power grid and its frequency are essential for the safe operation of existing and future power grids. Much previous research was focused on large continental areas, while small systems, such as islands are less well-studied. These natural island systems are ideal testing environments for microgrid proposals and artificially islanded grid operation. In the present paper, we utilise measurements of the power grid frequency obtained in European islands: the Faroe Islands, Ireland, the Balearic Islands and Iceland and investigate how their frequency can be predicted, compared to the Nordic power system, acting as a reference. The Balearic Islands are found to be particularly deterministic and easy to predict in contrast to hard-to-predict Iceland. Furthermore, we show that typically 2–4 weeks of data are needed to improve prediction performance beyond simple benchmarks.

1. Introduction

An increasingly interdisciplinary community is contributing to the understanding of power grid dynamics and stability, including dynamical modelling, simulations and machine-learning-based forecasts, see e.g. [1] for a recent review. The energy system is a very important and challenging research area as it constitutes a complex system with many interacting entities, e.g. its energy markets, generators, consumers, different control areas and control mechanisms.

In this article, we investigate the dynamics and predictability of the power grid frequency in five grids of very different sizes. The grid frequency is critical to ensure a safe and reliable supply of electricity due to its central role in power system control [2]. The electric frequency of the alternating current power grids mirrors the current balance in supply and demand: An excess of demand effectively draws from the kinetic energy stored in rotating turbines and slows these down. Vice versa, an excess of generation cannot be stored directly in the grid and thereby leads to an increased frequency (unless batteries, pumped hydropower or other storage options are set in operation to restore the power balance). Critically, the frequency has to be controlled tightly using numerous control mechanisms [2–4], as otherwise consumers or generators have to be disconnected. Desired frequency ranges depend on the synchronous area, i.e. the connected power grid sharing the same frequency, and range from $f = 50 \pm 0.2$ Hz [5] to $f = 50 \pm 0.05$ Hz [6] for the areas considered in the present paper. With an increasing share of volatile renewable generation, in particular wind and solar power, the total inertia of the power system decreases, while fluctuations overall increase [7, 8]. This further amplifies the need for accurate forecasts and a solid understanding of power systems.

Islands and their power grids are particularly interesting for two reasons: First, there exist many natural islands which are unconnected to any other power grid or will only be connected via high-voltage-directed-current (HVDC) links, still ensuring decoupled frequency dynamics [9]. Small island grids typically have

lower inertia and display larger frequency deviations than large integrated networks [10]. Hence they may provide valuable insights for variable grid operation. Second, microgrid proposals [11–13], which suggest establishing a local cell of the power grid, are receiving much attention. Depending on the specific scenario, cells might be mostly autonomous, have their own market [14] or only decouple during emergencies to stabilise the grid and avoid cascading failures [15, 16].

Fully modelling the short-term frequency dynamics is very challenging due to its combined stochastic-deterministic nature and the numerous external influences [10, 17, 18], suggesting the use of data-driven approaches [19] and the usage of machine learning to predict time series as well as to understand and control energy systems [20–23]. These data-driven approaches complement modelling-based approaches [10, 24] as they do not make any assumptions about the governing dynamical equations but still provide forecasts and explanations of the system.

Any data-based forecasting method depends on the availability of data with a sufficient quality covering a sufficiently long period. Unfortunately, such data are still not easily available. Even research projects, such as the GridEye/FNET initiative [25], are not always open, which limits their value for the research community. In recent work, we have pushed for more open data of power grid frequency measurements [26–28], including coverage of power grids on islands. Still, with only limited measurement devices available, some questions arise: Are previously developed methods which forecast the power grid frequency [20] applicable to islands? How much data are necessary to develop data-driven approaches that can outperform simple benchmarks? How can additional information, e.g. from transparency platforms [29], be integrated into frequency forecasts?

To answer these questions, the remaining paper is structured as follows. We start with an overview of the data obtained from the Nordic system, the Faroe Islands, Ireland, the Balearic Islands and Iceland in section 2. Next, we provide an introduction to the nearest-neighbour predictor used for forecasting in section 3.1. Then, we investigate the performance of the predictor, compared to benchmarks and as a function of available data in section 3.2 to then discuss the possibility of extending the predictor by including additional data in section 3.3. We close with a discussion in section 4.

2. Data overview

In the present article, we analyse the power grid frequency time series from four different European islands in addition to the Nordic synchronous area. These islands are the Faroe Islands, Ireland, the Balearic Islands and Iceland, see also figure 1, showing good geographical spread and a difference in size. We consider time series from 6 days (Faroe Islands) up to 450 days (Balearic Islands) for our analysis. The five regions (Nordic and the islands) are also very different in terms of their energy mix, regulations and interconnections:

The Nordic power system includes Norway, Sweden and Finland as well as the Eastern part of Denmark, a total of approximately 24 million people. The energy mix for the Nordic electricity supply consists of a significant share of renewable energy, in particular wind and hydropower, with a steady increase in recent years [30].

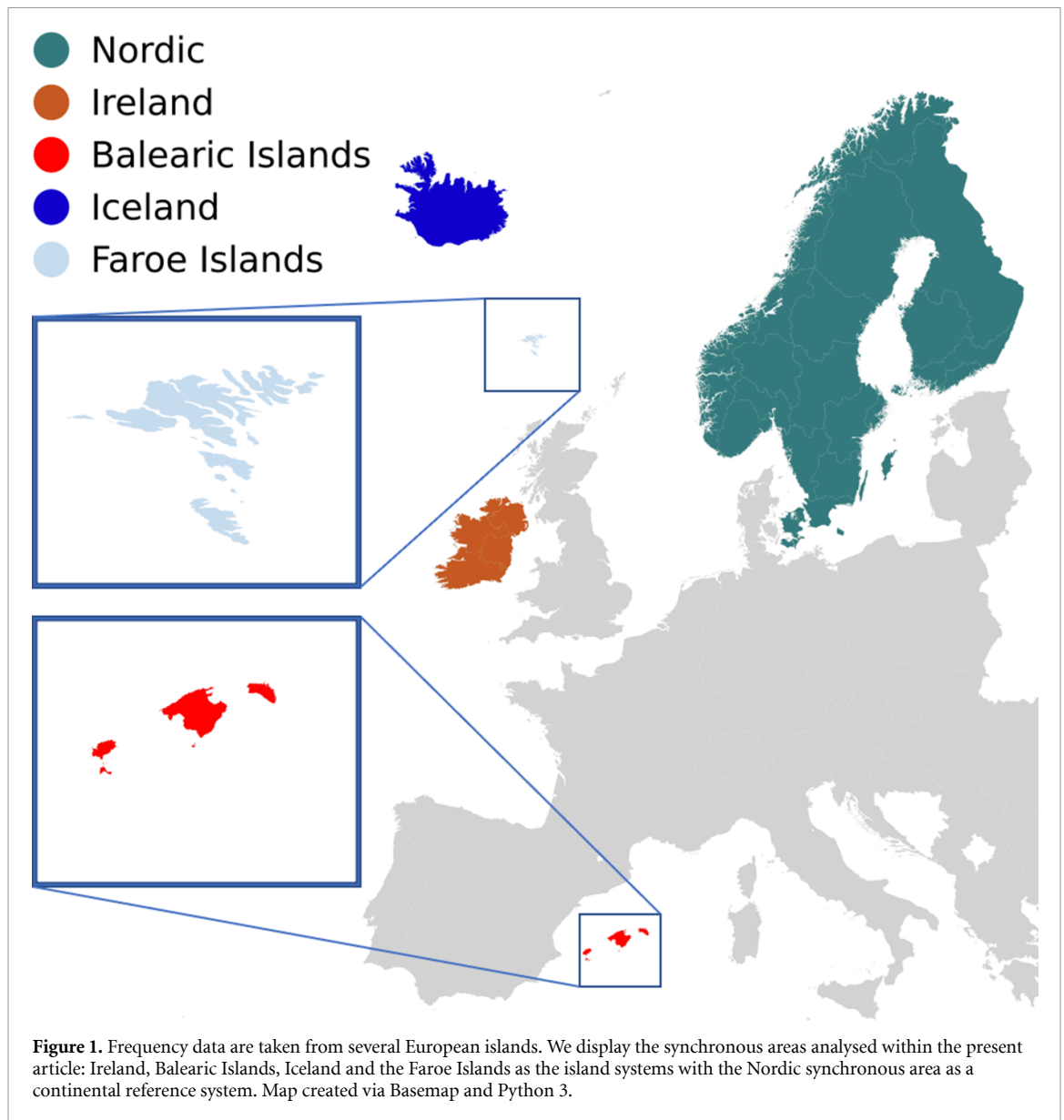
The Faroe Islands are an archipelago of 18 major islands in the Atlantic Ocean, with approximately 50 000 inhabitants. The power generation relies mostly on fossil fuels [31].

The Irish power system covers both the Republic of Ireland and Northern Ireland, with about 6.8 million inhabitants in total. Ireland's generation is tailed towards a high share of wind energy and in 2020 renewable electricity generation reached 43% of annual generation [32]. Ireland is connected to Great Britain via HVDC cables [33].

The Balearic Islands, belonging to Spain, are located in the Mediterranean Sea with a total population of about 1.2 million. Generation relies mostly on fossil fuels such as gas (combined cycle gas turbines) and diesel engines leading to a share in generation of more than 70% [34]. Furthermore, an HVDC cable connecting the islands to mainland Spain is an important factor in the electric system.

Finally, Iceland is located in the North Atlantic Ocean and has roughly 300 000 inhabitants. Its main source of electricity generation is hydropower, making up 75% of the total generation. In addition, Iceland's volcanic activities allow extensive geothermal power generation. These cheap energy sources have attracted energy-intensive industries such as ferrosilicon, aluminium smelters and data centres [35].

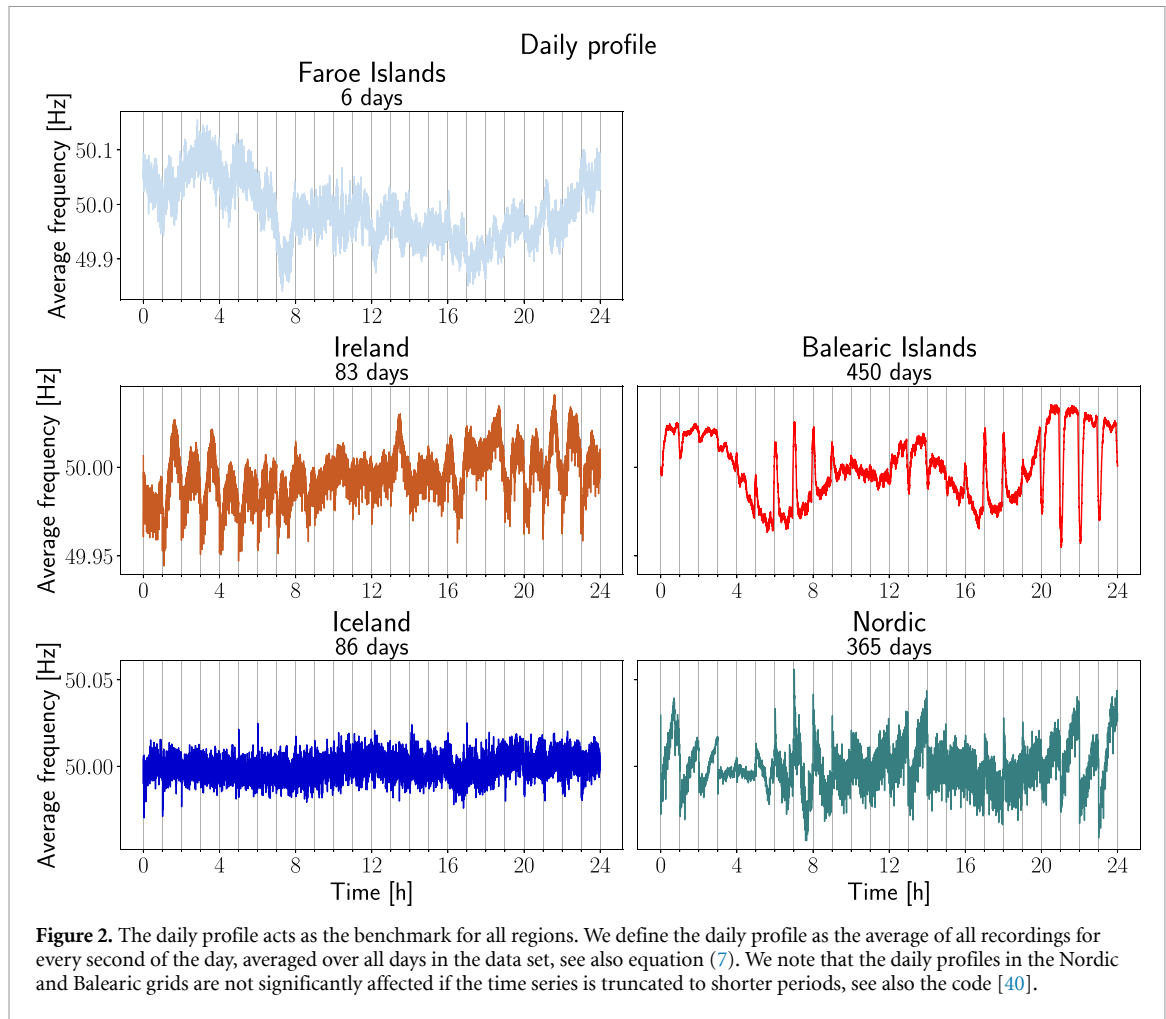
The data for the Nordic grid are based on measurements provided by the Finish operator Fingrid [36], see also [26–28]. Meanwhile, the time series of the islands are obtained via our measurement device, the electrical data recorder (EDR), which has been developed at KIT. The EDR records the voltage with a sampling frequency of 25 kHz and assigns a standard time to each data point based on a GPS recorder. Then, the frequency is extracted based on zero-crossings of the voltage signal, see [27, 37–39] for details. Finally, the



frequency data are pre-processed to identify acquisition interruptions and to remove outliers. Unfortunately, some single devices had some longer measurement outages, leading to up to 25% of missing data related to the installed time period in the Balearic Islands and Ireland, respectively, but typically less than 1% for the other locations, see also code for details [40].

To obtain an initial impression of the frequency dynamics in the different regions, we compute the daily profiles in figure 2, noting stark differences. The daily profile is the average frequency for each second of a day, averaged over all days in the data set, defined in detail in the next section. The daily profile in the Faroe Islands displays the largest deviations and seems to be somewhat unpredictable. This is in part due to the short time series available (6 days) and might further be linked to the small population size. Meanwhile, Ireland, the Balearic Islands and the Nordic grid all display clear patterns with the frequency jumping up and down at specific hours. These jumps are associated with typical power dispatch actions and load and generation ramps at full hours [41, 42]. Finally, the Icelandic daily profile is mostly flat and does not display clear patterns, likely because the dominant hydropower and geothermal power plants are easily controllable on the generation side.

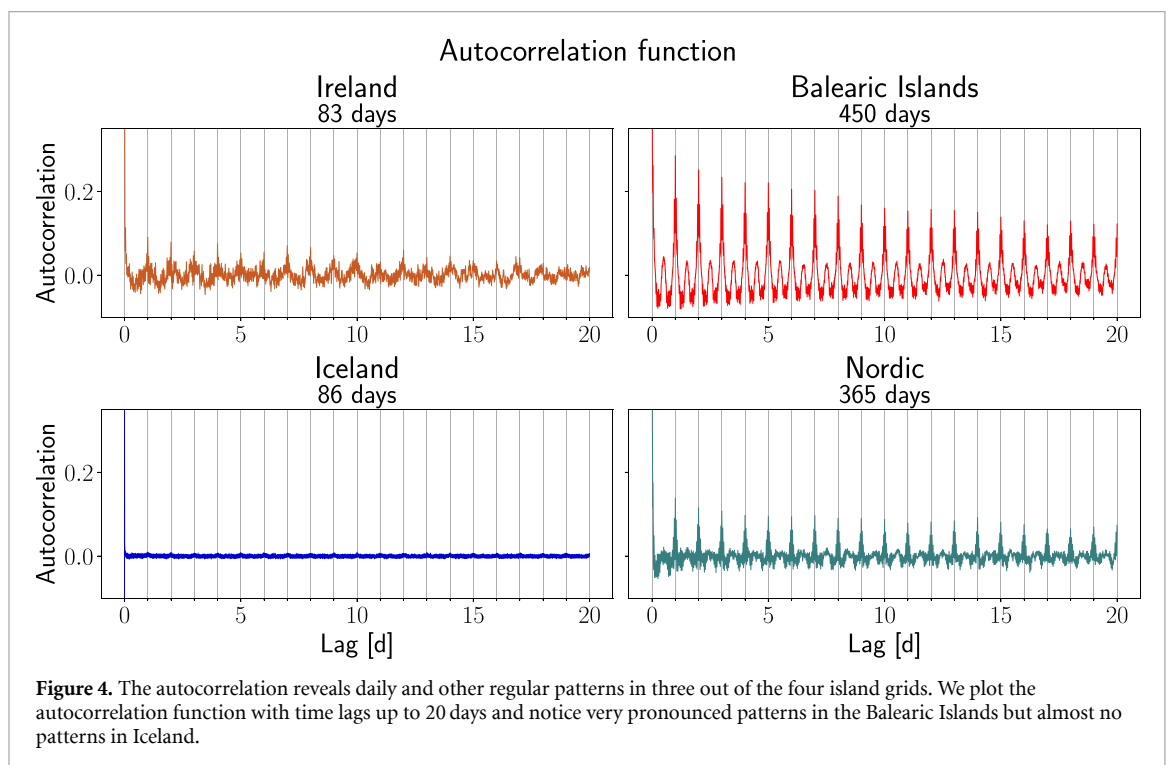
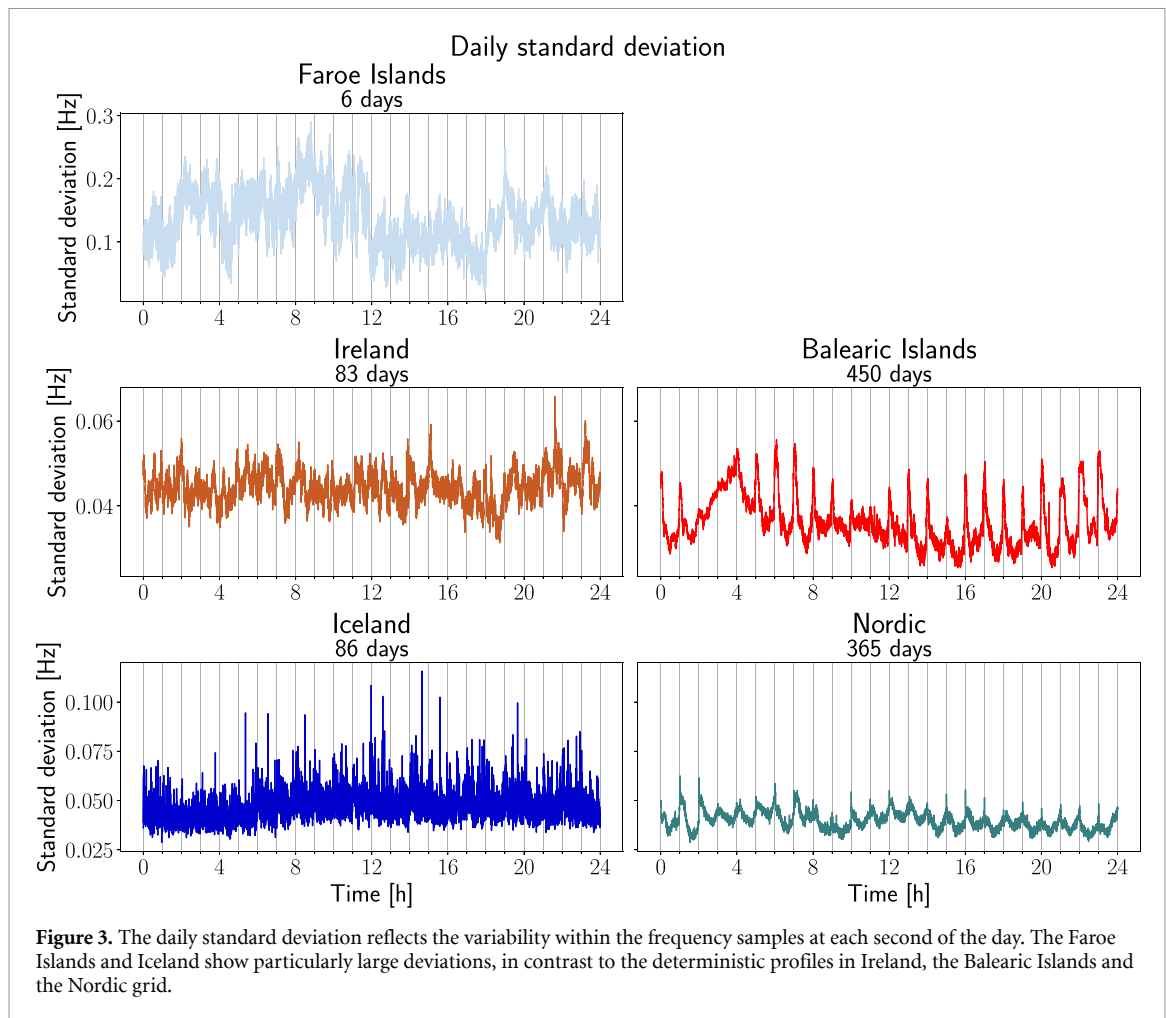
While the daily profile (mean daily frequency values) provides insights into what typical frequency values to expect, the daily standard deviation reveals the spread of frequency values over the day, see figure 3. Technically, instead of computing the mean value, we compute the standard deviation for each second of a day. Analysing the standard deviation confirms a previous observation from the daily profile: The Balearic

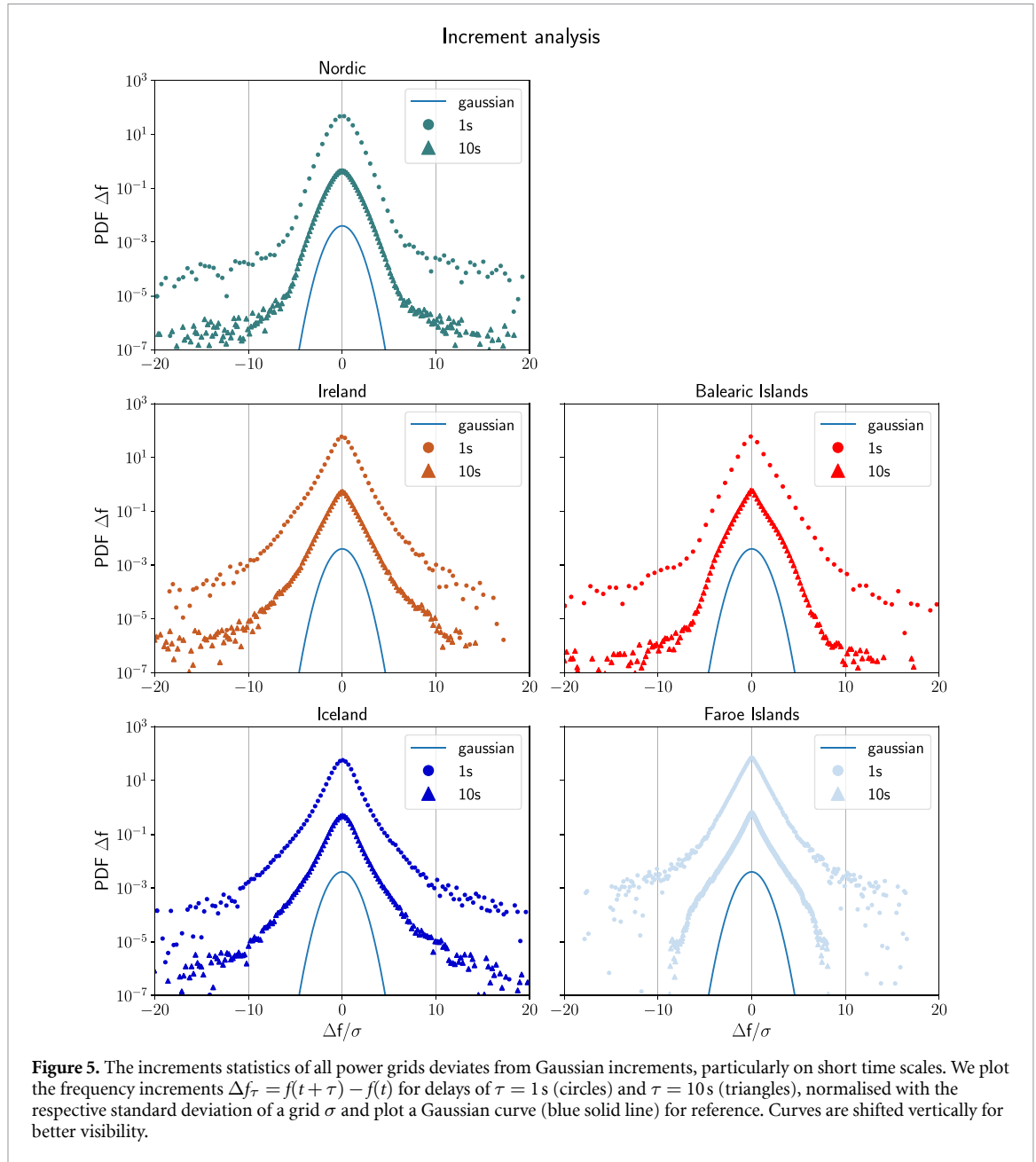


Islands, the Nordic and to a slightly lesser extent the Irish systems show clean curves with regular distinct peaks at the hourly dispatch intervals. This is consistent with earlier observations: Larger deterministic frequency excursions at the start of an hour are complemented with smaller but random fluctuations in between [10, 17, 42]. Meanwhile, the Faroe Islands show little structure and the largest absolute standard deviation, consistent with the large frequency deviations observed earlier. Iceland is special in that the standard deviation is on average higher than in Ireland, the Balearic Islands or the Nordic grid and occasionally, very large values of the standard deviations are observed due to extreme frequency deviation of the order of Hertz, instead of deviations of tens or maybe one hundred mHz as would be observed in the other synchronous areas.

Complementing the daily profile and standard deviation, we analyse how much history is preserved in the trajectories by computing the autocorrelation decay [17]. For a generic stochastic time series, we might expect an exponential decay of the autocorrelation [43] and we do indeed observe such a (quick) decay in all time series, see figure 4. For all grids except Iceland, we also observe regular peaks in the autocorrelation on a daily basis, reinforcing the idea that the frequency follows daily patterns. Iceland is special in that the frequency displays almost no correlation and seems to be much more stochastic than the frequency in the other regions. Note that the Faroe Islands are omitted due to their short time series.

Finally, we also investigate the increment distribution, see figure 5. The increments are given as $\Delta f_\tau = f(t + \tau) - f(t)$. If the system dynamics followed a simple stochastic process, e.g. an Ornstein–Uhlenbeck [10, 43], the increments should resemble a Gaussian distribution on all time scales τ . In practice, we observe that for smaller τ values, the deviations from Gaussianity are more pronounced, while for larger values of τ , the increments become more Gaussian [26, 44]. We observe the same effect in figure 5: All synchronous areas display deviations from Gaussianity on short time scales ($\tau = 1$ s, circles), while the statistics approach a more Gaussian shape for longer time scales ($\tau = 10$ s, triangles). Ireland and even more so Iceland still display strong non-Gaussian distributions for $\tau = 10$ s.





3. Forecasting frequency on islands

3.1. Introducing the weighted nearest neighbour (WNN) predictor

To forecast the power grid frequency, we expand upon an earlier paper [20], which investigated numerous aspects of frequency predictability with a WNN predictor. This earlier paper [20] focused on the three largest synchronous areas in Europe, for which several years of data are available. In contrast, measurements from islands, investigated in the present article, cover a few weeks or months up to 450 days in the case of the Balearic islands. The WNN algorithm identifies similar patterns in the past to forecast future dynamics, see figure 6 for an illustration.

Let us specify the details of the prediction process: For each synchronous area, the time series is split into chunks of equal length which are then used to find similar chunks (neighbours) from the past. More formally, we define the pattern vectors:

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

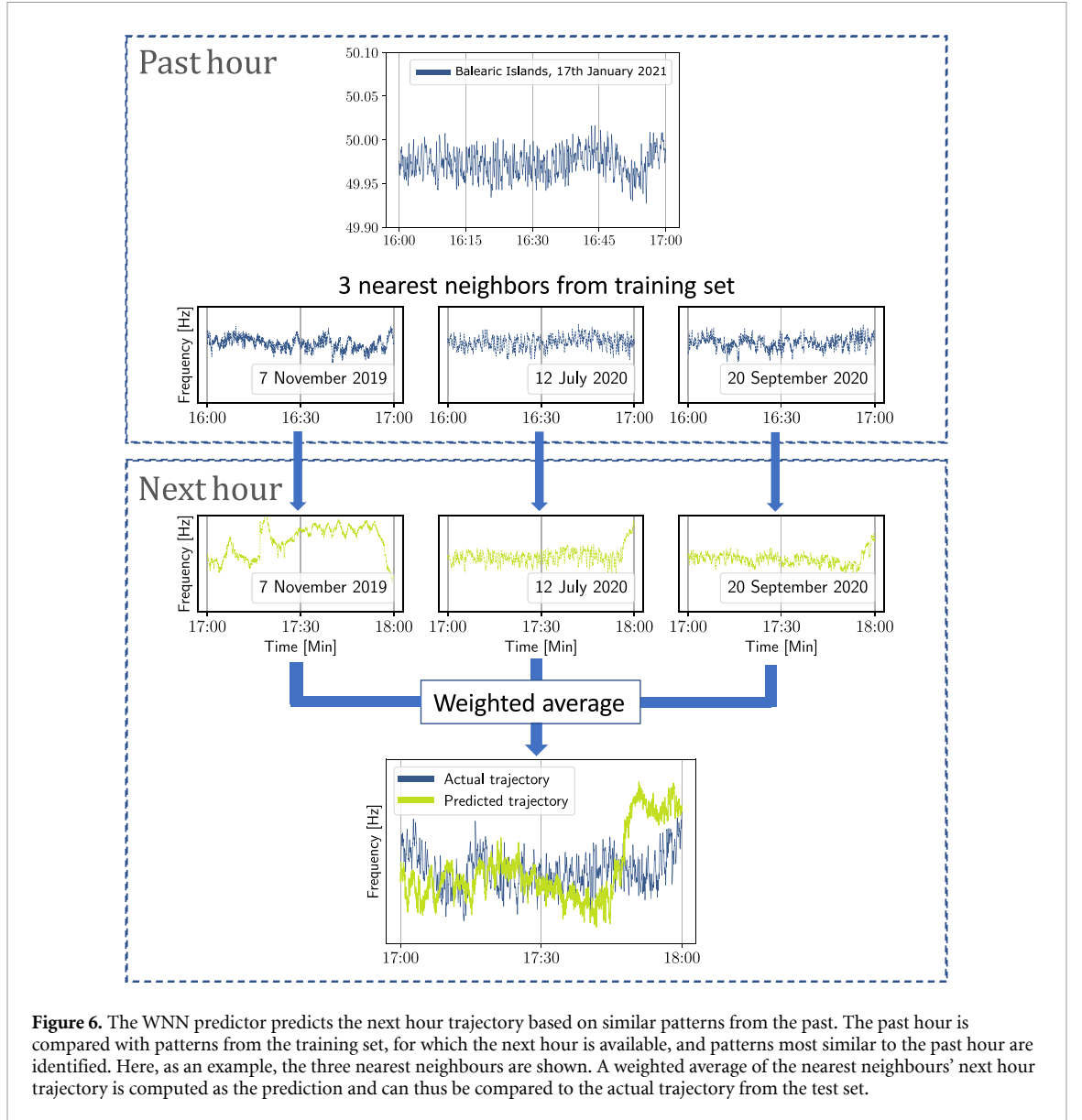


Figure 6. The WNN predictor predicts the next hour trajectory based on similar patterns from the past. The past hour is compared with patterns from the training set, for which the next hour is available, and patterns most similar to the past hour are identified. Here, as an example, the three nearest neighbours are shown. A weighted average of the nearest neighbours' next hour trajectory is computed as the prediction and can thus be compared to the actual trajectory from the test set.

where t_0 is the last time point for which we have data in our initial pattern \mathbf{F}_0 . Furthermore, the time delay τ is identical with the time resolution $\tau = 1\text{s}$ and we choose $\gamma = 3600$ data points so that $\gamma\tau = 3600\text{s}$ covers one hour, a central time scale for market dynamics and regular grid patterns [42]. Hence, each pattern \mathbf{F}_n is a vector with 3600 frequency measurements and the entire time series is covered by combining non-overlapping sets as:

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

and the index n is used to distinguish between different patterns. Given the past hour pattern as \mathbf{F}_0 , the next hour of the frequency time series is predicted by searching the set \mathcal{F} for the most similar patterns to this initial pattern. We implement this search by computing the Euclidean distance between the initial pattern and all patterns in \mathcal{F} :

$$d(\mathbf{F}_n) = \|\mathbf{F}_n - \mathbf{F}_0\|. \quad (3)$$

The patterns in \mathcal{F} are then sorted based on these distances in increasing order, thus resulting in a new set of patterns. This set only contains the k most similar patterns to \mathbf{F}_0 given as

$$\mathcal{N}_k = \{n_1, n_2, \dots, n_k | \mathbf{F}_{n_j} \in \mathcal{F}\}. \quad (4)$$

The final prediction $f_{\text{wnn}}(t_0 + \Delta t)$ with time steps $\Delta t \in \{1 \text{ s}, 2 \text{ s}, 3 \text{ s}, \dots, 3600 \text{ s}\}$ is performed by a weighted average of the k nearest neighbours' next hour trajectories:

$$f_{\text{wnn}}(t_0 + \Delta t) = \frac{1}{\sum_{j=1}^k \alpha_j} \sum_{j=1}^k \alpha_j f(t_0 - n_j \gamma \tau + \Delta t). \quad (5)$$

Note that by the construction of \mathcal{F} , we restrict our search for frequency patterns to the same hour, e.g. we predict the frequency at 9 pm only using training data from 9 pm as well, see also [20, 40]. The weighting α decreases with greater distance between the patterns, since close neighbours should have a stronger influence on the prediction than more distant ones:

$$\alpha_j = \frac{d(\mathbf{F}_{n_k}) - d(\mathbf{F}_{n_j})}{d(\mathbf{F}_{n_k}) - d(\mathbf{F}_{n_1})}. \quad (6)$$

The number of nearest neighbours k is a hyperparameter of the WNN predictor, which has to be optimised. Within this paper, we utilise the adaptive- k approach [20], which optimises k for every time step as follows: The optimal k -value is given by the lowest $\text{MSE}_{\Delta t}(f_{\text{wnn}})$ of the predictor for each Δt , i.e. we obtain an adaptive- k vector with 3600 values. The adaptive- k is smoothed with a sliding window of 1 min to avoid sudden jumps and noise for any time steps, see [40] for details.

As a benchmark, we compare the WNN performance with the daily profile, which effectively is an unweighted $k = \infty$ predictor (including all patterns) defined as:

$$f_{\text{dp}}(t_0 + \Delta t) = \frac{1}{|\mathcal{F}|} \sum_{n \in \mathcal{F}} f(t_0 - n \gamma \tau + \Delta t). \quad (7)$$

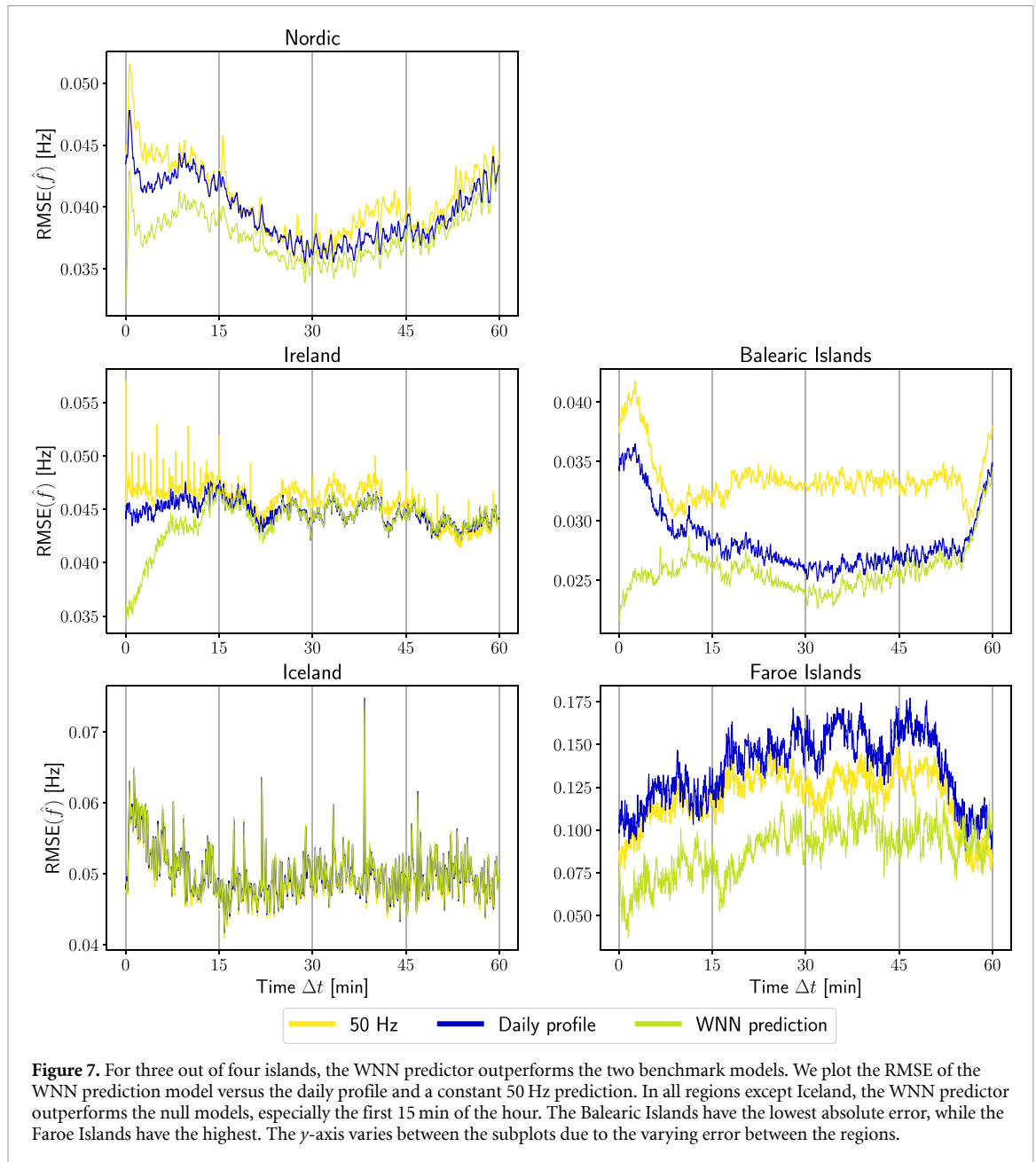
3.2. Performance of the WNN predictor

Let us now investigate whether the WNN predictor can outperform common benchmarks on islands, as it did in continental regions [20]. We consider three prediction models: first: a naive and constant 50 Hz predictor, as this is the typical mean value for each hour, second: the daily profile, which takes all patterns into account equally and third: the WNN, which selects fewer patterns and weights them to obtain a forecast. For all predictors, we use about 70% of the data for training (i.e. data to select patterns from or data to compute the daily profile), 15% as validation (to determine the optimal number of nearest neighbours k) and 15% as a test (to evaluate the performance), see the code [40] for details.

The WNN predictor typically outperforms these benchmarks even when little data are available, see figure 7. We compute the root-mean-squared error (RMSE) for each predictor for each minute, averaged over all hours. Consistent with the daily profiles, we note that both benchmarks display a large error at the start of an hour, as dispatch actions such as generation or load ramps lead to large deviations in the frequency. Especially at the beginning of an hour, the WNN predictor is more precise than the benchmarks, leveraging its specific information from the previous hour. The further away the prediction is evaluated from the start of the hour, the less specific information is contained in the previous patterns and the WNN approaches the daily profile behaviour, as also seen for continental regions [20]. There are two more interesting observations: First, even with only 6 days of training and test data available, the WNN outperforms the daily profile and 50 Hz predictor in the Faroe Islands. Second, all predictors fare approximately equal for Iceland, where the grid frequency time series shows almost no characteristic patterns but large stochastic deviations. Intuitively, power grids should be easier predictable, the more their trajectories follow deterministic patterns, compared to stochastic influences. Let us discuss this in terms of increment statistics: Large and particularly non-Gaussian increments signal many erratic (random) jumps, while smaller increments signal more regular dynamics. Indeed, we observed for Iceland clear non-Gaussian increments in figure 5. Meanwhile, we also observe non-Gaussian increments for Ireland, for which both the daily profile and WNN predictor yield lower forecasting errors. Hence, increment statistics might provide some possible explanations why certain areas are harder to predict than others but further comparisons and metrics will have to be explored in the future.

In many situations, we will not have access to years or months of data, due to a large variety of potential reasons: Measurement time is limited, large parts of the data are corrupted, the specific grid is very new, either because the grid is newly built or major changes were made, such as the expansion of a synchronous area [45] or a change in regulations. Regardless of what limits the access to data, we need to understand how well the WNN predicts the frequency with shorter periods of training data available.

Interestingly, there is no general threshold of how many weeks of data are necessary for the WNN to outperform the daily profile, see figure 8. There, we systematically investigate the prediction quality as a



function of the length of available data by computing the RMSE for the constant 50 Hz prediction, the daily profile and the WNN. The latter two predictors are based on data ranging from one week up to 6 months (where available) and always utilised an 80-20 split for training and validation, hence reporting the validation RMSE. Again, the Faroe Islands are excluded as less than one week of data are available. We typically note that for very little data, the constant prediction is the best, followed by the daily profile and the WNN coming in last. If measurement times are too short, the unspecific, average prediction is superior and the WNN restricts itself to finding patterns in its very limited training set, leading to a bad performance. Vice versa, if sufficient data are available, the WNN correctly identifies characteristic patterns and is capable to make the best predictions.

Notably, this transition from ‘unspecific information being best’ to ‘specific historic patterns being best’ depends on the power grid at hand: The Balearic islands with their very deterministic frequency profile, only require 2 weeks for the WNN to outperform the daily profile, compared to about 4 weeks necessary in Ireland, while in Iceland even with 2 months of data, the constant 50 Hz prediction is still better than daily profile or the WNN predictor.

3.3. Additional features

So far, we have only used historic frequency information to forecast frequency time series. However, there are an increasing amount of energy system data available, so we ask: How can additional information, such as

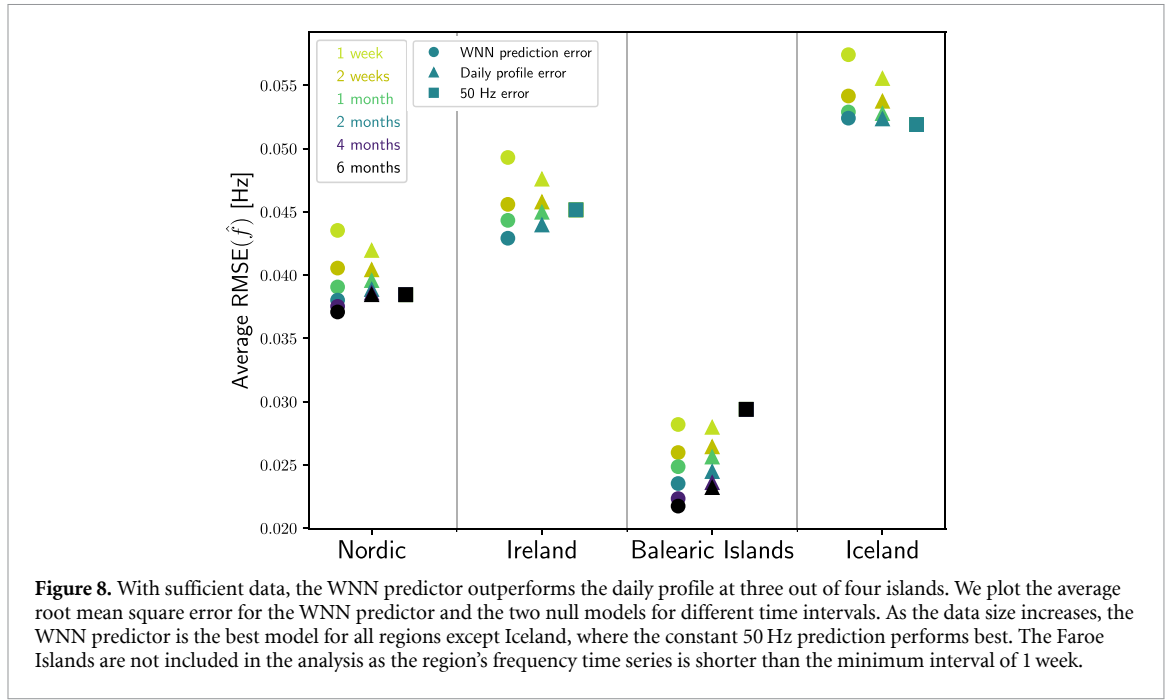


Figure 8. With sufficient data, the WNN predictor outperforms the daily profile at three out of four islands. We plot the average root mean square error for the WNN predictor and the two null models for different time intervals. As the data size increases, the WNN predictor is the best model for all regions except Iceland, where the constant 50 Hz prediction performs best. The Faroe Islands are not included in the analysis as the region's frequency time series is shorter than the minimum interval of 1 week.

the generation of coal power plants or volatile renewable generation (wind and solar power), be integrated into the frequency prediction, and how do they influence the performance of the WNN predictor?

To answer this question, we utilise additional information on the Balearic energy system, obtained from the transmission system operator of Spain, Red Eléctrica de España, and their real-time electricity demand and generation tracker [46]. It seems intuitive that (frequency) patterns of the power system will differ based on the total load or the mix of generation types. Mathematically, we extend the WNN predictor as follows: Let a_{origin} be the original additional time series. To compensate for the large deviations in the typical values of a_{origin} , we carry out a min-max scaling:

$$a = \frac{a_{\text{origin}} - a_{\min}}{a_{\max} - a_{\min}}, \quad (8)$$

where a is the transformed data point and a_{\max} and a_{\min} are the maximum and minimum values of the original time series. Next, the additional feature is split into non-overlapping patterns \mathbf{A}_n , similar to \mathbf{F}_n from equation (1):

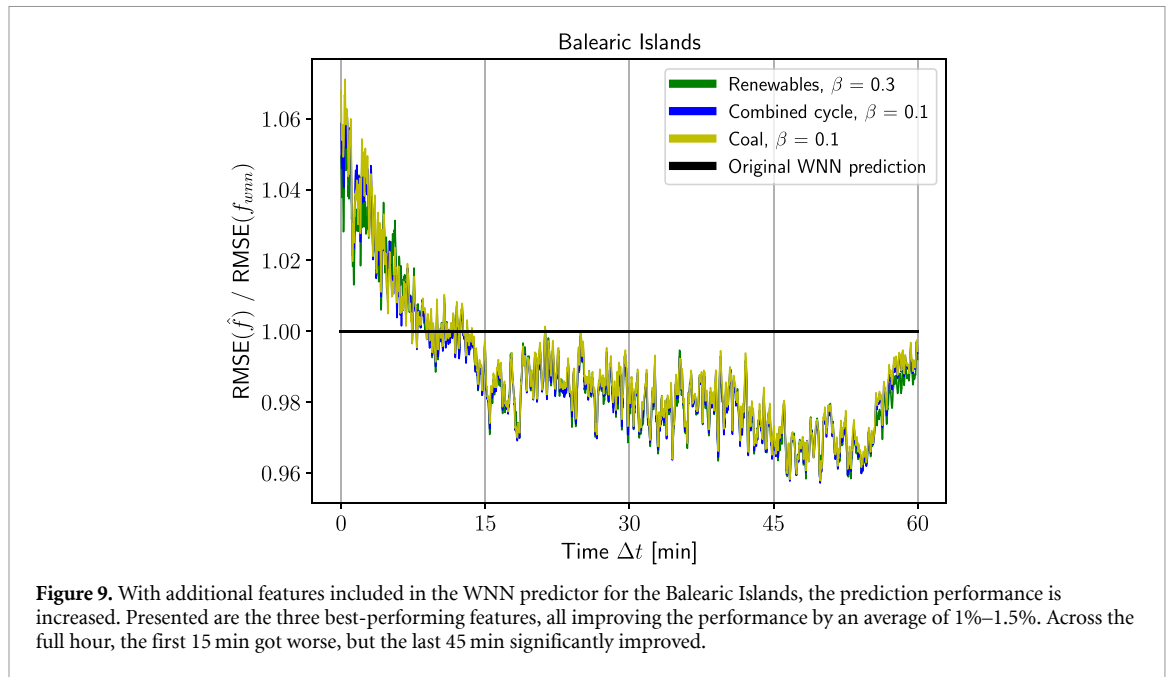
$$\mathbf{A}_n = \begin{pmatrix} a(t_0 - (n+1)\eta\lambda) \\ a(t_0 - (n+1)\eta\lambda + \lambda) \\ a(t_0 - (n+1)\eta\lambda + 2\lambda) \\ \dots \\ a(t_0 - n\eta\lambda - \lambda) \end{pmatrix}, \quad (9)$$

where η is the number of additional data points in each pattern, and λ is the time resolution. $\eta\lambda$ is thus the prediction window size and must be the same size as the frequency patterns to represent the same specific time window of 3600 s, matching \mathbf{F}_n . Hence, $\eta\lambda \stackrel{!}{=} \gamma\tau$. When the vectors \mathbf{A}_n are concatenated with \mathbf{F}_n , the result is new vectors given as:

$$\mathbf{G}_n = \begin{pmatrix} \mathbf{F}_n \\ \mathbf{A}_n \end{pmatrix}. \quad (10)$$

As the extended model is also limited to searching through patterns from the same time of the day, \mathcal{G} is a new set similar to \mathcal{F} from (2). To now choose the nearest neighbours in terms of distance, the previous distance computation equation (3) is extended to:

$$d(\mathbf{G}_n) = \|\mathbf{F}_n - \mathbf{F}_0\| + \beta \cdot \|\mathbf{A}_n - \mathbf{A}_0\|, \quad (11)$$



where $\|\cdot\|$ denotes the Euclidean distance, and β is a tunable weight to adjust the influence of the additional feature. With new distances, the weighting α from (6) is now computed as:

$$\alpha_j = \frac{d(\mathbf{G}_{n_k}) - d(\mathbf{G}_{n_j})}{d(\mathbf{G}_{n_k}) - d(\mathbf{G}_{n_i})}. \quad (12)$$

The weights are then used as in equation (5) to forecast the frequency time series.

Additional information might improve performance slightly for certain time periods, see figure 9. To obtain these results, we optimised the hyperparameter β via a grid search $\beta \in 0.1, 0.2, \dots, 1.5$. Including the information about actual renewable generation accounted for the best performance improvement, closely followed by improvements by including data from combined cycle or coal [47] generation, all of which lead to a reduction in RMSE of more than 1%. Interestingly, the additional features decreased the performance during the first 15 min, while it improved after that. Without additional features, the predicted trajectory has its best performance at the start of an hour and then deteriorates since the information of the previous hour is slowly being invalidated by additional, unforeseen events, see also [20]. The additional feature balances this effect: In the first few minutes, the additional feature dilutes the specific information, leading to a larger RMSE, while in the later parts of the hour, the additional information is more useful. Further research will be necessary to clarify the exact effects of additional features included in the WNN predictor, especially when applying it to other islands, such as Iceland, for which unfortunately no public data on demand and generation were available when writing this article. Moreover, in our current implementation, we use the actual generation for the hour for which we forecast the frequency. Thereby, we mix past and future information. Future research could rely on forecasts or scheduled generation, which are not always publicly available but might be available to system operators wishing to implement the WNN predictor.

4. Discussion and conclusion

The power grid frequency in islands may be predicted via a WNN predictor [20]. Compared to the continental areas considered [20] and the Nordic area analysed in the current article, the four investigated islands displayed more extreme properties: The daily profile in the Balearic islands is the most pronounced and regular behaviour we have observed so far, while the frequency dynamics in Iceland show no characteristic patterns but strong stochastic deviations, making it particularly tricky to predict. We also note some dependency on the population size and the composition of the power generation and consumption: the Faroe Islands and Iceland are the hardest to predict. These two islands are the least populated areas and have very special generation and demand patterns, influenced e.g. by large industry in Iceland. Meanwhile, the very good predictability of the Balearic Islands might at least be partially explained by its strong interconnection with the Continental European Synchronous Area and the predominance of combined cycle production with its fast control capabilities [48]. Overall, we have shown that the WNN predictor

outperforms the daily profile even when only limited data are available in three out of the four island grids. Furthermore, we have demonstrated how additional time series, such as generation, could be included as additional information when forecasting the frequency.

Concluding, we make two contributions in this article: First, we demonstrate with a simple nearest-neighbour predictor that the power grid frequency is predictable using machine learning in islands if sufficient data are available. Second, we provide a tool that could be useful for the planning and operation of islanded grids, let these be natural islands or islanded microgrids [12]. Transmission system operators (TSOs) could use these forecasts to estimate the stability conditions for the upcoming hour and both TSOs and balancing power provider could use forecasts to estimate control needs for future time intervals: Large frequency deviations will require additional control to be activated to stabilise the power system. Notably, almost no prior knowledge about the system at hand is necessary nor do we require many assumptions to apply the WNN. Even the difference between weekends and weekdays is implicitly included in the WNN: Given sufficient data, it will select suitable patterns. Furthermore, the WNN could be used as a diagnostic tool, e.g. to analyse systematic offsets in the frequency or external perturbations and thereby detect anomalies.

Further research is necessary to fully characterise to which extent power grids are predictable and which external factors should be included. Also, further ML methods, beyond nearest-neighbour techniques could be considered [49], e.g. to handle any concept drift in long-term forecasting or to provide further monitoring capabilities.










Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files) [28].

Acknowledgment

This project has received funding from the Helmholtz Association under Grant No. VH-NG-1727, the Program ‘Energy System Design’ and the Helmholtz Association’s Initiative and Networking Fund through Helmholtz AI. DG and PC acknowledge financial support from Maria de Maeztu program CEX2021-001164-M of MCIN/AEI /10.13039/501100011033/.

ORCID iDs

Heidi S Nygård  <https://orcid.org/0000-0002-1639-7634>
Damiá Gomila  <https://orcid.org/0000-0002-3500-3434>
Pere Colet  <https://orcid.org/0000-0002-5992-6292>
Ralf Mikut  <https://orcid.org/0000-0001-9100-5496>
Richard Jumar  <https://orcid.org/0000-0001-6854-4678>
Heiko Maass  <https://orcid.org/0000-0002-8365-6042>
Uwe Kühnapfel  <https://orcid.org/0000-0002-2218-3229>
Veit Hagenmeyer  <https://orcid.org/0000-0002-3572-9083>
Benjamin Schäfer  <https://orcid.org/0000-0003-1607-9748>

References

- [1] Witthaut D, Hellmann F, Kurths J, Kettemann S, Meyer-Ortmanns H and Timme M 2022 *Rev. Mod. Phys.* **94** 015005
- [2] Machowski J, Lubosny Z, Bialek J W and Bumby J R 2020 *Power System Dynamics Stability and Control* (New York: Wiley)
- [3] Modig N, Eriksson R, Ruokolainen P, Ødegård J N, Weizenegger S and Fechtenburg T D 2022 Overview of frequency control in the nordic power system (available at: www.epressi.com/media/userfiles/107305/1648196866/overview-of-frequency-control-in-the-nordic-power-system-1.pdf) (Accessed 25 April 2022)
- [4] Tosatto A, Dijokas M, Weckesser T, Chatzivasileiadis S and Eriksson R 2021 Sharing reserves through HVDC: potential cost savings in the Nordic Countries *IET Gener. Transm. Distrib.* **15** 480–94
- [5] Landsnet 2020 Tjónistýring (available at: www.landsnet.is/arsskyrslur/arsskyrsla-2021/frammistoduskysrsla/tidnistyring-og-spennugaedi/) (Accessed June 2022)
- [6] ENTSOE 2021 Operational limits and conditions for mutual frequency support over HVDC
- [7] Ulbig A, Borsche T S and Andersson G 2014 *IFAC Proc. Volumes* **47** 7290
- [8] Milano F, Dörfler F, Hug G, Hill D J and Verbič G 2018 *2018 Power Systems Computation Conference (PSCC)* (IEEE) pp 1–25
- [9] Pillai J R and Bak-Jensen B 2010 *IEEE PES General Meeting* (IEEE) pp 1–8 (available at: <https://ieeexplore.ieee.org/document/5589729>)
- [10] Gorjão L R, Anvari M, Kantz H, Beck C, Witthaut D, Timme M and Schäfer B 2020 *IEEE Access* **8** 43082
- [11] Parhizi S, Lotfi H, Khodaei A and Bahrnamirad S 2015 *IEEE Access* **3** 890
- [12] Lidula N and Rajapakse A 2011 *Renew. Sustain. Energy Rev.* **15** 186

- [13] Lagrange A, de Simón-Martín M, González-Martínez A, Bracco S and Rosales-Asensio E 2020 *Int. J. Electr. Power Energy Syst.* **119** 105865
- [14] Palizban O, Kauhaniemi K and Guerrero J M 2014 *Renew. Sustain. Energy Rev.* **36** 428
- [15] Teymouri F, Amraee T, Saberi H and Capitanescu F 2017 *IEEE Trans. on Smart Grid* **10** 1735
- [16] Schäfer B, Witthaut D, Timme M and Latora V 2018 *Nat. Commun.* **9** 1
- [17] Schäfer B, Beck C, Aihara K, Witthaut D and Timme M 2018 *Nat. Energy* **3** 119
- [18] Schäfer B, Timme M and Witthaut D 2018 *2018 IEEE PES Innovative Smart Grid Technologies Conf. Europe (ISGT-Europe)* (IEEE) pp 1–5 (available at: <https://ieeexplore.ieee.org/abstract/document/8571793>)
- [19] Vorobev P, Greenwood D M, Bell J H, Bialek J W, Taylor P C and Turitsyn K 2019 *IEEE Trans. Power Syst.* **34** 3098
- [20] Kruse J, Schäfer B and Witthaut D 2020 *IEEE Access* **8** 149435
- [21] Hossain E, Khan I, Un-Noor F, Sikander S S and Sunny M S H 2019 *IEEE Access* **7** 13960
- [22] Yang S-G, Kim B J, Son S-W and Kim H 2021 *Chaos* **31** 123127
- [23] Kruse J, Schäfer B and Witthaut D 2021 *Patterns* **2** 100365
- [24] Kyesswa M, Çakmak H, Kühnapfel U and Hagenmeyer V 2020 *IET Renew. Power Gener.* **14** 4010
- [25] Chai J, Liu Y, Guo J, Wu L, Zhou D, Yao W, Liu Y, King T, Gracia J R and Patel M 2016 *2016 Power Systems Computation Conf. (PSCC)* (available at: <https://ieeexplore.ieee.org/abstract/document/7540946>)
- [26] Rydin Gorjão L, Jumar R, Maass H, Hagenmeyer V, Yalcin G C, Kruse J, Timme M, Beck C, Witthaut D and Schäfer B 2020 *Nat. Commun.* **11** 6362
- [27] Jumar R, Maaß H, Schäfer B, Gorjão L R and Hagenmeyer V 2021 arXiv:2006.01771
- [28] Gorjão L R, Schäfer B and Hassan G 2021 Power-grid frequency databasepermalink (available at: <https://power-grid-frequency.org/>)
- [29] ENTSO-E 2022 Transparency Platform (available at: <https://transparency.entsoe.eu/>)
- [30] Research N E 2020 The Nordics: share of renewables has increased (available at: www.nordicenergy.org/figure/share-of-renewables-has-increased/) (Accessed 30 March 2022)
- [31] Tróndheim H M, Niclasen B A, Nielsen T, Silva F F D and Bak C L 2021 *IEEE Open Access J. Power Energy* **8** 23
- [32] SONI 2022 What is the DS3 Programme? (available at: www.soni.ltd.uk/_uuid/a6f0ce76-c5a9-4120-8b53-cac2daa99840/ds3-programme/) (Accessed 7 April 2022)
- [33] Pütz S, Schäfer B, Witthaut D and Kruse J 2022 (available at: <https://energyinformatics.springeropen.com/articles/10.1186/s42162-022-00241-4>)
- [34] Curto D and Trapanese M 2018 *2018 IEEE Int. Conf. on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/ICPS Europe)* pp 1–6 (available at: <https://ieeexplore.ieee.org/document/8493876>)
- [35] Orkustofnun—National Energy Authority 2022 *Electricity* (available at: <https://nea.is/hydro-power/electric-power/nr/69>)
- [36] Fingrid 2022 Nordic power system frequency measurement data (available at: <https://data.fingrid.fi/en/dataset/frequency-historical-data>)
- [37] Maaß H, Çakmak H K, Suess W, Quinte A, Jakob W, Stucky K-U and Kuehnepfel U G 2013 *IEEE Trans. Instrum. Meas.* **62** 2384
- [38] Maaß H, Çakmak H K, Bach F, Mikut R, Harrabi A, Süß W, Jakob W, Stucky K-U, Kühnapfel U G and Hagenmeyer V 2015 *EURASIP J. Adv. Signal Proces.* **2015** 14
- [39] Förstner E, Jumar R, Tanrikulu O D, Maaß H, Kühnapfel U and Hagenmeyer V 2022 *The 2022 Int. Conf. on Smart Grid Synchronized Measurements and Analytics—SGSMA* (available at: <https://ieeexplore.ieee.org/document/9806021>)
- [40] Onsaker T L Code for predicting islanded power grid frequency (available at: <https://github.com/thlonsaker/predictability-of-island-power-grids.git>)
- [41] Kruse J, Schäfer B and Witthaut D 2021 *2021 IEEE Int. Conf. on Communications, Control and Computing Technologies for Smart Grids (SmartGridComm)* (IEEE) pp 133–9 (available at: <https://ieeexplore.ieee.org/document/9632335>)
- [42] Weißbach T and Welfonder E 2009 *2009 IEEE/PES Power Systems Conf. and Exposition* (IEEE) pp 1–6 (available at: <https://ieeexplore.ieee.org/document/4840180>)
- [43] Gardiner C W et al 1985 *Handbook of Stochastic Methods* vol 3 (Berlin: springer)
- [44] Anvari M, Gorjão L R, Timme M, Witthaut D, Schäfer B and Kantz H 2020 *Phys. Rev. Res.* **2** 013339
- [45] Böttcher P C, Gorjão L R, Beck C, Jumar R, Maass H, Hagenmeyer V, Witthaut D and Schäfer B 2022 (available at: <https://pubs.rsc.org/en/content/articlehtml/2022/ya/d2ya00150k>)
- [46] de España. R E 2022 Islas Baleares - Electricity demand tracking in real time (available at: <https://demanda.ree.es/visiona/baleares/baleares/total>) (Accessed 8 February 2022)
- [47] Eléctrica R 2022 Singular features of the Balearic system (available at: www.ree.es/en/about-us/ree-2-minutes) (Accessed 8 February 2022)
- [48] Martínez-Barbeito M, Gomila D and Colet P 2021 *ENERGY 2021, 11th Int. Conf. on Smart Grids, Green Communications and IT Energy-Aware Technologies* vol 2021 pp 13–18
- [49] Lim B and Zohren S 2021 *Phil. Trans. R. Soc. A* **379** 20200209