Norwegian University
of Life Sciences

Master's Thesis 2022 30 ECTS
Faculty of Science and Technology

Pattern-based Prediction of Islanded Power Grid Frequency

Thorbjørn Lund Onsaker
Environmental Physics and Renewable Energy

Preface

This thesis marks the end of my five years as a student at the Norwegian University of Life Sciences (NMBU). Student life at NMBU has given me academic and social knowledge and skills as valuable tools as I take on new challenges.

Working with this thesis and the curriculum through five years of studying Environmental Physics and Renewable Energy has been exciting and educational. Knowledge about the challenges facing the world community from a climate and energy perspective and methods for contributing to future solutions is valuable knowledge I now possess. Equally valuable is having found out what I am not interested in, which has made it possible to make good choices among further opportunities in life.

First and foremost, I would like to thank my two supervisors, Heidi Samuelsen Nygård at NMBU and Benjamin Schäfer at KIT (Karlsruhe Institute of Technology). Weekly meetings and close follow-ups have meant that advice, knowledge and encouragement never were far away. I would also like to extend my thanks to Richard Jumar and Johannes Kruse for providing up-to-date data and advice during the process. In addition, Stian Teien deserves an appreciation for contributing knowledge of programming and data science to overcome some of the challenges I faced.

Lastly, I would like to thank all fellow students I have shared my everyday life with, both at NMBU and through exchanges to Perth and Freiburg. I hope our paths cross again.

Ås, May 2022

Thorbjørn Lund Onsaker

Abstract

As part of achieving the climate goals from the Paris Agreement set by a united world community, the transition from non-renewable energy sources and electrification of transport and industrial sectors is central. The power grid faces new challenges as emerging renewable energy sources such as solar and wind fluctuate more than the more controllable traditional coal, gas, oil, and hydropower. In addition to more fluctuations on the supply side, an increase in emerging consumers like electric cars and data centers strengthens the need for good tools to maintain grid stability. Therefore, forecasting models and analysis to investigate both large and decentralized power systems' dynamics are necessary to maintain control and reliability.

Expanding upon a pattern-based prediction model called the Weighted-nearest neighbors (WNN) predictor, this thesis investigates the predictability of the power grid frequency for different European islanded power grids through several approaches. The selected islands include Ireland, the Balearic Islands, Iceland, and the Faroe Islands, with the Nordic region as a basis for comparison. The WNN predictor is successfully applied to all regions and performs 60 min predictions better than average daily profiles except for Iceland with its stochastic behavior. The Balearic Islands are the most deterministic region with precise predictions, while Ireland performs slightly worse than the Nordic region. The Faroe Islands exhibit similar performance to Iceland, but with significantly less data available. With varying population and geographical size, the regions cover the range of possible future grids, consisting of larger synchronous areas to small isolated island systems and microgrids.

All regions exclusively show that predictions improve with more data available. The predictor outperforms daily profiles with about a month of data available and with even less for more predictable regions. When electricity generation time series are included in an extended model approach, the performance slightly increases for parts of the predicted hour for several features, despite low time resolution and partially poor quality of the additional data. This suggests that there is further information to be gleaned from other power grid time series to improve the prediction of the power grid frequency.

Sammendrag

Som et ledd i å nå klimamålene fra Parisavtalen satt av et samlet verdenssamfunn, står overgangen fra ikke-fornybare energikilder og elektrifisering av transport- og industrisektorer sentralt. Kraftnettet står overfor nye utfordringer ettersom økende kraftproduksjon fra fornybare energikilder som sol og vind varierer mer enn regulerbare tradisjonelle kilder som kull, gass, olje og vannkraft. I tillegg til mer uregulerbar kraftproduksjon, styrkes behovet for gode verktøy for å opprettholde nettstabilitet gjennom blant annet økende energibehov til elbiler og satsning på datasentre. Derfor er prognosemodeller og analyser for å undersøke både store og desentraliserte kraftsystemers dynamikk nødvendig for å videreføre kontroll og pålitelighet.

Denne oppgaven tar utgangspunkt i, samt modifiserer, en eksisterende mønsterbasert prediksjonsmodell kalt Weighted-nearest neighbors (WNN), og undersøker hvorvidt strømnettets frekvens for ulike europeiske øybaserte strømnett er predikerbar. Oppgaven tar for seg øyene Irland, Balearene, Island og Færøyene, med Norden som sammenligningsgrunnlag. 60 min predikeringer med WNN-modellen gir bedre resultater enn gjennomsnittlig daglige frekvensprofiler for alle undersøkte regioner, bortsett fra Island med sine stokastiske trekk. Balearene er den mest deterministiske regionen med presise prediksjoner, mens Irland presterer litt dårligere enn nokså gjennomsnittlige Norden. For Færøyene vises lignende resultater som Island, men med betydelig mindre data tilgjengelig. Med varierende befolkning og geografisk størrelse er regionene en naturlig representasjon av framtidens potensielle kraftsystemer som kan variere fra kontinentale synkronområder til små, isolerte øysystemer og mikronett.

Alle regioner viser utelukkende at prediksjoner forbedres med mer tilgjengelig data. Generelt predikerer WNN-modellen bedre enn gjennomsnittlig daglige frekvensprofiler fra og med omtrent én måned med data tilgjengelig, og med enda mindre data for de mest predikerbare regionene. Til slutt inkluderes data for elektrisitetsproduksjon fra ulike energikilder i en utvidet versjon av modellen, og forbedrer resultatet for deler av den predikerte timen ytterligere, til tross for lav tidsoppløsning og delvis dårlig kvalitet på produksjonsdataen. Dette tyder på at det er ytterligere informasjon å hente fra annen kraftnettrelatert data for å forbedre prediksjonen av strømnetts frekvens.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AC **Alternating Current** - *Flow of electric charge that cyclically changes magnitude and direction.*

ARMA **Autoregressive Moving Average** - *Analysis model for time series that predicts based on past values with a lagged moving average.*

CE **Continental Europe** - *Here used as an abbreviation for Europe's largest synchronous area.*

DC **Direct Current** - *One-directional flow of electric charge of constant magnitude.*

EDR **Electrical Data Recorder** - *A phasor-measurement-unit-like device used for measuring frequency deviation.*

ENTSO-E **European Network of Transmission System Operators for Electricity** - *European organization that brings together 35 countries' TSOs for continental cooperation on power exchange.*

EU **European Union** - *Union of 27 European member states for economic and political cooperation.*

FCR **Frequency Containment Reserves** - *Reserves that activate after a few seconds to stabilize regular imbalances in the power system after a significant frequency deviation.*

FFR **Fast Frequency Reserves** - *Reserves that are rapidly activated when other frequency reserves fail to stabilize and regain balance of the power system.*

FRR **Frequency Restoration Reserves** - *Reserves taking over for the FCR to restore the frequency to its nominal value.*

HVAC **High-Voltage Alternating Current** - *High voltage system used to transfer cyclically direction-changing current over long distances.*

HVDC **High-Voltage Direct Current** - *High voltage system used to transfer one-directional current over long distances.*

KNN **K-Nearest Neighbors** - *A supervised machine learning algorithm used for regression and classification problems.*

MSE **Mean Squared Error** - *The average of the squared difference between predicted and actual values.*

NaN **Not a Number** - *A data type in programming representing an undefined value.*

PV **Photovoltaics** - *Devices that convert light into electricity by exploiting the generation of current and voltage occurring for certain materials when exposed to light.*

REE **Red Eléctrica de España** - *The Spanish company responsible for operating the Spanish power grid (TSO).*

RMSE **Root Mean Square Error** - *The square root of the MSE and a standard way of measuring a model's prediction error.*

RSC **Regional Security Coordinator** - *Organization owned by TSOs to provide service across the TSOs' geographical boundaries.*

SD **Standard Deviation** - *An average of a data set's variability and a measure of its dispersion.*

TSO **Transmission System Operator** - *An entity that owns and operates the power grid and transmission of electricity within a geographical area.*

WNN **Weighted-Nearest Neighbors** - *KNN-based prediction model where the nearest neighbors are weighted in the prediction.*

1 Introduction

1.1 Background and Motivation

On 12 December 2015, the Paris Agreement was signed by representatives from a large majority of the world's countries with the aim to limit global warming to a maximum of 2°C, preferably 1.5°C, compared to the pre-industrial era. The Intergovernmental Panel on Climate Change (IPCC) highlights in its report *Climate Change 2021* that human influence, with great certainty, is the main reason for the sea-level increase, increased carbon levels in the atmosphere, global retreat of glaciers and other dramatic environmental and biological changes [1]. The legal obligations of the Paris Agreement state that all participants should follow long-term strategies for reducing greenhouse gas emissions and take actions to reduce this global trend. In response to these long-term strategies, countries worldwide have set targets for the future to contribute toward these common goals. Taking Norway as an example, political and economic actions have been implemented. This involves a cap on greenhouse gas emissions, pollution fees, and financial support schemes to transition to electrification or emission reduction [2].

Norway has also set common goals through collaboration with the European Union (EU), aiming to reduce greenhouse gas emissions by at least 55% by 2030 and create an economy with net-zero greenhouse gas emissions by 2050 - a climate-neutral future [3]. In order to achieve the goals, significant changes must be carried out within transport, agriculture, industry, finance and innovation. Among the major transitions is the change from non-renewable carbon-based energy sources such as coal, oil and fossil gas to renewable and sustainable energy production [3]. In particular, the EU highlights wind, hydro and solar energy as essential contributors to green energy production. The share of Europe's electricity generated from renewable energy sources rose to 38% in 2020, overtaking fossil fuels as the EU's leading source of electricity for the first time [4]. The commenced electrification of transport and industry combined with renewable energy sources strengthens the need for a robust, flexible and reliable transmission system.

The electric power system has traditionally been powered by nuclear, coal, gas and hydropower; quite controllable energy sources in terms of controlling production for demand and supply in the grid. However, solar and wind generate electricity periodically instead of on-demand, and the increased share of these energy sources in the energy mix creates new challenges for grid stability. This is caused by higher and more unpredictable power peaks, fluctuation in production, and a significant variation on the consumer side through electric heating, electric vehicles and emerging electricity consumers like data centers. Additionally, in some places, it can be a disadvantage or challenging to transmit the generated power over long distances due to geographical spread, such as on an isolated island. Hence, smaller decentralized grids are also part of the future power systems, and a deeper understanding of their dynamics and characteristics is crucial to keep them operational and reliable.

Diagnostic tools and control systems are essential to cope with the various grid stability challenges. In addition, forecasts play an essential role, providing valuable information through analysis and predictions. Such qualified assumptions about, e.g. energy price,

amount of power to be generated, or expected consumption will make both control and generation of electricity easier to plan and manage, both in smaller and greater systems [5]. With Artificial Intelligence (AI) and machine learning, both the selection and quality of diagnostic tools and models have gained momentum, and distinct models are being used for various features or time scales [6] [7]. This thesis will utilize one specific prediction algorithm to analyze and investigate the predictability of various island power systems' grid frequency based on frequency time series data.

1.2 Previous Research and Limitations

In terms of power systems, forecasting is highly relevant for everything from system control to energy prices. This includes the day-ahead market for electricity trading, load forecasts or specific renewable generation type forecasts [8]. The topic of this thesis will be power grid frequency, expanding on the previous research from the paper *Predictability of Power Grid Frequency* published by Kruse et al [9]. Here, the authors utilized a machine learning-based frequency predictor called Weighted-nearest neighbors (WNN) [10], further explained in Sections 2.3.3 and 3.2, to perform data analysis and predictions of the power grid frequency in three central synchronous areas in Europe: Nordic, Continental Europe CE and Great Britain. The study was based on public and easy accessible frequency recordings from 2015 to 2018. Unlike analyzing large European synchronous power grids, this thesis will focus on previously unavailable frequency data collected from European islands and smaller isolated power grids, with more recent frequency recordings from the Nordic synchronous area as a comparison. In [9], the Nordic region was somewhat in between relative stochastic Great Britain and the more deterministic CE area regarding predictability and frequency patterns. While using the developed prediction algorithm and some of the analysis methods from the mentioned paper, this thesis will also focus on the predictions' data size-dependency. Furthermore, the prediction model will be extended by adding features like selected electricity generation data to the analysis in addition to frequency.

As high-resolution frequency data is not as easily accessible and available for smaller grids, the data has been collected manually by Jumar et al. with an Electrical Data Recorder (EDR), a phasor-measurement-unit-like device, usually plugged into conventional sockets in the specific regions [11]. Thus, the data may vary in quality, consistency and duration.

1.3 Research Questions

So far, the WNN predictor has only been applied to large power systems. By application to smaller isolated grids, the research could provide valuable information about trends or predictability of smaller power systems. This is highly relevant as many power systems operate without energy exchange with other regions due to isolated geographical locations. In addition, as microgrids emerge, grids that are fully or partially decoupled from greater power systems in periods might increase [12]. This applies to both isolated areas like the Norwegian island Svalbard and airports or hospitals where microgrids are applicable

[13]. Hence, this thesis investigates both the frequency of smaller islands' power grids and the prediction performance dependency of the amount of data available, as this might be limited for smaller regions. As power systems are highly complex and driven by many factors, including power system data in addition to the frequency in the predictions is relevant. Therefore, the thesis provides and tests an extension of the prediction model and includes different electricity generation sources.

Based on the former research and developed algorithm mentioned above, the research of the thesis can further be divided into three main questions:

1. How does the WNN predictor perform when applied to time series frequency data from smaller island power systems?

2. How does the WNN predictor performance depend on the amount of frequency data available and utilized for forecasting in the regions?

3. How can the additional electricity generation time series be integrated into the frequency prediction, and how do they influence the performance of the WNN predictor?

In addition to the Nordic synchronous area, the islands that will be investigated are Ireland, the Balearic Islands, Iceland and the Faroe Islands.

2 Theory and Background

2.1 Electric Power Systems

An electric power system is a network of electrical producers, distributors and consumers. Together, these three main components depend on each other to deliver or receive electricity for industry, work, communication, transport, and everyday life. The electricity generators have conventionally been huge power plants powered by fossil fuels like coal and gas, later also powered by nuclear power. On the renewable side, hydropower has been used for electricity production since the late 1800s, in addition to a massive increase in wind and solar power over the last 20 years [14]. Except for photovoltaic (PV) cells that convert energy from the sun to electricity, power plants' main principle and work is to exploit the thermal or mechanical energy of wind, water, and fuels to drive vast turbines. These turbines drive generators that convert rotational energy into electric energy. This energy must be transmitted to consumers and consumed at the time of generation, as the power grid cannot store the energy. However, both use and research on battery technology and other energy storage solutions have grown in recent years [15]. With this technology, electric energy can be stored so that it can be converted back when needed. This is not utilized on a large scale for current power grids but may play a vital role in the future [16].

The backbone of the power system is the power grid. Consisting of electrical infrastructure like transmission lines, masts, transformers and sensors, the power grid ensures the instant flow of electricity from generation to consumers. The power grid is a critical infrastructure at the level of water supply, public health, and transportation in modern society [17]. The power is transmitted across great distances, both on land and by sea, across national borders.

Power can be transmitted as either alternating current (AC) or direct current (DC). With AC, the current switches direction a certain number of times during the second, in addition to a constantly changing magnitude, in contrast to the constant magnitude and direction for DC. AC usually has the form of a sine wave, and the number of oscillations per second is denoted as the utility frequency in a transmission system, measured in Hertz (Hz). Frequency in the European power grids are operated at 50 Hz, while 60 Hz is the standard in the USA. The importance of frequency will be further explained in Section 2.3.

Today, most transmission systems are based on three-phase AC, but high-voltage direct current (HVDC) systems are also highly important. This applies, among other things, to submarine cables. Due to increased capacitance in the conductors subsea, AC needs to be supplied by more reactive power, which is not the case for DC [18]. In short, this means that DC reduces power losses over greater distances, particularly under water. HVDC is also used to connect distinct synchronous areas (Section 2.1.1). Converters and inverters convert between AC and DC and enable power transmission between regions separated by DC cables.

2.1.1 Grid Structure and Operation

To manage the transmission, development, control and other vital tasks of the power grid, every country has one or more entities serving as Transmission System Operator (TSO). A TSO is responsible for transmission on a national or regional level, meaning the longest transmission lines operated at the highest voltages [19]. The TSO also handles the interconnecting cables to other countries. The structure and management of the power grid may vary among countries. However, the key role is always to be the link between generators and consumers that ensures a stable flow of electricity. Taking Norway as an example, the grid is divided into three main sections. The TSO, Statnett, is responsible for the transmission grid, operated between 300 and 420 kV, as well as the interconnectors to, e.g. Germany and Sweden. Connected to the transmission grid is the regional grid. Usually between 132 and 33 kV, this grid is owned by both private companies or municipality and county authorities. While some large industries, hospitals or airports directly get their electricity from the regional grid, this grid is again connected to the distribution grid. This is the lowest level grid with a voltage from 22 kV to 230V, where the customers are primarily households and industry. Figure 2.1 roughly illustrates the structure of the Norwegian transmission system from electricity generation to consumption.

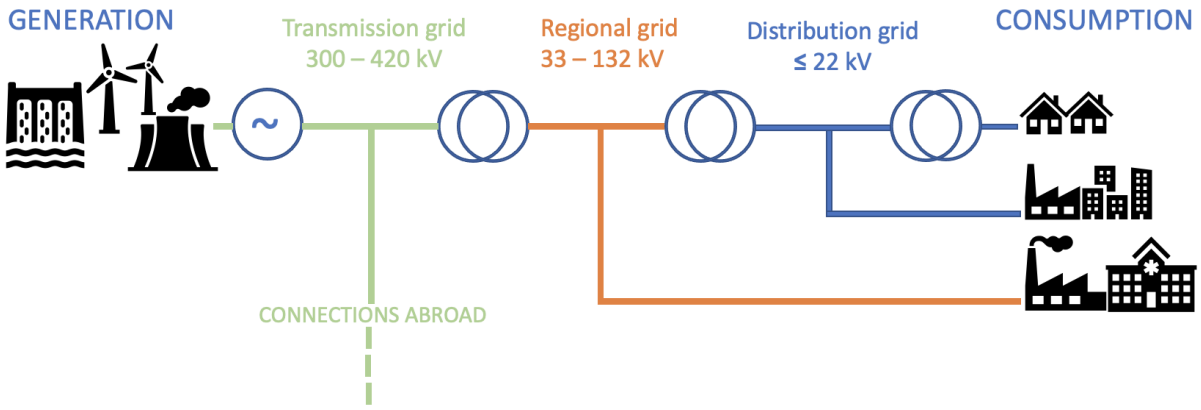


Figure 2.1: *Roughly illustrates the Norwegian transmission system with electricity generation, consumption and three levels of distribution grids. The transmission grid includes interconnectors to other countries and synchronous areas.*

The utility frequency of the Norwegian transmission system is synchronized with neighboring countries' power grids, and thus constitutes a synchronous area. The advantages of synchronous areas are many: By operating the grid at the same frequency, power can easily be transmitted through AC across regions and borders without conversion. Load can be levelled out, as one region might produce an overload of power and transmit this to another region with a power deficit. Thus, less controllable renewable resources such as renewable generation from wind and solar can also easier be distributed to a region in need of power instead of causing an over-generation problem. In Europe, there are five main synchronous areas, with the synchronous area of Continental Europe (CE) as the biggest one. The five European synchronous areas are shown in Figure 2.2. A national TSO governs most countries' power system, but within the synchronous area, a Regional Security Coordinator (RSC) provides information and strategies for the current TSOs as a platform for cooperation [20].

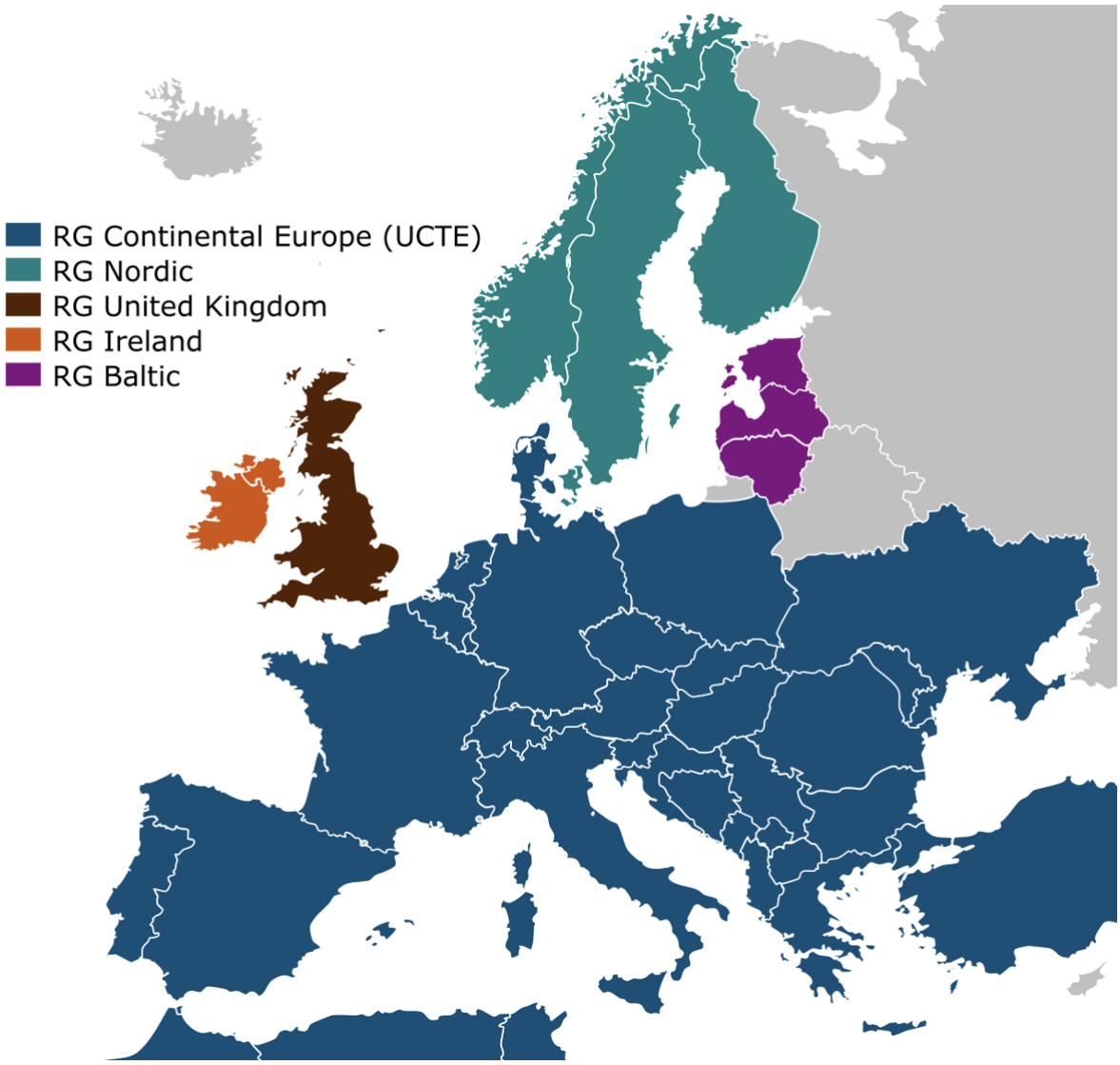


Figure 2.2: *The different European synchronous areas* [21]. *The Continental European (CE) synchronous area also includes the non-European countries Algeria, Morocco and Tunisia through an AC submarine cable from Gibraltar* [22]. *In addition, Ukraine and Moldova are synchronized with CE on a trial basis.*

2.1.2 ENTSO-E

The European Network of Transmission System Operators for Electricity (ENTSO-E) plays a key role in ensuring stable transmission operation and cooperation across borders and synchronous areas in Europe. The association gathers 39 TSOs from 35 European countries to coordinate operations and management of the European grid [20]. The ENTSO-E thus plays a decisive role in developing and shaping the future power systems to contribute to the European Green Deal's environmental goals for 2050 (Section 1.1). In addition to renewable conversion, the ENTSO-E's interests are optimized social welfare across Europe, and transparency and credibility are strictly necessary for cooperation on critical infrastructure for the whole of Europe. Therefore, the ENTSO-E provides public information and data such as generation, transmission and balancing data on their Transparency Platform [23].

2.1.3 Frequency and Stability

Frequency and stability are two closely related terms when discussing electric power systems. Power system stability is described as a system's ability to regain equilibrium after being exposed to sudden unexpected events and disturbances. Power system stability can be divided into three types: power angle, frequency and voltage stability. Power angle stability reflects a power system's connected synchronous machines' ability to synchronize after a sudden disturbance, and depends on the machines' ability to maintain balance between mechanical and electromagnetic torque. Voltage stability is more load-related, and refers to how well the power system maintains a sufficient voltage at each node of the system after a sudden disturbance [24]. Lastly, frequency stability reflects the system's ability to maintain its frequency within a given range after unforeseen events. The topic throughout this thesis is mainly relevant to the frequency part of stability. As voltage and power angle are related to individual nodes and generators, they are measured more on a local level, while frequency is a more global property of the whole synchronous area. The topic of this thesis will address the frequency part of stability.

Figure 2.3 expresses how electricity generation and consumption must be balanced to maintain stability. The frequency drops below 50 Hz when more power is consumed than generated. Opposite, the frequency rises above 50 Hz when there is too much power in the grid compared to how much is consumed. Whether at home or in industry, most electrical devices are designed to operate at 50 Hz (in Europe). These devices might get damaged if the frequency deviates more than 1% from the reference, or it might cause widespread disruptions. Large generating units will also be disconnected from the grid at high frequency deviations. If this occurs, the situation can escalate and lead to blackouts [25].

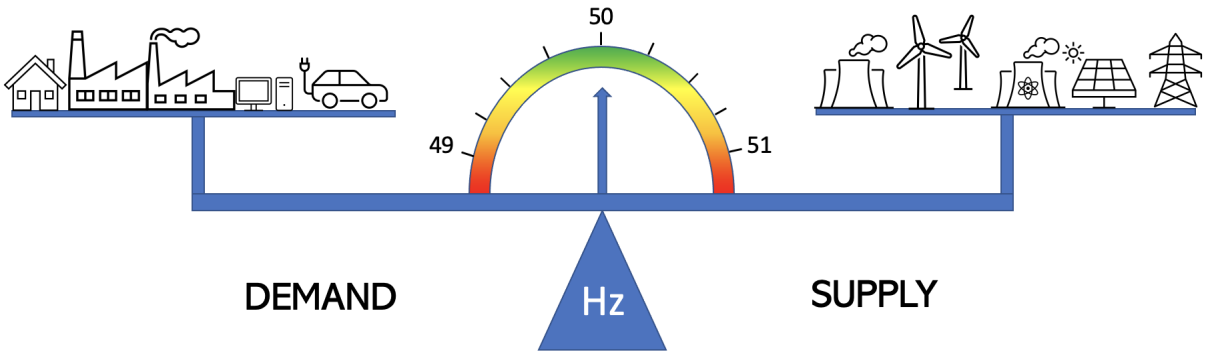


Figure 2.3: *System frequency expresses the balance between supply and demand in the power grid. At 50 Hz, the system is in equilibrium, but a sudden change in one of the sides will shift the frequency and may potentially damage operating machines and industry if not stabilized quickly.*

When the frequency is held at exactly 50 Hz, the system is in equilibrium. The frequency is said to be 50 Hz, but in reality, the frequency fluctuates at all times as demand and supply continuously change. As an example, the measured power grid frequency from the Spanish island Mallorca for a three hour period is shown in Figure 2.4. Here, the frequency stays close to the nominal frequency at 50 Hz but fluctuates continuously.

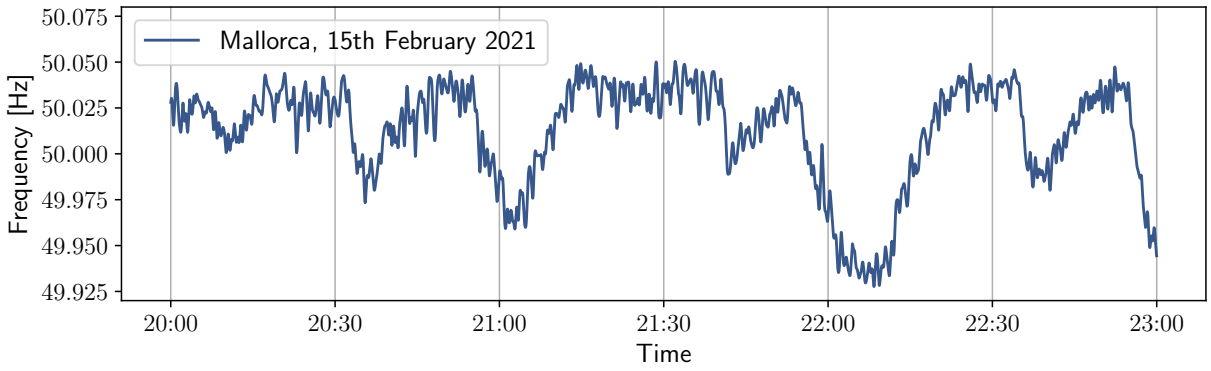


Figure 2.4: *A typical pattern showing the fluctuations in the power grid frequency for three hours measured on Mallorca, the largest island of the Balearic Islands. The frequency constantly varies around 50 Hz.*

2.1.4 Control and Reserves

The first contributor to resist the frequency change is the system kinetic energy, often referred to as inertia. With much inertia in the system, it takes more to change the frequency. This is because rotating machines and generators in the system start to absorb or deliver electric energy to compensate for lost generation or consumption. This is done by a change in the rotational speed of the machines, and the phenomenon, often called inertial response, can counteract frequency change a few seconds while control elements perceive the imbalance.

The amount of inertia in the grid has been critically reduced in several systems due to the transition from fossil fuels to renewables. Nuclear, coal and hydro are examples of sources with a high level of inertia due to generation by rotation of a turbine. On the other side, HVDC import, together with renewables such as PV and wind connected to the grid through inverters, provides electricity with no rotating masses in the system, thus low inertia levels. However, to still get a sufficient inertial response, much research has been done on synthetic inertia [26] [27], but a great focus is also on the later-acting frequency reserves.

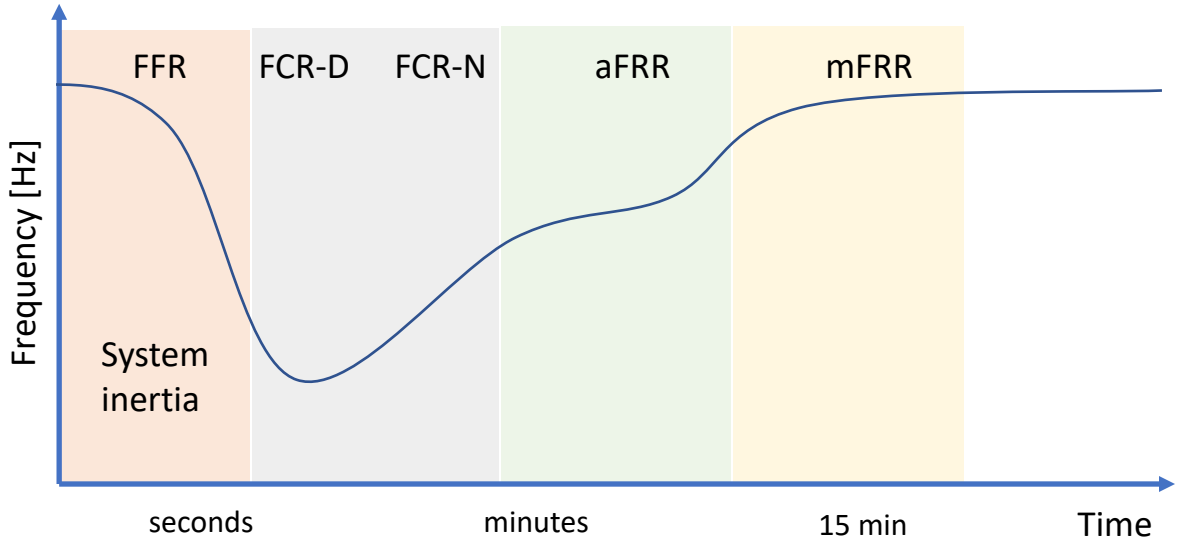


Figure 2.5: *After a sudden frequency drop, different frequency reserves are activated to regain balance in the power system. The fast frequency reserves (FFR) and the system inertia act the first seconds before the relative slower reserves act to take the frequency back to 50 Hz. Inspired by* [28].

To further control the frequency imbalance after the inertial response, primary, secondary, and tertiary control exists. With less inertia in the grid, these reserves are of even higher importance. Taking the Nordic power system as an example, the three control levels shown on a timeline after a frequency drop in Figure 2.5 consist of different types of frequency reserves:

- **Frequency Containment Reserves (FCR)** are the reserves that activate to stabilize regular imbalances. Activated automatically, the FCR-N (Normal) continuously work to counteract the frequency fluctuations under normal conditions. This is performed by an increase of generation power of each generator to resume the rotational speed lost in the inertial response to regain balance. The FCR-D (Disturbance) work to stabilize when rather large power imbalances or disturbances occur [25]. FCR is activated after just a few seconds to around a half minute and is thus the primary control.

- **Frequency Restoration Reserves (FRR)** are the reserves taking over for the FCR to restore the frequency to 50 Hz. Thus, the FCR can restore their capacity. These reserves are also divided into two: The aFRR (automatic) activate continuously, while mFRR (manual) are activated by the TSOs. This includes asking for start-up of generators not currently operating or requesting consumers to reduce load. FRR have a slower response time than the FCR, ranging from a full activation time for aFRR of 2 to 5 min, and are thus the secondary reserve. mFFR activate up to 15 min and are referred to as the tertiary reserve.

- Lastly, the **Fast Frequency Reserves (FFR)** constitute the rapidly activated reserves when the other reserves fail to stabilize and regain balance after large frequency drops. These reserves are often necessary when there is a small amount of kinetic energy in the system, and the risk of imbalance is significant. FFR activate

in seconds when the frequency reaches thresholds like e.g. 49.5 Hz. The amount of necessary FFR is dynamic, as it is related to the amount of system kinetic energy. With short-term forecasts, the kinetic energy is predicted for each day to know how much FFR, which are distributed across the Nordic TSOs, are needed [25].

As the frequency reserves need to be reliable and quickly activated, few renewable sources are suitable due to their fluctuating behavior controlled by the weather. To achieve the transition and goals of more renewable electricity generation in the future European power systems, precise predictions and understanding of system dynamics are essential to improve planning and management of frequency reserves, in addition to synthetic inertia [29]. The ENTSO-E has operation guidelines to ensure reliable amount and operation of reserves. There is even an FCR cooperation between multiple TSOs in the ENTSO-E that represent a market for the exchange of FCR in Europe, in addition to planned markets for other frequency reserves [28].

2.2 Selected Regions

This thesis will focus on the power grid frequency from four different European islands in addition to the Nordic synchronous area. These islands are Iceland, Ireland, the Balearic Islands and the Faroe Islands. The regions' location and prevalence are shown in Figure 2.6, showing good geographical spread and a difference in size. The islands are also very different in terms of present interconnectors and energy exchange with other countries, partly due to the various locations. These specific regions have been chosen since frequency measurements have been recorded here with the EDR [11]. Except for the Faroe Islands, all the regions are operated by TSOs bundled in the ENTSO-E. In the following sections, further details about the individual power systems follow, rounded off with some key information and differences highlighted in Table 2.1.

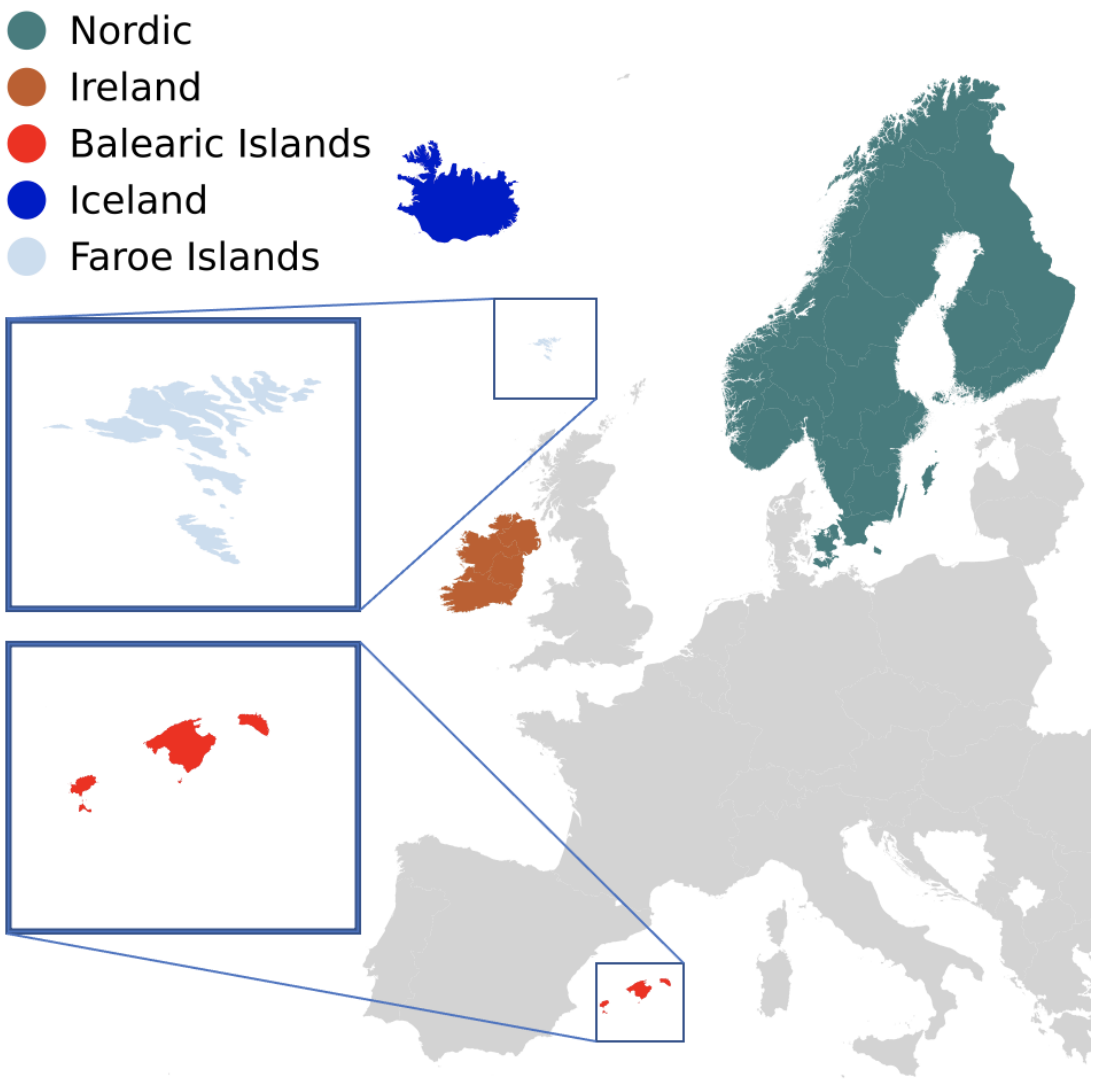


Figure 2.6: *Location of the regions with frequency time series to be analyzed and predicted throughout the thesis. Mallorca is the biggest island of the Balearic Islands, located in the middle of the archipelago, with Menorca to the east and Ibiza and the smallest Formentera located west.*

2.2.1 The Nordic Power System

The Nordic power system denotes the power system consisting of Norway, Sweden and Finland, as well as the eastern part of Denmark, involving Zealand and Bornholm as the most prominent islands. The region comprises almost 24 million people. As mentioned, Statnett is the Norwegian TSO, while Fingrid, Svenska Kraftnät and Energinet are the TSOs of Finland, Sweden and Denmark, respectively [30]. With a joint office in Copenhagen called the Nordic RSC, the TSOs of the four countries are supported with maintenance and grid operations. The power flow between the countries is secured by numerous high-voltage alternating current (HVAC) and HVDC cables on land or subsea between different regions of the countries. The Nordic RSC has set internal targets for the frequency to be outside the standard range of 50 ± 0.1 Hz for less than 10 000 min per year [31].

The energy mix for the Nordic electricity supply consists of a significant share of renewable energy, with a steady increase in recent years [32]. In Norway, more than 98% of the electricity generation comes from hydropower, in addition to some wind and fossil gas. Hydropower is also a significant contributor in Sweden and Finland, but here, nuclear power accounts for the main share of electricity production with about 40% and 30% respectively [33]. In Denmark, wind power is the dominant generation type, accounting for more than 45% of the electricity generation. As solar and wind power is also steadily boosting in the whole Nordic region, all four countries have a renewable share of electricity generation of more than 50%. By looking at the low-carbon electricity generation, thus including nuclear power, the Nordic region exceeds 80% [34].

The Nordic synchronous area is heavily connected to CE through links to, e.g. western Denmark, Netherlands, Germany and Poland. In addition to being linked to CE, the region has interconnectors to several other synchronous areas. For example, Norway has an HVDC cable to the UK, Sweden has an interconnector to the Baltic region, and Finland is connected to Russia [19].

2.2.2 The Irish Power System

The Irish power system describes the integrated electricity sector of the Republic of Ireland and Northern Ireland, both of which make up the island of Ireland. Northern Ireland is a part of the United Kingdom located northeast on the island, while the rest constitute the Republic of Ireland. The state-owned company EirGrid is the TSO of the Republic of Ireland, while SONI is the TSO of Northern Ireland. The population of the region is about 6.8 million. The region will be denoted as *Ireland* throughout the thesis, thus referring to the synchronous area consisting of the Republic of Ireland and Northern Ireland.

With windy weather due to the island's rough location in the Atlantic sea, wind is the second-biggest electricity provider after fossil gas. Ireland has an installed capacity of almost 5.6 GW of wind power, which is more than the average system demand, in addition to some solar and hydropower [35]. Through the program "Delivering a Secure, Sustainable Electricity System" (DS3), the two TSOs on the islands set a 40% instantaneous renewable electricity generation goal for 2020 to work towards the transition to a more renewable power system. The goal was achieved, and has later been raised and achieved at 75% in March 2022. In 2020, the renewable electricity generation reached 43% of annual generation [36]. EirGrid and SONI have set a target of keeping the frequency within the range 50 ± 0.1 Hz limit 94% of the time [37].

The Irish power system is connected by two 500 MW HVDC cables with a length of 64 km and 261 km to the British synchronous area. In 2026, Ireland will be connected via another cable to CE, namely the *Celtic Interconnector*, subject to deferral. The planned subsea interconnector is a cooperation with the french TSO, *Réseau de Transport d'Electricité*, to further secure the electricity supply as well as strengthen the energy integration with CE. The HVDC cable would have a length of about 575 km and a capacity of 700 MW [38].

2.2.3 The Balearic Power System

The Balearic Islands are an archipelago belonging to Spain, located in the Mediterranean Sea east of mainland Spain. The population of the whole archipelago is about 1.2 million, with a noticeable amount of tourists from all over Europe during the holiday season, as tourism is the archipelago's main industry. The majority of the inhabitants live on the largest island Mallorca, while the rest of the population live on the other three main islands called Menorca, Ibiza and Formentera. The spanish TSO, Red Eléctrica de España (REE), is also the grid operator at the Balearic Islands [39]. Renewables and nuclear power make up more than 60% of the main energy sources used to satisfy the need for electricity in mainland Spain, with wind as the main renewable contributor. Despite this high national renewable share, the distribution across different Spanish regions varies greatly, among others in the Balearic islands. Here, fossil fuels such as gas (combined cycle gas turbines) and diesel engines dominate the electricity production with more than 70%, followed by waste and a small share of PV solar power [40]. REE does not state a specific frequency deviation target for the Balearic Islands, but the synchronous area of CE, including mainland Spain and thus requirements that REE follows, is supposed to be operated within the range of 50 ± 0.05 Hz [41].

Before 2016, the Balearic islands' main grid consisted of two subsystems: Mallorca and Menorca were connected with AC cables, and the same was Ibiza and Formentera. This was due to the distance between the different islands, making stability and reliability difficult without interconnections [39]. The submarine cable now connecting the two former systems between Mallorca and Ibiza is one of the world's longest and deepest AC sea cables. The Balearic Islands are also connected to mainland Spain by a 244 km long HVDC cable, operating at 250 kV with a capacity of 400 MW. The cable is called COMETA, an abbreviation for COnexión MEditerránea Transporte Alta tensión. The cable was built to improve the electrical supply to the, at that time, two isolated power grids on the islands and made it possible to import energy corresponding to almost 20% of the annual electrical consumption [40]. With a maximum depth of 1485 m, COMETA is one of the deepest HVDC cables in the world [42].

2.2.4 The Icelandic Power System

Iceland is a European country and island located north in the Atlantic sea with roughly 300 000 inhabitants. The main source of electricity production is hydropower, making up 75% of the total generation. Iceland's location on the volcanic Mid-Atlantic Ridge has made it possible to utilize geothermal energy, and most households are heated with this energy form. Geothermal power stations are also generating electricity, and an increase in geothermal generation in recent years, now accounting for the last 25% of Iceland's electricity, makes all of Iceland's electricity renewable. In 2014, 80% of all electricity generated in Iceland was used by energy-intensive industries such as ferrosilicon and aluminium smelters [43].

The TSO of Iceland is Landsnet, a company owned by different Icelandic city- and state-owned companies. Founded in 2003, Landsnet has operated the national grid and administered its development and system operations for almost 20 years. The grid consists of

more than 3000 km of transmission lines, where most of the grid is operated between 30 and 220 kV [**?**]. The island's isolated location has made interconnections difficult, but the proposed subsea HVDC cable *IceLink* between Iceland and Great Britain has led to European cooperation becoming relevant. The possible 800-1200 MW cable will potentially become the world's longest interconnector with a length of 1000-1200 km. However, due to political disagreements and events like Brexit, neither an operation start date nor a confirmation that the cable is realistic has officially been stated [44].

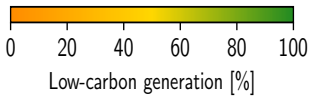
Landsnet monitors frequency and voltage quality in the entire transmission system year-round, where frequency values are recorded automatically every two seconds. Landsnet has set internal targets for the average value of frequency over a 10s period, where the frequency must be within 50 Hz $\pm$ 0.2 Hz 99.5% of the time [45].

2.2.5 The Faroese Power System

The Faroe Islands are an archipelago north in the Atlantic sea, located between Iceland, Norway and Scottland, without any gas or electricity connections to neither mainland Europe nor the British Isles. The Faroe Islands are a self-governing overseas administrative division of Denmark. SEV, a municipality-owned company, is both the leading electricity generator and distributor. With a population of roughly 50 000 spread across 18 major islands, the power system consists of seven isolated grids. Six of these belong to isolated islands and are powered by fossil fuel powerplants, while the last grid is the main grid that supplies a majority of the islands, including the capital Tórshavn, and more than 90% of the archipelago's electricity. Rough weather and wind due to the islands' location make the Faroe Islands exposed to blackouts, and the annual blackouts ratio is thus several times higher than in CE [46].

The main grid is operated at 60 kV, whereas 10 and 20 kV systems cover most towns and outlying islands. Around 60% of the electricity is produced by fuel and gas oil in 4 main plants. The Faroe Islands have seen rapid growth in renewable energy sources in the last years as SEV aims towards a 100% renewable electricity sector by 2030. The renewable share of installed capacity is now approaching 50% and generated about 40% of the electricity in 2021, driven mainly by wind and hydropower, in addition to some biogas power [46]. The frequency measurements from the Faroe Islands utilized in this thesis originate from the main grid.

Table 2.1: *The five examined regions vary significantly in population and leading electricity generation sources. Some of the regions are national regions; others are composed of several countries and are operated by multiple TSOs. Data retrieved from [34].*

Region	**Population**	**Country / Territories**	**TSO(s)**	**Main electricity generation sources** 
Nordic	24 mill	Norway, Sweden, Finland, Denmark (eastern parts)	Statnett, Svenska Kraftnät, Fingrid, Energinet	Hydro, nuclear, wind
Ireland	6.8 mill	Republic of Ireland, Northern Ireland	EirGrid, Soni	Gas, wind
Balearic Islands	1.2 mill	Province of Spain	Red Eléctrica de España	Gas, oil
Iceland	300 000	Iceland	Landsnet	Hydro, geothermal
Faroe Islands	50 000	Constituent country of Denmark	SEV	Oil, wind, hydro

2.3 Forecasting

Forecasting is simply about predicting the future and can provide valuable information about forthcoming events to respond appropriately or act on this basis. Historical time series data of price, frequency, temperature or other parameters are analyzed and used to predict the future development or behavior of a system. Cyclical patterns, seasonality or trends, in addition to time horizon, resolution and quality of the data, influence how accurate predictions can be achieved. Forecasting is widely used in many fields, such as finding trends in financial markets, weather forecasting and predicting energy prices or consumption [5] [8]. For example, precise forecasts of load or generation in the power system can reduce the amount of frequency reserves needed through more accurate system control, thus increasing reliability and reducing costs [25]. Hence, it is advantageous to have accurate forecasting methods from a security and economic perspective.

2.3.1 Forecasting Methods

Within the field of machine learning, a large selection of forecasting methods and algorithms applied to time series data exists. When focusing on forecasting quantities in the energy system, several current methods have been successfully applied to problems. A kernel regression model and an Autoregressive Moving Average (ARMA) model were used to forecast the power grid frequency with a month of data from three American power grids [10]. The non-parametric kernel regression technique tries to find a non-linear relationship between feature and target values through estimation of a random variable. The ARMA model predicts future values based on past values and past prediction error and smooths past time series with a lagged moving average. Other methods used for frequency time series forecasting include Bayesian network [47], Deep neural network or Lasso estimation [8], popular for electricity price forecasting. Additionally, modern deep learning techniques like Multilayer Perceptron (MLP) and Recurrent Neural Networks (RNN) are emerging within time series forecasting with the use of neural network [6]. Another popular branch within forecasting is pattern-based methods, explained in the next section.

2.3.2 State-of-the-art Pattern-based Forecasting

Several studies and forecasting methods are based on recognising similar recurrent patterns and using those to predict the future outcome, known as pattern-based forecasting. A former study by Martínez-Álvarez et al. introduced an algorithm called Pattern Sequence-based Forecasting to forecast time series of electricity prices and demand, obtaining competitive results [48]. Here, a clustering method was first used to group and label the data. Then, the predictions were performed for varying time windows based on the clustering results, where similar patterns from the data were found to compute an average as the prediction. Another study explained the K-nearest neighbors (KNN) algorithm approach, further explained in Section 2.3.3, to forecast the electricity demand for the next 24 hours [49]. Here, the most similar load trajectories from the past were used to compute a weighted average that makes up the prediction. The predictions showed lower forecasting error than a tree-based linear prediction model it was compared to.

The already mentioned paper by Kruse et al. investigated numerous aspects of frequency predictability with the WNN predictor [9]. The predictions were based on a weighted average of similar historical frequency trajectories. Firstly, the power grid frequency was predicted for the next 60 minutes. Secondly, the influence of the chosen starting time within the hour for prediction was investigated. Thirdly, the impact of the prediction window size was investigated, from the original 60 min window to as short as 15 min. As this was done for three greater synchronous areas with several years of data, similar analysis of smaller power systems with a variable amount of data is of high interest to deepen the understanding of smaller power grids' frequency predictability.

2.3.3 K-Nearest Neighbors Algorithm

K-nearest neighbors, usually abbreviated KNN, is one of the more traditional machine learning algorithms compared to popular emerging deep learning algorithms [6]. The algorithm learns from labelled training data, called supervised learning, meaning that the prediction result can be compared to the actual value and thus optimize the predictions by tuning hyperparameters. KNN is used for regression and classification problems and has many advantages that makes it useful for prediction. KNN is classified as a lazy learner algorithm, meaning it stores the given data set (training data) to search through it every time of prediction. This contrasts with other machine learning algorithms where models train on a given data set, producing a prediction function.

The concept of the algorithm is based on feature similarity, meaning that the algorithm looks through the training data to find similar values or series to the object to predict. The neighbor selection is usually determined by computing the euclidean distance between the object to predict and every object in the training set. Other distance metrics that are widely used are Manhattan, Minkowski and Hamming distance [50]. The number of neighbors, k, decides how many of the most similar values or patterns to include in the prediction decision. When used for regression, the prediction will usually be based on an average of the k nearest neighbors. Similarly, for classification problems, the algorithm decides that the unknown data point is the same type as the majority of its neighbors. A visualization of how two unknown data points are classified with KNN is shown in Figure 2.7.

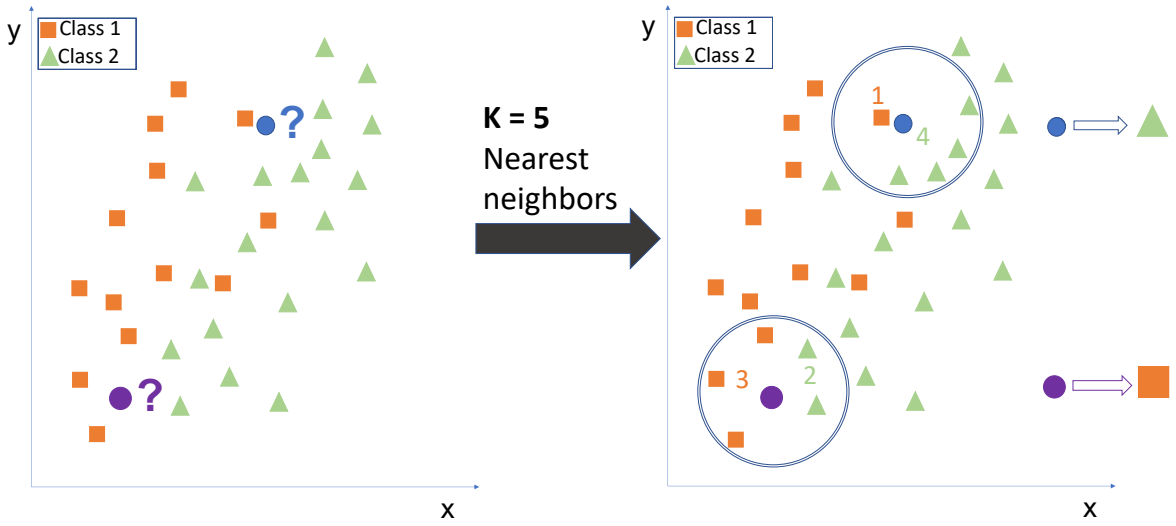


Figure 2.7: *An example of how KNN classification works. Two data points of unknown class are to be predicted.* k *is chosen to be 5, meaning the five nearest neighbors of each point will decide the class. The blue dot has four neighbors of class 2 (triangles) and only one of class 1 (squares). This data point is therefore set to class 2. for the purple dot, it is the opposite, and it becomes class 1 due to an excess of orange squares as neighbors.*

For most machine learning problems, data sets contain several features. These distinct features are not necessarily of the same size or fall within the same range for numerical values. For example, the price of a house and the number of rooms in the house are two

different features that will be different orders of magnitude. When euclidean distance is calculated on such a set, the price of the house would be dominant in the distance calculation. Therefore, feature scaling is essential. By scaling the features to a given domain or a specific mean, they can easier be compared and utilized as a whole. In Section 3.5, min-max scaling will be explained and applied, but other scaling methods include absolute scaling, standardization and robust scaling [51].

Throughout this thesis, a moderation of the KNN algorithm will be used. This is the Weighted-nearest neighbors (WNN), where the prediction is computed by a weighted average over the k nearest neighbors. This weighting can be self-defined and, if desired, optimized for the specific task. A detailed walk-through of the algorithm follows in Section 3.2.

3 Methods

3.1 Data Overview

3.1.1 Pre-Processing

The frequency recordings retrieved from the specific regions have a time resolution of 1 second, except for the data from the Nordic grid, which has a resolution of 0.1 seconds. Therefore, these data are re-sampled to 1 second by computing the mean of the frequencies within every second. The measurement recordings took place between 2019 and 2022 and varied in length. The start and end days of each frequency recording are shown in Table 3.1.

Table 3.1: *The start and end date of the frequency measurements made in each region.*

Region	**Frequency measurement recording periods**	
	Start	**End**
Nordic	1 January 2021	31 December 2021
Ireland	4 November 2021	22 February 2022
Balearic Islands	29 September 2019	10 March 2021
Iceland	5 November 2021	30 January 2022
Faroe Islands	3 November 2019	10 November 2019

For all regions, isolated measurements differing more than 50 mHz from the next and previous measurements have been filtered out, as well as series where the frequency, according to the measurements, is utterly stable for more than 15s with an accuracy greater than 10 µHz. Measurements above 51 Hz or below 49 Hz were also filtered out as they are extreme outliers. In practice, no data were filtered out but instead set to Not a Number (NaN) for the given timestamp to keep the time series complete. For short intervals up to 6s with NaNs, the data points were filled with the last valid measurement, while for longer intervals, the NaNs were kept and thus skipped in further analysis and predictions. The python code for analysis and processing of the frequency time series is available on Github [52].

3.1.2 Distribution and Increment Analysis

The frequency measurements for the distinct regions are expected to be based around the reference at 50 Hz (from Section 2.3). Whether the frequency exhibits large deviations or fluctuations from the reference over time is unknown and may provide further information about the grid dynamics. To get a first overview of the frequency measurements, distribution histograms for each time series were plotted. The histograms show how the frequency deviates from the reference frequency of 50 Hz. With a logarithmic probability scale, the distributions should exhibit an inverted parabola if the fluctuations are Gaussian distributed [53].

In addition, the short time change in frequency can be investigated by analyzing the increments in the time series. The frequency increment Δf_τ for a given time lag τ is given as

$$\Delta f_\tau = f_{t+\tau} - f_t. \tag{1}$$

The probability density function for the frequency increments was plotted with $\tau = 1$s, which is the lowest possible increment as it equals the time resolution of the time series, as well as a greater time lag of $\tau = 10$s.

3.1.3 Frequency Patterns

After analyzing the different frequency time series, additional information about the frequency dynamics could be found by exploring trends and repetitive patterns within certain time intervals. As a first analysis of the regions' frequency time series, the autocorrelation function for each time series was investigated for two different time lags. Autocorrelation describes how a given time series corresponds to a lagged version of itself for a given time interval. The autocorrelation function tells how the data points are related to data points in the time series at different delays. The resulting trajectory will be in the domain $[-1, 1]$, where -1 represents a perfect negative correlation, and 1 represents a perfect correlation. At 0, the function reveals that a time series shows no correlation for the given lag. The time lags used for the frequency time series were up to 20 days and 100 min to capture possible daily or hourly correlation.

In [9], The daily recurrent frequency pattern, denoted the daily profile, was highlighted as important for the three greater European synchronous areas investigated. The daily profile is further mathematically defined and used as a null model in Section 3.2 and is the average frequency for each specific time of the day for the whole time series. Thus, this null model provides an idea of what specific frequency to expect the most.

The standard deviation SD for the entire day and for an hour addressed how large frequency deviations from the daily profile to expect. The daily SD was found by computing the SD for each second of the day for all days in the time series. The hourly SD was calculated similarly, but now the focus was on the SD of each second of the hours in the time series, ignoring what specific hour of the day it is.

3.2 The Prediction Process

The WNN method predicts a frequency pattern based on patterns from the past, similar to the recent pattern $\mathbf{F}_0$ ending at the prediction start t_0. The computational process for the prediction in this section is the same as the one used and described in *Predictability of Power Grid Frequency* [9]. The python code for predictions is available on Github [52]. This process will be used as a whole in Section 4.3, while modifications and extensions to the model are presented in Section 3.4 and 3.5. Common to all three sections is that the different data sets were split into three subsets. These are training, validation, and test sets commonly used in machine learning approaches [54] [55]. The size of the subsets varied due to the varying length and quality of the time series and the different approaches.

Mainly, the selected data were split such that the training set made up 60-70% and the validation and test set about 15-20% each.

3.2.1 Patterns and Prediction

After the pre-processing, the cleaned data were split into patterns with a specific time delay τ and number of data points γ:

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

As this thesis only focuses on 60 min predictions and the time resolution of all the frequency time series is 1s, the pattern length is thus $\gamma\tau = 3600$s. Hence, $\mathbf{F}_n$ is a vector with 3600 data points. For most predictions, the model was constructed to only look for similar patterns within the same hour of the day unless stated otherwise. This splits the set $\mathbf{F}_n$ into non-overlapping sets, one for each wanted time slot of the day, given as

$$\mathcal{F} = \{\mathbf{F}_n | \exists i \in \mathbb{N} : n\gamma\tau = i \cdot 24\text{h}\}. \tag{3}$$

The set $\mathcal{F}$ is searched to find the most similar patterns to the initial pattern. This is done by computing the euclidean distance between the initial pattern and all patterns in $\mathcal{F}$, producing distances

$$\text{d}(\mathbf{F}_n) = \|\mathbf{F}_n - \mathbf{F}_0\|. \tag{4}$$

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

$$\mathcal{N}_k = \{n_1, n_2, \ldots, n_k | \mathbf{F}_{n_j} \in \mathcal{F}\}. \tag{5}$$

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

$$f_{wnn}(t_0 + \Delta t) = \frac{1}{\sum_{j=1}^{k} \alpha_j} \sum_{j=1}^{k} \alpha_j f(t_0 - n_j\gamma\tau + \Delta t). \tag{6}$$

The weighting α used in this WNN method decreases with greater distance between the patterns and is calculated by:

$$\alpha_j = \frac{\text{d}(\mathbf{F}_{n_k}) - \text{d}(\mathbf{F}_{n_j})}{\text{d}(\mathbf{F}_{n_k}) - \text{d}(\mathbf{F}_{n_1})}. \tag{7}$$

The weights cause the closest neighbors to be more decisive in the prediction than the k nearest neighbor, as more similar patterns are expected to be of greater interest than less similar ones. The weight α_k is 0 due to $d(\mathbf{F}_{n_k}) - d(\mathbf{F}_{n_k}) = 0$. Thus the predictions are in

practice based on the $k-1$ nearest neighbors rather than the k nearest. A visualization of the WNN prediction process is shown in Figure 3.1.

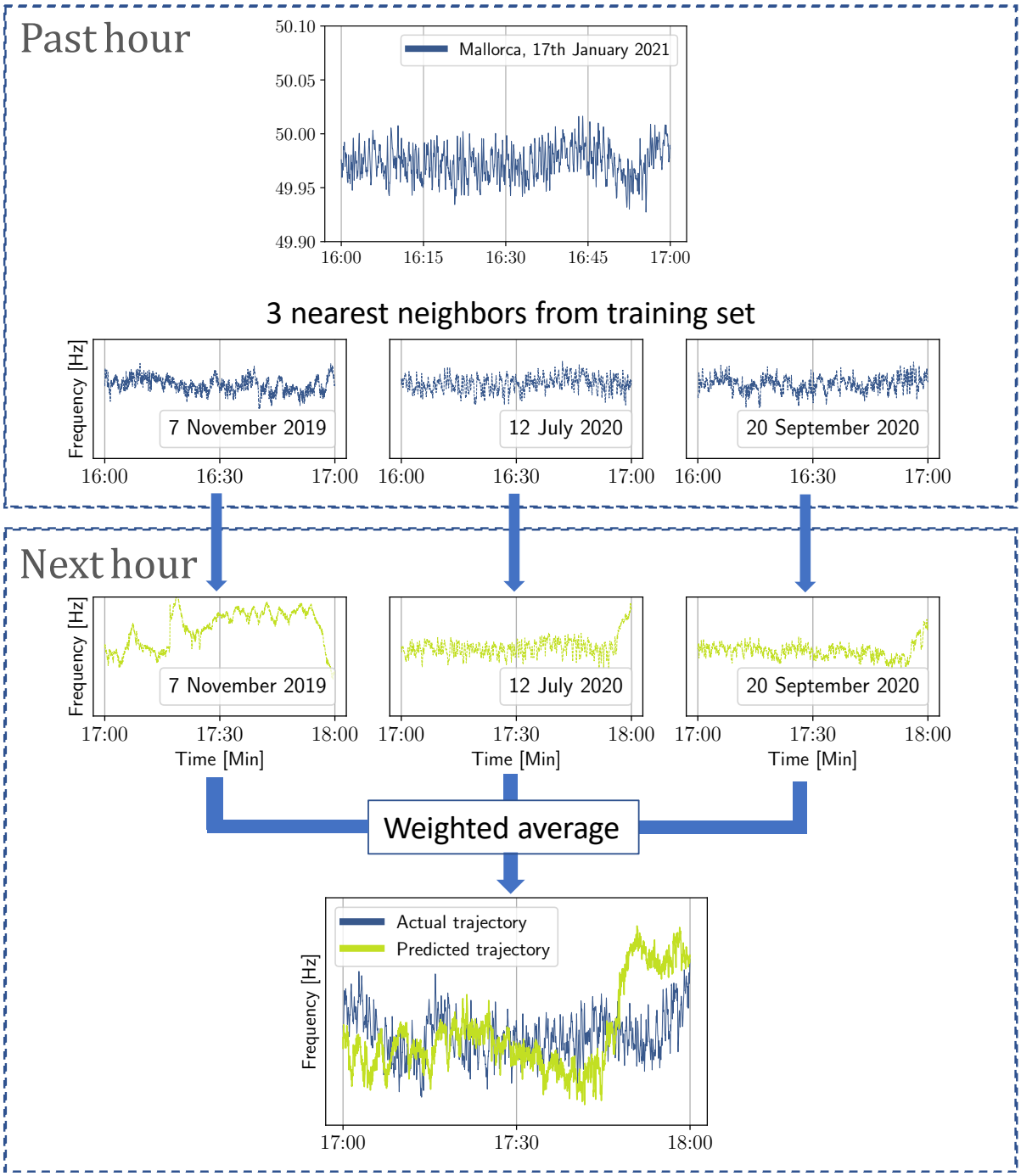


Figure 3.1: *The WNN predictor predicts the next hour trajectory based on similar patterns from the past. The past hour is matched against previous patterns from the past to find the most ones. Here, as an example, the three nearest neighbors are found. A weighted average of the neighbors' next hour trajectory are computed as the prediction and can thus be compared to the actual trajectory from the test set.*

3.2.2 Performance Measures and Optimization

Both for optimization and performance evaluation of the WNN predictor, suitable performance measures are important. The mean squared error (MSE) is used for optimization of the model, and evaluates the time-dependent MSE of the model for every second of the hour:

$$\mathrm{MSE}_{\Delta t}(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{f}(h_0^i + \Delta t) - f(h_0^i + \Delta t)\right)^2. \tag{8}$$

Here, N is the number of hours used from the time series, and vary depending on the length of the training data used for the prediction. h_0^i counts the hours after the beginning of the training data.

To further optimize the model, the hyperparameter k is tuned. In [9], two different k-approaches, *fixed-* and *adaptive-k*, were used. The *fixed-k* approach is common for both regression and classification KNN methods, where a specific k is chosen for the task [56] [57]. In this thesis, only the *adaptive-k* will be used for predictions. This is due to predictions with it from the greater European synchronous areas gave slightly better results than *fixed-k*, especially in the first minutes [9]. The *adaptive-k* is the optimal k-value giving the lowest $\mathrm{MSE}_{\Delta t}(f_{wnn})$ of the predictor for each Δt. Hence is *adaptive-k* a vector with 3600 values. The *adaptive-k* is smoothed with a sliding window of one minute to avoid sudden jumps and noise for any time steps. Throughout the thesis, the selection of k was in practice done by performing a grid search to find the optimal k from the set $\mathcal{K} = \{1, 2, 3, ..., K\}$. K is the number of full days in the training data, as it is the maximum number of possible neighbors. For the longest time series, K was set to 300 as a maximum.

3.2.3 Benchmarking

To benchmark the performance of the model, it was compared to the two null models introduced in [9]. These null models are easily interpretable, yet highly relevant for frequency predictions. The first null model is just a constant frequency of 50 Hz, which would be the desired frequency at any time (Section 2.1.3):

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

The second null model is the daily profile (Section 3.1.3), where the average over all patterns from the set $\mathcal{F}$ is computed with uniform weighting:

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

This is equivalent to using uniform weights for the WNN predictor and letting k be the number of patterns in the training data. The actual benchmarking was done by computing the root mean square error (RMSE) for the WNN predictor and the two null models:

$$\mathrm{RMSE}(\hat{f}) = \sqrt{\mathrm{MSE}_{\Delta t}(\hat{f})}. \tag{11}$$

In addition to RMSE, the uncertainty in the predictions were addressed through the SD of the nearest neighbors from the set $\mathcal{N}_k$:

$$\sigma_{\Delta t} = \sqrt{\langle f(t_0 - n\gamma\tau + \Delta t)^2 \rangle - \langle f(t_0 - n\gamma\tau + \Delta t) \rangle^2}. \quad (12)$$

Here, $\langle \cdot \rangle$ represents the average over all $n \in S_k$. By plotting the time dependent standard deviation with the predicted trajectory, a prediction error can be visualized.

3.3 Predictions

The first application of the WNN predictor was to perform predictions for the full time series for all regions. As the time series from the Faroe Islands consists of less than a week of data, these predictions were based on all patterns from the past and not just the specific hour of the day (from (3)). The grid search for the Faroe Islands was thus performed on the set $\mathcal{K} = \{1, 2, 3, ..., 24 \cdot K\}$. Apart from this, the process was solely the prediction process from Section 3.2.

3.4 Size Dependence

The second approach to examine was the data size dependence of the predictor. As the results from the full time series predictions in Section 4.3 are not directly comparable between the regions due to the different amounts of data available, a study with equal time series lengths is valuable. It can also reveal whether it will make sense to use the WNN predictor for predictions or analysis with limited data available. Six different time intervals were chosen, ranging from one week to six months. As some of the regions' frequency recordings did not last six months; these regions were naturally only predicted for short enough periods. The exact time intervals predicted for each region are shown in Table 3.2. Regardless of the time interval, all predictions were tested on the last week of each full time series. Here, a month is considered to always consist of 30 days, while a week naturally is seven days. The prediction results were compared to the two null models from Section 3.2 for all time intervals.

Table 3.2: *To investigate the size dependence of the WNN predictor, the frequency time series from the different regions were split into different time intervals. The table lists the different time intervals p tested for each region, ranging from one week to six months. As the time series from the Faroe Islands is shorter than a week, it was not included in these predictions.*

Prediction time intervals						
Time period, p:	Weeks		Months			
Region:	1	2	1	2	4	6
Balearic Islands	x	x	x	x	x	x
Nordic	x	x	x	x	x	x
Ireland	x	x	x	x	-	-
Iceland	x	x	x	x	-	-
Faroe Islands	-	-	-	-	-	-

To ensure a given time interval p from Table 3.2 is robust against rare trajectories or events that might influence the results, especially for short time scales, 25 different samples $\mathcal{S}_n$ for each p are used for the analysis. The samples are randomly chosen from the frequency time series, with the last week of the series left out as the test set. $\mathcal{S}_n$ thus starts at a random hour, day and time of the year. The exception is for Ireland with $p = 2$ months. Here, the pre-processing in Section 3.1.1 reduced the frequency time series to slightly more than two months of valid data, excluding the test set. $\mathcal{S}_n$ is thus chosen by starting the period with an increment of 4 hours from the start of the time series to achieve some variety without containing too many NaNs. For all regions, $\mathcal{S}_n$ is split into train and validation sets for every p and make up 80% and 20% of p, respectively. When the subsets are created, a last check is performed to secure the sets against consisting of a large share of invalid data. When cut into non-overlapping intervals of one hour, less than 10% of the chunks are allowed to contain NaNs. If the subsets do not pass the last check, they are rejected, and a new random period is chosen.

The hyperparameter optimization and performance evaluation is similar to Section 3.2 and is performed equally for all time intervals p: First, a specific *adaptive-k* for each $\mathcal{S}_n$ is found, followed by predictions. To make an average performance evaluation of all $\mathcal{S}_n$, the RMSE is first calculated for each sample individually. Furthermore, an average RMSE can be calculated for both the WNN predictor and the null models:

$$\mathrm{RMSE}(\hat{f}) = \frac{1}{25} \sum_{n=1}^{25} \mathrm{RMSE}_n(\hat{f}). \tag{13}$$

The result is now an average error trajectory for the entire hour. To easier compare the WNN predictor and the two null models for each p, a simple average value for the whole hour was computed:

$$\mathrm{RMSE}_{avg}(\hat{f}) = \frac{1}{3600} \sum_{n=1}^{3600} \mathrm{RMSE}(\hat{f}). \tag{14}$$

3.5 Inclusion of Additional Power System Quantities

3.5.1 Additional Data

So far, frequency trajectories were predicted only based on frequency. Now, additional feature trajectories are added to the frequency patterns. As frequency reflects the balance between generation and consumption in often highly complex power systems, multiple features influence the overall system balance. Firstly, as mentioned, electricity demand and load patterns are important. Secondly, these patterns are again influenced by several features. For example, the electricity generation from renewable energy sources like solar or wind depends on the weather: Windy days increase wind power generation, and cloudy days reduce the solar generation. With low temperatures, electricity demand might increase if many households have electric heaters. In Figure 3.2, the electricity generation from six central energy sources in the Balearic Islands is shown for a random week. Each source exhibits its distinct generation pattern and might potentially correlate or share trends with the frequency trajectories.

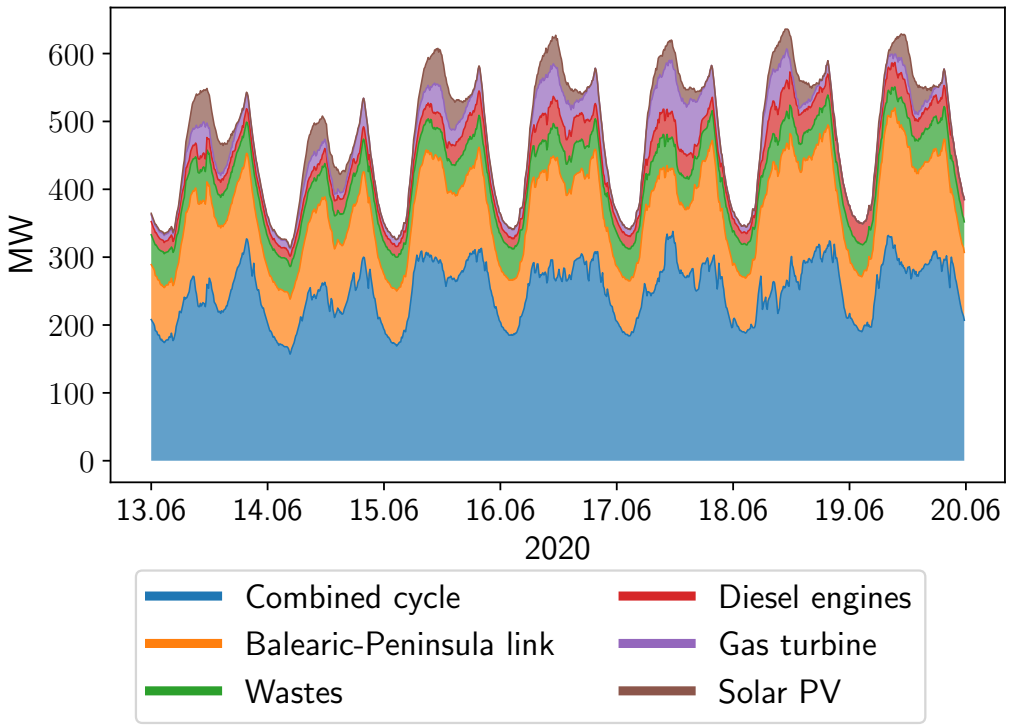


Figure 3.2: *The electricity generation mix of a power system consists of different energy sources. Presented are the six largest sources for a randomly chosen one-week period in the Balearic Islands [58]. This also includes the HVDC import of power from mainland Spain. As the sources contribute to the balance between load and generation, the different trajectories might provide additional information relevant for frequency predictions.*

In this approach, the predictions are still based on a weighting of the nearest frequency neighbors. The new concept is that additional features will influence the choice of nearest neighbors: Rather than choosing the most similar frequency patterns, other patterns might potentially be chosen instead because their additional feature trajectories were similar. The features were not predicted themselves but were added to provide additional information to the target values, which is the frequency. The additional features that were tested for the extended predictor were all generation or import. However, the extension is designed so that all types of data such as demand, temperature or precipitation time series can be tested as well. The additional features included different types of generation, such as renewables and coal. The Balearic Islands generation data were downloaded from REE's real-time electricity demand and generation tracker [58], and the Irish data were downloaded from the ENTSO-E Transparency Platform [23], as well as from Eirgrid's system information chart [59]. All additional data had a significantly lower sampling rate than the frequency measurements.

3.5.2 Extending the Algorithm

If the additional feature time series has the same time resolution as the frequency data, the feature patterns can be added to the data set and the data set scaled as stated in Section 2.3.3. As this is not the case for the available data used in this thesis, the data must be

included in the WNN predictor differently. The additional feature data points will instead be added as an extension to the vectors $\mathbf{F}_n$. The additional data are scaled with min-max scaling before they are split into patterns and added to the frequency set. This is because the frequency time series always have a mean around 50 and a relatively low standard deviation, but the additional features might vary largely in order of magnitude, both between and within the different time series. The min-max scaling transforms the data to fall in the domain $[0, 1]$, and a data point a is transformed by the following formula:

$$a_{new} = \frac{a - a_{min}}{a_{max} - a_{min}}. \tag{15}$$

Here, a is the original data point, a_{new} is the transformed data point and a_{max} and a_{min} are the maximum and minimum values in the time series, respectively. furthermore, the additional feature is split into non-overlapping patterns $\mathbf{A}_n$, similar to $\mathbf{F}_n$ from (2):

$$\mathbf{A}_n = \begin{pmatrix} a(t_0 - (n+1)\eta\lambda) \\ a(t_0 - (n+1)\eta\lambda + \lambda) \\ a(t_0 - (n+1)\eta\lambda + 2\lambda) \\ \dots \\ a(t_0 - n\eta\lambda - \lambda) \end{pmatrix}, \tag{16}$$

where a is the additional feature data point, 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 60 min. 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:

$$\mathbf{G}_n = \begin{pmatrix} \mathbf{F}_n \\ \mathbf{A}_n \end{pmatrix}. \tag{17}$$

As the extended model is also limited to search through patterns from the same time of the day, $\mathcal{G}$ is a new set similar to $\mathcal{F}$ from (3). To now choose the nearest neighbors in terms of distance, (4) is extended to

$$\mathrm{d}(\mathbf{G}_n) = \|\mathbf{F}_n - \mathbf{F}_0\| + \beta \cdot \|\mathbf{A}_n - \mathbf{A}_0\| \tag{18}$$

where $\|\cdot\|$ denotes the euclidean distance, and β is a tunable weight to adjust the influence of the new additional feature. With new distances, the weighting α from (7) is now computed as

$$\alpha_j = \frac{\mathrm{d}(\mathbf{G}_{n_k}) - \mathrm{d}(\mathbf{G}_{n_j})}{\mathrm{d}(\mathbf{G}_{n_k}) - \mathrm{d}(\mathbf{G}_{n_1})}. \tag{19}$$

Hence, the additional features also directly influence the frequency prediction. As in the original prediction process, the patterns are sorted based on their distances. This again results in the set $\mathcal{N}_k$, now given as

$$\mathcal{N}_k = \{n_1, n_2, \dots, n_k | \mathbf{F}_{n_j} \in \mathcal{G}\}. \tag{20}$$

Here, the additional feature vectors $\mathbf{A}_n$ are removed, and only the frequency part of the vectors in $\mathcal{G}$ are kept in $\mathcal{N}_k$. As only the frequency trajectories are predicted in the model, the additional feature has accomplished its job by influencing the selection of neighbors and α, and is no longer of interest. Therefore, the prediction and benchmarking are now performed in the same ways as in the original process from Section 3.2.

3.5.3 Optimization

The new hyperparameter β must be optimized to choose how the influence of the neighbor selection should be distributed between frequency similarity and generation similarity. As there now are two hyperparameters, k and β, grid search was used to optimize both. The optimization was performed by a grid search on the set of weights $\mathcal{W} = \{0, 0.1, 0.2, ..., 1.5\}$. 1.5 was set as the upper limit after multiple attempts with several features where the $\text{MSE}_{\Delta t}(f_{wnn})$ clearly increased after passing 1. For every weight in $\mathcal{W}$, a grid search on the set $\mathcal{K}$ like in Section 3.2.2 was performed. This resulted in an optimal weight β, which is fixed for all data points in $\mathbf{A}_n$, and a new *adaptive-k* that gave the lowest overall $\text{MSE}_{\Delta t}(f_{wnn})$. 0 is included in $\mathcal{W}$ to include the possibility that ignoring the additional feature in the neighbor selection might be the optimal choice.

4 Results and Discussion

4.1 Data Analysis

4.1.1 Pre-Processing

During cleaning of the data, the time series from the Balearic Islands and Ireland contained a significant amount of bad data that were set to NaNs. The main problem with the data were that the exact same frequency was logged for longer periods, resulting in periods with invalid data. For the other three regions, the time series had negligible to no bad recordings and could be used as a whole during the data analysis. The length of the original time series and the new length represented as full days when NaNs are ignored are shown in Table 4.1. During the WNN prediction process, a time series is split into non-overlapping chunks of one hour (Section 3.2). When a chunk contains too many NaNs, it is ignored during the process. This applies if there are NaN periods longer than 6s in the chunk of one hour, which was the maximum allowed NaN length from Section 3.1.1, meaning all 3600 values are ignored. This again results in a new share of data not being utilized, represented as *hourly chunks* in Table 4.1. The percentage not utilized data only significantly increased due to this process for the Nordic region. This is because the few invalid measurements from the Nordic data appeared as shorter error series here and there as opposed to the long invalid periods for the Balearic Islands and Ireland, thus affected several hourly chunks.

Table 4.1: *When pre-processed, the frequency time series for some regions are reduced due to invalid values set to NaN. The table shows each region's original time series length, given in full days. Also included is the percentage NaNs after data cleaning, reducing it down to a new time series length for either full time series or hourly chunks.*

Information about frequency time series			
Region	Original time series [days]	Percentages NaNs full series / hourly chunks* [%]	New time series full series / hourly chunks* [days]
Nordic	365	0.1 / 5.6	365 / 344
Ireland	110	24.6 / 24.7	83 / 83
Balearic Islands	529	14.9 / 15.2	450 / 448
Iceland	86	0.0 / 0.0	86 / 86
Faroe Islands	6	0.1 / 0.7	6 / 6

*Amount of data not used in Section 4.3 due to removed chunks containing NaNs.

The clear difference between the length of the time series is worth noting, varying from 450 to six days after pre-processing. With more data, the reliability of an analysis and predictions increases, as the time series then easier capture trends, has a wider purview and more data to train prediction models on. With about 15% invalid measurements, the time series from the Balearic Islands was significantly reduced. The quality of the Irish measurements were the worst, where one in four measurements on average were invalid.

4.1.2 Frequency Distribution

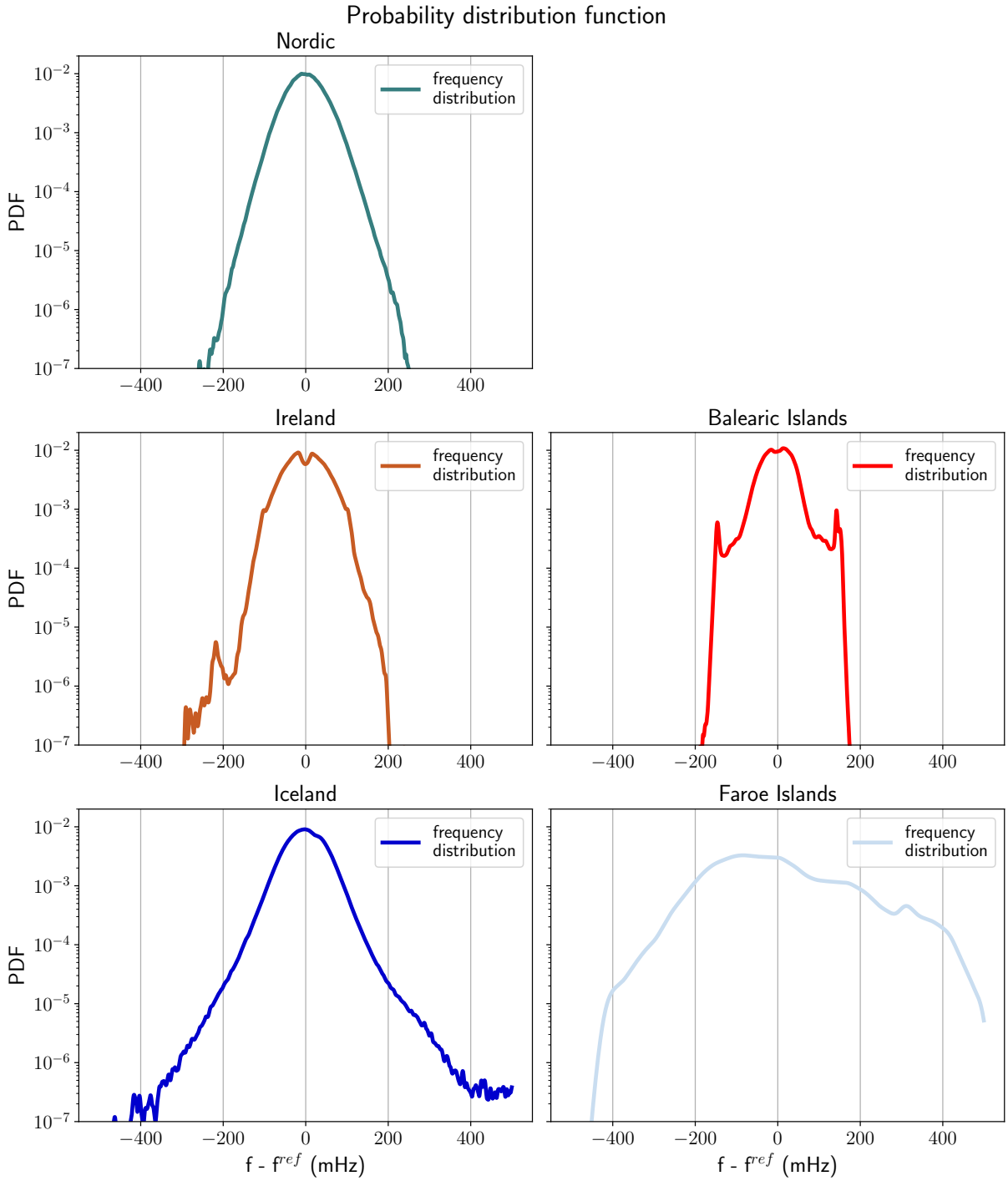


Figure 4.1: *The probability distribution function of the frequency time series. With a logarithmic y-axis, the histograms should show an inverted parabola for Gaussian distributed data. The regions tend to deviate more from Gaussian characteristics as the population decreases.*

When the now cleaned frequency recordings were examined, the probability density functions of the time series exhibited huge variations. The plots in Figure 4.1 show the

frequency distribution for the five regions as the frequency deviation from the reference frequency at 50 Hz. As the largest in terms of population, the Nordic region shows an almost perfect Gaussian distribution (Inverted parabola in Figure 4.2). Ireland also exhibits some Gaussian characteristics but has a dip around 50 Hz and a slight rise in probability for lower frequencies below 49.8 Hz. The Balearic Islands have local peaks when deviation approaches ± 150 mHz, but is then cut off. The two smallest islands, Iceland and the Faroe Islands, have heavier tails and a significantly greater probability for higher frequency deviations.

The frequency deviation tends to increase more as the region's population decreases. In addition, the two regions that deviate most from Gaussian distribution are the two that have no interconnections to other regions. The region that is most dependent on interconnection, the Balearic Islands, shows no greater frequency deviation than 200 mHz. The Balearic Islands are also the region with the least share of renewable energy in its electric power system and are mainly driven by highly controllable generation types and a stable import of power from mainland Spain.

When studying the increment analysis of the frequency time series, the regions exhibited clear differences as for the probability distribution functions. In Figure 4.2, the frequency increments divided by the SD of the given lag are plotted together with a Gaussian distribution. The three plots are shifted in the y-direction for visibility. Nordic is the only region similar to a Gaussian distribution, with a higher probability for greater frequency increments for the 1s lag than 10s. The remaining four regions show non-Gaussian characteristics for both lags, with slightly heavier tails for the 1s increment than 10s. Especially Iceland and the Faroe Islands deviate the most from the Gaussian curve, like the findings in Figure 4.1.

Ireland, which had a similar frequency distribution as Nordic in Figure 4.1, is now exhibiting heavier tails for both lags. Wind power as a steadily increasing share of the Irish power system could describe the difference from the more controllable hydro and nuclear-driven power system in the Nordic region. Through its wind power, Ireland is one of the synchronous areas in Europe with the lowest level of inertia [29]. For the Faroe Islands, the 1s lag distribution has significantly heavier tails than for 10s, which is cut off to Gaussian characteristic for greater frequency increments. The low population in the two smallest regions make the power systems more vulnerable to outages of large individual power plants or industry due to the small system size [60].

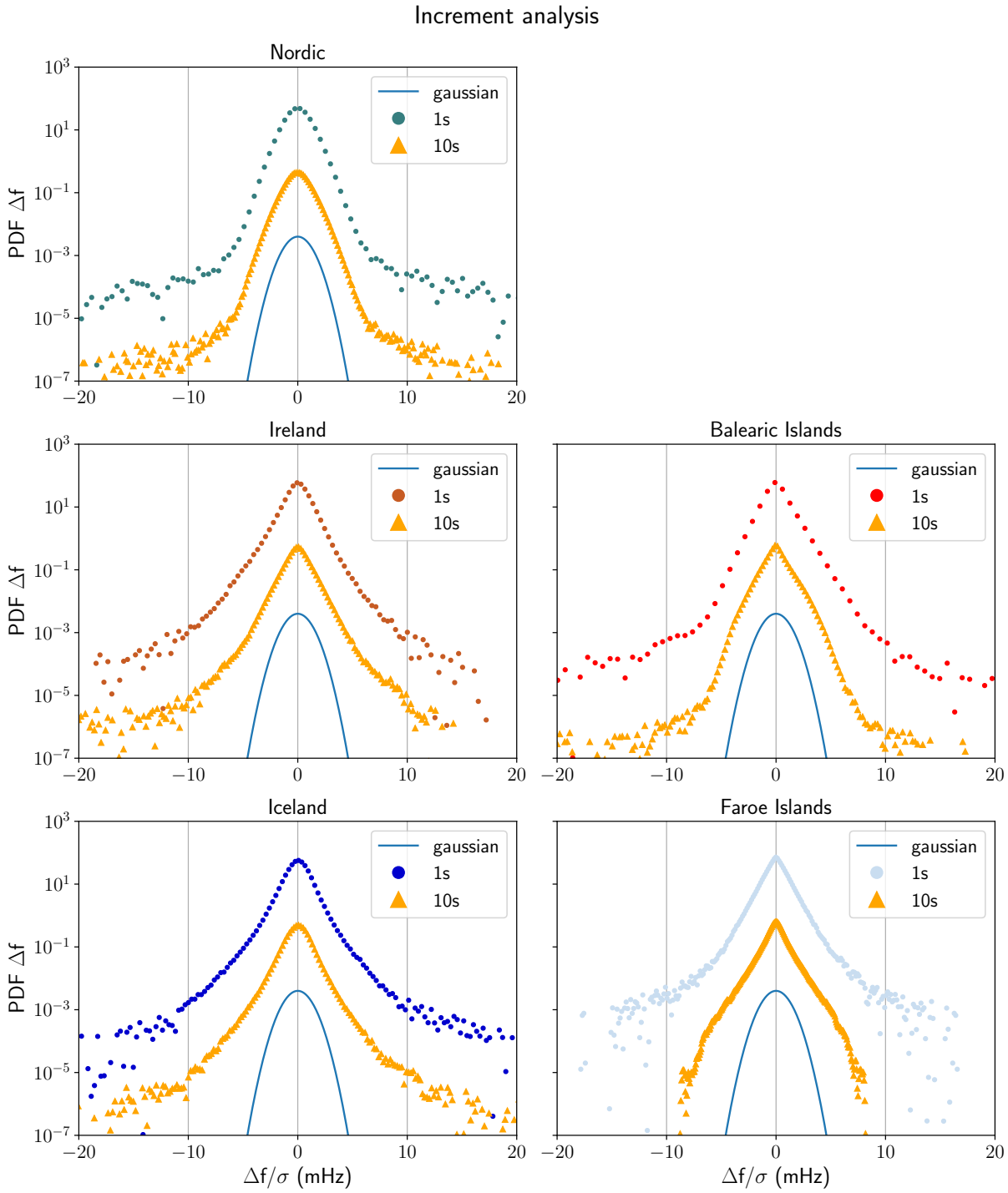


Figure 4.2: *The frequency increments are divided by the standard deviation for 1s and 10s lags. Together with the increment plots is a Gaussian distribution plotted as a comparison. All regions exhibit more Gaussian characteristics for the 10s lag than 1s. For visibility, the curves are shifted in the y-direction. Inspired by* [53].

4.1.3 Comparison with TSOs' Frequency Targets

The frequency time series also made it possible to compare the frequency distribution to the internal frequency targets set by some of the TSOs, given in Section 2.2. Iceland

showed heavy tails in Figures 4.1 and 4.2, but less probability for substantial frequency deviations for a 10s lag. As Landsnet investigate the frequency by averaging over a 10s period with a 2s time resolution to check the frequency quality; this will positively affect staying within the wanted range. For the 86 days used in this analysis, the frequency stays within the range 50 ± 0.2 Hz 99.88% of the time and thus achieves its annually goal of 99.50%. To compare the Faroe Islands' frequency deviations with the Icelandic target, the frequency stays within the same range 85.83% of the time.

A frequency quality report from Fingrid in 2021 showed that the Nordic region's goal of less than 10 000 min outside of the standard frequency range was reached in 2020 for the first time in the period 2015-2020 [31]. Here, the calculation was based on the valid data from each year, with a sample interval of 1s. When performing the same calculations on the frequency time series from 2021, The frequency deviates from the standard range for 10 875 min, thus not reaching the target. Nevertheless, when comparing this result with the data from the five years from [31], it is the second-best year after 2020. These frequency distribution calculations also strengthen the credibility of the measurement data as they are fairly consistent with the TSOs' calculations.

Ireland's target of staying within its 50 ± 0.1 Hz limit 94% of the time was achieved in 2020, reaching 99.67% [37]. For this time series, which involves months in 2021 and 2022, the frequency stays within the range 97.29% of the time when the original 1s time resolution is used. The time resolution for the basis of calculation is not stated in [37] and could be some of the reasons for the difference in percentage between the published number from 2020. Moreover, the high percentage of invalid measurements in the Irish time series must again be pointed out as a possible source of error. The Balearic Islands only fulfilled the deviation target for CE 84.51% of the time. Nevertheless, as seen in Figure 4.1, the frequency never exceeds 50 ± 0.2 Hz, which ENTSO-E states as "the target for maximum steady-state frequency deviation after dimensioning incident" for CE [41].

4.2 Frequency Patterns

4.2.1 Autocorrelation Function

The autocorrelation function with a time lag for up to several days can highlight whether there are recurrent daily patterns in the frequency. In Figure 4.3, the different regions vary significantly. The Balearic Islands show clear peaks every 24 h, in addition to less pronounced yet significant peaks every 12 h. 24 h peaks also apply to the Nordic and to some extent Ireland. Here, the peaks are somewhat broader around the full hour. Lastly, Iceland stands out as very stochastic, with an autocorrelation function close to 0 for all lags. This suggests that strong daily patterns and recurrences will be challenging to find for Iceland and, on the other hand, very likely for the Balearic Islands. Nevertheless, minor peaks occur every 24 h for Iceland, thus affirming that there, to some extent, are daily actions influencing the frequency here as well.

The amount of data available after pre-processing and utilized for the analysis are additionally to Table 4.1 included in Figure 4.3-4.7 as a reminder of the different basis for each region.

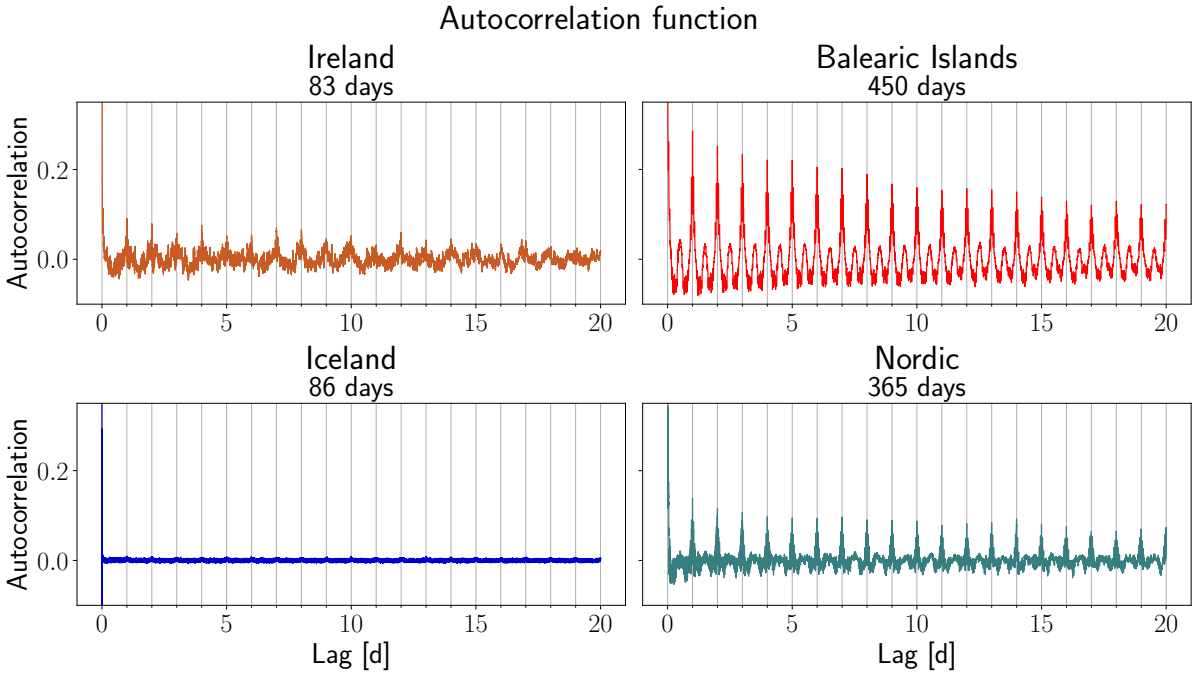


Figure 4.3: *The autocorrelation function with time lags up to 20 days reveals recurrent patterns occurring every 24 h for the Balearic Islands and Nordic and some less pronounced ones for Ireland. As a difference, Iceland shows no signs of correlation with a flat function around 0 for all lags.*

Figure 4.4 shows an autocorrelation function with a lag up to 100 min to examine correlation for shorter lags. The Faroe Islands have a relatively strong correlation for all lags, but with only about six days of data, it is limited whether this applies to a longer frequency time series. Nevertheless, it is clear that there is a correlation for different lags, and recurrent frequency patterns do occur. Again, Iceland stands out as very stochastic, with close to no correlation for any lags. The other regions exhibit a peak around the 60 min lag, suggesting that there are hourly trends in addition to the daily recurrent patterns. For the Nordic region, the peak is seen after exactly 60 min, immediately followed by a dip. In contrast, the peak is slightly bigger and broader around the full hour for Ireland and the Balearic Islands.

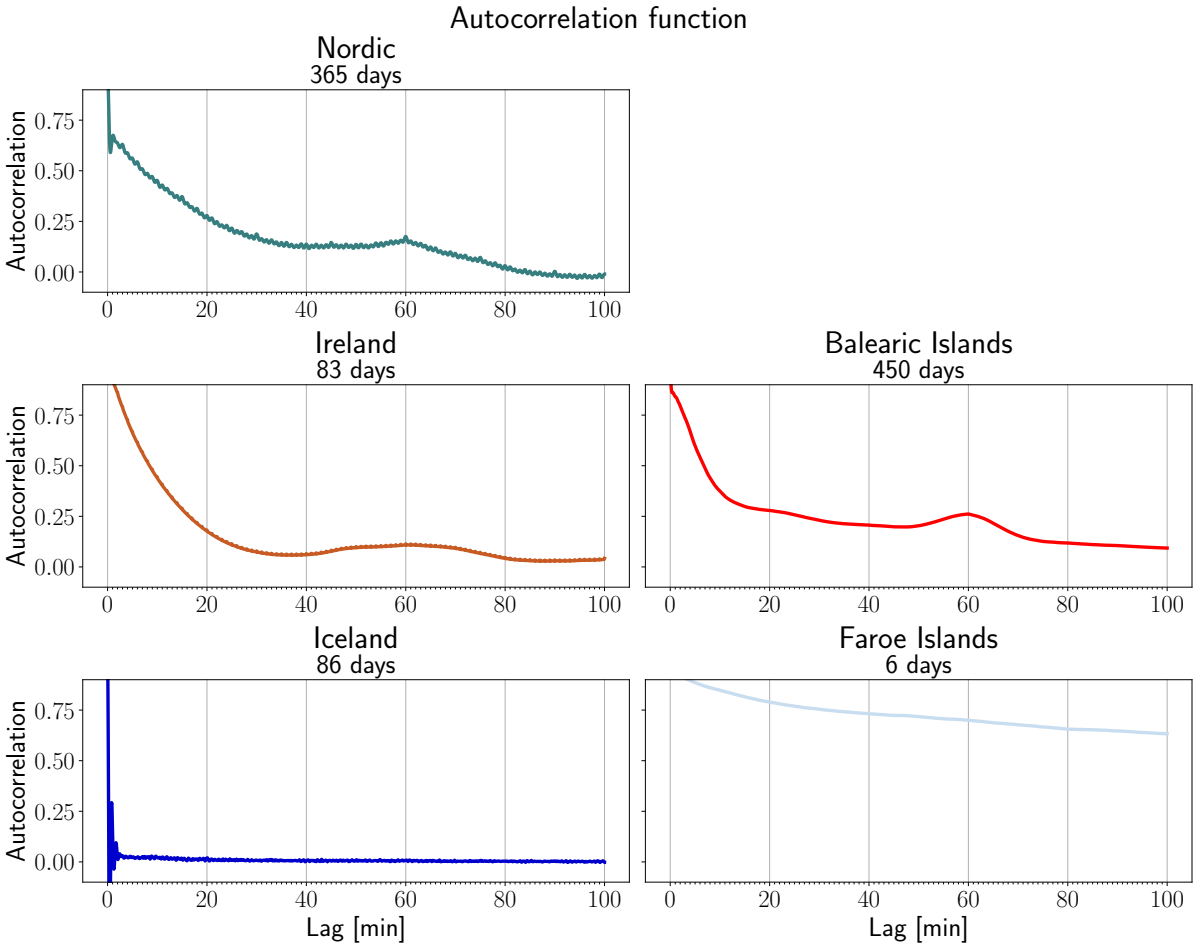


Figure 4.4: *With a shorter time lag, the autocorrelation function reveals hourly frequency correlation. With significant peaks after 60 min for Ireland, Nordic and the Balearic Islands, Iceland does again show stochastic characteristics with no correlation.*

4.2.2 Daily Profile

The daily profile, which from Section 3.1.3 denotes the average of all frequency recordings for each specific time of the day, is given for each region in Figure 4.5. In Figure 4.5-4.7, the Faroe Islands is plotted alone due to its short frequency time series, resulting in greater deviations and fluctuation on the y-axis. The Nordic region thus shares the y-axis with the Icelandic region.

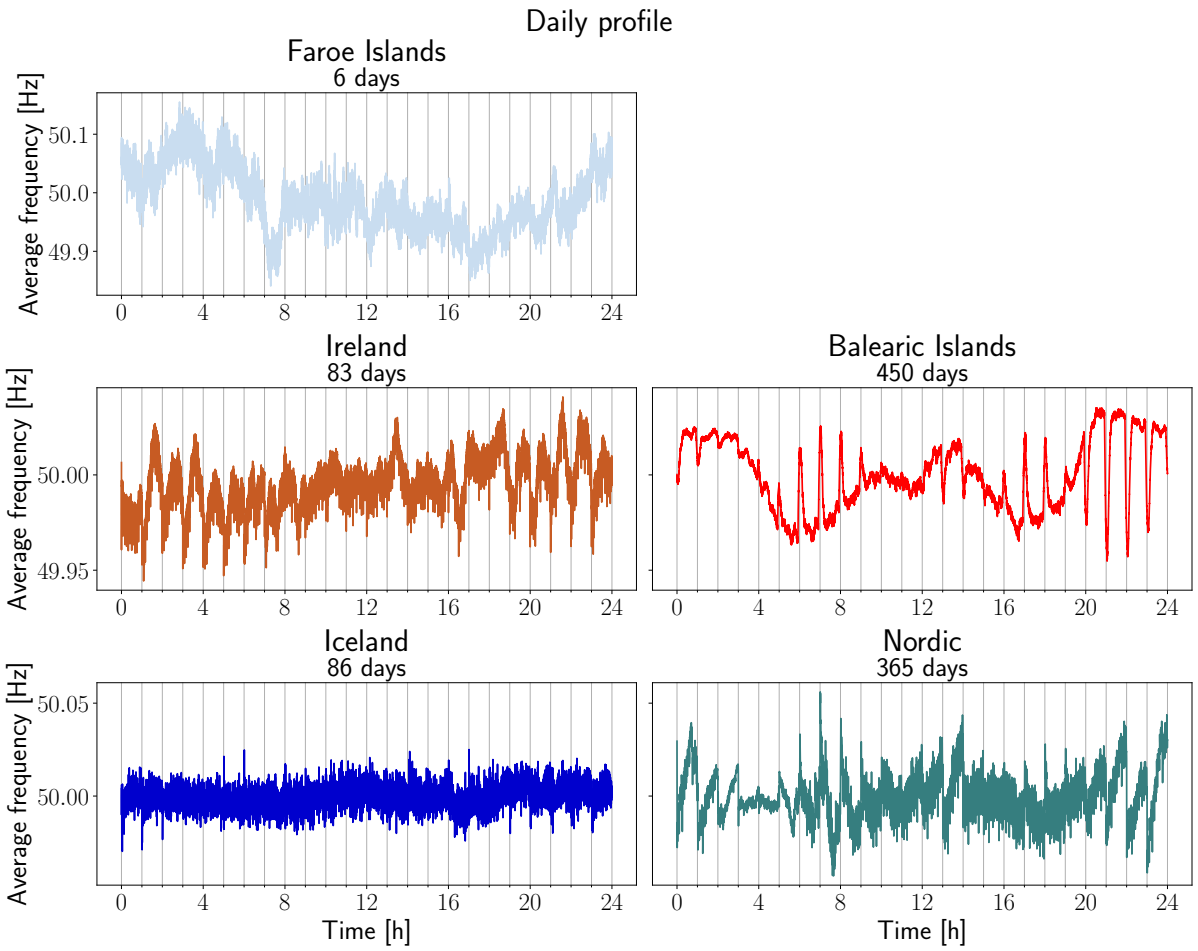


Figure 4.5: *The average of all recordings for every second of the day, denoted as the daily profile. While the Balearic Islands exhibit a clear daily pattern, the Faroe Islands and Iceland frequency profiles are the most stochastic. Nordic and Ireland are somewhat in between, with some clear peaks and valleys at the beginning of hours.*

The daily profile of the Balearic Islands is by far the most apparent pattern. Here, the frequency is generally higher during evening and night, with local valleys at the beginning of each hour. The frequency is slightly lower during morning and afternoon but now has local peaks at the beginning of the hour. The Nordic and Irish regions also exhibit some recurrent patterns through local peaks or valleys at the beginning of each hour. The Faroe Islands also show frequency valleys during morning and afternoon, in addition to some peaks at the full hours. As this time series only consists of six full days of data, the daily profile averages of only six or seven values. This results in a fluctuating profile which may not necessarily represent how a possible profile evolves for longer periods. Iceland stands out as very stochastic in relation to the remaining regions. At the beginning of some hours, there are some peaks and valleys, but without any daily pattern or tendency. Nevertheless, it is notable that the Icelandic daily profile varies in a smaller range than for the other regions.

4.2.3 Daily Standard Deviation

The next pattern analysis was the daily SD, reflecting the variability within the frequency samples of the same time of the day. The plots are shown for all regions in Figure 4.6. For the Balearic Islands and Nordic, the SD has clear peaks at every full hour, also showing the lowest variability in the middle of each hour. The peaks support hourly trends in the frequency behavior also pointed out through the daily profile. Similar applies to Ireland, but with less apparent peaks as these occur in the middle of the hours as well. Iceland is again the most stochastic pattern, with random peaks throughout the day and no specific tendency to occur. As the Faroe Islands' time series is significantly shorter than the others, the SD is not surprisingly by far the greatest. Nevertheless, the Faroe Islands also exhibit some peaks around the full hours.

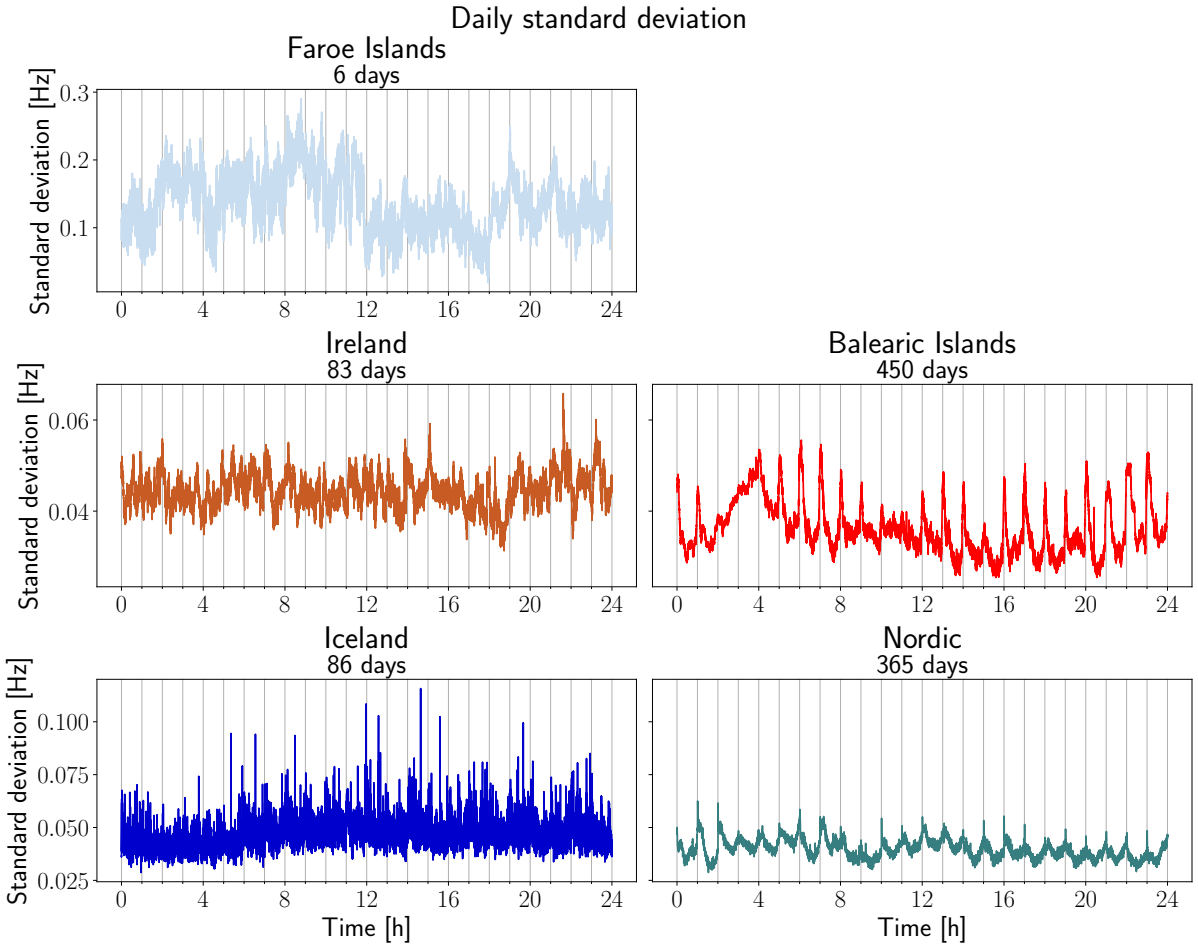


Figure 4.6: *The daily standard deviation reflects the variability within the frequency samples at each second of the day. The standard deviation has clear peaks at every full hour for the Balearic Islands and Nordic. The same applies to Ireland, but with less apparent peaks as these also occur in the middle of the hours. Iceland is again the most stochastic pattern, with random peaks throughout the day.*

4.2.4 Hourly Standard Deviation

Some of the hourly trends seen in Figure 4.6 could be examined further when studying the hourly SD. In Figure 4.7, the hourly SD for all regions are presented. The full hour peaks are again seen for the Balearic Islands and the Nordic, where the SD peaks after a few minutes. An early hourly peak for Iceland, which was difficult to perceive in Figure 4.6, is also revealed. Again, the frequency from the Faroe Islands has the greatest variability. Ireland exhibit a more evenly distributed variability, but smaller peaks at the boundaries and after 30 min could be somewhat related to the 30 min window of the Irish intraday electricity market [61].

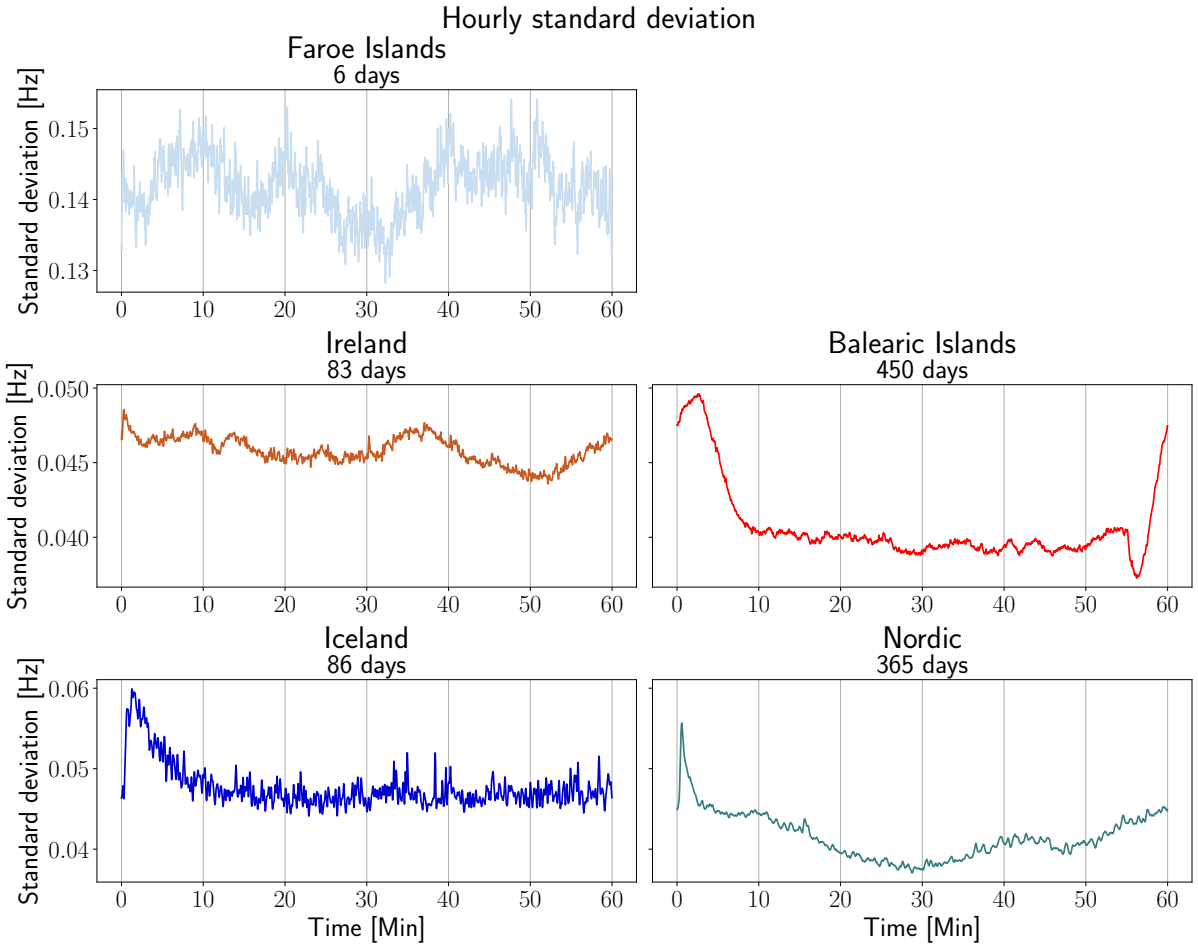


Figure 4.7: *The hourly standard deviation reflects the variability within the frequency samples at each second of the hour. the Balearic Islands and Nordic have their greatest variability at the beginning and end of the hour, while Iceland peaks after the first minutes. The same applies to Ireland, but with less apparent peaks as these occur in the middle of the hour as well.*

4.2.5 Contributors to Repetitive Patterns

Periodically recurrent patterns and correlation can be explained by several power system related factors. Firstly, the electricity load usually shows a daily cycle with peaks during mornings and afternoons and the least electricity during the night [62]. This in-

troduce regular daily dynamics. There are higher stochastic contributions for the regions with high SD, like the Faroe Islands and Iceland. In contrast, the other regions' dynamics are governed by mostly deterministic effects exhibited by clear daily profiles and lower SD.

Secondly, electricity generation is being scheduled through different market mechanisms. The day-ahead market takes place at noon every day to determine the electricity price and volumes within specific bidding zones for every hour of the following day [63]. Most countries operate with one-hour windows, but 30 and 15 min are also used. These market activities introduce regular uncertainty seen as peaks in the SD. Additionally, the intraday market exists, among other things, to perform corrective actions on the possible deviation between the day-ahead market clearing and actual operation. The balance from the day-ahead market is based on forecasts, and a change in weather may cause less generation or more consumption, or an industrial company might change its expected consumption. The Intraday market takes place the same day and usually closes an hour before the electricity delivery hour. There is no electricity market for the Balearic islands, and REE is responsible for scheduling an economical and technical dispatch of generation for one-hour intervals [39].

Lastly, the balancing markets for frequency reserves from Section 2.1.4, controlled by the TSOs. These markets ensure the final system balance and work down on the minute scale close to the electricity delivery hour. The different markets contribute to the balance of the power system, and, e.g. the change in electricity price and generation volumes between each hour traded on the day-ahead market will potentially increase the variability for the first period of these hour-intervals. This causes generation to follow more periodic patterns, while consumption continuously fluctuates based on the demand in industry and households. These potential imbalances, particularly at the trading boundaries, can describe the variability peaks seen for several regions in the previous figures.

4.3 Predictions

The time series were manually split into a train, validation, and test set for the full frequency predictions. Due to the amount of NaNs in some of the time series, the different series were split such that the train set represented roughly 70% of the time series, and the validation and test set represented roughly 15% each. All sets start at the beginning of an hour and end at the end of an hour. The percentage size of each set and the specific start and end times are shown in Table 4.3 for all regions. The following subsections first present the best and worst predictions before presenting overall performances compared to the null models to better understand the accuracy of the WNN predictor.

Table 4.2: *The time series are split into train, validation, and test sets for the frequency predictions. The table displays the different timestamps for the start and end of each set, in addition to what percentage the set makes up of the full time series*

Time information about prediction subsets				
Region	**Subset**	**Start time**	**End time**	**Percentage**
Nordic	Train	2021-01-01 00:00:00	2021-08-31 23:59:59	66.6%
	Validation	2021-09-01 00:00:00	2021-10-31 23:59:59	16.7%
	Test	2021-11-01 00:00:00	2021-12-31 23:59:59	16.7%
Ireland	Train	2021-11-04 17:00:00	2022-01-18 23:59:59	68.3%
	Validation	2022-01-19 00:00:00	2022-02-05 23:59:59	16.3%
	Test	2022-02-06 00:00:00	2022-02-22 23:59:59	15.4%
Balearic Islands	Train	2019-09-29 00:00:00	2020-09-30 23:59:59	69.6%
	Validation	2020-10-01 00:00:00	2020-12-31 23:59:59	17.4%
	Test	2021-01-01 00:00:00	2021-03-10 23:59:59	13.0%
Iceland	Train	2021-11-05 09:00:00	2022-01-02 23:59:59	67.6%
	Validation	2022-01-03 00:00:00	2022-01-16 23:59:59	16.2%
	Test	2022-01-17 00:00:00	2022-01-30 23:59:59	16.2%
Faroe Islands	Train	2019-11-03 22:00:00	2019-11-08 08:59:59	69.0%
	Validation	2019-11-08 09:00:00	2019-11-09 08:59:59	15.5%
	Test	2019-11-09 09:00:00	2019-11-10 08:59:59	15.5%

4.3.1 Best Predictions

In Figure 4.8, the best predicted hour from the test set is shown for all regions. The figure presents the predicted frequency trajectory, as well as the daily profile and the test (actual) trajectory. Additionally, the prediction error is shown given as one SD within the maximum number of nearest neighbors. The best predictions exhibited considerable variation in terms of accuracy between the regions.

The Nordic region's best prediction is a smoothed version of the test series where the hour's average trajectory is captured, but not minutes with local peaks and valleys. To some extent, the prediction for Ireland also acts as a smoother trajectory of the test series and follows it better than the daily profile. The Balearic Islands have the best performing prediction: The WNN predictor followed the test series accurately, both at the boundaries where the frequency exhibits a small drop in the first minutes and a steady increase in the last five minutes, and the middle of the hour. The shaded prediction uncertainty ($\pm$ one SD) is therefore, not unexpected, the smallest for the Balearic Islands with about 27 mHz on average, while one SD for Ireland and the Nordic lie between 40 and 50 mHz on average throughout the hour.

For the two least populated regions, the predictions are not that accurate. For Iceland, both the WNN predictor and the daily profile are based around 50 Hz as a noisy

50 Hz predictor. They do not capture what appear to be large stochastic fluctuations spread across the hour but reflect well an approximate average of the trajectory of about 50 Hz. Thus, the prediction is similar to Ireland and the Nordic predictions in terms of uncertainty. For an average of the whole hour, the WNN prediction for the Faroe Islands is good but does not capture the main several minutes fluctuations after 15 and 30 min. The daily profile also misses these frequency jumps. The prediction uncertainty is also significant as it is greater than 100 mHz on average.

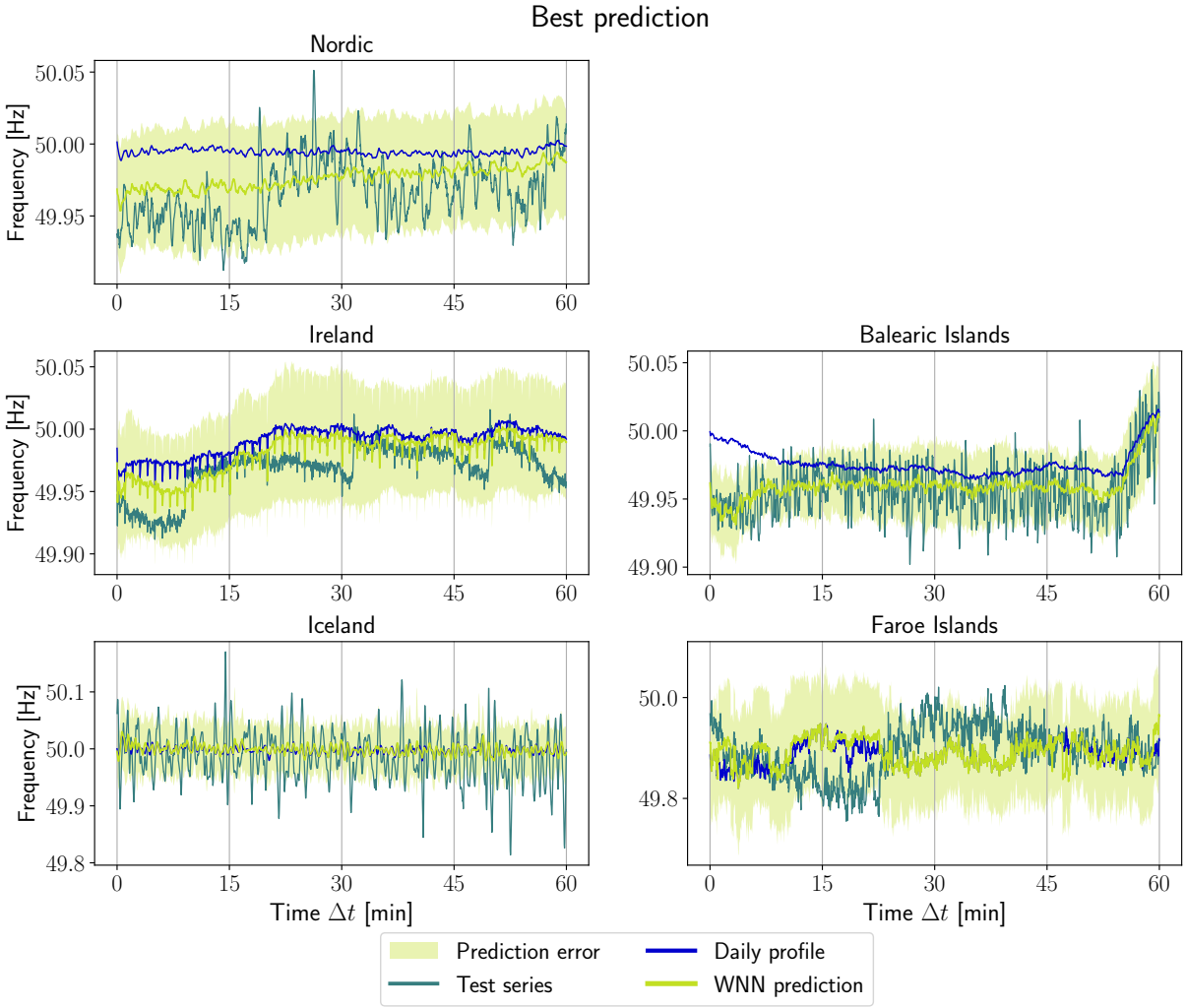


Figure 4.8: *To better understand how well the WNN predictor can perform, the best performing hour for each region is presented. Its prediction is compared to the daily profile and the test series. The prediction error is given as one SD within the maximum number of nearest neighbors in the adaptive-k. The chosen trajectories are based on the relative error RMSE(f_{wnn})/RMSE(f_{50}).*

The predicted and daily profile frequency for Ireland exhibited systematic valleys in the trajectory in Figure 4.8. These occur every minute and can also be seen less pronounced for the test series. Whether this reflects a specific behavior in the Irish power grid, a pre-processing or measurement error from the raw data from the EDR, or some other circumstances is unknown and should be investigated further. This pattern is also present in the trajectories for the worst prediction in Section 4.3.2.

4.3.2 Worst Predictions

For the worst predicted hours, the WNN predictor misses the test series differently for each region. Both models expected a frequency increase for Nordic, the Balearic Islands and the Faroe Islands, where the WNN predictor assumed the highest frequency. In reality, the frequency instead dropped a bit. For Ireland, the opposite happened. Here, the frequency was steady above 50 Hz the whole hour, but the predictors predicted some fluctuations, mainly below 50 Hz. Nevertheless, the WNN followed the daily profile also here, even with a more accurate prediction in the first minutes. Again, Iceland looks similar to its best prediction exhibited as a noisy 50 Hz signal, with stochastic peaks not captured by any predictor throughout the hour. Furthermore, on average, the prediction uncertainty is approximately identical to the best predictions' uncertainty for all regions.

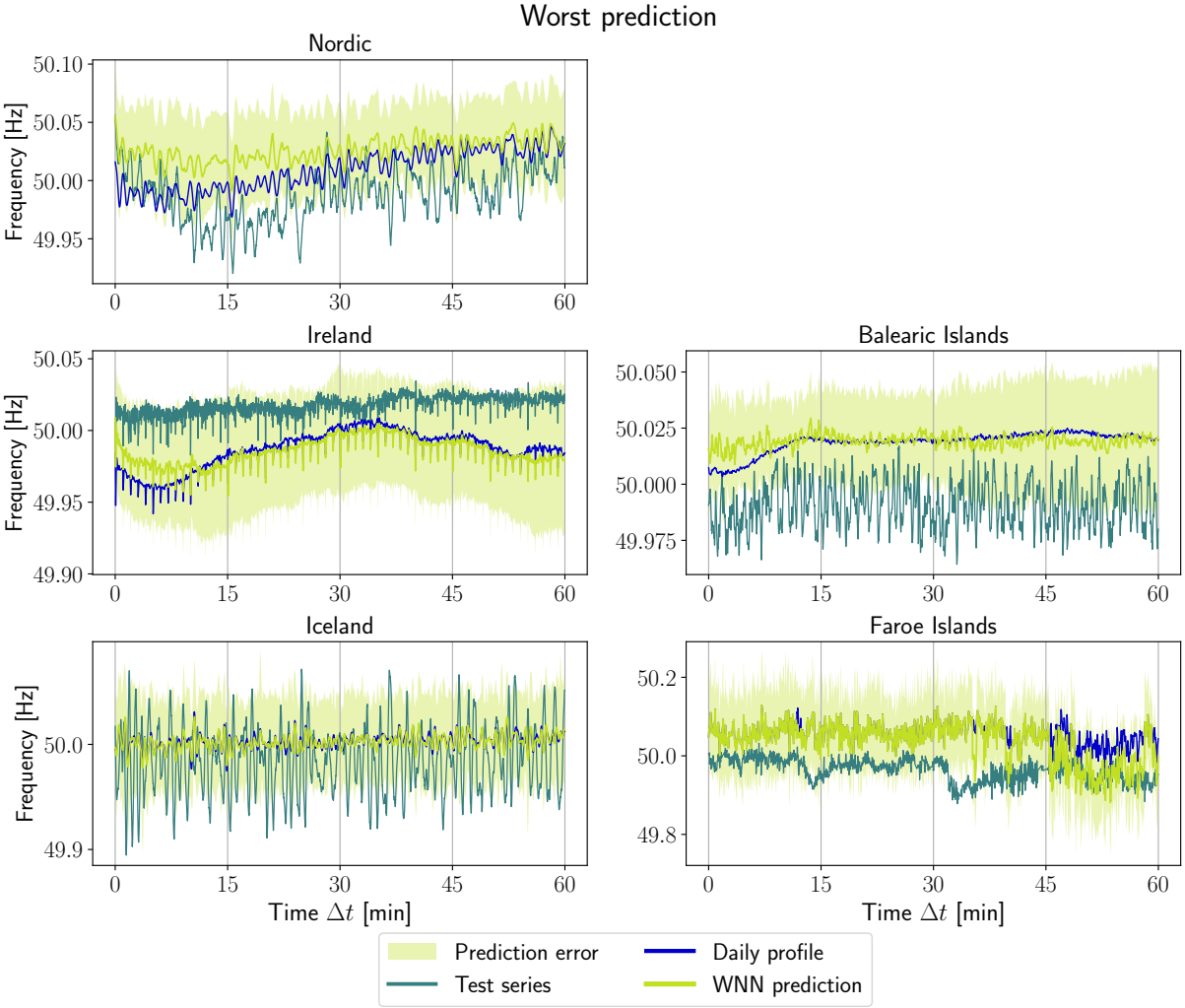


Figure 4.9: *In contrast to the best predicted hours from Figure 4.8, the worst predictions are presented here. The basis for selection as well as the prediction error is the same as for the best predictions.*

4.3.3 Overall Performance

The performance of the WNN predictor was compared to the performance of the two null models from Section 3.2. In Figure 4.10, the root mean square error (RMSE) of all models are presented for each region, reflecting the predicted error given in Hz. In all regions but Iceland, the WNN predictor outperformed the daily profile and the 50 Hz model. In general, the null models are clearly outperformed the first 15 minutes, but all three models later converge towards each other as the error also tends to increase.

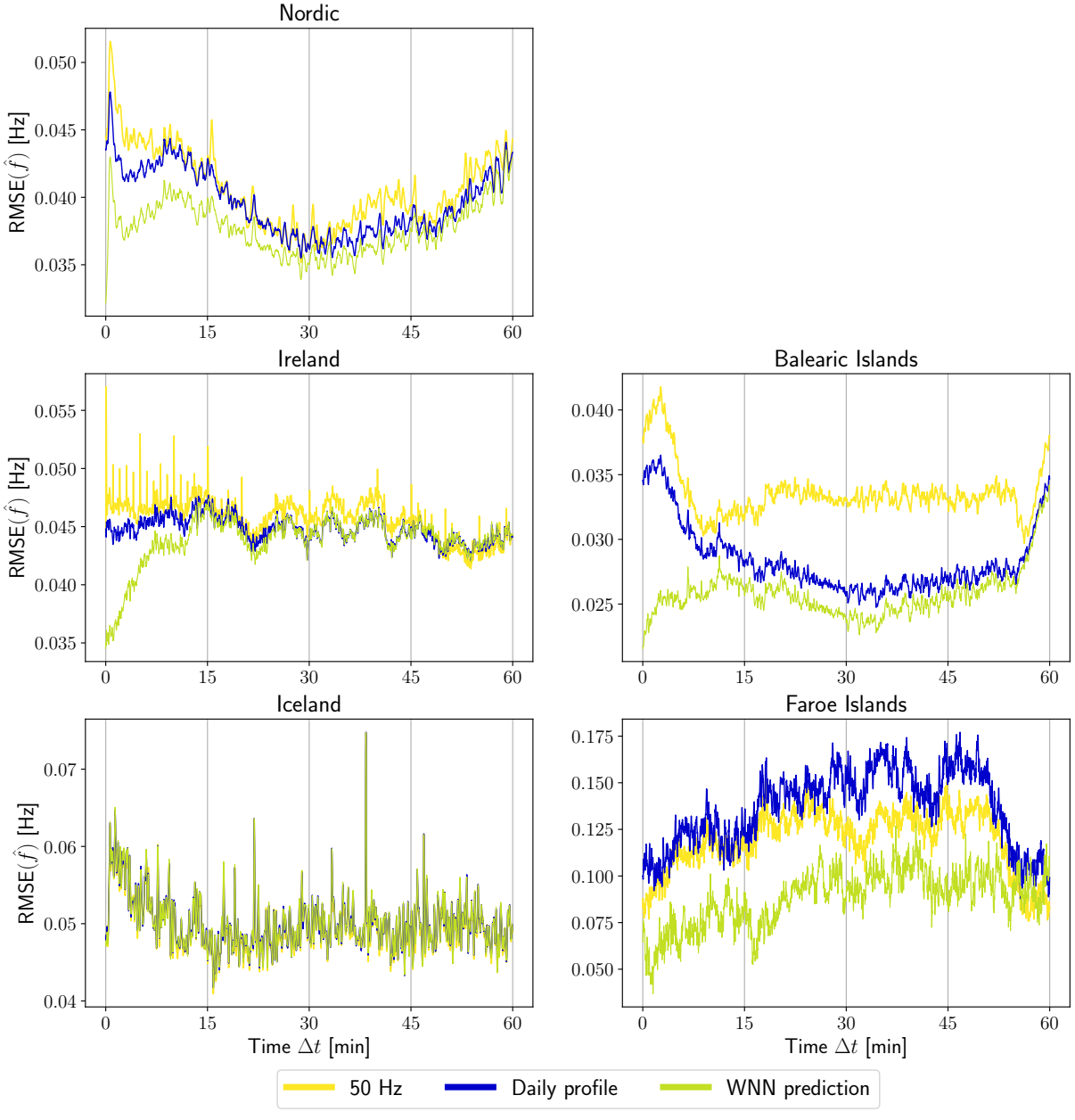


Figure 4.10: *Root Mean Square Error of the WNN prediction model and the two null models. In all regions except Iceland, the WNN predictor outperforms the null models, especially the first 15 minutes of the hour. the Balearic Islands have the lowest error, while the Faroe Islands have the highest. The y-axis varies between the subplots due to the varying error between the regions.*

All regions' performance curves are for large parts of the hour similar to the hourly SD from Figure 4.7. The highest variability, mainly at the beginning of the hour, is reflected in the RMSE peaks for the performance. However, for Ireland and the Balearic Islands, the WNN predictor predicts with low error also here as specific frequency trajectories provide accurate information about the first minutes in the future. This suggests that different trading and dispatch patterns are regular, and thus information is already present in historical data. For the Nordic, the trading information is less known and more stochastic, while for Iceland and the Faroe Islands, stochastic processes are dominant.

Iceland is the major exception in predictability, where the frequency's clear stochastic behavior, exhibited by the previous results, makes the prediction error similar for all models. Corresponding with the increased early-hour variability, the performance of all three models has a higher error for this period. The Faroe Islands has the greatest prediction error for all models. Here, however, the WNN prediction error performs 30-40% better than the daily profile for most of the hour, which is a significant improvement. Considering these predictions are based on all hours of the day, the WNN prediction has an advantage over the daily profile. It is here only based on the six historical trajectories, i.e. an hour for each of the six days of data.

Even though the study of the Faroe Islands frequency time series is very limited due to the short time range, it provides a valuable understanding of smaller regions. With about 50 000 inhabitants, it is a typical population size of other isolated islands, cities or areas that potentially can be or are operated as microgrids. Furthermore, as the WNN predictor performed significantly better than the daily profile, it is clear that the frequency exhibit deterministic behavior with patterns and recurrent trends. Thus, the model is also highly relevant for analysing microgrids or other smaller power systems in the future.

4.3.4 Optimal Number of Nearest Neighbors

The optimized *adaptive-k* reveals how the number of historical frequency patterns needed to achieve the best predictions changes throughout the hour. In [9], The *adaptive-k* for the synchronous areas in Great Britain, CE, and the Nordic increased throughout the hour. In Figure 4.11, the *adaptive-k* is plotted relative to K, which is the upper limit for k as it represents the number of patterns in the daily profile. The number of neighbors tends to increase throughout the hour for all regions. At the beginning of hours, the historical frequency trajectories provide more information than later, and there is thus a need for fewer patterns in the prediction. Throughout the hour, the frequency patterns become more unspecific as the prediction is further into the future, and k increases. The Faroe Islands are not included as all hours of the day are potential nearest neighbors.

Iceland exhibits the least apparent increase but still has slightly lower numbers in the first 15 minutes. For Iceland and Ireland, the optimal k is many times at maximum, meaning close to all available patterns were used during the weighted average prediction. Hence, it is reasonable to assume that by having more frequency measurements for a longer time frame, the *adaptive-k* might increase even more, especially towards the end of the hour. Eventually, however, with an increased amount of frequency data, it will get

to a point where specific frequency patterns are preferred over an average of all. Thus k moves away from K, seen for the Balearic Islands and the Nordic region.

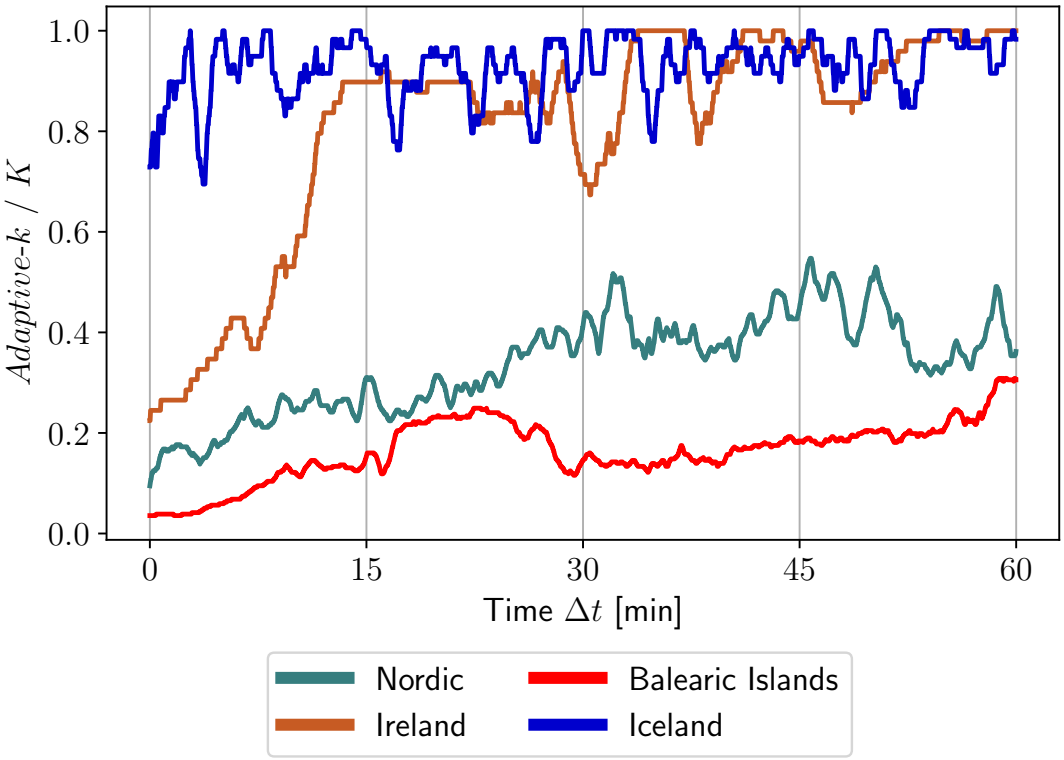


Figure 4.11: *The optimal number of nearest neighbor relative to K is the number of neighbors in the daily profile, and thus the upper bound for k. The adaptive-k tends to increase throughout the hour for all regions. The Faroe Islands are left out as the adaptive-k is based on all hours of the day as possible neighbors.*

4.4 Size Dependence

4.4.1 Overall Performance

For small power systems and regions, obtaining high-quality data or longer time series of measurements might involve high costs. Hence, it is interesting to investigate how much data is necessary to obtain reliable predictions. The change of performance associated with different amounts of data is valuable information for assessing how much data can or should be used for accurate predictions, depending on the data basis. In addition, the size dependence approach also lays a basis for comparing the prediction performance of each region with equal lengths of frequency time series. From 25 random frequency time series samples for each time interval p, each region's average performance is presented in Figure 4.12. This applies to both the two null models and the WNN prediction. The 50 Hz prediction error is the same for all p as it does not depend on historical data. The Faroe Islands were not included in this analysis as the shortest interval to investigate of one week is longer than the length of the frequency time series recorded in the region.

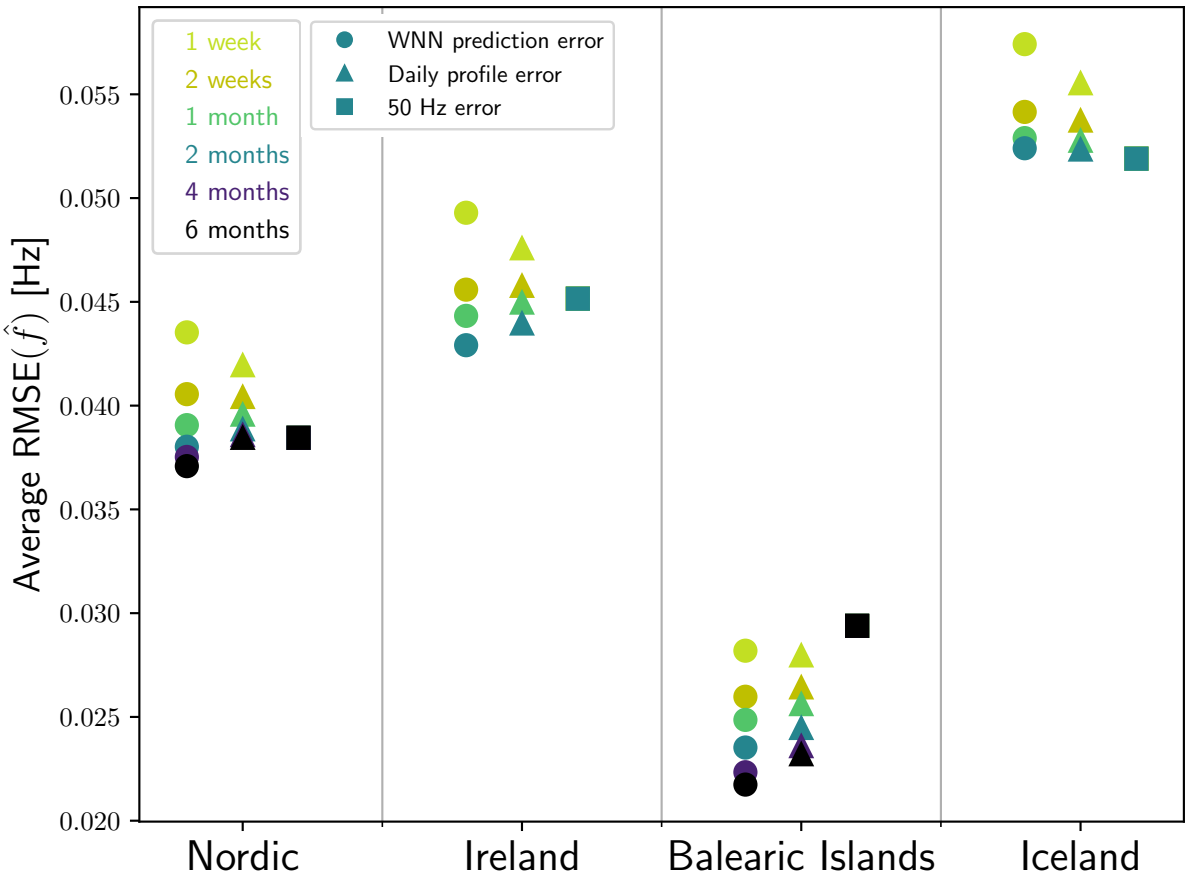


Figure 4.12: *Average root mean square error for the WNN predictor and the two null models for different time intervals. As the data size increases, the WNN predictor is the best model for all regions except Iceland, where the constant 50 Hz prediction performs best. The Faroe Islands are not included in the analysis as the region's frequency time series is shorter than the minimum interval of one week.*

The WNN predictor improved performance for all regions when the time interval increased. Except for Iceland, the WNN predictor mainly outperformed the daily profile and 50 Hz predictor when two weeks to a month of data were utilized, thus serving as a boundary for when the daily profile is outperformed. Even though Iceland is very stochastic, the improvement for increasing time intervals exhibits that there, to some extent, is specific information to gain from a greater amount of data even here. The average performance with only one week of data for the Balearic Islands is significantly better than all other regions' performances, regardless of the time interval. The daily profile and WNN predictor also outperform the 50 Hz for all intervals. This emphasizes the deterministic and highly predictable frequency trajectories of the Balearic Islands. The daily profile also performs well, implying significant recurrent patterns and trends.

With limited data available for the regions, only one week of the frequency time series was used as the test set. One week period for testing may be too short to yield conclusive results as whether this week experienced unexpected and rare events or adequately represent a typical week is unknown. This thesis has not examined whether predictability might change related to seasons due to, e.g. different electricity generation mixes. Thus, the

test week's predictability may depend on when in the year it is chosen. One week was also found to be too short for selecting specific information, as the daily profile outperformed the WNN predictor for this time interval. Additionally, the performances for Ireland and Iceland for $p = 2$ months are an average of 25 very overlapping $\mathcal{S}_n$, as these cleaned time series had a length of only between two and three months.

4.4.2 Performance Improvement

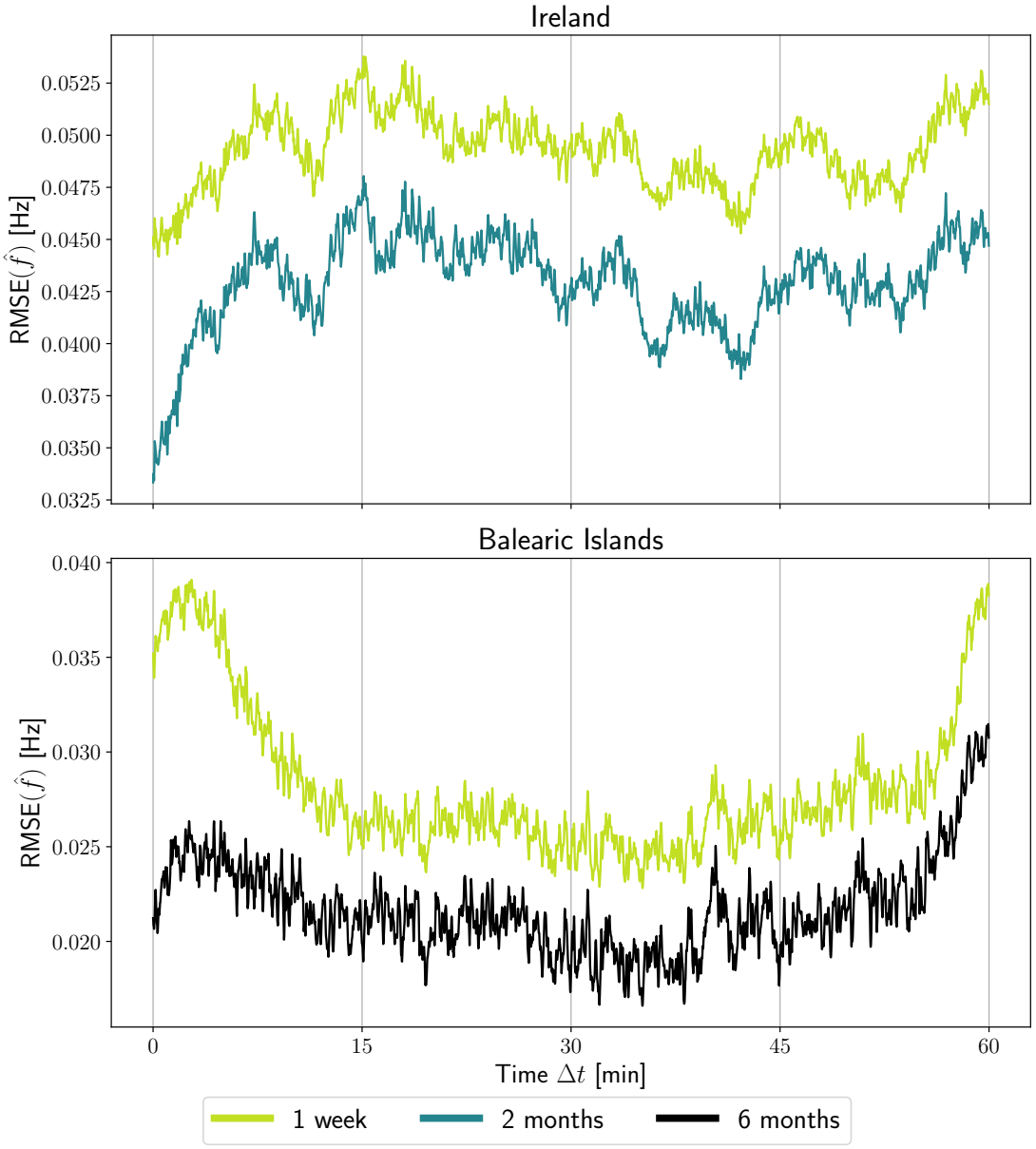


Figure 4.13: *The first minutes exhibit the best improvement from the shortest to longest time interval throughout the full hour. The RMSE of the Balearic Islands and Ireland is here shown for one week as the shortest interval, and two months (Ireland) and six months (Balearic Islands) as the longest.*

The average RMSE has proven to be reduced when the model is trained on more data, and especially the first 5 min were significantly improved compared to the null models in Figure 4.10. To further investigate how much the amount of historical data improves the predictions for distinct periods of the hour, the shortest and longest prediction interval errors were plotted for two of the regions in Figure 4.13. Here, Ireland and the Balearic Islands are chosen as they both showed great improvement in the first 5 min of the hour in Figure 4.10 and represent different time series lengths. From one week to two months for Ireland, the greatest error reduction is only seen for the mentioned first 5 min with an improvement of 20% on average. The rest of the hour has a relatively constant performance improvement of 14% on average. For the Balearic Islands, this early-hour improvement extends to almost 15 min at 40% improvement on average, until the performance improvement is close to constant at about 26% for the rest of the hour. Hence, using more frequency data gives great returns in better prediction results, especially for the first minutes.

4.5 Extended Model

4.5.1 Additional Features

For the extended WNN approach, different electricity generation sources were tested. Firstly, wind generation in Ireland was tested. This included both the total of onshore and offshore generation and only the onshore wind power alone. Wind power was chosen as it forms a central part of Ireland's electricity generation. Offshore wind constitutes less than 1% of the total wind electricity generation in Ireland. However, with numerous projects planned and a target of 5 GW installed in 2030, offshore wind will be a significant contributor in the near future [64]. Secondly, six different electricity generation sources were tested for the Balearic Islands. This includes all renewable energy sources as one feature, as renewables generally accounted for a small share of the Balearic Islands' electricity production, yet highly relevant to investigate as it is less controllable than other sources. In addition, the total generation was also tested as one feature. The distinct additional features that were tested and their time series resolution are given in Table 4.3. Presented is also the result of the optimization of the new hyperparameter β.

Table 4.3: *The WNN predictor is extended to include additional features. The table includes the specific features tested for the Balearic Islands and Ireland, as well as the time resolution of the time series and the resulting optimized weight β. For all features except all wind generation in Ireland, β was optimized to be 0 when tested on the validation set.*

Additional features			
Region	**Generation/Import type**	**data time resolution [min]**	**Optimal weight β**
Ireland	Onshore wind	30	0
	All wind generation	15	0.4
Balearic Islands	Renewables	10	0
	Diesel engines	10	0
	Gas turbine	10	0
	Coal	10	0
	Combined cycle (gas + steam turbines)	10	0
	HVDC Cable (from mainland Spain)	10	0
	All generation (incl. import/export)	10	0

4.5.2 Relative Performance Ireland

The quality of the generation data from the ENTSO-E transparency platform limited the analysis for the onshore wind generation. The data contained a significant amount of NaNs in the period the frequency measurements were recorded. As the quality of the frequency measurements already reduced the amount of data utilized in the predictions, a new share of data was now omitted. The now shorter time series and basis for prediction may thus have influenced the results.

As the only feature, all wind generation in Ireland got a non-zero optimal weighting of the additional feature during the nearest neighbors selection when tuned on the validation set for the extended WNN predictor. The wind generation data downloaded from the region's TSOs contained no invalid measurements and could be used as a whole for the prevailing period. Figure 4.14 presents the RMSE of the extended model with the Irish wind generation included when applied to the test set relative to the original prediction result for Ireland from Figure 4.10. The result reveals no further improvement of the predictions but rather a worse average performance error of 1-2%. Nevertheless, for a few minutes, e.g. around 45 minutes, the extended model performs up to 2% better.

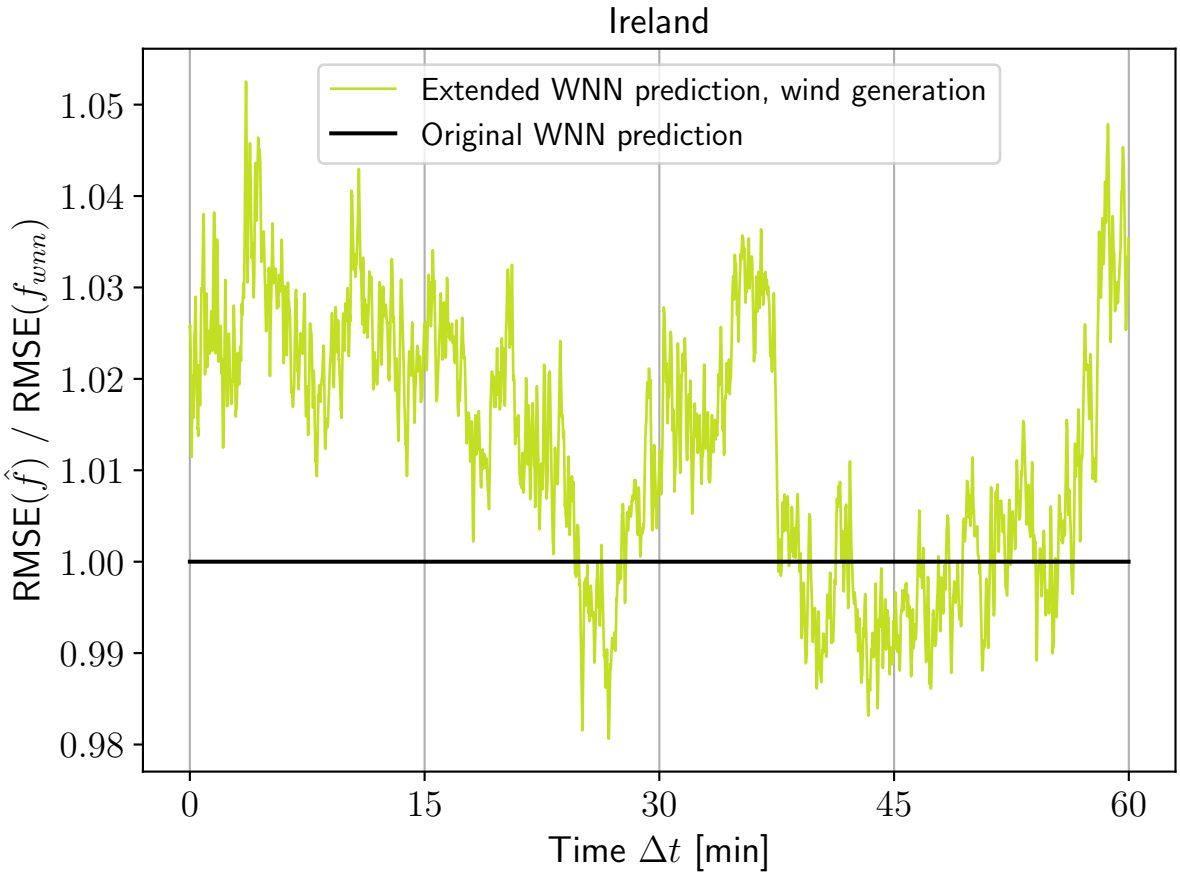


Figure 4.14: *The extended WNN model only got an optimal non-zero weight β for wind generation in Ireland. When evaluating the prediction, the extended predictor performed 1-2% worse on average. Here seen as the performance relative to the original WNN predictor, the extended model still predicts up to 2% better for selected minutes.*

4.5.3 Relative Performance Balearic Islands

When the hyperparameter β was tuned during the optimization on the validation set, 0 was found to be the optimal weighting of the additional feature euclidean distance when selecting nearest neighbors. This implies that the frequency is best predicted when only the frequency trajectories themselves are taken into account, which is thus the performance from Figure 4.10. Therefore, the optimization was performed again, excluding the zero weight to investigate whether a non-zero weight might positively influence the predictions anyway. The new optimal non-zero weights and the associated *adaptive-k* were then used for predictions on the test set. Regarding data quality, the generation data downloaded from REE contained negligible to no invalid measurements and were utilized as a whole.

The result of predictions on the test set with the additional features from Table 4.3 showed that all features performed better than the original WNN predictor for at least one non-zero weight on the test set. Renewables accounted for the best performance improvement of about 1.5%. Also combined cycle and coal improved the predictions by more

than 1%. The performance of the three best features relative to the original WNN prediction error is shown in Figure 4.15. Presented is also the β used to weight the influence of the additional features.

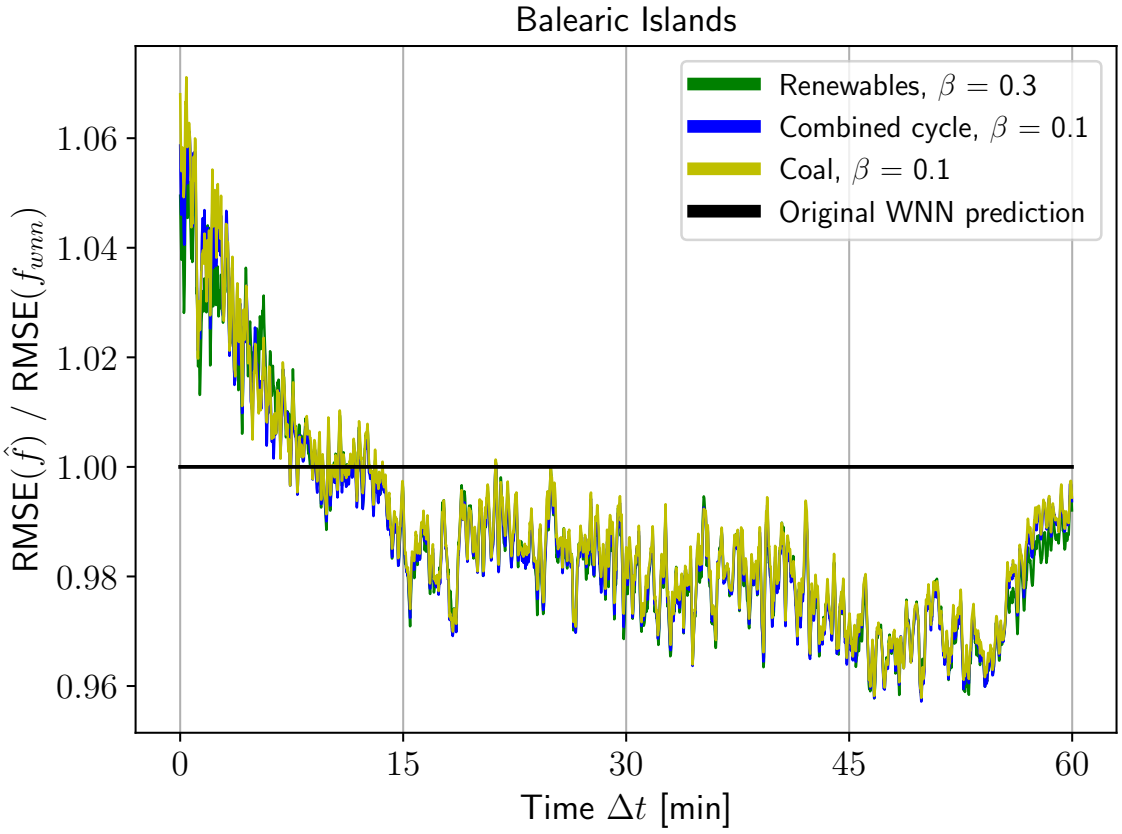


Figure 4.15: *With additional features included in the WNN predictor for the Balearic Islands, the prediction performance was increased. Presented are the three best performing features, all improving the performance by an average of 1-1.5%. Across the full hour, the first 15 min got worse, but the last 45 min significantly improved.*

Applicable to all the Balearic features was that they, on average, performed clearly worse in the first 15 min, as seen by three of the features in Figure 4.15. For the rest of the hour, the extended model now improves the performance by more than 2% on average. From the previous results from [9] and from Figure 4.11, fewer frequency trajectories are typically needed in the first minutes of the prediction. Here, the additional features disturb the already specific information provided by a few frequency trajectories, thus increasing the error. Further into the future, i.e. the rest of the hour, the additional features now provide information as less information is available through the increasing number of frequency trajectories. This could suggest that an adaptive β could lead to further enhancement, although the increasing computational effort during optimization of hyperparameters must then be taken into account.

The analysis of coal as additional feature is especially interesting as coal power plants from 2019 started to be shut down and replaced by mainly combined cycle gas turbines as the main electricity generation source in the Balearic Islands [65]. During the test set, which is all recordings from 2021, coal was not used for electricity generation at all. Thus,

for predictions, frequency trajectories from hours where coal was not used during the train set had an advantage during the nearest neighbors selection as the euclidean distance for the coal data would be 0. A study by Martínez-Barbeito et al. found that the frequency deviates less after the transition from coal to combined cycle and less frequently reaches the level of activation of frequency reserves, despite coal's former high contribution to system inertia [65].

4.5.4 Model Extension Discussion

The extended model reveals that distinct additional features related to the frequency fluctuations do not necessarily provide clear additional information about the frequency trajectory. For most features, the frequency trajectory is best predicted during the optimization when only the past frequency trajectories are taken into account, but with a slightly increased performance for distinct weights on the test set. Nevertheless, it is conceivable that if the additional feature data had a better time resolution, more information could be provided. With a time resolution of 10, 15 and 30 min, there are 600, 900 and 1800 frequency measurements per additional feature measurement, respectively. During these measurements, the frequency fluctuates a lot, and correlating fluctuations could have been seen in the additional features with a higher time resolution. Another approach could also have been to extend the additional features to consist of 3600 values by, e.g. linear interpolation between the excising values. The additional feature would then consist of mainly interpolated values but then, to some extent, realistically reflect the electricity generation of a more controllable source like hydro or coal. However, especially fluctuating sources like wind or solar would rarely follow linear or step-wise generation levels, and the new information would be misleading for the frequency prediction. Additionally, with two hyperparameters tuned, *adaptive-k* and β, the model might also suffer from slightly overfitting when trained on the validation set.

Another potential reason the additional features did not significantly improve the full hour prediction performance is the amount of frequency data. As the frequency time series for the Balearic Islands extends from September 2019 to March 2021, there are not enough data for the test and validation set to consist of one or more full years. The same applies to Ireland with less than three months of data. Thus, seasonal patterns like electricity generation from solar energy are not potentially revealed and will not be present in one or more of the sets. This applies, among other things, to the coal feature, where no electricity was generated from coal during the test set period, as opposed to being a large share of generation during the validation set. Thus, the optimization process on the validation set would not reflect the appropriate coal contribution for the other subsets, resulting in erroneous hyperparameters. However, with multiple years of data, the predictor could be trained and tested on all seasons and a greater chance of including seasonal patterns that influence frequency.

One challenge that was addressed during the model extension was the varying magnitude of each additional feature, both between the features and within each feature over time. The frequency constantly fluctuates close to 50 Hz, but the features exhibit significant variation and depend on import, weather, electricity price or consumption patterns. This was attempted to be solved by introducing min-max scaling. This partly solves the

problem with varying amounts of electricity generation from a feature. However, a new challenge occurs when a new electricity generation source is integrated into the power system. As an example, a future wind farm could potentially be installed and connected to the power grid of the Balearic Islands. Firstly, the historical training data in the model would not contain any information about wind generation. To take advantage of potential information provided by the wind generation, the model now needs new data for a longer time period to be optimized for the specific feature and gain sufficient historical data. Secondly, wind generation would, to some extent, be a substitute for other electricity generation sources, e.g. the replacement of coal for the Balearic Islands. With a heavy reduction of coal electricity generation, the potential information provided by the coal trajectories would now be reduced or change compared to before, thus affecting the predictor. The dynamic of the power grid could change as found in [65], and thus the frequency behavior and correlation with generation sources utilized as additional features in the extended model would also change.

This model extension only involved testing electricity generation sources as additional features. In the highly complex modern power systems, numerous factors influence the overall system balance and behavior [66]. The electricity price is an example of one related feature that would be interesting to test for the extended model. The design of the extended part of the model makes this possible without changes. The feature weighting β can be optimized for each specific feature. Also, as already seen in Table 4.3, the model can handle time series of different time resolutions. The model is thus applicable to different types of power grid relevant time series.

5 General Discussion

In general, the power system in the Balearic Islands is by far the most deterministic system of the five investigated regions. The WNN predictor predicted the frequency with low error and is a highly relevant predictor for this power system due to its clear repetitive patterns even for shorter time scales. The low prediction error and some of the frequency patterns from Section 4.2 are similar to those obtained by the study of frequency time series from CE [9]. Several factors could describe the correlating behaviors. The Balearic Islands are only connected to CE, in contrast to, e.g. the Nordic region with interconnectors to multiple synchronous areas or Iceland without connections. In addition, the Balearic TSO REE also operates the power grid in mainland Spain. Mainland Spain is a part of the Continental European synchronous area. Hence, REE acts in accordance with the same regulations and structures as other system operators in CE through the RSC or ENTSO-E.

Predictability seems to be strongly related to population size and interconnections. The most populated regions' power systems tend to be more deterministic than the less populated and thus have a lower prediction error. In Iceland, the frequency has relative large excursions. This is mainly caused by changes in large power-intensive loads such as aluminium smelters or unforeseen events in large generating units, occurring at random times. Large operators make up a significant proportion of small systems of Icelandic or Faroese size and thus have a considerable influence on the balance of these power systems. Moreover, the ageing Icelandic transmission system often reaches its limits due to the increasing demand from industry and electrification, causing bottlenecks in the transmission system [60]. This makes the Icelandic power grid frequency challenging to predict. The Balearic Islands are the exception in this smaller-systems trend with precise predictions; Interconnection to CE increases reliability through the possibility of rapidly increased power exchange to counteract an imbalance [67]. Also Ireland and Nordic are connected to other synchronous areas and benefit from the balancing possibilities through import and export; an absent opportunity for the isolated Icelandic and Faroese grids.

The distinct regions' predictability also depends on the electricity generation sources. The Balearic Islands have the greatest share of non-renewable energy sources in the power system among the investigated regions. With gas and oil power plants, the electricity generation is more controllable than the more wind power based regions of Ireland and, to some extent, the Faroe Islands. There is thus less inertia in the Irish and Faroese power grid, which increases the chance of greater frequency fluctuations. In Iceland, the frequency has relative large excursions related to low levels of inertia through geothermal electricity generation and the mentioned stochastic dynamics introduced by the consumer side.

For power systems with a high share of solar or wind, irregular electricity generation contributes to frequency fluctuations. As these energy sources can not be planned, generation rely on forecasts of, e.g. wind conditions to predict the power output for a period. This makes balancing the load and generation more complicated and might cause more frequent fluctuations. By 2025, all countries in the EU and other European countries like Norway will switch to 15 min windows for the day-ahead trading. A higher time resolution on the bids and volumes might provide more information and interaction between

forecasts and actual generation. Hence, accurate first-minutes predictions by specific frequency trajectories seen in this thesis might occur multiple times through trading intervals during the entire hour.

Overall, the frequency measurements utilized during the thesis were of good quality, further confirmed by comparison with some TSOs calculations of frequency deviations. The exception is uncertainty about the measurements from Ireland, as several of the results showed unknown systematic frequency patterns in addition to about 25% invalid measurements. Also, data from the ENTSO-E transparency platform were of partly poor quality, and there is a potential for improvement in increasing the applicability of this data for public use.

The WNN predictor can contribute to frequency control and analysis in multiple ways. As a forecasting method, the WNN predictor can provide guidance for TSOs. With accurate predictions, the balancing and management of the slower frequency reserves such as mFRR can be more accurately scheduled with respect to needed capacity and bids. Conversely, with very poor predictions, TSOs may be aware earlier that a fluctuation unknown to former patterns is taking place. It could lead to deviations requiring further reserves to be activated in the near future. An advantage of KNN-based forecasting is that it does not need to be retrained when new frequency time series data are added over time. The data are rather added to the training set, thus increasing the basis for further predictions. Hence, the WNN predictor can be utilized with both a little and a great amount of data, only with the need to update the optimal k as the data size increases.

As performed in this thesis, the predictor reveals the predictability of power grids' frequency, thus also highlighting the deterministic or stochastic behaviors power grids exhibit. This can be used to choose other suitable prediction or analysis methods for the specific power grids. When the predictor performs badly it may indicate that the power grid experience some abnormal behavior or special events unknown to former trends. Thus, the predictor can be utilized by TSOs for event detection. With poor performance over time, the predictor reflects or points out a change in the dynamics of the power system. This can be caused by a recent integration of a greater share of renewable energy generation or interconnectors to other regions being put into or out of use.

6 Conclusion and Further Research

6.1 Conclusion

In this master's thesis, the already established WNN predictor was successfully applied to frequency time series from different European islands. The pattern-based prediction model was previously used to analyze power grids and predict the power grid frequency of three large European synchronous areas. The main goal of this thesis was to investigate smaller islanded power grids with a limited amount of frequency measurements available. Frequency measurements have been recorded with a sampling rate of 1 second in Ireland, the Balearic Islands, Iceland and the Faroe Islands and were used to analyze these regions. The research question of the thesis was mainly divided into three to examine various aspects.

Firstly, The WNN predictor was successfully applied for all investigated regions and performed approximately equal or better than two null models when comparing the models' root mean square error (RMSE). 60 min frequency predictions based on a weighted average of historic frequency trajectories were performed with all available data for each region after a data cleaning process to investigate the predictability of island power grids. The Nordic region was included as a basis for comparison, as this region in previous research with the WNN predictor performed somewhat in between the more deterministic synchronous area of Continental Europe and the more stochastic Great Britain [9].

Secondly, the size dependence of the WNN predictor was investigated. Predictions were performed with time intervals ranging from one week to two or six months, depending on the amount of data available for each specific region. Through an average of 25 samples, the results explicitly show that more data give better results as all islands improved their performance for each increasing time interval. Even Iceland, as the most stochastic region, improved its prediction results when a longer time series was utilized, but the WNN was still outperformed by both null models. For the other regions, the WNN predictor outperformed the two null models when the utilized time series length was between two weeks and one month. Thus, this seems to be a turning point in time series length as more specific frequency trajectories started to provide better information than an average of all the historical trajectories.

The first two parts laid a basis to describe each region's predictability. The Balearic islands exhibited deterministic characteristics as frequency was predicted with low error. Ireland performed reasonably similar to the Nordic region, with some uncertainty associated with its frequency measurements. Iceland exhibited very stochastic behavior with a relatively high error and the 50 Hz prediction as the best model. With little data available, the Faroe Islands had a poor performance due to a lack of sufficient data, but the analysis showed some deterministic characteristics to investigate further. Interconnections with other synchronous areas seem to play a significant part in a region's predictability as to have a possibility of power exchange for system balance. The power system's size is also influential as single actions by large operator or generation units influence the balance in a small system more decisive. Additionally, the worst-performing power systems tend to have less system inertia through, e.g. higher share of renewable electricity generation.

Lastly, one extension of the prediction model was designed and tested. This included the influence of additional features to select the nearest frequency neighbors. The features tested included wind power generation in Ireland and renewables and fossil-fuel generation in the Balearic Islands. The features were included in the prediction model by extending the hourly frequency trajectories with the associated generation measurements of different time resolution. By tuning a new hyperparameter β, the impact a new feature should have on frequency prediction could be optimized. During the optimization process, only wind generation in Ireland got a non-zero weighting, but the predictor's performance was not improved when applied to the test series. However, the features tested for the Balearic Islands all improved the performance by around 1% for at least one β, even though the optimization did not result in a non-zero weighting. This suggests that there is information to be gleaned from other power grid time series to improve the prediction of power grid frequency further.

6.2 Further Research

Numerous things can contribute to further research or extension involving the WNN predictor for predicting island power grid frequency in the future. A primary limitation of the thesis was the availability of the frequency data. For greater synchronous areas like the Nordic region or CE, high-resolution frequency data is usually easily accessible through the TSOs web pages. On the other hand, data availability for smaller regions or islands is limited. Not all system operators publicly share this type of data, or there is a lack of logging of good quality or high-resolution data. While most of the time series in this thesis ranged from a week to a few months, data for multiple years would improve the research and strengthen the basis for predictions.

During the thesis, one approach was designed and tested to extend the model, involving the influence of an additional power generation feature's time series in selecting the nearest frequency trajectory neighbors. This method was strictly limited by the low time resolution of the additional data and the quality, especially from the data from the ENTSO-E. Optimally, this data should have the exact, or at least more similar, time resolution as the frequency data to learn about the different generation time series, which is more common in KNN applications with several features. As a potential change, additional features could be included in the predictor differently. This includes a different influence of the prediction rather than the nearest neighbor selection, or several features could be included at once. The additional feature could also be weighted differently, e.g. creating an adaptive weight similar to the *adaptive-k*, taking into account a risk of overfitting and increased computational effort. In addition, other types of power grid relevant time series could be added instead of generation types, such as weather data, load or electricity price.

To extend the research of this thesis in the future, there are multiple ways to move forward in terms of other prediction models. As the different frequency series ranged from deterministic to very stochastic behaviour, there are a lot of other machine learning algorithms and prediction methods suitable for time series forecasting in addition to the KNN algorithm the WNN predictor is based on. This involves, among other methods, Multilayer Perceptron (MLP) and Recurrent Neural Networks (RNN) [6]. More stochastic approaches could be considered for the stochastic behavior exhibited in Iceland.

References

[1] [V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Climate change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. 2021. doi:10.1017/9781009157896.

[2] Klima og miljødepartementet. Klimaendringer og norsk klimapolitikk. https://www.regjeringen.no/no/tema/klima-og-miljo/innsiktsartikler-klima-miljo/klimaendringer-og-norsk-klimapolitikk/id2636812/. [Online; accessed 20-April-2022].

[3] C. Fetting. The European Green Deal, ESDN Report. https://www.esdn.eu/fileadmin/ESDN_Reports/ESDN_Report_2_2020.pdf, December 2020. [Online; accessed 12-February-2022].

[4] Agora Energiewende and Ember. The European Power Sector in 2020: Up-to-Date Analysis on the Electricity Transition. https://static.agora-energiewende.de/fileadmin/Projekte/2021/2020_01_EU-Annual-Review_2020/A-EW_202_Report_European-Power-Sector-2020.pdf, January 2021. [Online; accessed 24-February-2022].

[5] Jorge González-Ordiano, Simon Waczowicz, V. Hagenmeyer, and Ralf Mikut. Energy forecasting tools and services. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8:e1235, 2017. doi:10.1002/widm.1235.

[6] Jason Brownlee. *Deep Learning for Time Series Forecasting*. Jason Brownlee, 2018.

[7] Jorge Ángel González Ordiano, Lutz Gröll, Ralf Mikut, and Veit Hagenmeyer. Probabilistic energy forecasting using the nearest neighbors quantile filter and quantile regression. *International Journal of Forecasting*, 36(2):310–323, 2020. doi:10.48550/arXiv.1903.07390.

[8] Jesus Lago, Grzegorz Marcjasz, Bart De Schutter, and Rafał Weron. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293:116983, 2021. doi:10.1016/j.apenergy.2021.116983.

[9] Johannes Kruse, Benjamin Schäfer, and Dirk Witthaut. Predictability of Power Grid Frequency. *IEEE Access*, 8:149435–149446, 2020. doi:10.1109/ACCESS.2020.3016477.

[10] Woorim Bang and Ji Won Yoon. Forecasting the Electric Network Frequency Signals on Power Grid. In *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1218–1223, 2019. doi:10.1109/ICTC46691.2019.8939676.

[11] Richard Jumar, Heiko Maaß, Benjamin Schäfer, Leonardo Rydin Gorjão, and Veit Hagenmeyer. Database of Power Grid Frequency Measurements. *arXiv preprint arXiv:2006.01771*, 2021. doi:10.48550/arXiv.2006.01771.

[12] N.W.A. Lidula and A.D. Rajapakse. Microgrids research: A review of experimental microgrids and test systems. *Renewable and Sustainable Energy Reviews*, 15(1):186–202, 2011. doi:10.1016/j.rser.2010.09.041.

[13] Alexis Lagrange, Miguel de Simón-Martín, Alberto González-Martínez, Stefano Bracco, and Enrique Rosales-Asensio. Sustainable microgrids with energy storage as a means to increase power resilience in critical facilities: An application to a hospital. *International Journal of Electrical Power & Energy Systems*, 119:105865, 2020. doi:10.1016/j.ijepes.2020.105865.

[14] Office of Energy Efficiency & Renewable Energy. History of hydropower. https://www.energy.gov/eere/water/history-hydropower, 2022. [Online; accessed 10-March-2022].

[15] Haisheng Chen, Thang Ngoc Cong, Wei Yang, Chunqing Tan, Yongliang Li, and Yulong Ding. Progress in electrical energy storage system: A critical review. *Progress in Natural Science*, 19(3):291–312, 2009. doi:10.1016/j.pnsc.2008.07.014.

[16] Ahmed Zayed AL Shaqsi, Kamaruzzaman Sopian, and Amer Al-Hinai. Review of energy storage services, applications, limitations, and benefits. *Energy Reports*, 6:288–306, 2020. SI:Energy Storage - driving towards a clean energy future. doi:10.1016/j.egyr.2020.07.028.

[17] Cristina Alcaraz and Sherali Zeadally. Critical infrastructure protection: Requirements and challenges for the 21st century. *International Journal of Critical Infrastructure Protection*, 8:53–66, 2015. doi:10.1016/j.ijcip.2014.12.002.

[18] Joseph Song-Manguelle, Maja Harfman Todorovic, Song Chi, Satish K. Gunturi, and Rajib Datta. Power Transfer Capability of HVAC Cables for Subsea Transmission and Distribution Systems. *IEEE Transactions on Industry Applications*, 50(4):2382–2391, 2014. doi:10.1109/TIA.2013.2291934.

[19] Statnett. Tall og data fra kraftsystemet. https://www.statnett.no/for-aktorer-i-kraftbransjen/tall-og-data-fra-kraftsystemet/, 2022. [Online; accessed 31-March-2022].

[20] ENTSO-E. ENTSO-E Mission Statement. https://www.entsoe.eu/about/inside-entsoe/objectives/#our-mission, 2022. [Online; accessed 31-March-2022].

[21] Wikimedia Commons. File:ElectricityUCTE.svg — Wikimedia Commons, the free media repository. https://commons.wikimedia.org/w/index.php?title=File:ElectricityUCTE.svg&oldid=642643853, 2022. [Online; accessed 22-March-2022].

[22] Power Engineering International. Sharing the power. https://www.powerengineeringint.com/world-regions/europe/sharing-the-power/, 2004. [Online; accessed 22-March-2022].

[23] ENTSO-E. Actual Generation per Production Type. https://transparency.entsoe.eu/generation/r2/actualGenerationPerProductionType/show, 2022. [Online; accessed 22-March-2022].

[24] Jan Machowski, Zebigniew Lubosny, Janus W. Bialek, and james R. Bumby. *Power system dynamics stability and control*. John Wiley & Sons Ltd, 2020.

[25] Niklas Modig, Robert Eriksson, Pia Ruokolainen, Jon Nerbø Ødegård, Simon Weizenegger, and Thomas Dalgas Fechtenburg. Overview of Frequency Control in the Nordic Power System. https://www.epressi.com/media/userfiles/107305/1648196866/overview-of-frequency-control-in-the-nordic-power-system-1.pdf, 2022. [Online; accessed 25-April-2022].

[26] Paul Denholm, Trieu Mai, Richard Wallace Kenyon, Benjamin Kroposki, and Mark O'Malley. Inertia and the Power Grid: A Guide Without the Spin. May 2020. doi:10.2172/1659820.

[27] E Ørum, M Kuivaniemi, M Laasonen, AI Bruseth, EA Jansson, A Danell, K Elkington, and N Modig. Future system inertia, ENTSO-E. Technical report, Brussels, Tech. Rep, 2015.

[28] Next Kraftwerke. What is Frequency Containment Reserve (FCR)? https://www.next-kraftwerke.com/knowledge/frequency-containment-reserve-fcr#why-do-we-need-the-fcr, 2022. [Online; accessed 25-April-2022].

[29] Robert Eriksson, Niklas Modig, and Katherine Elkington. Synthetic inertia versus fast frequency response: a definition. *IET Renewable Power Generation*, 12, 09 2017. doi:10.1049/iet-rpg.2017.0370.

[30] Nordic Regional Security Operator. About Nordic RSC. https://nordic-rsc.net, 2022. [Online; accessed 07-April-2022].

[31] Fingrid. Frequency quality analysis 2020. https://www.fingrid.fi/globalassets/dokumentit/fi/kantaverkko/suomen-sahkojarjestelma/frequency_quality_analysis_2020.pdf, 2021. [Online; accessed 02-April-2022].

[32] Nordic Energy Research. The Nordics: Share of renewables has increased. https://www.nordicenergy.org/figure/share-of-renewables-has-increased/, 2020. [Online; accessed 30-March-2022].

[33] World Nuclear Association. Nuclear Power in Sweden. https://world-nuclear.org/information-library/country-profiles/countries-o-s/sweden.aspx, 2022. [Online; accessed 30-March-2022].

[34] Data from BP Statistical Review of World Energy, Ember Global Electricity Review (2022) and Ember European Electricity Review (2022). Share of electricity from low-carbon sources. https://ourworldindata.org/energy/country/finland?country=FIN~NOR~DNK~SWE~ISL~FRO~IRL, 2022. [Published online at OurWorldInData.org; accessed 04-May-2022].

[35] Wind Energy Ireland. Facts & Stats. https://windenergyireland.com/about-wind/facts-stats, 2022. [Online; accessed 23-February-2022].

[36] SONI. What is the DS3 Programme? https://www.soni.ltd.uk/__uuid/a6f0ce76-c5a9-4120-8b53-cac2daa99840/ds3-programme/, 2022. [Online; accessed 07-April-2022].

[37] SONI and EirGrid. All-Island Transmission System Performance Report 2020. http://www.eirgridgroup.com/site-files/library/EirGrid/All-Island-Transmi

ssion-System-Performance-Report-2020.pdf, 2021. [Online; accessed 11-April-2022].

[38] EirGrid Group. What is the Celtic Interconnector? http://www.eirgridgroup.com/the-grid/projects/celtic-interconnector/the-project/, 2022. [Online; accessed 28-March-2022].

[39] Red Eléctrica de España. Singular features of the Balearic system. https://www.ree.es/en/about-us/ree-2-minutes, 2022. [Online; accessed 08-February-2022].

[40] Domenico Curto and Marco Trapanese. A Renewable Energy mix to Supply the Balearic Islands: Sea Wave, Wind and Solar. In *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I CPS Europe)*, pages 1–6, 2018. doi:10.1109/EEEIC.2018.8493876.

[41] ENTSO-E. Operational Limits and Conditions for Mutual Frequency support over HVDC. Technical report, ENTSO-E, 02 2021.

[42] Ardelean M and Minnebo P. HVDC Submarine Power Cables in the World: State-of-the-Art Knowledge. 2015. doi:10.2790/023689.

[43] Orkustofnun National Energy Authority. Electricity. https://nea.is/hydro-power/electric-power/nr/69. [Online; accessed 10-March-2022].

[44] Askja Energy. Icelink interconnector in operation by 2025? https://askjaenergy.com/2018/04/17/icelink-in-operation-by-2025/, 2018. [Online; accessed 09-March-2022].

[45] Landsnet. Tíðnistýring. https://www.landsnet.is/arsskyrslur/arsskyrsla-2019/islenska/frammistoduskyrsla/tidnistyring-og-spennugaedi/, 2020. [Online; accessed 09-March-2022].

[46] Helma Maria Tróndheim, Bárdr A. Niclasen, Terji Nielsen, Filipe Faria Da Silva, and Claus Leth Bak. 100% Sustainable Electricity in the Faroe Islands: Expansion Planning Through Economic Optimization. *IEEE Open Access Journal of Power and Energy*, 8:23–34, 2021. doi:10.1109/OAJPE.2021.3051917.

[47] Hannan Ma and Husheng Li. Analysis of Frequency Dynamics in Power Grid: A Bayesian Structure Learning Approach. *Smart Grid, IEEE Transactions on*, 4, 03 2013. doi:10.1109/TSG.2012.2226066.

[48] Francisco Martinez Alvarez, Alicia Troncoso, Jose C. Riquelme, and Jesus S. Aguilar Ruiz. Energy Time Series Forecasting Based on Pattern Sequence Similarity. *IEEE Transactions on Knowledge and Data Engineering*, 23(8):1230–1243, 2011. doi:10.1109/TKDE.2010.227.

[49] Alicia Troncoso, José Riquelme, Jose Martinez-Ramos, Jesús Santos, and Antonio Gomez-Exposito. Influence of kNN-Based Load Forecasting Errors on Optimal Energy Production. volume 2902, pages 189–203, 12 2003. doi:10.1007/978-3-540-24580-3_26.

[50] Janett Williams and Yan Li. Comparative Study of Distance Functions for Nearest Neighbors. pages 79–84, 01 2008. doi:10.1007/978-90-481-3660-5_14.

[51] Analytics Vidhya. Feature Scaling Techniques in Python – A Complete Guide. https://www.analyticsvidhya.com/blog/2021/05/feature-scaling-techniques-in-python-a-complete-guide/, 2021. [Online; accessed 29-March-2022].

[52] https://github.com/thlonsaker/predictability-of-island-power-grids.git.

[53] Leonardo Rydin Gorjão, Richard Jumar, Heiko Maass, Veit Hagenmeyer, G. Cigdem Yalcin, Johannes Kruse, Marc Timme, Christian Beck, Dirk Witthaut, and Benjamin Schäfer. Open database analysis of scaling and spatio-temporal properties of power grid frequencies. *Nature Communications*, 11(1):6362, 2020. doi:10.1038/s41467-020-19732-7.

[54] Davide Chicco. Ten quick tips for machine learning in computational biology. *BioData Mining*, 10(1):35, 2017. doi:10.1186/s13040-017-0155-3.

[55] Jason Brownlee. *Long Short-Term Memory Networks With Python*. Routledge, 2017.

[56] Subhash Chandra Gupta and Noopur Goel. Performance enhancement of diabetes prediction by finding optimum K for KNN classifier with feature selection method. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 980–986, 2020. doi:10.1109/ICSSIT48917.2020.9214129.

[57] Zhongheng Zhang. Introduction to machine learning: k-nearest neighbors. *Ann Transl Med*, 4(11):218, Jun 2016. doi:10.21037/atm.2016.03.37.

[58] Red Eléctrica de España. Islas Baleares - Electricity demand tracking in real time. https://demanda.ree.es/visiona/baleares/baleares/total, 2022. [Online; accessed 08-February-2022].

[59] EirGrid Group. All Island - Actual and Forecast Wind. https://www.eirgridgroup.com/how-the-grid-works/system-information/, 2022. [Online; accessed 20-April-2022].

[60] Freyr Hardarson, Birkir Heimisson, Sigurgeir Björn Geirsson, Jón Arnar Emilsson, Íris Baldursdóttir, and Valur Knútsson. Theistareykir Geothermal Power Plant, Challenges in a Weak Electrical Grid. *GRC Transactions*, 42, 2018. URL: https://publications.mygeoenergynow.org/grc/1034049.pdf.

[61] SEMOpx. Intraday Continuous Market. https://www.semopx.com/markets/ict-market/, 2022. [Online; accessed 06-May-2022].

[62] Navneet Singh, Asheesh Singh, and Manoj Tripathy. A comparative study of BPNN, RBFNN and ELMAN neural network for short-term electric load forecasting: A case study of Delhi region. *9th International Conference on Industrial and Information Systems, ICIIS 2014*, 02 2015. doi:10.1109/ICIINFS.2014.7036502.

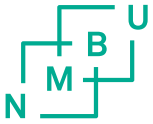
[63] Statnett. 15 minutters avregning og energimarkeder. https://www.statnett.no/for-aktorer-i-kraftbransjen/systemansvaret/kraftmarkedet/15-minutters-avregning-og-energimarkeder/, 2022. [Online; accessed 06-May-2022].

[64] Government of Ireland. Programme for Government: Our Shared Future. https://www.gov.ie/en/publication/7e05d-programme-for-government-our-shared-future/. [Online; accessed 11-May-2022].

[65] María Martínez-Barbeito, Damiá Gomila, and Pere Colet. Data Analysis of Frequency Fluctuations in the Balearic Grid Before and After Coal Closure. https://dspace.uib.es/xmlui/bitstream/handle/11201/158554/562267.pdf?sequence=1&isAllowed=y, 2021. [Online; accessed 06-May-2022].

[66] Johannes Kruse, Benjamin Schäfer, and Dirk Witthaut. Revealing drivers and risks for power grid frequency stability with explainable AI. *Patterns*, 2(11):100365, nov 2021. doi:10.1016/j.patter.2021.100365.

[67] Andrea Tosatto, Matas Dijokas, Tilman Weckesser, Spyros Chatzivasileiadis, and Robert Eriksson. Sharing Reserves through HVDC: Potential Cost Savings in the Nordic Countries, 2020. doi:10.48550/ARXIV.2001.00664.



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway