Norwegian University
of Life Sciences

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

# Applying a Top-Down Estimation of Inertia in the Nordic Power System

Lars-Petter Andersen
MSc. Environmental Physics and Renewable Energy

# Preface

This Master's Thesis concludes my Master of Science (MsC) degree in Environmental Physics and Renewable Energy at the Norwegian University of Life Sciences (NMBU). The work was performed under supervision of Heidi S. Nygård within the Faculty of Science and Technology. The idea of the thesis came from Statnett SF, and I would like to thank Espen Aronsveen and Simon Weizenegger for sharing insights and thoughts at the start of the thesis. I would also like to thank Kristian W. Høiem for being available to answer questions when it was time to conclude the work.

A special thanks goes to my supervisor. Thanks for believing in me Heidi, and help me getting through the work required to hand in this thesis. Certainly, this thesis would have been laid on a shelf if not for your help and support along the way.

Thanks to my family, especially my mom, Anette F. Andersen, for believing in me and for all the emotional support you have given along the way. Thanks to my uncle and aunt, Einar and Birgitte for giving me an office I could use at the end of this period, and for making sure I didn't starve in the process. Thank you Christiane and Jørgen for proof reading the thesis for me.

This thesis did show me that things can take unexpected turns, and that life most certainly isn't always a bed of roses. Yet I look back at an interesting period where I have gained knowledge within topics regarding power system stability, but also research in general. And indeed, the work has been a great help to gain insight in my own working methods - I have found many areas of improvement!

Fredrikstad, February 4th, 2022
Lars-Petter Andersen

# Abstract

The penetration of Variable Renewable Energy Sources (VRES) in the power system creates new challenges for the Transmission System Operators (TSOs). In the Nordic Power System (NPS), consisting of Norway, Sweden, Finland and eastern Denmark, all electricity is generated at the same frequency. When the system is in balance, meaning production and consumption is on the same level, the frequency is 50.00 Hz. It is crucial that the frequency remains stable. If the total load exceeds total generation, the frequency falls. Too much generation, and frequency rises. Conventional generation plants are connected to the power grid by means of rotating machines driven by a rotor, generating electricity at the desired frequency. Once in motion, the kinetic energy stored in the rotor will keep the generator running, even if the mechanical force driving it disappears. This effect is called inertia and is an essential inert part of the stability of a power system.

VRES are not connected to the grid by rotating machines but through Power Electronics (PE), and do not contribute to the system's inertia. The increasing share of renewable energy sources leads to decreased power system inertia, and frequency stability becomes a concern. Therefore, the estimation of power system inertia has been of focus in the NPS in recent years.

In order to secure frequency stability in the Nordic synchronous area, a new fast power reserve will be introduced. Fast frequency reserves (FFR) will complement the existing primary reserves for disturbances (FCR-D). To plan for future operation, the aim is to forecast the system inertia such that the instantaneous frequency minimum caused by the loss of the largest generator can be assessed, and sufficient fast reserves are procured.

Since 2015, the Nordic TSOs have estimated and stored the value of system inertia by calculating the sum of inertia contributions from synchronous machines connected to the grid. This value is closely related to the amount of generation in the power system. This thesis applied a top-down estimation using linear models with production data as explanatory variables to estimate system inertia. Data from 2018 and 2019 on production and inertia were collected. Data from January 2018 to June 2019 were used to train models, and the models' performance was tested on data from July and August 2019. On the Nordic level, these models received a Mean Average Percentage Error of 2 % on the test set, revealing potential of using this approach to forecast inertia by using production forecasts per production type.

The Nordic estimation of inertia level is a sum of estimations from Norway, Sweden, Denmark and Finland, so models were also tested using data from each country. This revealed that the Norwegian inertia estimation is already a top-down estimation. It was

also found that the top-down estimation model is inaccurate in Denmark. The results from Sweden and Finland indicated that conventional power production, such as nuclear, thermal and hydro, was relevant to models while VRES production was irrelevant.

# Sammendrag

Fremveksten av variable fornybare energikilder i kraftsystemet er grunnlag for nye utfordringer for kraftsystemets systemansvarlige (TSO - transmission system operators). I det nordiske kraftsystemet (NPS - nordic power system) som består av Norge, Sverige, Finland og øst-Danmark genereres all elektrisitet med samme frekvens. Når systemet er i balanse er produksjon og forbruk på samme nivå, og frekvensen er 50,00 Hz. Det er avgjørende at frekvensen forblir stabil. Hvis total last er større enn total produksjon, faller frekvensen. For mye produksjon fører til at frekvensen øker. Tradisjonelle kraftverk er koblet til kraftsystemet gjennom roterende maskiner som blir drevet av en rotor, som produserer elektrisitet med ønsket frekvens. Når maskinene er i gang, vil den kinetiske energien som er lagret i form av rotasjonsenergi i rotoren holde generatoren i gang, selv om en andel av kraften som driver rotoren faller bort. Denne effekten kalles inertia, og er en avgjørende egenskap som påvirker stabiliteten i kraftsystemet.

Fornybare energikilder er ikke koblet til kraftsystemet med roterende masse, men gjennom kraftelektronikk, og bidrar ikke til inertia i kraftsystemet. Den økende andelen av fornybare energikilder fører derfor til at mengden inertia er synkende, og dette øker bekymringen om frekvensstabilitet. Derfor har estimering av kraftsystemets inertia vært et fokusområde i NPS de siste årene.

For å sikre frekvensstabilitet i det nordiske synkronområdet, vil en ny rask reserve bli introdusert. Raske frekvensreserver (fast frequency reserves, FFR) vil fungere som en ekstra effektrespons som et komplement til den eksisterende primærreserven for driftsforstyrrelser (FCR-D). For å kunne planlegge systemdriften er det et mål i NPS å lage et prognoseverktøy slik at frekvensminimum som kan oppstå på grunn av utfall av den største produksjonsenheten kan beregnes, og tilstrekkelig mengde med effektreserve kan bestilles.

De nordiske systemoperatørene har siden 2015 estimert mengden inertia i systemet, ved å beregne summen av bidrag av rotasjonsmasse fra alle synkrontilkoblede maskiner til enhver tid. Denne verdien er nært knyttet til mengden produsert effekt i kraftsystemet. I denne oppgaven ble en ovenfra-og-ned estimering tilpasset, for å estimere systemets inertia basert på lineære modeller som benytter produksjonsdata som forklaringskolonner. Data om inertia og produksjon fra 2018 og 2019 ble samlet inn. Data fra Januar 2018 til Juni 2019 ble brukt til å trene lineære modeller, og modellenes ytelse ble testet på data fra Juli og August 2019. På et nordisk nivå fikk disse modellene et gjennomsnittlige prosentavvik på 2 % i testperioden, som tyder på at slike modeller kan benyttes i prognosemodeller ved å benytte predikerte produksjonsverdier per produksjonstype.

Den nordiske estimasjonen av inertia er en sum av estimasjoner fra Norge, Sverige,

Danmark og Finland. Modeller for hvert av disse landene ble derfor også testet. Disse avslørte at den norske inertiaestimasjonen allerede er en ovenfra-ned estimasjon. Videre avslørte resultatene at en ovenfra-ned estimasjon er unøyaktig i Danmark. Resultatene fra Sverige og Finland indikerte at konvensjonelle produksjonstyper, som kjernekraft, termiske kraftverk og vannkraftverk har forklaringspotensiale i ovenfra-ned modeller, mens fornybare energikilder ikke bidrar i slike modeller.

# Contents

# List of Figures

# Abbreviations

**FFR** Fast Frequency Reserves

**aFRR** automatic Frequency Restoration Reserve

**mFRR** manual Frequency Restoration Reserve

**HVDC** High Voltage Direct Current

**LR** Linear Regression

**MAE** Mean Average Error

**MAPE** Mean Average Percentage Error

**MSE** Mean Squared Error

**NAG** Nordic Analysis Group

**NPS** Nordic Power System

**PE** Power Electronics

**TSO** Transmission System Operator

**VRES** Variable Renewable Energy Source(s)

# Chapter 1

# Introduction

## 1.1 Background

*The future is electric*, is the motto of the Norwegian state-owned enterprise Statnett, the transmission system operator (TSO) in Norway. Indeed, electricity is the backbone of the technological growth mankind has experienced in the last hundred years, and electricity demand will only continue to rise as more and more sectors are electrified. There are many appealing characteristics of electricity. It is an excellent energy carrier with great convenience in terms of transmission and conversion. Further, the energy quality of electricity is very high, meaning that electricity can be converted with high efficiency to other energy forms, such as heating, lighting or kinetic energy (e.g. making an electric vehicle move) [1].

Electrical power is distributed through power systems. At the very beginning of the history of electricity, a standard power system consisted of individual generators connected to matching loads, usually for lighting purposes [1, 2, 3] Since then, power systems have increased in size, motivated by various technical, social and economical factors. Interconnection of neighbouring utilities has usually led to improved system security and economy of operation. The security of supply increases as the utilities from different regions can assist each other. The economy of operation increases because a combined larger system needs less generating reserve capacity than two smaller systems. Interconnection also permits the utility always to take advantage of the most economical power sources. Since the beginning, these benefits have been recognised, and power systems around the world continue to strengthen transmission interconnections. An example of this are the high voltage direct current (HVDC) links that tie the Nordic Power System (NPS) and the synchronous grid of Continental Europe together.

In 2020 and 2021, two new HVDC links were connected: NordLink, from Sirdal to Wilster, Germany, and North Sea Link from Kvilldal to Blythe, England. In Statnetts net development plan [4], Statnett points out that these new HVDC links may lead to a more complex system operation due to higher variability of power flow and more intraday variations.

Nowadays, in modern society, we take electricity for granted and use it for every purpose in our daily activities. Electricity is widely used for industrial, domestic and leisure purposes. Since electricity plays such an essential role in our everyday life, the operation of the electrical power systems must remain reliable, no matter the challenges that occur. This has been, and will continue to be a huge task, as the interconnected power system is a large system of enormous complexity.

Industrialization and urbanisation have resulted in severe environmental pollution of our planet. A significant contributor to this pollution is the release of Carbon Dioxide ($CO_2$) to the atmosphere. Consequently, the authorities and political leaders in different countries have developed strategies and created plans on how to decrease $CO_2$ emissions. These strategies mainly focus on two solutions. We can either use less energy, by reducing our demand or making our energy consumption more efficient. To succeed, we must most likely do both. An example of reducing overall energy usage is electrification since electrical technology is more energy-efficient than fossil. In addition, traditional fossil energy production is being replaced by clean electricity from renewable energy sources (RES) and is expected to decrease from the current mix of about 66 % to around 31 % by 2050, according to Bloombergs New Energy Outlook 2019 [5]. In this outlook, wind and solar will supply almost 50 % of world electricity in 2050, while in Europe, wind and solar is expected to account for 80 %.

### New challenges for the power systems

The changes in system configuration impact the operating conditions, creating new challenges that the System Operators (SOs) must face and handle. Despite having an overall positive impact on the environment, the intermittent and distributed nature of the new power plants is forcing the system closer to its safe operational limits. Whereas production capability traditionally has been located in large power plants, with a high degree of control on production capacity, the future will see a higher share of uncontrollable, highly volatile production, forcing SOs to measure and model system components more accurately. Further, the traditional measures to ensure safe operation might need to be replaced by means of new technology. Balancing the system in the future will require faster, more accurate actions.

To ensure a reliable power system operation and find ways to withstand severe disturbances, SOs are conducting various analyses regarding power system stability. Power system stability is by Kundur et al. in [6] defined as *"the ability of an electric power system, for a given initial operating condition, to regain a state of operating equilibrium after being subjected to a physical disturbance, with most system variables bounded so that practically the entire system remains intact"*. Further definitions and classification of power system stability can be found in [6] and [2]. The classification from [6] is shown schematically in Figure 1.1. The focus point of this thesis is related to frequency stability.

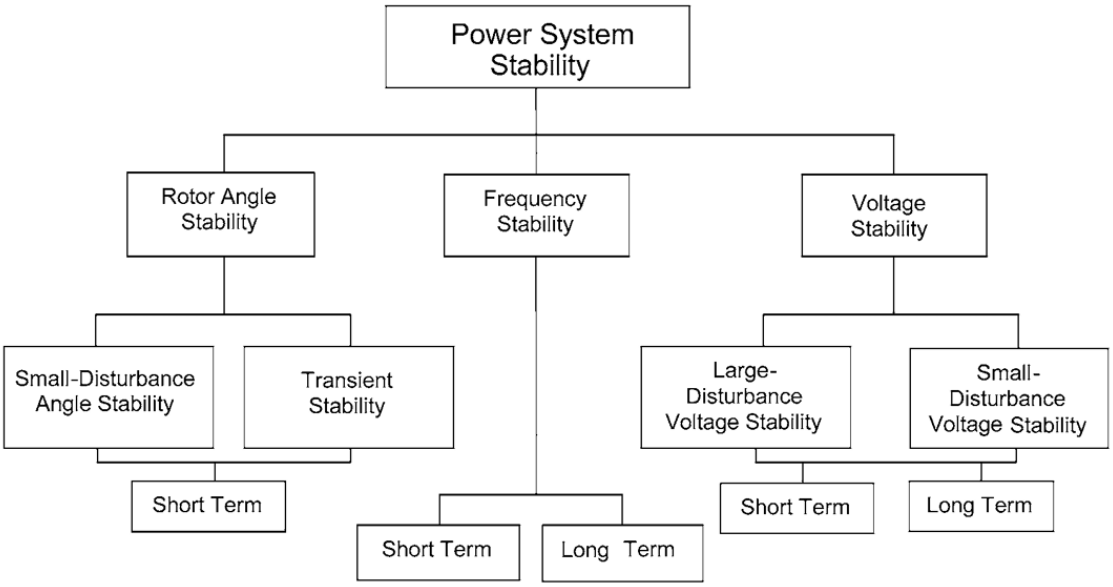



**Figure 1.1:** *Classification of power system stability. The figure is a copy from Kundur et al. [6].*

There are still no good ways to store electricity. Surely, household batteries, and electric vehicle (EV) batteries are doing their purpose, but there are no efficient and economical ways to store large amounts of electrical energy on a larger scale. Therefore, in a power system, the amount of active power generated and imported (hereafter called generation) should always equal the active power consumed and exported, including system losses (hereafter called load). The system operators must ensure that this balance is perfect, and the measure of this balance is the system frequency. The frequency of power systems in Europe has a nominal value of 50 Hz.

In the case of an active power imbalance between generation and load, the frequency will deviate from its nominal value. The frequency of the electric power system should remain in a specific range, close to its nominal value, to ensure that the system components are operating in safe conditions. The normal frequency band in the NPS is 49.9 to 50.1 Hz. If frequency falls below or rises above the normal frequency band, SOs will initiate measures. Initially, this involves using the Frequency Restoration Reserves (FRR) to increase/decrease generation.

In more critical events, for instance right after a disturbance, one can expect the frequency to drift outside the normal operating limits. In these cases, the different power grids define their own frequency limit values and usually list specific actions that should be taken if the frequency drops below, or rises above the set limits. These actions are designed by proper authorities to reduce the risk of damaging system components and prevent a system collapse. Examples of such actions are the disconnection of system components, such as loads in the case of undergeneration, and even power plants in the case of overgeneration. The disconnection of system components could lead to undesired events, such as further uncontrolled system separation, cascade effects or even worse, a total system blackout. Even though there are not many recorded cases in recent history, the impact can be huge; In London, on August 9th, 2019, approximately 1

million customers lost power for 15 to 45 minutes due to a chain of unfortunate events in the power system [7]. In 2003, on September 28th the entire Italian power system blacked out, affecting around 45 million people. Complete restoration of the power system took 19 hours [8].

A power system is frequency stable when it maintains a frequency within the allowed frequency band after a disturbance causing an imbalance between generation and load [6]. As indicated in Figure 1.1, frequency stability includes both short term and long term components. Short term frequency stability refers to a time frame of up to tens of seconds, usually caused by fast frequency deviations occuring due to momentary power imbalance. Long term refers to frequency instabilities that may range from tens of seconds to minutes and could be caused by inadequate amount of reserves to keep the power balance at the nominal frequency. The focus point of this thesis is related to short term frequency stability.

There are three significant contributors or factors regarding short term frequency stability:

- Amount of active power imbalance
- Available reserves
- System inertia

The active power imbalance refers to the momentary difference between generation and load after a disturbance[1]. The larger this active power imbalance is, the greater the instantaneous frequency deviation the power system will suffer. SOs must regard the maximum instant frequency deviation in an N-1 event when planning the future operation. This means that the power system should remain stable even with the loss of one component. In the Nordic Power System, the greatest size of a single disturbance[2] is 1,450 MW, and it is caused by the disconnection of the power plant Oskarshamn 3 [9]. The dimensioning incident defines the required reserve capacity [10], and it is perhaps not so surprising to see that the new HVDC interconnections to Germany and UK have a capacity of 1,400 MW [11, 12].

The power system inertia determines the frequency behaviour the first few seconds after a disturbance. In the event of a disruption causing a power deficit, e.g. the tripping of a central power plant, the immediate response is the release of kinetic energy[3] stored in the rotating masses of the synchronous machines connected to the grid. This effect is referred to as the system's resistance to change. All rotating masses synchronously connected to the system contributes to this resistance with their kinetic energy. As a result of the immediate release of kinetic energy, the synchronous machines slow down, causing the system frequency to decline, creating an underfrequency event. On the contrary, if a power surplus occurs, the inertial response is the absorption of kinetic energy by rotating masses synchronously connected to the grid. In this case, the machines will begin to rotate faster, driving the system frequency up creating an overfrequency event.

[1] In this thesis, when referring to a disturbance, it means the sudden disconnection of a significant component, such as the trip of a big distribution transformer, a HVDC link or an entire power plant. The small load changes that are continuously happening are not considered.

[2] Also called dimensioning or reference incident in literature.

[3] In this thesis, the terms *kinetic energy* and *inertia* are used interchangeably.

A graphical representation of this relationship between inertia, generation and load is presented in Figure 1.2. The impact inertia has on frequency behaviour is analogous to a spring, where the spring's greater stiffness refers to a higher amount of system inertia and vice versa. According to *Future system inertia* [13], *"power system inertia is the ability of a power system to oppose changes in system frequency due to the resistance provided by rotating masses"*. Because the generation mix between operating days and hours change, the system inertia will also change.

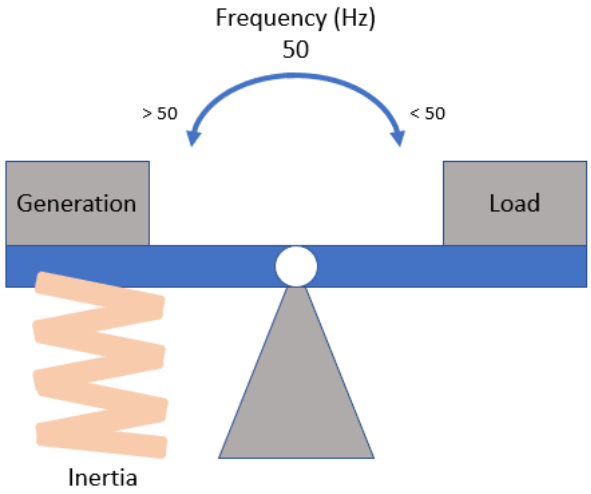



**Figure 1.2:** *Relationship between inertia, generation and load. Inertia is represented by a spring. Adapted from [14].*

Traditionally, overfrequency events are much easier to solve than situations leading to underfrequency. Firstly, the NPS's dimensioning component for an overfrequency incident in the NPS is much lower than an underfrequency event.[4] Secondly, overfrequency situations can be met by simply decreasing generation. This is easily solved by having power plant governors reduce the generator's output, or by having more power transferred out of the power system via HVDC links. Because of this, overfrequency events do generally not normally affect system performance, and hence, the literature and studies on the topic are mostly related to underfrequency events. Underfrequency events require the increase of generation. If there is no way of increasing generation, load has to be cut. This is known as Under Frequency Load Shedding (UFLS), leading to power not being delivered to end-users.

In general, a power system with a lower amount of kinetic energy connected to rotating masses (hereafter called inertia) will be less resilient to frequency deviation. This is because such a system can extract less power from the rotating masses. Effectively, the frequency will fall lower and faster. Having as much inertia as possible in the power system is therefore desired. Previously, inertia connected to power systems has been an "ignored" effect, as most of the generation has originated from big synchronous generators. With the shift of generation to devices connected to the system through power electronics, there is a tendency for system inertia to be reduced, and inertia has become a focus area. In 2013, the Nordic Analysis Group (NAG) initiated the project "Future system inertia". This project aims to improve knowledge of power system behaviour related to system inertia. One of the main interests was to know the amount of system

[4]This will change when the new HVDC cables are finished, but is not a focus of this thesis.

kinetic energy, to plan and maintain system operation as efficiently and securely as possible. NAG has published two reports, *Future system inertia 1* [13] and *2* [15].

## 1.2 Motivation

With system inertia decreasing, imbalances will cause larger instantaneous frequency deviation. Low inertia situations arise in the summer period, often at nighttime. When NordLink, the new subsea cable connecting the Norwegian and German electricity markets, is commissioned, it is expected that the number of low inertia hours will increase. A system having low inertia and at the same time having large dimensioning incidents are leading to an increased risk of UFLS in case of a sudden N-1 imbalance. Currently, in this situation, the Nordic System Operators (SO) is managing this risk by limiting the magnitude of the highest possible frequency deviation [16]. During the summer of 2019, this was achieved by having Oskarshamn 3, the largest power plant connected to the Nordic system, reduce its active power output by 100 MW when inertia fell too low [16]. Having nuclear plants reduce their power output is an imperfect solution, as nuclear plants are driven under strict restrictions. Having them reduce the production takes time, and is sometimes not even possible. If the highest dimensioning fault in the power system is the trip of a HVDC cable, another measure is to limit the capacity of said cable, which is an unwanted intervention in the power market.

Therefore, the Nordic SOs aim to implement very fast activated reserve power, activated in a fault event. The NAG has concluded that implementation of Fast Frequency Reserves (FFR) is the best measure to meet challenges related to low inertia situations. FFR will work as a primary response when system frequency falls below a set threshold. [17]. In order to decide the required volume of FFR, good prognosis tools of system inertia have to be implemented.

## 1.3 Scope

This study will focus on system inertia in the NPS. The Nordic TSOs have since 2015 estimated the kinetic energy of the Nordic power system (Norway, Sweden, Finland and eastern Denmark). This estimation is based on real-time telemetry of individual generators. In "*Ensuring future frequency stability in the Nordic synchronous area*" by *Eriksson et. al* [18], this is called a bottom-up approach. In the same paper, a top-down approach is proposed, estimating inertia based on production per generation type and using a generic inertia constant.

This thesis aims to implement and study a top-down estimation of system inertia. This work will use a combination of visual inspection and data-driven modelling with linear models to study the top-down approach. The models and results will be compared to the historical estimations done by the TSOs with a bottom-up method. This is a step in the current work done by the Nordic TSOs on improving the inertia forecasting tool. The idea of this thesis came from Statnett, where Landssentralen already has a prototype forecasting tool running. However, this thesis is not linked to this prototype but is a free-standing exploratory analysis. Because future information on generating units connected to the system is unknown prior to operation, a bottom-up method is unsuitable for forecasting system inertia. However, forecasts of generation/load are generated

with a 7-day horizon, meaning that a top-down estimation model potentially can be used to forecast system inertia with the same time horizon.

The main goal for the work of this thesis is:

**M** Develop linear models using production data that can estimate the amount of kinetic energy in the Nordic Power System.

The first step to achieve this goal is to create a dataset of historical values of inertia and load/generation, and after that train models using subsets of the dataset to estimate system inertia. The focus is to estimate model accuracy, to explore which features are most relevant, and to compare the differences when adding or reducing constraints on the models.

As this is a master thesis with limited time, the analyses are performed on historical data. Keep in mind that all data used is real historical data, and all analyses are performed in order to potentially increase the model accuracy of forecasts. Forecasting load and consumption in the NPS is done by various companies and service providers, and is NOT a focus in this thesis. In this essence, the results of models created in this thesis can be looked at as "best case" models, with perfect forecasts of generation per production type and price area.

# Chapter 2

# Theory

This chapter provides an overview of the theoretical background concerning the power system relevant to this thesis, including power system frequency and stability. The last section in the chapter contains theory regarding supervised learning and linear regression models, upon which the methods of the thesis is built.

## 2.1 The Nordic Power System

The Norwegian power system is a part of the Nordic Power System (NPS), consisting of Norway, Sweden, Finland and the eastern part of Denmark (corresponding to elspotarea DK2, Figure 2.1.). The frequency is close to uniform within this area, with a nominal frequency of 50 Hz, and a normal frequency band between 49.9 Hz and 50.1 Hz. Through HVDC links, NPS is further connected to the Baltic countries and the continent. Different production mixes characterise the energy systems of the Nordic countries. In the following, the data source is the *International Energy Agency's* data and statistics page on electricity, [19]. In 2020, Norwegian power plants produced 154 TWh, with hydropower accounting for 92 % and 6 % coming from wind power. Towards 2050 the total generation is expected to rise to around 200 TWh, mainly driven by an increased share of wind power plants (25 %), according to Statnetts long-term market analysis [20]. Swedish power plants produced 163 TWh electric energy in 2020, with a combination of mainly hydro- and nuclear power, accounting for 44 % and 30 % respectively. The remaining production is wind power (17 %) or thermal power plants. Statnett expects the share of wind power in Sweden to increase from 23 TWh/y to 85 TWh/y towards 2040, while nuclear power plants will be decommissioned [20]. Finlands yearly production was 69 TWh in 2020, with a mix of power sources; Wind (12 %), Hydro (23 %), Nuclear (34 %) and regular thermal power plants using biofuel (16 %), coal (8 %) or natural gas (5 %). Denmark has mostly wind (57 %) and thermal power plants (39 %), producing 29 TWh in 2020.

#### Price areas in the Nordic Power System

The entire Nordic region is a single electricity market divided into different price areas/Elspot areas. A more extended discussion of the Nordic power market is out of scope in this thesis but is described in detail in [21]. Power within this market is traded in the Nord Pool power exchange [22]. The purpose of dividing the region into price

areas is to take into account limitations in the transfer capacities in the grid. Since November 2011, Norway has five price areas, Sweden has four. Finland operates with only one price area. Denmark is split into two price areas, and the Danish grid has two separated transmission systems. The western part of Denmark is price area DK1 and is connected to the synchronous grid of Continental Europe. The eastern part of Denmark is price area DK2 and is connected to the Nordic synchronous grid. Therefore, in this thesis, only data from DK2 is of interest. The approximate location and borders of each price area are shown in Figure 2.1.

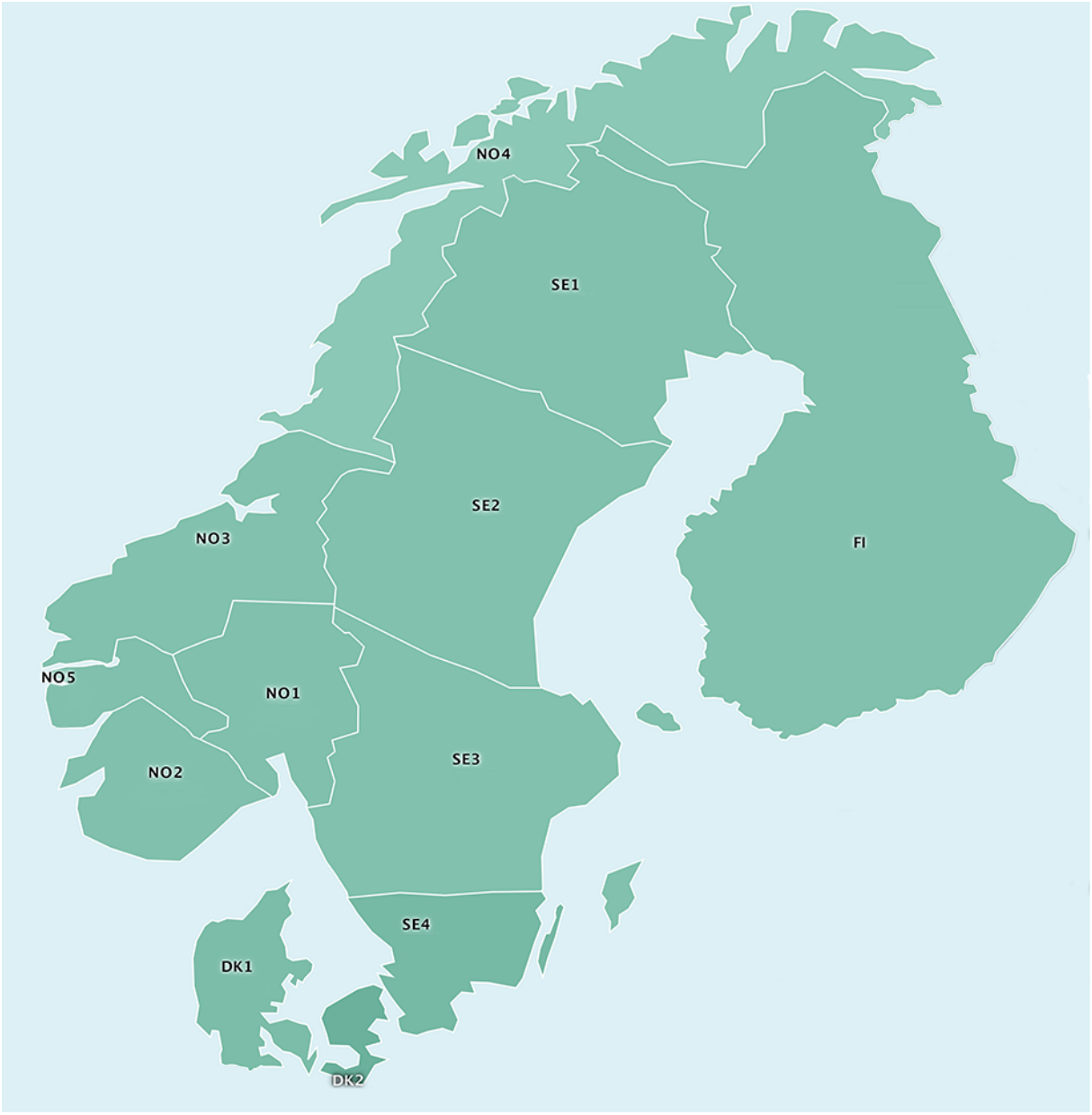



**Figure 2.1:** *Illustration of the Nordic Power System, divided into Elspot areas. Figure is adapted from [23].*

## 2.2 Frequency

As mentioned in the introduction, the frequency of a power system can be considered an index that represents the balance between generation and load. The frequency should stay between specified limits and close to its nominal value. In the NPS, this reference value is at 50.00 Hz. The TSOs continuously monitor the frequency to ensure a stable operation.[1] The total generation should be equal to the total consumption. The system frequency will stay at its nominal value if this criterion is met. Significant deviations can harm the components of the system, and cause interruptions to the electric supply of the customers. In *Definition and Classification of Power System Stability,* by Kundur et al. 2004 [6], frequency stability is defined as the system's ability to keep a steady frequency after an event leading to a significant deviation. The TSOs task is to operate the power system in such a way that equilibrium between generation and load can be restored, without unintentional loss of load.

Immediately following a large power imbalance, for example a trip of a production unit, the deficit in power will be delivered from all synchronously connected rotating masses. That is, power is extracted from the kinetic energy of the rotating masses, causing the rotation to slow down, and thereby the grid frequency decreases [2], [13]. Therefore, initially after a disruption, the generation or power delivered into the system remains the same as before the disturbance. This effect, called the *inertial response*, enables the turbine governors to react to the disturbance. The decreasing frequency will trigger Frequency Containment Reserves (FCR), that injects active power into the system. Eventually, the system will reach the maximum instantaneous frequency deviation, and the synchronously connected machines will begin to accelerate again while the FCR is covering the power deficit. The inertial and governor response (FCR) constitutes the *frequency response* of the power system. Figure 2.2 displays frequency behaviour considering different amounts of kinetic energy in the system. The solid lines represent a system with FCR, and the dotted lines represents a system without FCR. A higher amount of kinetic energy results in a slower frequency drop, with a higher minimum instantaneous frequency. For the cases without FCR, the frequency will continue to drop without recovering.

Frequency deviation may harm the system components, as these are designed to operate close to the system's nominal frequency. A drifting frequency is a risk mainly for the synchronous machines connected to the system, generators and synchronous motors, as the internal windings may experience irregular current flow and become overloaded. For this reason, synchronous generators are equipped with relays set to trigger in the events of over- or underfrequency conditions. If a generator disconnects because of underfrequency, this will cause further frequency drop since the imbalance between generation and load is amplified. A cascade effect - or even a blackout - can be provoked since the additional frequency drop may trigger more generators protection schemes. In such situations, system operators have to disconnect loads to prevent a system collapse or blackout and restore the system balance. This is called Under Frequency Load Shedding.

[1] A real time visualization is provided by Statnett on their webpage `https://www.statnett.no/en/`.

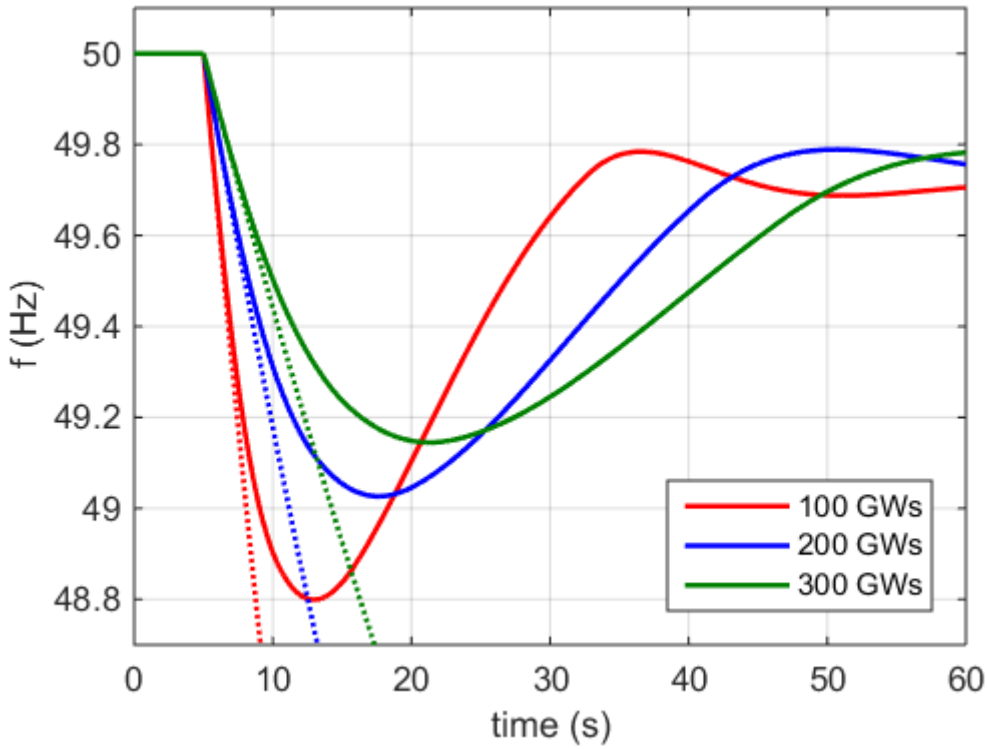



**Figure 2.2:** *The effect the amount of kinetic energy on frequency behaviour following a loss of production. Solid lines represents a system with FCR, and dotted lines represents a system without FCR. The figure is copied from Future System Inertia [13].*

## 2.3 Swing equation

To understand the inertial response of a power system, one should first study the swing equation. This equation explains the main physics behind the concept of frequency change. Derived from Newton's second law for rotation, the swing equation relates the acceleration or deceleration of a synchronous generator and turbine to the imbalance between mechanical- and electrical torque. For a single generator $G_i$, Newton's second law of rotation gives

$$J_i \frac{d\omega_{mi}}{dt} = \tau_{mi} - \tau_{ei}. \tag{2.1}$$

Here, $J_i$ is the total moment of inertia of the turbine, shaft and rotor given in [kgm$^2$], $\omega_{mi}$ is the mechanical angular velocity of the rotor in [rad/s], $\tau_{mi}$ is the mechanical torque in [Nm], and $\tau_{ei}$ is the electromagnetic torque in [Nm]. This equation is called the swing equation. It is possible to write this equation by means of power simply by multiplying by $\omega_{mi}$, so

$$\omega_{mi} J_i \frac{d\omega_{mi}}{dt} = P_{mi} - P_{ei}, \tag{2.2}$$

where $P_{mi} = \omega_{mi}\tau_{mi}$ is the mechanical input to the generator expressed in [W] and $P_{ei} = \omega_{mi}\tau_{ei}$ is the electrical power output of the generator, also in [W]. The inertia constant $H_i$ of a single generator $G_i$ is given by the following equation:

$$H_i = \frac{1}{2} \frac{J_i \omega_{mi}^2}{S_{ni}}. \tag{2.3}$$

In this equation, it is evident that $H_i$ is a function of $\omega_{mi}$. However, it is a reasonable assumption that even during frequency events, $\omega_{mi}$ does not deviate significantly from the generators rated mechanical velocity $\omega_{msi}$. This way, $H_i$ can be considered to be a constant given by

$$H_i = \frac{1}{2}\frac{J_i\omega_{msi}^2}{S_{ni}}. \tag{2.4}$$

The inertia constant is given in seconds [s] and represents the time that it takes to immobilize a generator rotating at synchronous speed, while no mechanical power is provided and still delivering electrical power. Depending on the type of generator, inertia constants are in the range of 1-10 seconds [3], [24].

## 2.4 Inertia of a power system

If inertia constants and rated apparent powers of individual turbine-generators are known, the inertia of an entire power system $H_{sys}$ can be calculated as

$$H_{sys} = \frac{\sum_{i=1}^{N} S_{ni}H_i}{S_{n,sys}} \text{ [s]}, \tag{2.5}$$

with $S_{n,sys} = \sum_{i=1}^{N} S_{ni}$ and $S_{ni}$ being the rated apparent power of generator $i$ in [VA] and $H_i$ being the inertia constant of the same generator $i$.

Synchronously connected motors to the power system is also contributing to system inertia and should be included in the same way as generators.

The TSOs in the NPS is not monitoring the system inertia in seconds, but are instead monitoring a closely related measurement, namely the kinetic energy of rotating masses connected to the system. System kinetic energy can be expressed in terms of energy [MWs] or [GWs] by writing Equation 2.5 as:

$$E_{k,sys} = S_{n,sys}H_{sys} = \sum_{i=1}^{N} S_{ni}H_i \text{ [GWs]}. \tag{2.6}$$

## 2.5 Frequency behaviour

There will always be power changes in consumption and production. Small power changes are only visible as noise in the frequency, because of the large number of rotating machines contributing with inertia in the power system. However, for larger power imbalances, the system frequency can deviate further from the nominal value. The change in frequency is depending on the actual power imbalance $\Delta P$, and on the amount of system inertia in that instant. The dynamic behaviour of every individual generator can be described using the swing equation:

$$H_i\frac{df_i}{dt} = \frac{f_n^2}{2S_{ni}f_i}\left(P_{mi} - P_{ei}\right), \tag{2.7}$$

where $f_i$ is the frequency of generator $i$, $f_n$ is the nominal frequency, $P_{mi}$ is the mechanical power of turbine-generator $i$, and $P_{ei}$ is the electrical power of generator $i$ [13].

Equation 2.7 shows that an imbalance between mechanical and electrical power of a generator will result in the change in frequency with time. This frequency derivative will be dependent on the imbalance $(P_{mi} - P_{ei})$, and inertia $H_i$ of the generator.

In a power system there will be a large number of generators, and a transmission network connecting generators to the load. Hence, the frequency drop due to an imbalance in one part of the system will not transmit uniformly. The generators with the smallest electrical distance to the fault location will be the first reactors. This is elaborated upon in *Future system inertia 1* [13], which also includes a worked example of the frequency response in a simplified network consisting of two generators and one load.

## 2.6 Ancillary Services

Ancillary services are operational reserves procured by the Transmission System Operators to balance supply and demand and maintain power quality. Examples of services include voltage control, black start capability and grid loss compensation. Since this thesis focuses on frequency control, this subsection will focus on the balancing market. As stated previously, the frequency in the power system is constantly changing. To adjust and fine-tune the balance between load and generation, the Nordic TSOs are buying reserves in balancing markets. The Nordic region divides the balancing market into

- Primary reserves - Frequency Containment Reserves (FCR)
- Secondary reserves - automatic Frequency Restoration Reserves (aFRR)
- Tertiary reserves - manual Frequency Restoration Reserves (mFRR)

Figure 2.3 illustrates the activation and response time of the three types of reserve, in addition to the inertial response. FCR and aFRR are activated automatically in response to measured frequency changes, and tertiary reserves (mFRR) are activated manually by the Nordic TSOs if needed [21], [25].

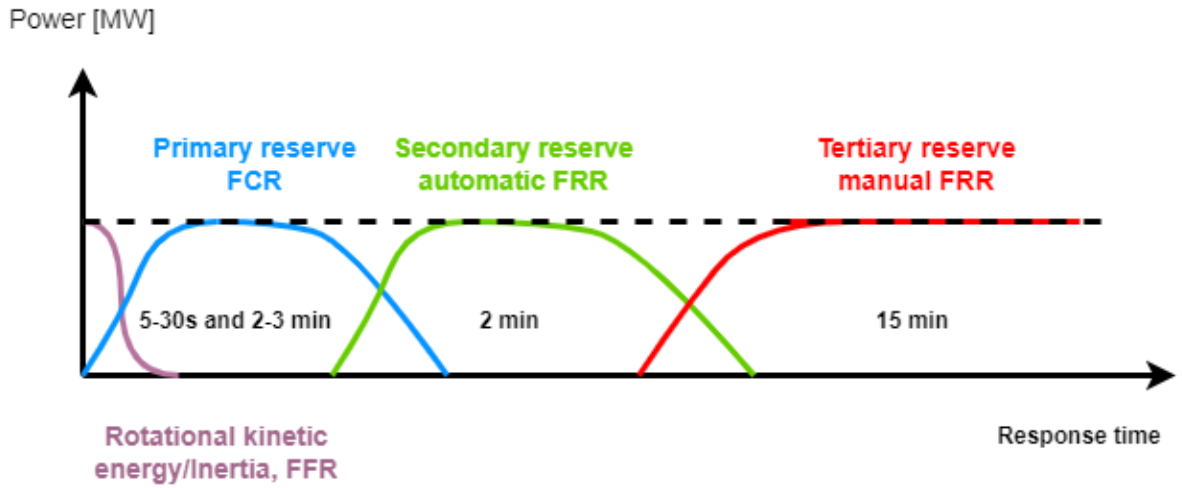



**Figure 2.3:** *Illustration of activation and response time of the balancing reserves in the Nordic Power System. Primary reserves (FCR), secondary reserves (aFRR) and tertiary reserves (mFRR).*

FCR, also called primary control, is the part of the control system that deals with momentary imbalances between production and consumption [26]. The inertial response of a power system is making the frequency of the power system change, and this change activates primary reserves, FCR. Reserve capacity that responds as expected is fundamental to maintain a safe operation of the power system.

### 2.6.1 FCR-N

FCR-N (Frequency Containment reserves for normal operation) is a symmetric power reserve, automatically activated with frequency deviation within the normal frequency band $50 \pm 0.1$ Hz. These reserves should be regulated out within 2 minutes. The total capacity of FCR-N in the NPS today is 600 MW, where 210 MW is located in Norway. Each subsystem must have 2/3 its own FCR-N, for each subsystem to be self-sufficient and able to be run in islanded operation until a major disturbance is cleared. The remaining part of FCR-N can be power exchanges between subsystems. In Norway and Sweden, hydropower is used exclusively as FCR-N [27].

### 2.6.2 FCR-D

FCR-D (Frequency Containment reserves for disturbances) is a continuation of FCR-N, activated when the frequency drops below 49.9 Hz. At the moment, this reserve is unsymmetric, only activated in under frequency events. The TSOs set the requirements for FCR-D:

- FCR-D should activate at 49.9 Hz and be fully activated at 49.5 Hz.
- Required FCR-D capacity should equal the largest possible imbalance, or *dimensioning fault*.
- Activation of FCR-D should not create other problems in the power system
- Each subsystem should be able to supply 2/3 of its FCR-D itself

In the case of a significant disturbance, the following requirements should be met:

- 50 % of FCR-D should be activated within 5 seconds after frequency drops below 49.5 Hz.
- 100 % of FCR-D should be activated within 30 seconds after frequency drops below 49.5 Hz.

### 2.6.3 Fast frequency reserves, FFR

Following the work on *Future system inertia 1* [13] and *2* [15], the Nordic Analysis Group concluded that the most efficient and cost-effective solution to mitigate low inertia situations is to implement a new type of reserve [28]. In such situations, FCR-D is not reacting fast enough.

Fast Frequency Reserves (FFR) will work as a power response, activating within a second after system frequency drops below a threshold. A pilot project by Statnett in 2018 [29], involved contributions from industry, pump storages and datacenters. The pilot was a success, and Statnett gained knowledge on the availability of FFR from different technologies.

Further demonstration projects were run in 2020 and 2021. Both years, Statnett procured FFR through two different contract options in the summer season (defined as May through September.)

- **FFR Profile**, in 2021 delivered from week 20 through week 36 (May 17 to September 12). This reserve capacity was procured during the entire weekdays at nighttime (between 10 pm until 7 am the following day). In addition, FFR Profile was procured the whole weekend).
- **FFR Flex** was bought on a short term prognosis, when forecasts showed that FFR need was greater than the amount already available through FFR Profile.

In 2020 Statnett bought a total volume of 27.2 MW FFR Profile, at a total cost of 4.6 MNOK. The recieved bids did not cover Statnetts demand of 126 MW FFR Profile, and no bids fulfilled the techical requirements of FFR Flex. [30]. In 2021 Statnett accepted bids of 51.18 MW FFR Profile at a price of 112 NOK/MW/hour. 68.3 MW FFR Flex was also bought at a uniform market price of 495 NOK/MW/hour [31].

## 2.7 Inertia online estimation

Previously, much work has been laid down to provide SOs with a tool to estimate system inertia real-time. The Nordic Analysis Group (NAG) did in *Future System Inertia* [13] argue that inertia estimation tools using measured frequency deviation during disturbances is challenging, and also only able to provide inertia estimation during actual disturbances. Because of this, the same authors in chapter 4 in the same paper [13] describe the implementation of a real-time inertia estimation currently used and shared between the Nordic TSOs.

Each TSO in the nordic cooperation have information about their own production, so the real-time estimation is carried out so each TSO is making their own real-time estimation of its own area. For an estimation of inertia at a nordic level, the estimation from each TSO is combined. Chapter 4 in [13] is describing, per area, how the real-time inertia estimation has been performed.

### 2.7.1 Calculation of kinetic energy

Kinetic energy in the NPS can be estimated by knowing the circuit breaker positions of production units. When a generator is connected, i.e. the circuit breaker is in closed position, it is assumed that the generator can contribute to the kinetic energy of the system. The kinetic energy $E_{sys}$ can then be calculated according to Equation 1.4 in [13], which is rewritten here as

$$E_{sys} = \sum_{i=1}^{N} S_{ni} H_i \text{ [GWs]}, \tag{2.8}$$

where the inertia constant $H_i$ and rated power $S_{ni}$ is retrieved from a generator register, and $N$ is the number of connected generators.

In terms of kinetic energy capacity of each country's power system, values are summed up in Table 2.1. If all available kinetic energy capacity is provided in the Nordic power system, the value is 390 GWs.

**Table 2.1:** *Available kinetic energy in each country in the NPS. Data from Future System Inertia 1 [13].*

| Area | Kinetic energy capacity (GWs) |
|---|---|
| Sweden (SE) | 170 |
| Norway (NO) | 100 |
| Finland (FI) | 90 |
| Denmark (DK2) | 30 |
| **SUM** | 390 |

### 2.7.2 Real time measurements

The terms *inertia* and *kinetic energy* is often used interchangeably in literature. Nordic TSOs each make real time inertia estimations in their own area, and the values are added up to get an estimation of the total kinetic energy of the power system. Fingrid is providing real time measurements of the total kinetic energy of the NPS[2]. Figure 2.4 shows data from Fingrid representing the total kinetic energy estimation in the NPS February 2020. There are clear variations in night/day, and also differences in week/weekday. (February 1, 2020 was a Saturday).


**Figure 2.4:** *Stored values of real time estimation of kinetic energy in the NPS, February 2020. The figure is a screenshot from [32]*

[2]Available at `https://www.fingrid.fi/en/electricity-market/electricity-market-information/InertiaofNordicpowersystem/` [32].

## 2.8 Supervised learning and linear models

The following section describes the theory behind the method used in this thesis. The main goal of this thesis is to develop linear models using production data that can estimate the amount of kinetic energy in the Nordic Power system.

### 2.8.1 Supervised learning

The common principle underlying all supervised machine learning algorithms for predictive modelling is as follows:

The machine learning algorithm can be described as a function ($f$) that maps input variables ($X$) to an output variable ($Y$).

$$Y = f(X). \tag{2.9}$$

Supervised learning learns this model based on training on known response values. Later, the trained model can be used to predict the outcome of previously unseen data. Figure 2.5 displays a typical roadmap for supervised machine learning problems. In the following, each step will be briefly explained.

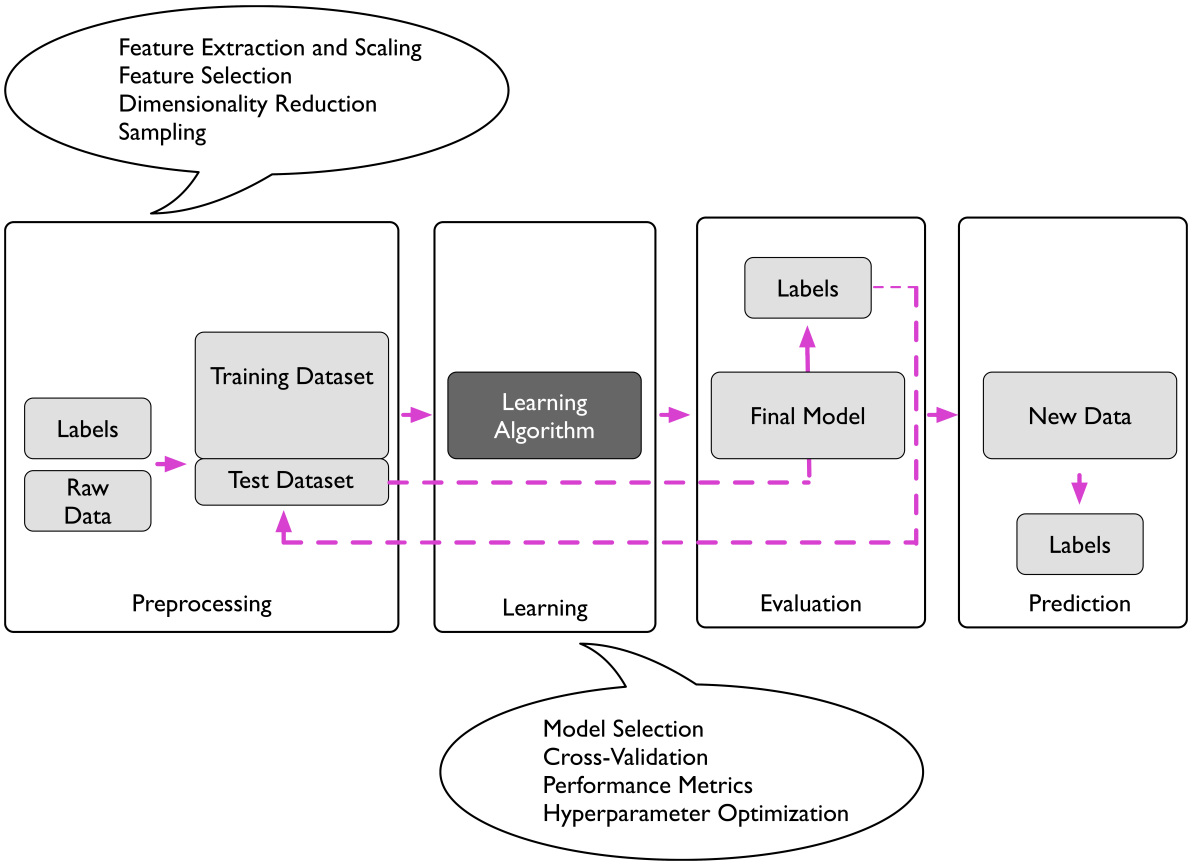



**Figure 2.5:** *Roadmap for a machine learning algorithm using supervised learning. The four boxes illustrate the essential steps in preprocessing and development of a machine learning system. The figure is reprinted with permission from [33].*

**Step 1: Preprocessing**

The first main box in Figure 2.5 is the first, and arguably the most important step when developing a machine learning model. Before a dataset can be input to a machine learning model, the data must be preprocessed to the correct format. This can involve *Feature Extraction and Scaling*, *Feature Selection*, *Dimensionality Reduction* and *Sampling*. After preprocessing, data is usually split into two parts - a training and test dataset. In the first learning step, the training set with the known outcome is input to the learning algorithm [33].

**Step 2: Learning**

The next step to perform after the preparation of data is to apply one or more machine learning models on the training dataset prepared in step one. Different techniques, as displayed in Figure 2.5 can be performed to fine tune the models.

**Step 3: Evaluation**

After the model is trained on the training dataset, it is time to compare the performance of the model on the test dataset. For this purpose, the chosen performance metric is applied. After this step, the model can be evaluated. The model, trained on the training dataset, is now applied to the test dataset, and predicts outputs that can be compared to the known actual values. If the performance is below expectations, the usual procedure is to go back to step 2 and further develop the model.

**Step 4: Prediction**

When the model's performance in step 3 is adequate, the model can be used for predicting new, unseen data - and the predictions can then be used for instance in forecasting.

### 2.8.2 The bias-variance trade-off

With access to training data, a crucial part of supervised learning is to obtain a good compromise between **bias** and **variance** [33].

**Bias** is the difference between the average prediction of a model and the correct value it is trying to predict. A model with a high bias pays very little attention to training data and oversimplifies the model. Such models always lead to high error on training and test data [34].

**Variance** is the variability of model prediction for a given data point or a value that tells how scattered the data is. A model with high variance pays a lot of attention to training data and does not generalize well on unseen data. As a result, such models perform very well on training data but can have high error rates on test data [34].

Mathematically, let's denote $Y$ as the variable to predict, and $X$ are the explanatory variables. Assumed there is a relationship between the two such that $Y = f(X) + e$, where $e$ is the error term, normally distributed with a zero mean. Modelling $\hat{f}(X)$ of $f(X)$, using linear regression or another modelling technique, the expected squared error at point $x$ is

$$Err(x) = E[(Y - \hat{f}(x))^2]. \tag{2.10}$$

which can be further decomposed as

$$\begin{aligned} Err(x) &= \Big(E[\hat{f}(X)] - f(x)\Big)^2 + E\left[\Big(\hat{f}(x) - E[\hat{f}(x)]\Big)^2\right] + \sigma_e^2 \\ Err(x) &= Bias^2 + Variance + IrreducibleError \end{aligned} \tag{2.11}$$

The first term is the squared bias, the amount by which the average of the estimation differs from the true mean. [35]. The second term is the variance, the expected squared deviation of $\hat{f}(x)$ around its mean. The last term is the variance of the target variable around its true mean. It can be regarded as noise in the data, and cannot be avoided no matter how well $f(X)$ is estimated, unless $\sigma_e^2 = 0$ [34], [35].

### Underfitting and overfitting

In supervised learning, the model can be characterized by how good it captures the underlying patterns in the data. This is visualized in Figure 2.6. The situation shown in Fig 2.6 a) shows a model that is too general and unable to capture the underlying pattern of the data. This is called **underfitting** [34]. Such models usually have high bias and low variance. This can happen when the amount of data available for training is low or trying to build a linear model on data that is nonlinear.

Models that capture noise along with the underlying pattern in the data are called **overfitting** [34]. This can happen when a model is trained a lot on a noisy dataset. Such models have low bias and high variance. Referring again to Fig. 2.6, the model shown in b) is very sensitive to noise or randomness in the training data.

Fig. 2.6 c) shows a good compromise between having high bias and high variance, and would likely be general enough to give good predictions if applied on previously unseen data.

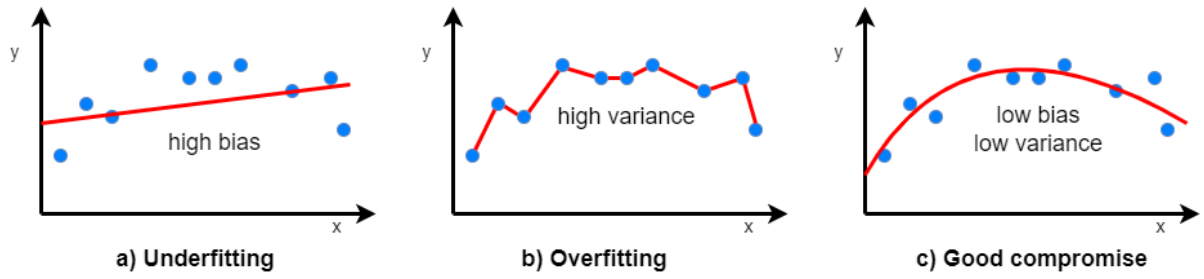



**Figure 2.6:** *Illustration of the concept bias-variance tradeoff. The model shown in* ***a)*** *is underfitting. It is too general to capture the underlying pattern in the data. In* ***b)*** *the model is too complex and is overfitting.* ***c)*** *shows a good compromise. The figure is recreated from [34].*

To summarize, the bias-variance tradeoff arises because a model that is too simple and has few parameters will have high bias and low variance. Making the model more complex, for instance by having more parameters might cause the model to have high variance and low bias. An algorithm can not simultaneously be more complex and less complex at the same time, and that is the reason for the bias-variance tradeoff [34].

### 2.8.3 Regression for predicting continuous outcomes

Regression analysis is the prediction of continuous outcomes. This thesis is an example of a regression problem. Given a number of explanatory variables, the goal is to predict a continuous response. Here we try to find a relationship between those variables that allows us to predict an outcome.

Figure 2.7 illustrates the concept of linear regression. Given a predictor variable $x$ and a response variable $y$ we fit a straight line to this data that minimizes the distance (most often the average squared distance) between the sample points and the fitted line. The intercept and slope learned from this data can be used to predict the outcome variable of new data.

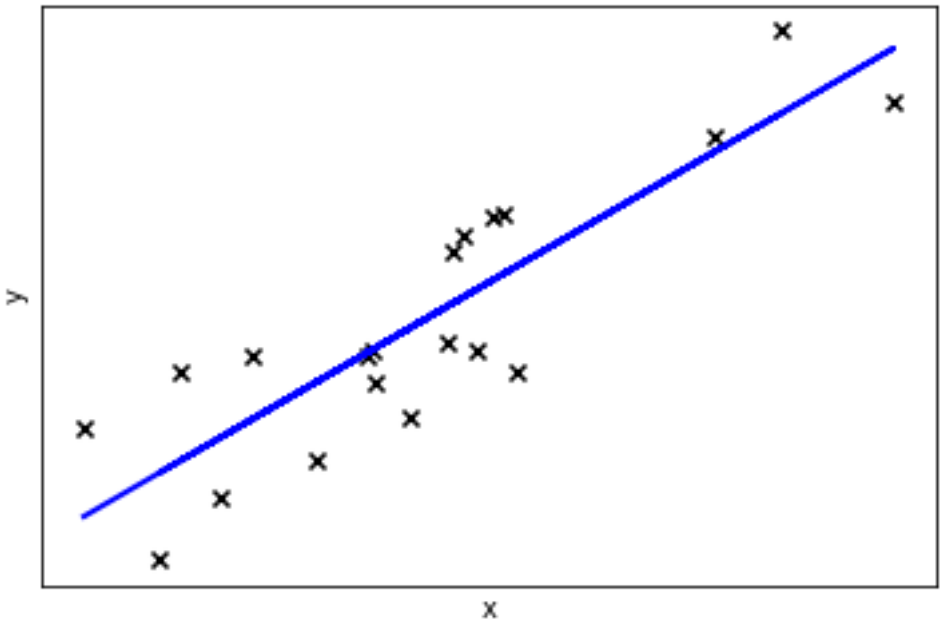



**Figure 2.7:** *Illustrating the concept of linear regression. Given a number of explanatory variables $x$ fit a straight line that minimizes the distance to the dependant variable.*

### 2.8.4 Regression algorithms

There are many common regression algorithms. Some are linear, others tree-based, and again other are memory-based. The common point is that the goal of all algorithms is to predict a response value based on a set of explanatory variables.

The first and most common algorithm to test is the linear regression, described below:

#### 2.8.4.1 Linear regression

A linear regression model assumes that there is a linear relationship between input and target variables. Multiple linear regression is a generalized version of the simple linear regression, and is described with the following equation,

$$\hat{y} = w_0 x_0 + w_1 x_1 + ... + w_m x_m = \sum_{j=0}^{m} w_j x_j = \sum_{j=0}^{m} w^T x, \tag{2.12}$$

where $\hat{y}$ is the target variable. $w_0$ is the $y$-axis intercept with $x_0 = 1$. $w_1$ to $w_m$ are unknown regression coefficients and $x_1$ to $x_m$ are the explanatory variables. The regression coefficients can be found using training data and apply the least squares approach so the residual sum of squares (RSS) is minimized [35].

$$\begin{aligned} RSS(w) &= \sum_{i=0}^{N} (y_i - \hat{y}_i)^2 \\ &= \sum_{i=0}^{N} \left( y_i - \sum_{j=0}^{m} w_j x_{i,j} \right)^2, \end{aligned} \tag{2.13}$$

where $i$ is the measurement number from 1 to the total number of measurements $N$ and $j$ is the variable number from 1 to $m$. It can be shown mathematically that the least square solution is

$$\hat{w} = (X^T X)^{-1} X^T y. \tag{2.14}$$

Finally, the target variables can be estimated with Eq. 2.12. In order to obtain a better bias-variance trade-off, it is often useful to apply different shrinkage methods, also called regularization. Such methods limits the regression coefficients by penalising their size [35].

#### 2.8.4.2 Ridge regression

Ridge Regression is one of the most popular approach to tackle the problem of overfitting [33]. A model trained with ridge regression will have coefficients that are restricted. The ridge coefficients minimise a penalised residual sum of squares as follows

$$w^{Ridge} = \underset{w}{\operatorname{argmin}} \, ||Xw - y||_2^2 + \alpha ||w||_2^2, \tag{2.15}$$

where the complexity parameter $\alpha \geq 0$ controls the amount of shrinkage. By increasing the value of $\alpha$, the regularization strength is greater - meaning the weights will be shrinked. A lower $\alpha$ indicate a low degree of penalisation which allows the regression coefficients a higher value. Worth noting is that the model is not shrinking the intercept term $w_0$.

#### 2.8.4.3 Lasso regression

Where the Ridge regression penalizes the squared sum of coefficients $\sum_{j=1}^{n} w_j^2$, Lasso penalises the absolute value of regression coefficients $\sum_{j=1}^{n} |w_j|$. Lasso regression can shrink some of the regression coefficients to zero by increasing the penalisation term $\alpha$. This makes Lasso useful as a supervised feature selection technique [33].

### 2.8.5 Performance metrics

In the evaluation of regression models, there are many scoring metrics that can be used. In this thesis the relevant metrics is the *mean squared error* (MSE) and *mean absolute error* (MAE). These are examples of scale dependant errors. These error metrics can be mathematically expressed as:

$$MAE = \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{n} \tag{2.16}$$

$$MSE = \sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n} \tag{2.17}$$

where $\hat{y}_i$ is the predicted value by the model, and $y_i$ is the true target value.

A lower MSE or MAE value indicate a better score.

The *mean absolute percentage error* is a percentage error metric that allows for easier comparison between datasets of different scales. It can be calculated as:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \tag{2.18}$$

# Chapter 3

# Data and Method

This chapter will present the methodology used in this thesis[1]. To analyse the relationship between the production level and the inertia level in the NPS, data were first acquired from relevant sources. As the following sections will show, the data quality was questionable, and because of this, considerable time was spent to ensure data going into the models was of as good quality as possible. Data from the NPS in the period January 2018 until December 2019 were collected, and it was decided to use data from Januar 2018 to June 2019 to train models and test the models on data from July and August 2019. Initially, modelling was done on the entire dataset, but based on the first results, a decision was made to split the data into regional parts, with separate analyses being performed for each country. The reasoning is that the estimation of system inertia is performed differently in the four countries.

Python [36] was used for all data processing, analysis and modelling, mostly with Jupyter notebooks [37], where the data were first unpacked into Pandas data frames. Pandas is a library that simplifies managing and manipulating numerical tables and time series in Python [38], [39]. Models were made using Scikit-learn "*a Python module integrating a wide range of state-of-the-art machine learning algorithms for medium-scale supervised and unsupervised problems*" [40]. Graphs and visualisations were created with Plotly [41], or by inbuilt tools in Pandas, built on top of the Matplotlib [42] library. Seaborn [43] was used to control figure aesthetics.

This chapter is split into four main sections: Section 3.1 describes how inertia data was collected and preprocessed. Section 3.2 describes how production data was collected and preprocessed. Section 3.3 explains the motivation for using linear models, and finally Section 3.4 describes how the linear models were applied on the data.

Currently, the production and consumption levels in the NPS is commonly stored in hourly resolution, and although one of the sources of data (Fingrid) offered a more fine-grained resolution, it was decided to use hourly resolution throughout the work.

[1]When reading this chapter, keep in mind that most of the work performed in this thesis, including data collection and preprocessing was done February through April 2020, while the thesis was submitted in February 2022.

## 3.1 Inertia data

Inertia, or *the kinetic energy of rotating masses connected to the power system* can be estimated by monitoring the real-time telemetry of individual generators. The Nordic TSOs have since 2015 estimated and stored this data, published openly on two different platforms, Fingrid Open data and Energi data service (maintained by Energinet). In addition, Statnett stores the data in its datawarehouse "Innsikt", in the premade report "*SD056 Inertia (time)*". However, this data is not openly accessible but requires user-access via Statnetts internal systems. A brief comparison between these three data sources is given in Table 3.1. Because of the expected amount of data preprocessing and early observations on data quality at the start of the thesis work, a decision was made to focus on data from January 2018 through December 2019. This section describes how this data was collected and preprocessed.

**Table 3.1:** *Inertia data sources.*

| Platform | *Innsikt* | *Fingrid Open data* | *Energi data service* |
|---|---|---|---|
| **Public access** | No | Yes [44] | Yes [45] |
| **Temporal resolution** | Hour | Minute | Hour |
| **Inertia by country** | Yes | No | Yes |
| **Data available from** | 2015-03-27 | 2015-03-27 | 2019-10-26 |

### 3.1.1 Collecting inertia data

The first step towards creating a dataset was to acquire the historical inertia estimates. As described in section 2.7, the Nordic TSOs estimate and store the current level of rotational kinetic energy, typically in units of GWs or MWs. Access to Statnetts storage of this data was acquired and Figure 3.1 displays the historical estimates of inertia as recorded by Statnett in the period January 2018 to December 2019. It is a time series with recorded hourly inertia values per country. As Figure 3.1 reveals, data from this source is partly missing. Specifically, data from the Swedish power system (InertiaSE) is recorded with the value 0 from the beginning of 2018 until June 5, 2018, while data from the danish power system (InertiaDE) is missing (having no value) during the summer of 2019, from March 5 until July 23. In addition, there are missing or zero values spread on the whole dataset. This can be observed as downward spikes in *Inertia sum*.

Due to the missing data from Sweden and Denmark, other sources of inertia data were considered. Initially, data downloaded through Fingrids data portal (*Kinetic energy of the Nordic power system - real-time data)* [44] looked promising, but during preprocessing, this source was also found to contain missing values and unclear transition from summertime to standard time. When discussing these matters with the listed contact person for the dataset, he offered to pull data directly from Fingrids main database as a workaround [46]. This data was found to be complete without many obvious inaccurate values. However, the data only had information about the total level of inertia in the NPS, not split by country. It was decided to combine the dataset obtained from Fingrids main database with the Innsikt data to create a complete dataset including the total Nordic level of inertia, and inertia split by country.


**Figure 3.1:** *Inertia as stored by Innsikt. The figure displays the kinetic energy in GWs for Denmark (blue), Finland (yellow), Norway (green) and Sweden (red). The purple curve is the aggregated total.*

As a side note, it is worth mentioning one more source of inertia data, namely the one published by Energinet in the open data platform *Energi data service*, named *Inertia, Nordic Synchronous Area* [45]. This dataset has stored values split by country since October 26, 2019 and uses the ISO 8601 standard [47] on the time and date-related data points. It is an ideal data source if one wants to look at data from 2020 and 2021. However, this is out of scope in this thesis, as the data source was discovered too late in the work process.

### 3.1.2 Preprocessing inertia data by combining several sources

Because of the missing data from Sweden and Denmark, it was decided to combine the time series from Fingrid, and the time series from Innsikt. When comparing these sources, they were not time-aligned. Following this, the data from all available sources of inertia data was cross checked, and this revealed that data from Innsikt was lagging by 1 hour compared to all other data sources.

Cross-checking data and visualizing also identified some irregularities in the time series that will be discussed below.

### 3.1.3 Identifying periods with missing values

#### Missing data

At this point in preprocessing, the two datasets from Fingrid (received 2020-02-04) and Innsikt were aligned hourly and adjusted to a common timezone, Coordinated Universal

Time (UTC). Further, to identify time periods with potential problems in the Fingrid dataset, the hourly change of inertia value was calculated and plotted. Figure 3.2 shows a visual representation of the resulting time series. This revealed that hourly changes in the interval [-25, 25] GWs is normal. All changes exceeding these, including the positive and negative spikes found in late 2018, were further analysed manually by looking at the higher resolution minute data on the Fingrid website [32]. The following time periods had missing data, and were consequently removed from the dataset.

- 2018-10-11 hour 8 and 9 UTC
- 2018-11-01 hour 15 to 23 UTC
- 2018-11-08 hour 15 to 2018-11-09 hour 10 UTC
- 2018-11-30 hour 10 to 12 UTC

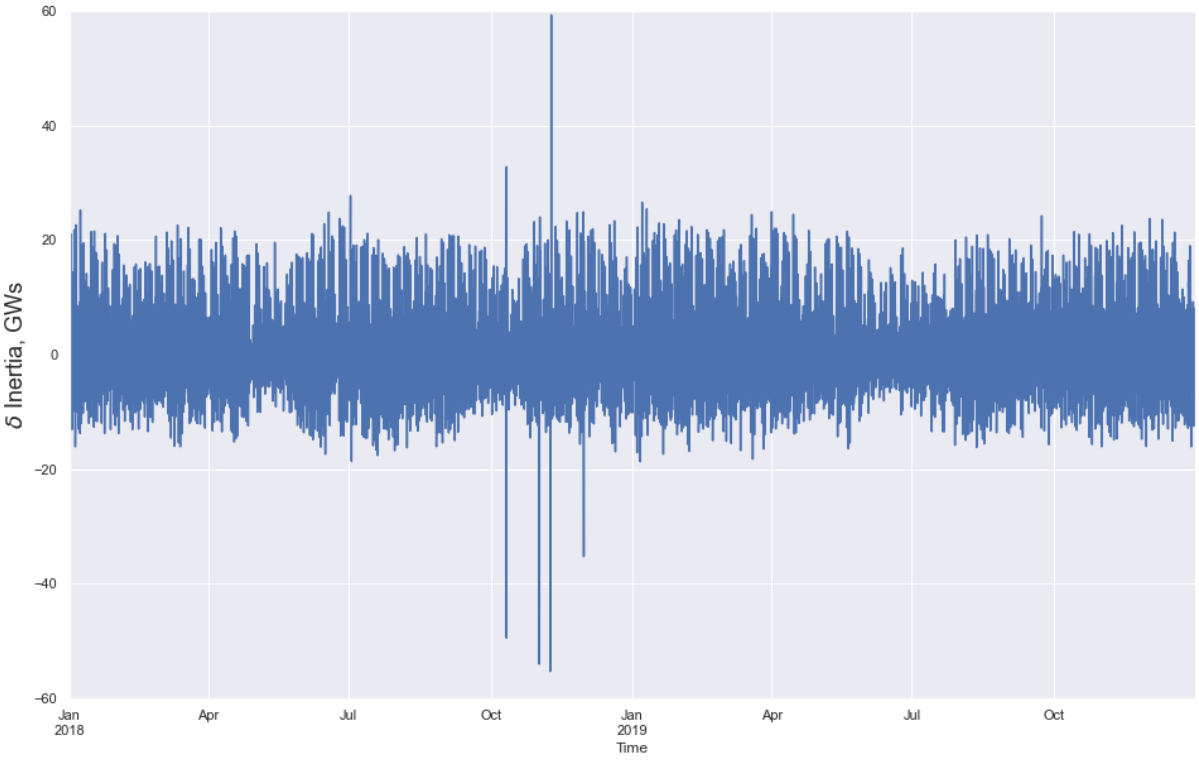



**Figure 3.2:** *The figure displays the hourly change between inertia values stored in the Fingrid data. Note the outliers in late 2018.*

These time periods were also investigated in the Innsikt data, often revealing missing data values from one or more countries, indicating that something has gone wrong in either the estimation or the storage of the values. The conclusion was to remove the data from these timeslots, preventing this data from being part of model training.

### Imputing missing swedish and danish data

The next step was to impute missing values in the hours where such imputations were possible. The total value for the inertia estimation in the Nordic countries *InertiaNordic* is a sum of the inertia estimations from Norway, Sweden, Denmark and Finland.

$$InertiaNordic = InertiaNO + InertiaSE + InertiaDE + InertiaFI$$

Because of this, during the two main periods with missing data from a single country[2], these missing values can be imputed by taking the value of the total kinetic energy, and subtract the values from the other countries. This will work for every hour where exactly one of the four national inertia estimations is missing.

For instance, the missing inertia values from the swedish data during the first months of 2018 can be estimated the following way:

$$InertiaSE = InertiaNordic - (InertiaNO + InertiaDE + InertiaFI)$$

Similarly, the missing danish data during 2019 can be imputed with

$$InertiaDE = InertiaNordic - (InertiaNO + InertiaSE + InertiaFI)\,.$$

These imputations were performed for every timestep where precisely one of the four countries had missing values and filled most of the missing values in the dataset. The remaining rows with missing data all included at least two missing measurements. Rather than using other imputing techniques such as forward or backward filling, it was decided to leave any remaining row with missing data out of modelling, as the number of rows including missing data were down to 30 out of a total of 17520 rows[3].

### 3.1.4 Final dataset of Nordic inertia

The result after this section was the creation of the file *NordicInertiaWithNaN.csv*. There are five variables in the inertia dataset, listed in table 3.2. The index column are timestamps, representing each hour in the year from 2018-01-01 00:00 to 2019-12-31 23:00. The top 10 rows of the dataset are displayed in Appendix A, Table A.1.

**Table 3.2:** *Columns in the dataset NordicInertiaWithNaN.csv. The column InertiaNordicGWs is the sum of the other four columns, and all units are in GWs.*

| Column name | Description |
|---|---|
| **InertiaNordicGWs** | Inertia of production units in the Nordic synchronous area |
| **InertiaDK** | Inertia of production units in Denmark (Price area DK2) |
| **InertiaFI** | Inertia of production units in Finland |
| **InertiaNO** | Inertia of production units in Norway |
| **InertiaSE** | Inertia of production units in Sweden |

[2]Essentially, swedish data in the first months of 2018, and danish data in the summer of 2019, referring to Fig. 3.1

[3]Number of hours in a year = 8760. Number of hours in 2 years = 17520

## 3.2 Production data

Data on production in the NPS is also available. As described earlier in section 2.1, the NPS consist of the power systems of Norway, Sweden, Finland and Eastern Denmark (Zealand/Sjælland), the price area referred to as DK2. Data on production per generation type and price area is openly available from the power systems of Sweden, Finland and Denmark, through the platforms listed in Table 3.3. Data on Norway's total production and consumption levels are publicly available on Statnett's home page[4], although not split into production per generation type or price areas. However, by user access to Statnett's internal datawarehouse "Innsikt", also Norwegian data split into generation type and price area were acquired.

**Table 3.3:** *Production data sources*

| Country | TSO | Dataplatform |
|---|---|---|
| Norway (NO) | Statnett | Innsikt |
| Sweden (SE) | Svenska Kraftnät | Elstatistik [48] |
| Finland (FI) | Fingrid | Fingrid Open Data [49] |
| Denmark (DK2) | Energinet | Energi Data Service [50] |

The work performed in collecting production data can be summed up in the following list and each point will be elaborated upon in the following subsections:

1. Identify data sources and download raw data on production per generation type
2. Do preprocessing, i.e. handle DST issues, rename columns, crosscheck time alignment with other sources of data
3. Combine all raw files into a single dataset

### 3.2.1 Collecting production data

Production data from all TSOs were downloaded via data sources listed in Table 3.3, and stored as raw files. During the work of combining all the separate data into a single dataset two main problems were encountered, in addition to the occasional missing data:

- File formats - each country store their data in different formats
- Uncertainties regarding timestamps and time alignment

To clarify how the formats and timestamps were interpreted, a short description of the raw files from each country is described below. In addition, the columns were renamed to a standard to simplify the making of models and comparing results. When processing and collecting production data, the end goal was to get a dataset with a similar format as the dataset on inertia, described in section 3.1.4, with the index column representing hours from 2018-01-01 00:00 to 2019-12-31 23:00.

For Norway, Sweden and Denmark, the data stored were hourly average production values stored in units of energy per hour (MW h/h). Finland stores a reading of the actual power output in units of power (MW) from each generation type every three minutes

[4] `https://www.statnett.no/`

in their database. The Finnish production data were therefore resampled to hourly average production. This was done by taking the average of every three-minute measurement within one hour, storing that value as the power output for that hour. This way, the Finnish production data could also be expressed as MWh/h.

### Norway

Norwegian data on production were acquired from Statnetts data warehouse Innsikt. The data is hourly average production values for each year in 2018 and 2019. The raw files were excel sheets from 2018 and 2019, with timestamps in the format DD.MM.YYYY-00, DD.MM.YYYY-01, where -00 was interpreted as the first hour in the day (production between 00:00 and 01:00), -01 is interpreted as hour two (01:00-02:00) and so on. The available production columns for each of Norway's five price zones were as named by Innsikt (English translation in parenthesis):

**Ukjent** (Unknown)

**Vann** (Water)

**Vind** (Wind)

**Termisk** (Thermal)

For price area NO5, there was no column named Vind (Wind).

By cross-checking with other data sources, the production data from Innsikt were confirmed to be stored in Norwegian local time (CET). Regarding daylight saving time (DST) issues, in the transition back to standard time in October, hour 2 is repeated. When investigating the data, the values for hour 2 across all columns appeared to have twice the value as the hours surrounding it. It was decided to solve this by deleting the abnormal values from the data, creating missing values ensuring that the rows would be removed when running models.

The columns were renamed to NO1-5 subscripts **Unkn** (Ukjent), **Hydr** (Vann), **Wind** (Vind) and **Thrm** (Termisk), and were copied to the final production dataset.

### Sweden

Various data from the swedish power system is published through the website of *Svenska kraftnät*. Excel sheets containing measured values on, among other things, production, consumption and power exchange, are published on the page "elstatistik"[5] [48]. In this thesis, production per area and production type are of interest, and the excel sheet containing this data was "*Statistik per elområde och timme*" (translated: "statistics per area and hour"). The data files containing the values from 2018 and 2019 were downloaded.

The timestamps in these files were in the format (DD.MM.YYYY 0:00, DD.MM.YYYY 1:00 ...) and were interpreted the same way as the Norwegian data. Because there was no clear indication of what time zone this time format represented, representatives from SVK were contacted, and they replied that the data is presented as "Swedish normal time (CET) not adjusted for daylight saving time" [51].

[5]The data is only accessible through the Swedish version of the website.

The data in "*Statistik per elområde och timme*" includes measured production values from each of the four price areas in Sweden (SE1-4). The unit is MWh in all columns. The available production columns are listed in Table 3.4, along with the name the columns are given in this thesis.

**Table 3.4:** *Available production columns from each of Swedens four price areas through the excel sheet "Statistik per elområde och timme".*

| Production column | English translation | Column name in thesis |
|---|---|---|
| Ospecificerad | Unspecified production | Unkn |
| Vattenkraft | Hydro power | Hydr |
| Vindkraft | Wind power | Wind |
| Kärnkraft (Only SE3) | Nuclear power | Nucl |
| Värmekraft | Thermal power | Thrm |
| Solkraft | Solar power | Solr |

The columns were renamed SE1-4 with subscripts as listed in Table 3.4. Then the columns were copied to the final production dataset, aligned with the norwegian data.

### Denmark

Energinet, the TSO of Denmark, has an excellent datahub called *Energi data service* [52], making energy data stored by Energinet available. Hourly data on production per production type were found in the dataset *Electricity Balance* [50]. DK2 production data from 2018 and 2019 were downloaded as excel sheets using the websites data filters. The metadata clarifies that the timestamps, both UTC and Danish local time follow the ISO 8601 standard [47]. Each row has an Hour UTC index which is the date and time interval shown in *UTC time zone*. 00:00 o'clock is the first hour of a given day (interval 00:00 - 00:59), 01:00 covers the second hour (interval) and so on.

The Danish production dataset had the same problem encountered with Norwegian data in the transition from summertime to standard time. The recorded measurements for the repeated hour appeared to be around twice the value of the preceding and trailing hours. It was again decided to delete the production data in these rows, creating missing values in rows that will later be excluded from modelling.

The available production columns in the Danish data are as follows, with units MWh per hour:

**Local Power Production** Sum of production from local Combined Heat and Power units (CHP).

**Offshore wind power** Electricity production from offshore wind power.

**Onshore wind power** Electricity production from onshore wind power

**Central production** Production from power plants registered as central power plants.

**Solar power production** Production of electricity from solar cells. To some extent, this production is estimated by Energinet.

These columns were renamed DK2_Thrm_Di[6], DK2_Wind_Offshore, DK2_Wind_Onshore, DK2_Thrm_In and DK2_Solr. Then the columns were copied into the final production dataset.

### Finland

The Finnish grid is not divided into price areas. Data on production per production type in Finland is accessible through the open data platform *Fingrid Open Data* [49]. As mentioned above, this data source has measurements in units of power (MW), recorded every 3 minutes, rather than units of energy (MW h) per hour. All data from Finland were downloaded via API, and the entire script is listed in Appendix B.

Fingrid open data [49] stores real-time measurements of production in datasets with a unique Variable ID. The available datasets of interest in this thesis are listed in Table 3.5, along with the column name that they are given. The script was used to collect available data from Fingrid open data between January 2018 and December 2019. The technical details of how the scripts load the available data into a Pandas data frame [38], [39] are not of great importance for the thesis. The documentation for Fingrids open data API states that timestamps are presented in UTC [53]. In order to transform the data from a power measurement every three minutes to hourly values, Pandas resample tool was used. By taking the mean of all measurements within an hour H, and use that mean value as the average power production in hour H the production data from Finland should be of equivalent format to the data collected from the other Nordic countries.

**Table 3.5:** *Overview of datasets acquired from Fingrid open data [49] used in this thesis.*

| Variable ID | Name of the resource | Column name in thesis |
|---|---|---|
| 188 | Nuclear power production | FIN_Nucl |
| 181 | Wind power production | FIN_Wind |
| 191 | Hydro power production | FIN_Hydr |
| 202 | Industrial cogeneration | FIN_Thrm_In |
| 201 | Cogeneration of district heating | FIN_Thrm_Di |
| 205 | Reserve power plants and small-scale production | FIN_Unkn |
| 192 | Electricity production in Finland | FIN_Prod_tot |

Finally, the columns were copied into the production dataset, aligned with the data from the other Nordic countries.

[6]The choice of name was based on the Finnish columns, where there also was two thermal production columns, named District and Industrial cogeneration.

### 3.2.2 Preprocessing of production data

Most of the preprocessing of production data were done regarding to time alignment. In addition, each country's production data was compared to data downloaded from Nordpool [54], both in order to affirm the time alignment, and to look for missing data. These investigations were brief and indicated that the collected data from Norway and Finland were time-aligned and complete. Data from Denmark could not be verified this way because Nordpool data did not include production per price area. Collected data from Sweden through "elstatistik" [48] did not appear to correspond to data from Nordpool. This will be discussed further in Chapter 4. Due to time constraints, it was decided to go forward with creating models based on the data that had been collected.

### 3.2.3 Final dataset of Nordic production data

The result of the data collection described above was that production data from the Nordic countries were collected in the CSV-file *NordicProductionWithNaN.csv*. The index column are timestamps, representing each hour in the year from 2018-01-01 00:00 to 2019-12-31 23:00. The contents of the production columns are as described in section 3.2.1 the total energy produced in hour $H$ in units of energy (MWh), split into generation type and price area. Price areas are denoted with NO1-5, SE1-4, FIN and DK2, and generation types are given subscripts as shown in Table 3.6.

**Table 3.6:** *Production columns collected in the CSV-file NordicProductionWithNaN.csv.*

| **Norwegian production columns** Total 19 columns | | **Danish production columns** Total 5 columns | |
|---|---|---|---|
| Unkn | NO1-5 | Thrm_In | DK2 |
| Hydr | NO1-5 | Thrm_Di | DK2 |
| Wind | NO1-4 (Not NO5) | Wind_Offshore | DK2 |
| Thrm | NO1-5 | Wind_Onshore | DK2 |
| | | Solr | DK2 |
| **Swedish production columns** Total 21 columns | | **Finnish production columns** Total 6 columns | |
| Unkn | SE1-4 | Nucl | FIN |
| Hydr | SE1-4 | Wind | FIN |
| Wind | SE1-4 | Hydr | FIN |
| Nucl | SE3 | Thrm_In | FIN |
| Thrm | SE1-4 | Thrm_Di | FIN |
| Solr | SE1-4 | Unkn | FIN |

## 3.3 Motivation for using linear models

As described in Section 2.4, in Eq. 2.6, rewritten here for convenience:

$$E_{k,sys} = S_{n,sys} H_{sys} = \sum_{i=1}^{N} S_{ni} H_i \text{ [GWs]} \qquad \text{(2.6 revisited)}$$

system kinetic energy $E_{k,sys}$ can be estimated if every generator $i$ connected to the system has a known value for the rated apparent power $S_{ni}$, and inertia constant $H_i$.

This is the basis for the kinetic energy online estimation in the NPS as presented in *Future system inertia* [13]. In "*Ensuring future frequency stability in the Nordic synchronous area*", an article by *Eriksson et. al.* [18] this way to estimate kinetic energy is presented as a bottom-up method. The authors then present the bottom-up method as a possible way to forecast the kinetic energy level up to D+1 (1 day forward), as that is the timeframe the TSOs receive production plans from producers. The main goal in this thesis is to investigate the top-down approach presented in the same article [18]. The top-down approach is based on production per generation type and using a generic inertia constant. Because forecasts of production per generation type are created on a longer horizon, *Eriksson et. al.* recommend a top-down forecasting approach for D-2 and longer time horizons. The main scope of this thesis is to investigate how closely such a top-down approach can estimate kinetic energy.

#### Estimating generic inertia constants with linear regression models

Using linear regression models with production data as explanatory variables to estimate system kinetic energy was discussed with representatives from Statnett [55]. The top-down approach can be mathematically expressed as

$$E_{k,sys} = \sum_{p} P_p H_p \text{ [GWs]}, \tag{3.1}$$

where $P_p$ is the production per production type, given in MWh/h, and $H_p$ is the generic inertia constant, in seconds $s$.

With the data collected in sections 3.1 and 3.2, Equation 3.1 can be further expanded:

$$E_{k,sys} = \sum_{n}^{N} \sum_{p}^{P} P_{np} H_{np}, \tag{3.2}$$

with $p$ being the type of production, $P$ is the set of all production types, $n$ being price area and $N$ is the set of all price areas. A summary of production types and price areas is given in Table 3.6.

The target, $E_{k,sys}$, is the inertia level, with data collected as described in section 3.1.

It was proposed by Statnett [55] first to apply first order regressor models, as described in section 2.8.1, on the data as a starting point, and to analyse the coefficients and performance metrics of these models to develop an understanding of how linear top-down models can estimate system kinetic energy.

As a starting point, it was decided to stick with linear regression models, as they are simple to implement and easy to understand. More complex models such as decision tree regression models and neural network models dramatically increase model complexity without necessarily improving the performance of estimation models.

The first model to test out is a simple first-order regressor where the model assigns a weight to every production type and price area. The interesting part of observing this model is looking at the coefficients given to each parameter and at the y-axis intercept, which indicates how much of the variations in data the model can pick up. Other more complex models may not give the possibility to extract this information. Neural network models, for instance, have extremely many coefficients in various layers and quickly becomes too difficult to analyse and extract information from. Another problem with neural networks is that they require the training and test set to be normalized in order to be able to train properly. This scaling means that the learned coefficients are hard to interpret, as opposed to the first-order regressor model where all inputs are physical sizes that are recognisable to those that use the models.

## 3.4 Models

Following sections 3.1 and 3.2, and motivated by 3.3, the work proceeded to the next step of developing linear models using subsets of the production data to estimate system inertia. These models were made both on a Nordic level and split into countries. As previously mentioned, data from January 2018 to December 2019 were collected. Based on the available data, it was decided to use data from January 2018 to June 2019 to train models and test models on data from July and August 2019. The choice of test set was not arbitrary; July and August are summer months, and the inertia level is lowest, and consequently most important to estimate correctly in this period.

There was no previous information about what a good model would be when going into modelling. For this reason, in order to create some basic understanding of the data, it was decided to create a benchmark estimation/prediction. The creation of these baseline models is described in section 3.4.1.

Following this, linear regression models using all Nordic production data to estimate the Nordic kinetic energy were created. This is described further in section 3.4.2. Models were also created per country because the estimation of system kinetic energy varies between countries.

### 3.4.1 Baseline/benchmark

It was decided to create baseline models to create a benchmark of model performance and develop a basic understanding of the data. These models were naive/persistence models applied to the testing data.

A naive/persistence forecasting model is a useful tool to forecast future values in a time series. For instance, the next value in a time series $t + 1$ can be estimated using the last stored value of the time series $t$.

In this thesis, persistence models with a 1h lag, 24h lag and 168h (equivalent to one week) were tested, meaning that values were "estimated" or "forecasted" using a value

from the previous hour, or the value for the same hour, the previous day or week.

The naive baseline model was applied on the test set of the Nordic data, **InertiaNordicGWs**, as well as all the individual data from each country. The performance metrics MAE, MSE, and MAPE were calculated for each of the baseline models, in addition to the max error.

### 3.4.2 Models to estimate total inertia

The column **InertiaNordicGWs** was imported as target variable $y$, and all available production columns from all Nordic countries were imported as explanatory variables $X$. Linear models as described in chapter 2.8.1 were created, using the training set of Inertia and production values (January 2018 to June 2019) to train models, and then applying the trained model on production values from the test set (July and August 2019), creating estimated inertia values $\hat{y}$. The models' performance metrics were then calculated, and the estimated values $\hat{y}$ were graphed together with the true target values $y$, to visualize the performance of the top-down models. The regression coefficients $w_j$ each correspond to a production column, and the relative sizes of these coefficients were also looked at.

The modelling approach was exploratory. The first model applied started off very basic without constraints, and further models applied included more and more constraints. Models without an intercept term were of interest from a physical point of view, as the zero intercept term has the physical interpretation as "kinetic energy level if there is no production". Non-negative least square models, forcing the coefficients $w$ to be zero or positive, were also tested. In the power system, every connected generator or power electronics delivering power to the grid is assumed to either contribute or not contribute with inertia, but no generating unit should lower the amount of kinetic energy.

The last models applied on the total Nordic data tested model performance if certain production columns were held out or aggregated together.

### 3.4.3 Models to estimate inertia per country

During modelling, it became clear that it was beneficial to look at each country individually. For each of the four countries in the NPS, inertia data were imported as target variable $Y$ and production data from that country were imported as explanatory variables $X$. Different linear models were then applied on the data, and the performance metrics on the training and test set were recorded and compared to the baseline models applied in that country. Further details for each country are included in the respective subchapters in Chapter 4, where the results are presented and discussed.

# Chapter 4

# Results and Discussion

This chapter presents the findings of the modelling on the data. As this thesis is of exploratory character, some of the discussion will follow the presented results. At the end of the chapter, there is a section with an overall discussion.

## 4.1 Linear regression models on the total Nordic dataset

The work started with exploring the results of the baseline models on the total inertia level of the NPS, followed by the training of linear models on the training data. Then the performance on the test set was recorded. For some models, the coefficients found by the linear models are discussed because their relative sizes can be used to identify if a model is overfitting (as explained in subsection 2.8.2). Initially, all columns were used as input to the models. Later, the models were tested where some columns were removed to investigate the impact on the overall model performance. Table 4.1 summarises the metric results of the models run on the Nordic level of inertia. The different model variations are described in the following, while the results of models run per country are shown in later subsections.

**Table 4.1:** *Results of models trained and tested using the Nordic level of inertia as target variable.*

| data: InertiaNordicGWs | | TEST DATA (2019-07 to 2019-08) | | | | TRAIN DATA (2018-01 to 2019-06) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Modelname** | **n_features** | **MSE** | **MAE (GWs)** | **MAX_E (GWs)** | **MAPE** | **MSE** | **MAE (GWs)** | **MAX_E (GWs)** | **MAPE** | intercept (GWs) |
| Naive_1h | 1 | 27.267 | 3.66 | 20.924 | 2.130 % | | | | | - |
| Naive_24h | 1 | 183.001 | 9.491 | 48.797 | 5.532 % | | | | | - |
| Naive_week | 1 | 102.339 | 8.113 | 33.821 | 4.716 % | | | | | - |
| lr_prod | 51 | 17.601 | 3.411 | 19.539 | 1.919 % | 7.543 | 2.069 | 28.996 | 1.082 % | 35.719 |
| lr_prod_forced0intercept | 51 | 28.662 | 4.34 | 17.144 | 2.461 % | 10.061 | 2.431 | 27.904 | 1.258 % | - |
| lr_prod_nnls | 51 | 18.672 | 3.551 | 13.352 | 1.997 % | 8.466 | 2.178 | 29.714 | 1.143 % | 30.059 |
| lr_prod_nnls_forced0intercept | 51 | 30.586 | 4.507 | 17.084 | 2.550 % | 10.752 | 2.506 | 28.105 | 1.303 % | - |
| lr_prod_noUnkn | 41 | 18.897 | 3.47 | 15.306 | 1.961 % | 10.241 | 2.416 | 28.845 | 1.263 % | 34.607 |
| lr_prod_noUnknSolr | 36 | 19.636 | 3.561 | 15.06 | 2.006 % | 10.357 | 2.424 | 29.079 | 1.266 % | 34.406 |
| lr_prod_noUSgrThm | 32 | 18.683 | 3.457 | 15.721 | 1.941 % | 10.659 | 2.459 | 29.160 | 1.285 % | 34.611 |
| lr_prod_noUSgrThm2 | 29 | 18.149 | 3.39 | 14.829 | 1.902 % | 11.314 | 2.552 | 28.600 | 1.334 % | 33.179 |
| lr_prod_noUSgrThm2nnls | 29 | 19.834 | 3.559 | 15.313 | 2.001 % | 11.524 | 2.600 | 28.936 | 1.358 % | 32.025 |

### 4.1.1 Baseline results - Nordic inertia

Figure 4.1 displays the total Nordic kinetic energy level (*InertiaNordicGWs*) estimated in the NPS, as well as two of the baseline models, the naive 1 hour and 24-hour model. The first nine days in the test set is displayed. With July 1, 2019 being a Monday, it can be observed that the *InertiaNordicGWs* time series has a repeating weekly seasonal pattern, with a higher inertia level on the weekdays compared to the weekends. It also has a daily pattern where the inertia level peaks in the morning and afternoon, with crests at nighttime. As shown in Figure 4.1 the **Naive_1h** model drawn in red is hard to differentiate from the actual inertia value **InertiaNordicGWs**. The Mean Absolute Percentage Error (MAPE) on the test data is 2.130 %. The **Naive_24h** model shown in light green deviates more from the target, especially at the start of the week and at weekends. On the test data this model achieves a MAPE of 5.532 %. Not shown in Figure 4.1 is the **Naive_week** model which achieved a MAPE of 4.716 % on the test set. Also, the MSE, MAE and max error of these models were recorded for the test set listed in Table 4.1.

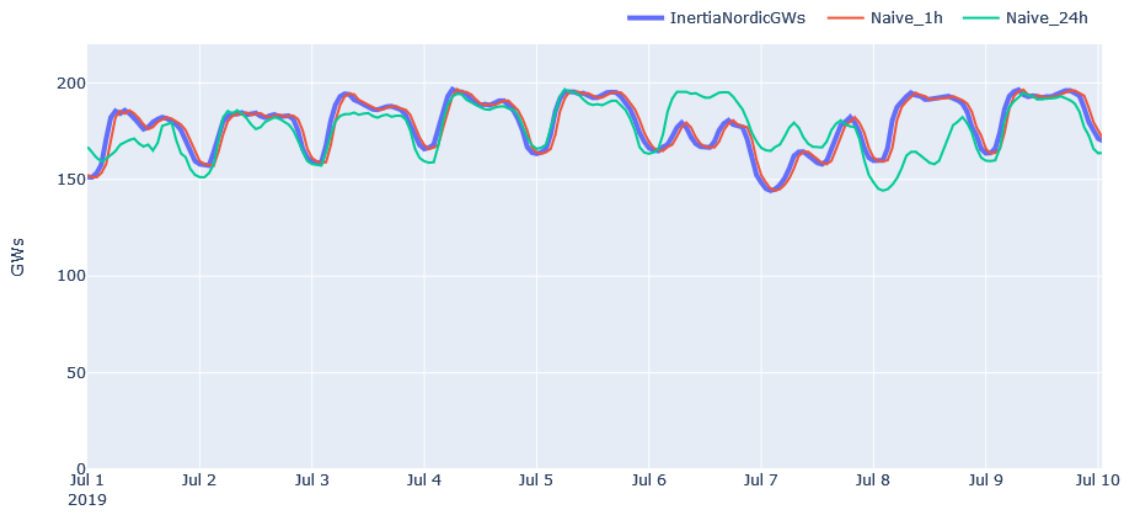



**Figure 4.1:** *Baseline models applied on the Nordic inertia time series. Naive 1h and 24h models are displayed along with the kinetic energy level in the NPS the first nine days in the chosen test set. July 1, 2019 was a Monday.*

#### Comment regarding the baseline models

The **Naive_1h** baseline model appears to be a good estimator. This is because the inertia level is not a very volatile time series but has a relatively smooth profile, especially in the day-hours. In those hours, estimating the inertia level next hour appears easy as the level will often remain the same. The areas where the 1h baseline model deviates the most are the morning and afternoon periods where the production level in the power system rises or falls significantly, thereby the inertia level.

Compared to the **Naive_1h** model, we expect the **Naive_24h** model to predict these ramping periods better. However, because of the difference between weekdays and weekends, the 24h model will generally fail to predict the inertia level on Saturdays and

Mondays. This is also seen in Figure 4.1, where July 6 is a Saturday, and July 8 is a Monday. Because Saturdays typically have a lower level of inertia than Fridays, the 24h baseline model will tend to overestimate the inertia level on Saturdays, exemplified with Figure 4.1. Mondays are the opposite, where using Sundays values will lead to an underestimation.

The **Naive_week** model should capture the weekly trend of a higher inertia level on weekdays and a lower inertia level at weekends, but will struggle more with weekly level changes. Table 4.1 indicate that the weekly model performs better than the 24h model.

### 4.1.2 Linear regression

The result of applying a linear regression model using all available production columns in the NPS (listed in Table 3.6) as explanatory variables is shown in Figure 4.2. This model is referred to as **lr_prod**. The blue line is the "true" target value, while the orange line is the estimated inertia value by the model **lr_prod**. At the start of the test period, the model appears to estimate the inertia value with high accuracy. Around July 17, the model underestimates the inertia level until around August 5. It then returns to a close fit for a week before again appearing to be underestimating the inertia level by roughly 5-10 GWs for the rest of August. The model appears to underestimate rather than overestimate the inertia level in the test period.

#### Metrics: lr_prod

The metrics are listed in Table 4.1. With a MAPE of 1.919 % on the test data, **lr_prod** appears more accurate than the best performing baseline model. Compared to the MAPE on the training data (1.082 %), there is an indication of a model overfitting on the training data.

The y-intercept of the model was found to be 35.719 GWs. The physical interpretation of this intercept is *inertia level if there is no production.* The coefficients $w$ are listed in Table C.1 (Appendix C), where most columns get a moderate value, with some outliers. Specifically, production columns with unknown production get high values, both with positive and negative signs. This is another indication of an overfitting model.

#### Interpretations

Already, a simple out-of-the-box linear model without constraint appears to outperform the baseline models. However, both the model coefficients and differences in metrics calculated on the training and the test set indicate an overfitting model.

In future work, these models are to be extended to forecasting models. Based on this, having columns with extreme coefficients can be problematic, as they can be potential sources of significant errors.

The next model to test was a model without an intercept term. From a physical point of view, it is expected that the inertia level in the system is 0 if there is no production.

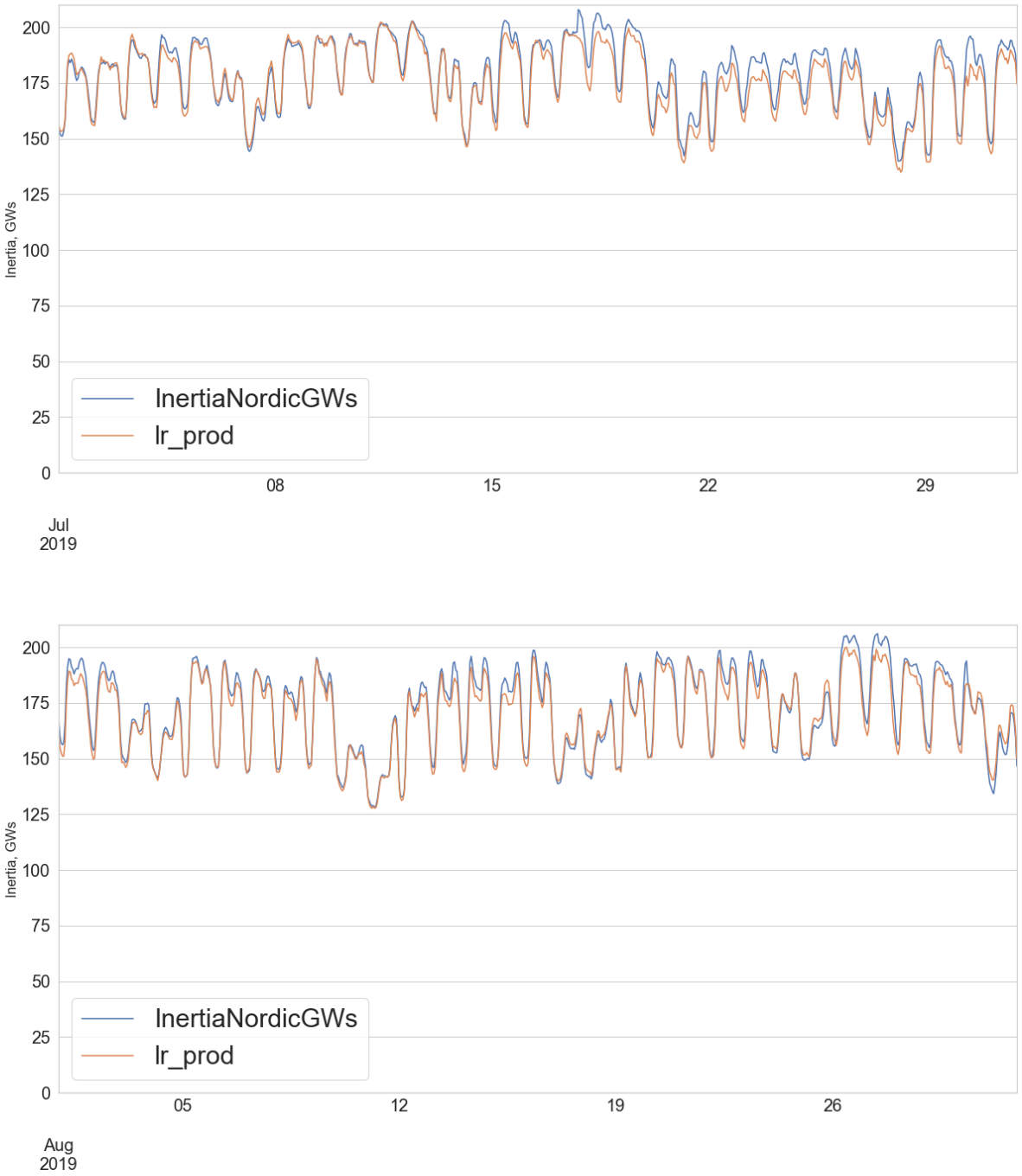



**Figure 4.2:** *Results of applying a linear regression model (lr_prod) on all production columns.*

### 4.1.3 Linear regression without intercept

Figure 4.3 displays the results of a linear regression model trained without an intercept term. This model is called **lr_prod_forced0intercept**. Also shown in the figure is the result of the previous model **lr_prod**. Visually, **lr_prod_forced0intercept** appears to follow the same pattern as **lr_prod**, and tends to underestimate inertia. The metrics are shown in Table 4.1, and confirms that **lr_prod_forced0intercept** is less accurate than **lr_prod** on both the train and test data.

The coefficients of the model **lr_prod_forced0intercept** can be found in Appendix C, Table C.2. This model also gave columns with unknown production significant positive and negative coefficients.

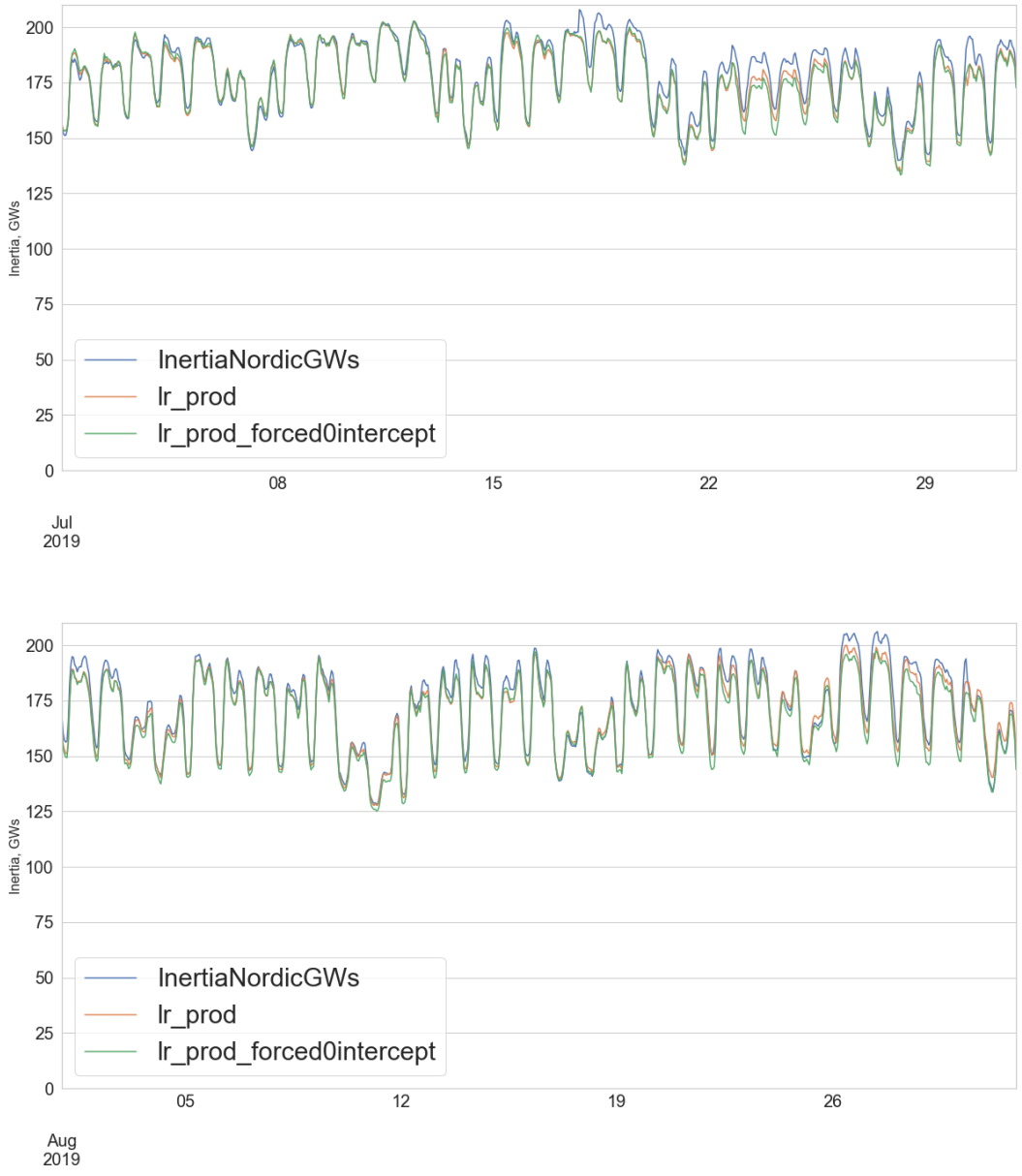



**Figure 4.3:** *Results of applying a linear regression model with (lr_prod) and without (lr_prod_forced0intercept) an intercept term on all production columns.*

### 4.1.4 Non-negative Linear regression

The previous models have allowed columns to get negative coefficients. The physical interpretation of this is that some production contributes to negative inertia in the power system. Theoretically, this does not make sense. Production should either contribute with no inertia, or with a positive amount. Therefore, the following models applied were non-negative least square models. Both a model with intercept, **lr_prod_nnls**, and without **lr_prod_nnls_forced0intercept** were tested. The metric results of these models can be found in Table 4.1.[1]

**lr_prod_nnls** had a lower intercept term than **lr_prod**, as well as a lower max error (13.352 GWs, compared to 19.539 GWs).

The coefficients of these models can also be found in Appendix C, Table C.3, and C.4. Here, several columns were given coefficients of 0, effectively meaning the model does not include these columns in the inertia estimation. For the most part, the columns that were given a 0 coefficient were from the production group wind and solar, as well as unknown production. Surprisingly, some thermal columns were also given 0 coefficients.

### 4.1.5 Further models on total data

All models described above used all available production columns, 51 in total, listed in Table 3.6. Having this many columns to train on increases the tendency of models to overfit, and based on the relative sizes of coefficients, and the difference in performance on training and test data, all models tested so far have been overfitting. In an effort to reduce model complexity, some models were trained where a subset of columns were held out. **lr_prod_noUnkn** is a model trained with all unknown production columns left out, and **lr_prod_noUnknSolr** also leave out all solar production columns. Referring to Table 4.1, these adjustments to the models do not alter the performance on the test set massively, although the number of input columns was reduced from 51 to 36.

Further simplifications were tested in the models **lr_prod_noUSgrThm**, which in addition to leaving unknown and solar columns out, grouped the Norwegian thermal production columns into a single column. **lr_prod_noUSgrThm2** was built with the same logic, giving the Swedish thermal columns the same treatment. These models increased the accuracy of estimation on the test set, and **lr_prod_noUSgrThm2** is the best performing model based on the MAPE of 1.902 % on the chosen test data.

Finally, the same columns input to **lr_prod_noUSgrThm2** were used in a non-negative least square model, which achieved a MAPE of 2.001 %.

### 4.1.6 General discussion - Nordic inertia models

All linear models tested performed better than the naive 24h and naive_week baseline models, and achieved around the same scores as the naive 1h baseline model, according to the metrics in Table 4.1. Removing columns assumed not to contribute with inertia,

[1]The line plots for further models on the Nordic inertia level appear similar to the line plots already shown in Figure 4.2 and 4.3 and are not included in this report.

such as solar production columns, did not appear to lower model performance on the test set.

Based on the results achieved so far, a top-down approach using production data appears to estimate the Nordic inertia level with a MAPE of around 2 %, in this specific case. However, the results are not tested on an independent test set, thus there is a high chance of information leakage and confirmation bias towards creating a good performing model on the test set. Therefore, the results of the models above should be viewed as an introductory feasibility analysis of the method.

Doing feature selection or fitting regularized models such as Ridge or Lasso models on the Nordic dataset are possible topics for further work. In this thesis it was decided to proceed by looking at each country individually.

## 4.2 Linear regression models for each country

### 4.2.1 Norway

A very strong correlation between the total power production and inertia level estimated by Statnett was found during preprocessing of Norwegian data. The values for inertia and total production in Norway for the total dataset in 2018 and 2019 are shown in a scatter plot in Figure 4.4. Each blue dot is the recorded hourly value of inertia and total production. The figure shows a clear linear relationship between the two variables, as the observations appear along a straight line in the diagonal. Some observations appear in straight horizontal lines in the figure, for inertia values 56 GWs, 66 GWs and 70 GWs. These values are assumed to be outliers caused by problems with the inertia estimation in Norway December 4 and 11, 2018, and October 9, 2019. These dates, the inertia estimations were observed to have several repeated entries in the data.[2] A dual-axis plot of Norwegian inertia (blue) and production (red) is shown in Figure 4.5. Subfigure **a)** displays the entire collected data period from January 2018 to December 2019, and indicates that the two time series of inertia and production are closely linked. Subfigure **b)** shows a zoomed-in view of May 2019, picked at random. The figure indicates that the Norwegian inertia estimation follows the total production level in Norway precisely, scaled with some constant. Both time series are seen to follow a yearly seasonal pattern in subfigure **(a)**, with a general higher level of production in the winter. A weekly and daily pattern can be observed in subfigure **(b)**, where weekdays have peaks in the morning and afternoons and crests at nighttime. At weekends, the level of production and inertia is generally lower. (May 7, 2019 was a Tuesday).

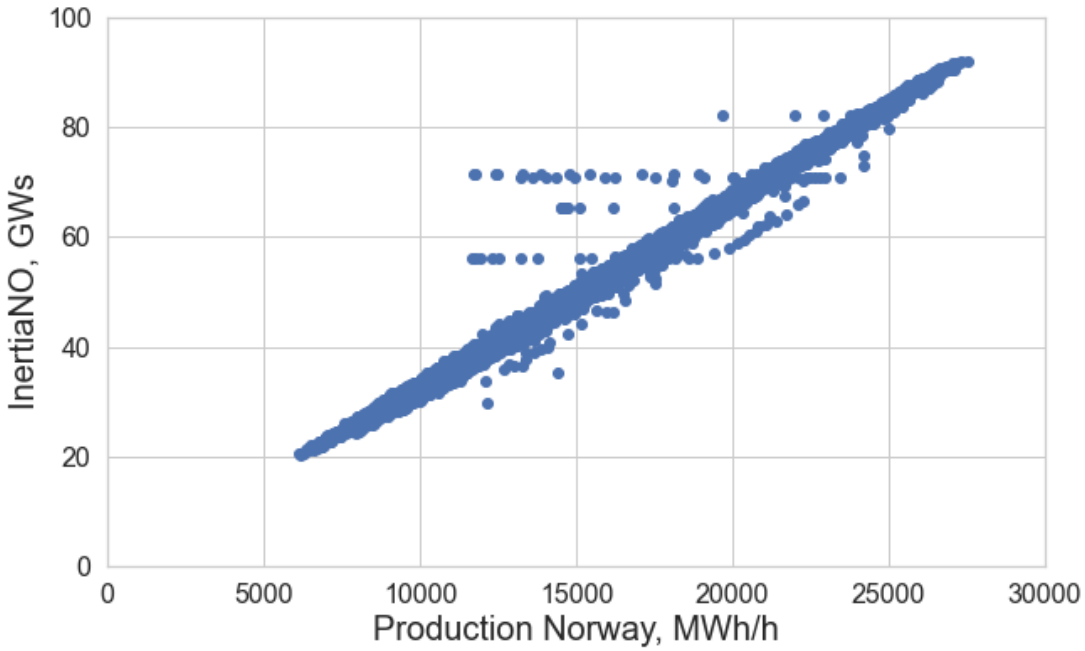
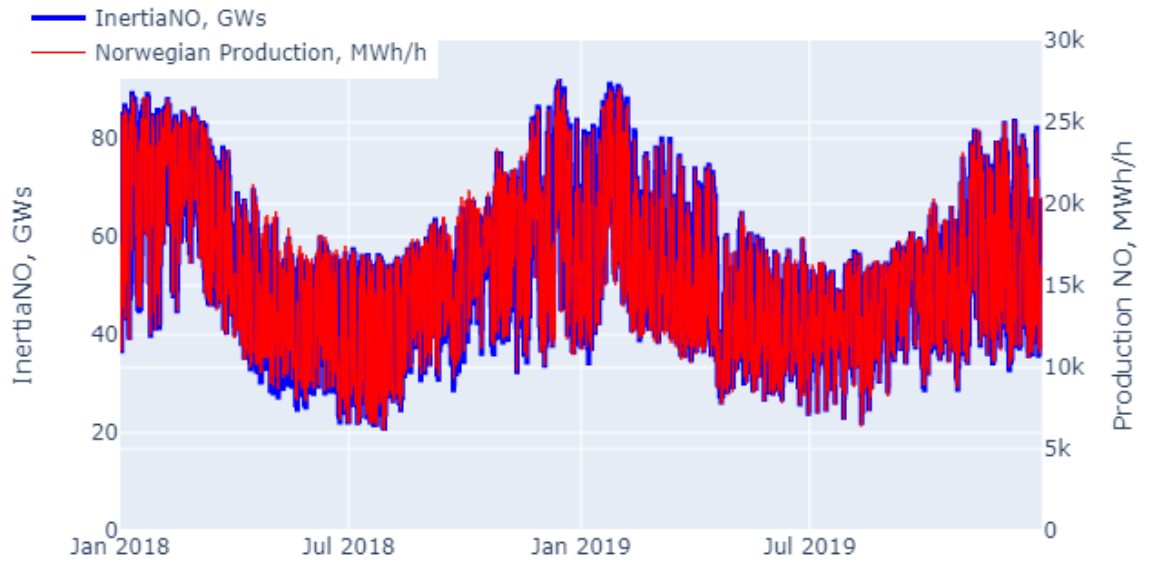
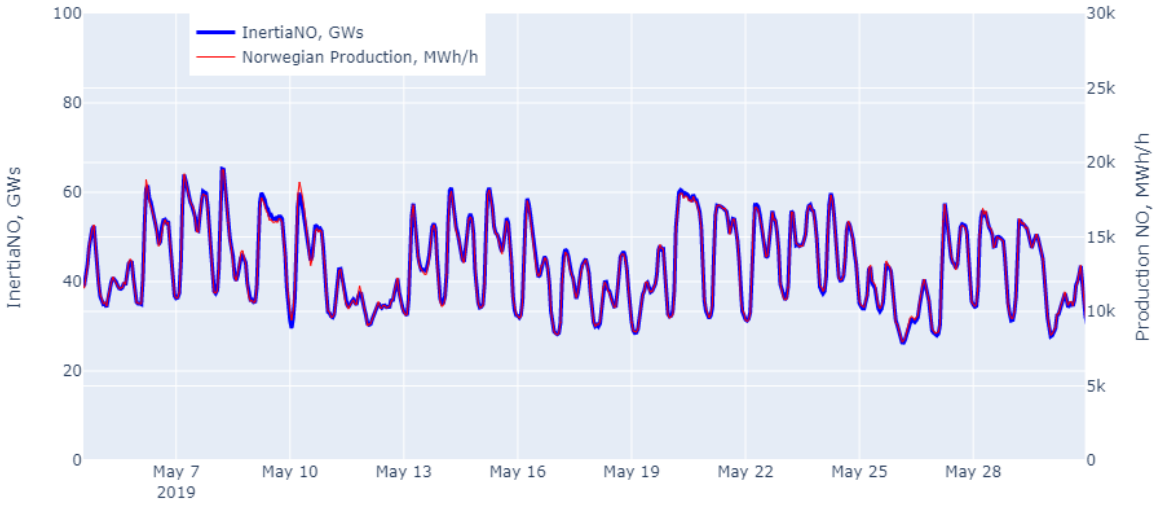



**Figure 4.4:** *Scatterplot of hourly measurements of Norwegian inertia InertiaNO, and total production in Norway.*

[2]These outliers were not removed from the dataset.




**(a)** *2018-2019*




**(b)** *Zoomed in. May 7, 2019 was a Tuesday*

**Figure 4.5:** *Comparing the Norwegian inertia level and the total production in Norway.*

#### 4.2.1.1 Linear models based on production level in Norway

The results of the baseline models, as well as the linear models that were fit on Norwegian data are shown in Table 4.2. Because of the strong correlation between total production and inertia level, it was decided to use aggregated production values to create the models in Norway. This reduced model complexity.

As seen in Table 4.2, the models **lr_only_total_prod** and **lr_prod_forced0intercept** only have one feature (column), the total Norwegian production aggregated across all generation types and price areas. The two models **lr_prod_pr_area** and **lr_prod_pr_area_forced** used 5 columns as input data. These columns were the aggregated production in each price area. The model coefficients for all these models are listed in Table 4.3. Lastly, a model was tested in which all available production columns in Norway were input to the model, **lr_prod**. The coefficients for this model can be found in Appendix C, Table C.5.

The results of **lr_prod_forced0intercept**, the best performing model on the test data according to MAPE metrics, is shown in Figure 4.6. The target variable *InertiaNO* shown in blue, and the model estimate shown in red follow each other closely, and the MAPE on the test set is 0.820 %. This model had no intercept term, and the coefficient value was 3.33, as listed in Table 4.3.

**Table 4.2:** *Results of models trained and tested on the Norwegian data.*

| data: InertiaNO | | TEST DATA (2019-07 to 2019-08) | | | | TRAIN DATA (2018-01 to 2019-06) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Modelname** | n_features | **MSE** | **MAE (GWs)** | **MAX_E (GWs)** | **MAPE** | MSE | MAE (GWs) | MAX_E (GWs) | MAPE | intercept (GWs) |
| Naive_1h | 1 | 7.453 | 1.933 | 12.394 | 4.361 % | | | | | - |
| Naive_24h | 1 | 42.651 | 4.481 | 24.291 | 11.061 % | | | | | - |
| Naive_week | 1 | 26.946 | 4.093 | 16.294 | 9.787 % | | | | | - |
| lr_only_total_prod | 1 | 0.485 | 0.599 | 3.412 | 1.463 % | 1.646 | 0.754 | 32.929 | 1.550 % | -1.67 |
| lr_prod_forced0intercept | 1 | 0.195 | 0.356 | 3.681 | 0.820 % | 1.844 | 0.819 | 32.381 | 1.638 % | 0 |
| lr_prod_pr_area | 5 | 0.339 | 0.463 | 3.66 | 1.150 % | 1.494 | 0.664 | 33.136 | 1.351 % | -1.962 |
| lr_prod_pr_area_forced | 5 | 0.206 | 0.351 | 3.533 | 0.846 % | 1.635 | 0.759 | 32.856 | 1.543 % | 0 |
| lr_prod | 19 | 0.319 | 0.429 | 4.309 | 1.002 % | 1.145 | 0.479 | 32.056 | 0.942% | -1.59 |

**Table 4.3:** *Model coefficients on Norwegian data.*

| | **lr_only_total_prod** | **lr_prod_forced0intercept** |
|---|---|---|
| *NO_Prod* | 3.43 | 3.33 |
| | **lr_prod_pr_area** | **lr_prod_pr_area_forced** |
| *NO1_Prod* | 3.13 | 2.66 |
| *NO2_Prod* | 3.28 | 3.33 |
| *NO3_Prod* | 3.56 | 3.41 |
| *NO4_Prod* | 4.17 | 3.79 |
| *NO5_Prod* | 3.21 | 3.27 |

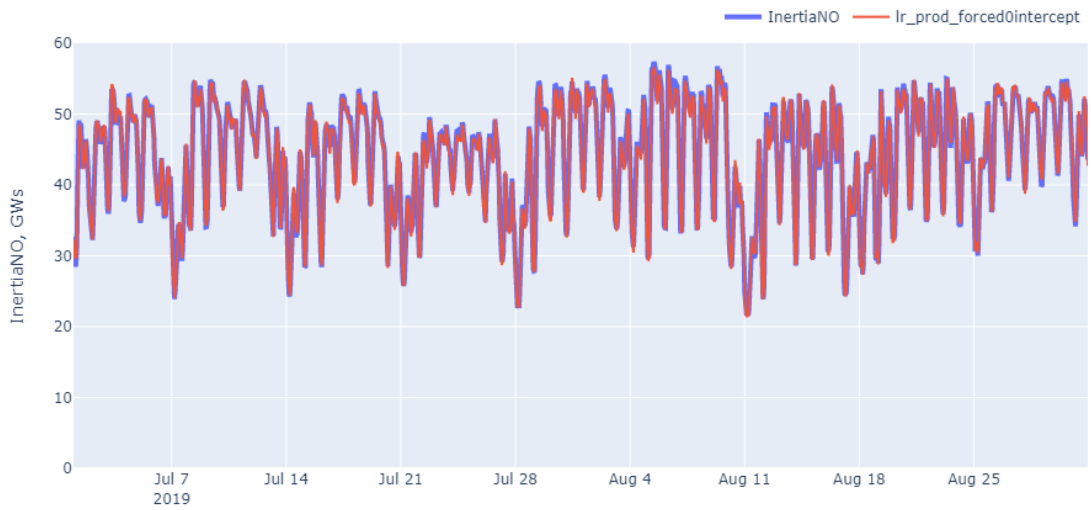



**Figure 4.6:** *Results of a linear regression model applied to the test data. The model is using the total Norwegian production as an explanatory variable to estimate the inertia level in Norway.*

#### 4.2.1.2 Discussion Norway

Although the outliers revealed in Figure 4.4 should have been removed from the training set, the linear models using production data appear to estimate the inertia level in the test set with high precision, outperforming the baseline models by a considerable margin.

Observed in Table 4.2, the linear models received a much higher max error in the train data than in the test set, and this is caused by the periods in the training set where the inertia estimation had several repeated entries. Because of time constraints, it was decided not to investigation these values further, as it is assumed that it would not improve the models.

Comparing the results and coefficients between **lr_prod**, which had 19 explanatory columns, and the other models in Norway, with 1 or 5 columns, it is clear that having more columns do not necessarily equal a better model. With 19 columns given to fit coefficients, **lr_prod** overfit to the training data by fitting high positive and negative coefficient to some columns, in this case, unknown production columns.

The Norwegian results compared to the results of the other Nordic countries are discussed further in section 4.2.5.

### 4.2.2 Sweden

The average Swedish contribution to the kinetic energy level in the NPS was 50 % during 2018 and 2019. Swedish inertia is displayed in blue along with the total level of Swedish production shown in red in Figure 4.7. The resolution of the data is hourly. Based on the figure, the Swedish inertia level intraday seems to vary within a 20 GWs band, being typically 100 GWs in the winter period (October to April), and lower in the summer period. Compared to a similar figure from Norway (Figure 4.5), the correlation between total production and inertia is lower, but there is still a clear tendency that the inertia level follow the production level.

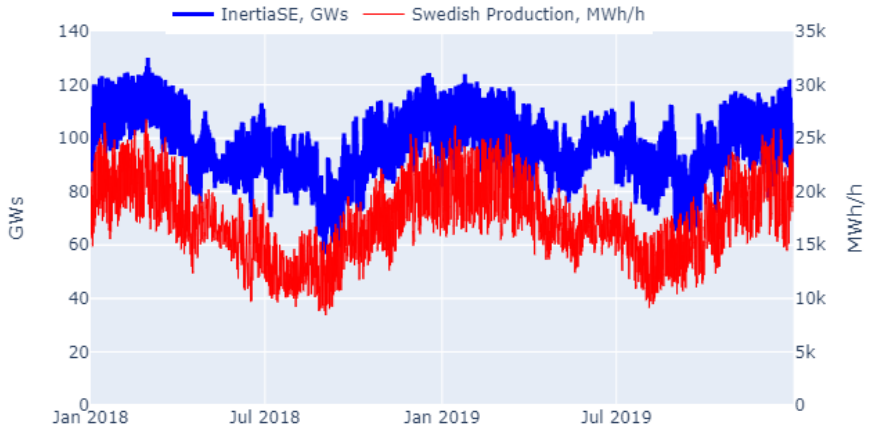



**Figure 4.7:** *Comparing the Swedish inertia level to the total production in Sweden.*

The production columns in Sweden are split into 21 columns, listed in Table 3.6. The structure of the work in Sweden was similar to the method regarding Nordic inertia, described in Section 4.1; first exploring the results of baseline models, then applying linear models, discussing the results including model coefficients. Table 4.4 summarises the metric results of the models run on Swedish data.

**Table 4.4:** *Results of models trained and tested on the Swedish data.*

| data: InertiaSE | | TEST DATA (2019-07 to 2019-08) | | | | TRAIN DATA (2018-01 to 2019-06) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Modelname** | n_features | **MSE** | **MAE (GWs)** | **MAX_E (GWs)** | **MAPE** | MSE | MAE (GWs) | MAX_E (GWs) | MAPE | intercept (GWs) |
| Naive_1h | 1 | 6.504 | 1.694 | 11.440 | 1.880 % | | | | | - |
| Naive_2h | 1 | 21.787 | 3.246 | 16.690 | 3.609 % | | | | | - |
| Naive_24h | 1 | 46.758 | 4.902 | 32.789 | 5.435 % | | | | | - |
| Naive_week | 1 | 58.772 | 6.315 | 21.521 | 7.010 % | | | | | - |
| lr_prod | 21 | 9.432 | 2.296 | 18.910 | 2.476 % | 5.328 | 1.731 | 17.143 | 1.801 % | 16.49 |
| lr_prod_forced0intercept | 21 | 11.642 | 2.725 | 13.548 | 3.019 % | 7.713 | 2.108 | 17.032 | 2.205 % | 0 |
| lr_prod_nnls | 21 | 11.296 | 2.587 | 12.988 | 2.783 % | 5.867 | 1.827 | 17.026 | 1.916 % | 16.94 |
| rr_prod_nnls | 21 | 12.336 | 2.773 | 12.775 | 2.971 % | 5.982 | 1.850 | 16.711 | 1.942% | 17.69 |

#### 4.2.2.1 Baseline results - Swedish inertia

Figure 4.8 displays the Swedish inertia level in a nine day period in the test set, August 2019. August 4 was a Sunday. Similar to the Nordic inertia level shown in Figure 4.1, the Swedish inertia level also has weekly and daily seasonal patterns. The inertia level peaks in the morning and afternoon and has crests at nighttime. The inertia level is lower at the weekends, with only minor variations throughout the day. The **Naive_1h** model gets a MAPE of 1.880 % on the Swedish test set (Table 4.4). In Sweden, the results of a 2h persistence model (**Naive_2h)** was also recorded as a baseline, achieving a MAPE equal to 3.609 % on the test data. The **Naive_24h** model is also shown in the figure and has the same weakness regarding weekends as discussed in Nordic data. (Overestimating inertia level on Saturdays and underestimating on Mondays). In Sweden, the 24h model do however estimate the inertia level better than the **Naive_week** model (MAPE of 5.435 % compared to 7.010 %).

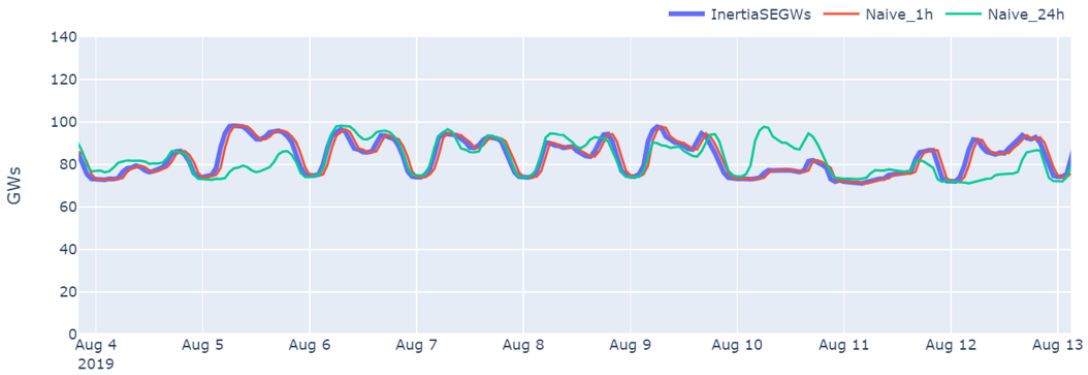



**Figure 4.8:** *Baseline models applied on the Swedish data, a week in August. (August 4 2019 was a Sunday).*

#### 4.2.2.2 Linear models - Sweden

The work proceeded to test linear models where all 21 production columns in Sweden (listed in Table 3.6) were used as explanatory variables. Figure 4.9 displays the results of **lr_prod** and **lr_prod_forced0intercept** on the Swedish test data. The target variable *InertiaSE* is shown in blue, the **lr_prod** estimation is shown in orange, and **lr_prod_forced0intercept** is the green line. For the first couple of weeks in July, the models estimate inertia with good precision, but July 17, the target value *InertiaSE* rises abruptly with around 10 GWs. This is not captured by the models, which continue to estimate at a level around 10 GWs less than the target value for one and a half day. Also, the last week in August, starting the 26th, the models seem to estimate a value of around 10 GWs lower than the target. Throughout the test period, the models tend to underestimate the inertia level.

The coefficients of both models are listed in Appendix C, Table C.6, and they indicate overfitting to the training set, where especially solar and unknown columns were given high positive and negative coefficient values. Another remark is that in both models, three out of four thermal columns got a negative coefficient.

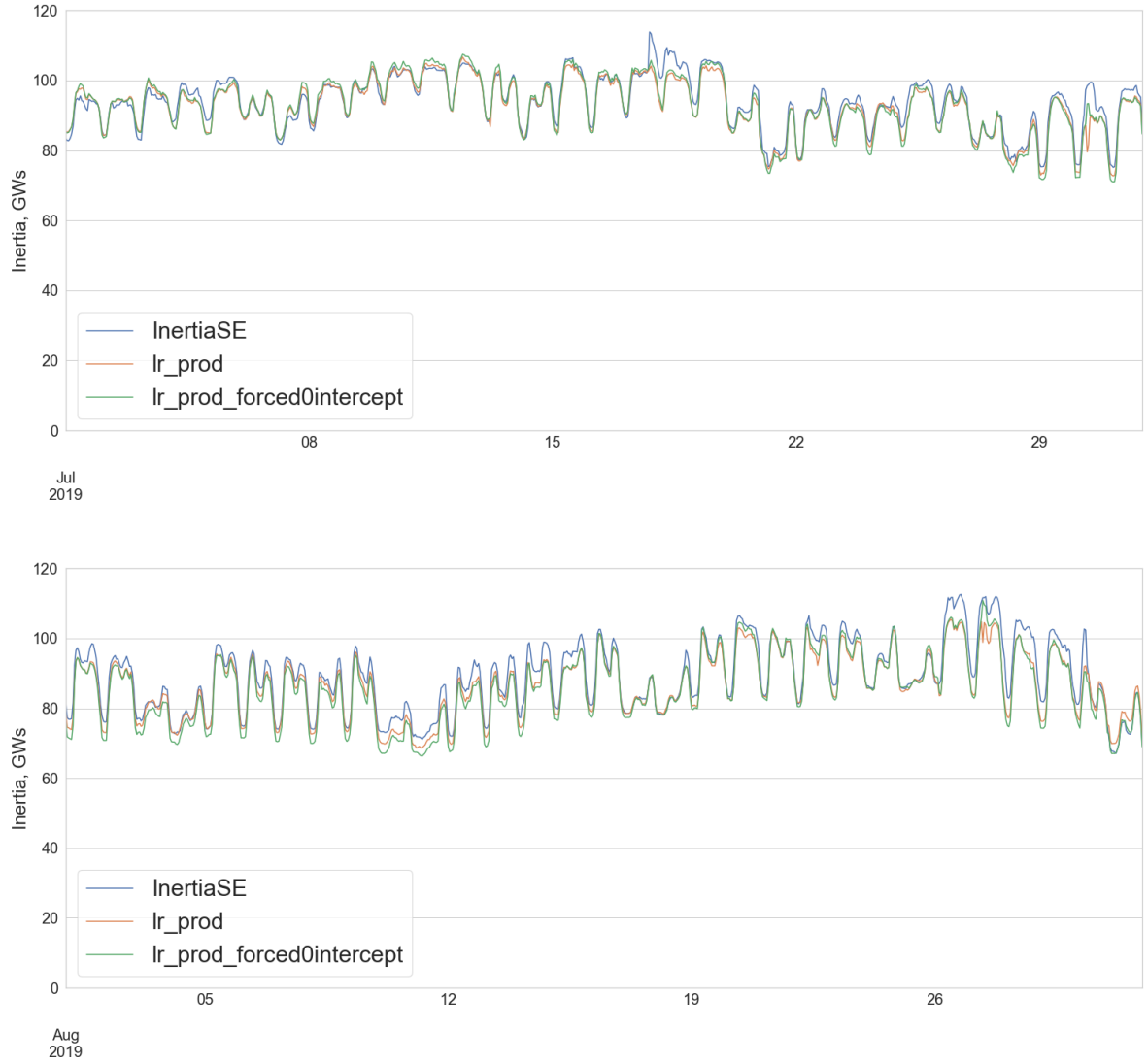



**Figure 4.9:** *Result of applying linear regression models on Swedish data. Both models use all production columns as explanatory variables, and* ***lr_prod_forced0intercept*** *is trained without fitting an intercept.*

#### 4.2.2.3 Non-negative Linear regression

As mentioned earlier, negative column coefficients do not make sense from a physical point of view. Therefore, models with a non-negative coefficient constraint were tested in Sweden.
**lr_prod_nnls** is displayed in Figure 4.10 in red, together with the other linear models described above. The figure reveals that the different models in Sweden are producing more or less the same output. The metrics from Table 4.4 reveals that **lr_prod_nnls** had a marginally higher MAPE on the test data than **lr_prod**.

The coefficients for **lr_prod_nnls** are listed in Appendix C, Table C.7. Only seven production columns got a non-zero coefficient; nuclear production, hydro production in price areas 1, 2 and 3 and solar production in price areas 1, 3 and 4. No thermal columns were given a non-zero coefficient. This was a surprising result as theoretically solar columns should not contribute to the inertia level, while thermal columns should.

The last model tested in Sweden was an extension of the NNLS model. A ridge regression model was applied to the Swedish data to shrink the coefficients, while keeping the constraint of non-negative coefficients. This model was called **rr__prod__nnls**, and was trained with regularization parameter $\alpha = 1$. Figure 4.11 compares this model to the unrestricted model **lr__prod** and shows that these two models produce estimations of inertia that are difficult to distinguish on the test set.

The coefficients for **rr__prod__nnls** are listed in Appendix C, Table C.8. The columns given a non-zero coefficient are the same as in the previous model **lr__prod__nnls**, nuclear production and hydro production in price areas 1, 2 and 3. In this model, solar production columns were given small coefficients.

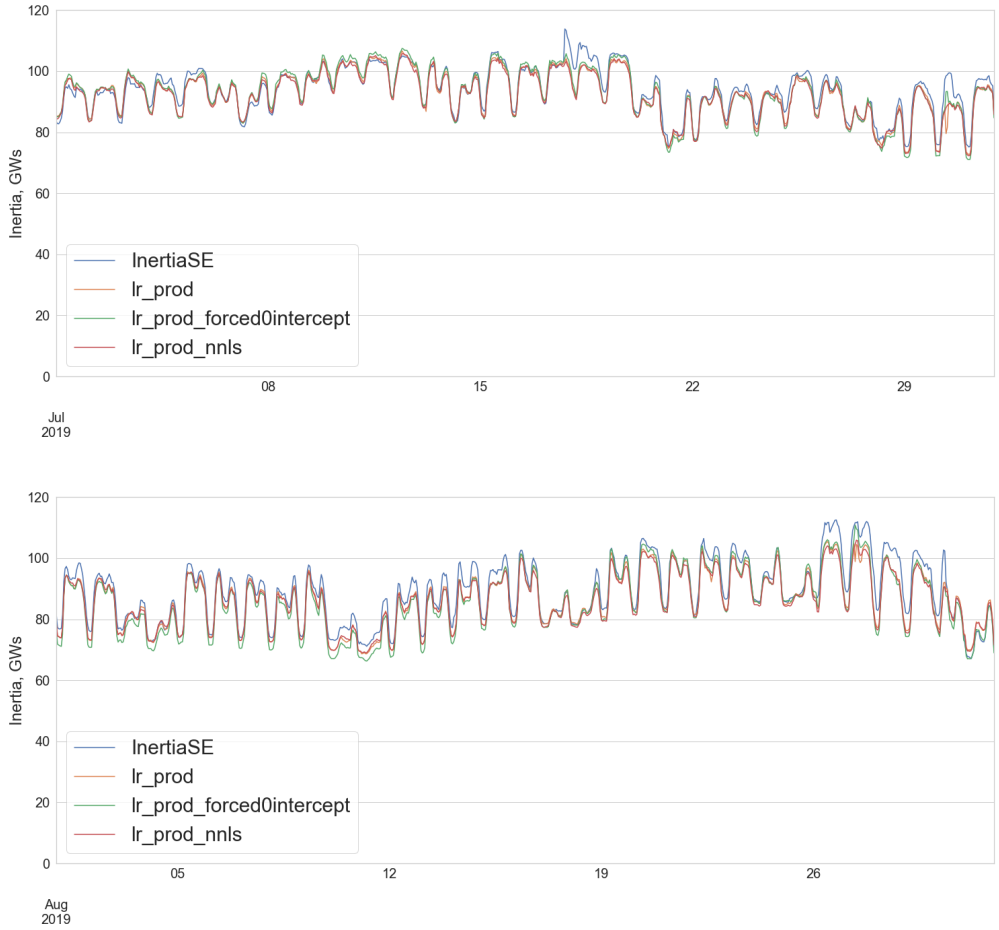



**Figure 4.10:** *Applying a non-negative linear regression model on Swedish data. The other Swedish models are also shown for comparison.*

#### 4.2.2.4 Discussion Sweden

The results of the Swedish models indicate that inertia can be estimated with a MAPE below 3 % by using only a subset of the available production columns; nuclear and hydro production columns. The contribution of solar production columns is minor. It is assumed that solar production columns are included in the model because the inertia level generally is higher at daytime. This is also the nature of solar production, which has values of 0 at night, and are positive at daytime.

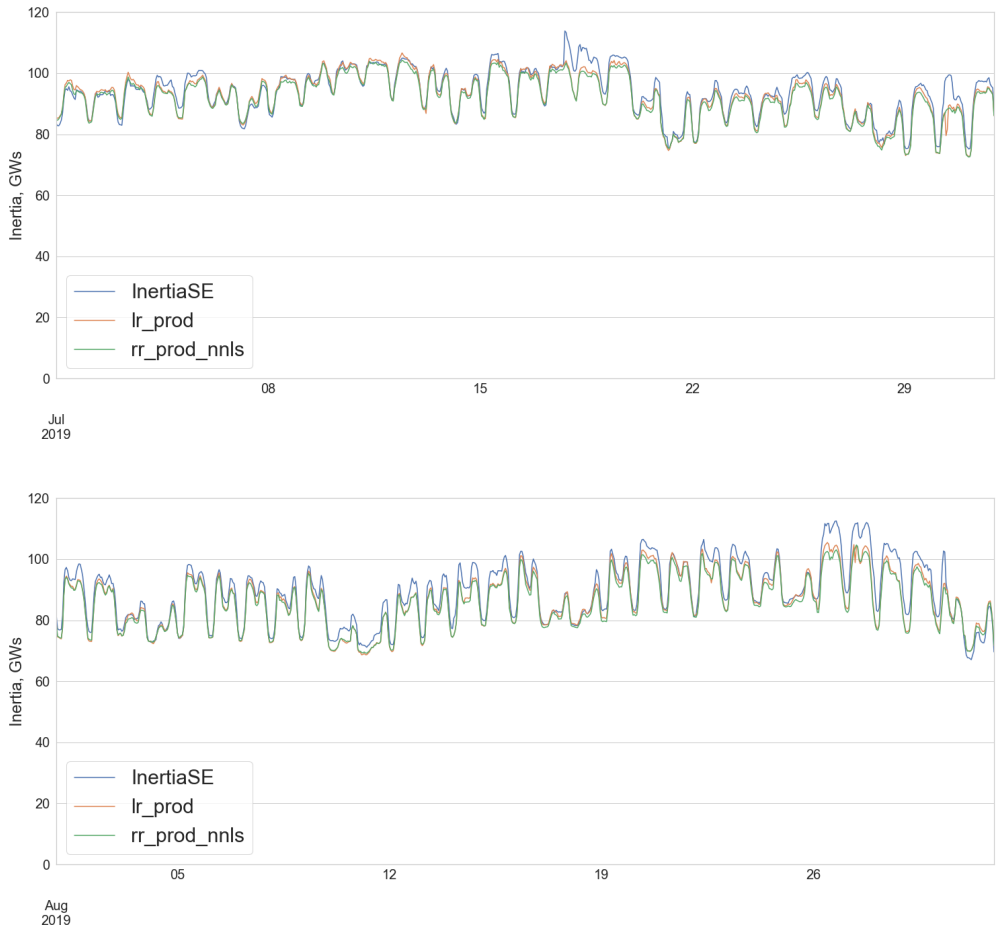



**Figure 4.11:** *Applying a ridge regression model on Swedish data.*

The exclusion of thermal columns in the NNLS models was of more significant concern, and a possible explanation of this can be the data collection and preprocessing of Swedish data. As mentioned in Section 3.2.2, the production values downloaded from SvKs "elstatistik" [48] did not correspond to historic market data from other data sources, such as Nordpool [54]. At an earlier point in the thesis work, considerable effort was made to ensure correct time alignment of Swedish data, including comparing the time series of production as recorded by SvK to production values downloaded from Volue [56], a leading forecast and consulting service provider in the Nordic Power System. When comparing production data from these sources, time alignment and volume discrepancies were found. When plotting comparable time series of production values stored by Volue and "elstatistik" in the transition from standard time to daylight savings time, some production columns were visually found to lose time alignment. The access to Volue data was unfortunately lost early in the thesis work, which is why models were applied to the downloaded data from "elstatistik".

It is heavily recommended that future work redo the data collection of Swedish data, perhaps using another data source, such as Volue [56]. Despite these concerns regarding Swedish data, all linear models tested on Swedish data were performing better than the 24h and week baseline models.

### Nuclear power production

Going into modelling, it was expected that nuclear power units were a significant contributor to the amount of kinetic energy in the system. Nuclear power units are often decoupled in the summer period for planned maintenance. Table 4.5 lists the nuclear unavailability for the summer period of 2019. Forsmark 3 ended its maintenance period on July 17, 2019. In Figure 4.9, the inertia level in Sweden *InertiaSE* is suddenly increasing by around 10 GWs this date. None of the tested models appears to pick up this sudden increase in inertia level that is assumed caused by the addition of Forsmark 3 to the grid.

A possible explanation could be that nuclear power units are ramping down their power output before they are decoupled, and vice versa, after reinsertion to the grid, the power output is ramped up. However, the inertia such power units contribute is available as long as they are connected synchronously to the grid. Because nuclear power plants usually operate at full capacity, this is a source of error in the top-down estimation. This could be why linear models do not "see" the inertia contribution of nuclear units before they are disconnected or after they are reconnected.

**Table 4.5:** *Major outages (>5 days) and maintenance dates of nuclear power in the NPS, summer 2019. Data on outages is available at Nordic Unavailability Collection System [57].*

| Name | Start | End | Duration |
|---|---|---|---|
| **Sweden** | | | |
| Forsmark 2 | 29.07.2019 | 13.08.2019 | 15 days |
| Forsmark 3 | 01.07.2019 | 17.07.2019 | 16 days |
| Ringhals 1 | 19.07.2019 | 18.08.2019 | 30 days |
| Ringhals 4 | 29.08.2019 | 15.10.2019 | 47 days |
| Oskarshamn 3 | 31.08.2019 | 24.09.2019 | 24 days |
| **Finland** | | | |
| Olkiluoto 1 | 02.06.2019 | 11.06.2019 | 9 days |
| Olkiluoto 2 | 02.05.2019 | 26.05.2019 | 24 days |
| Loviisa 1 | 08.09.2019 | 28.09.2019 | 20 days |
| Loviisa 2 | 19.08.2019 | 13.09.2019 | 25 days |

### 4.2.3 Denmark

The kinetic energy of online generation units in Denmark is contributing around 7-10 GWs. Compared to the total level of Inertia in the NPS, which is in the range of 140-220 GWs, the Danish contribution is close to 5 %. Before further details, the reader is also reminded that part of the test set in Denmark is imputed, as described in the last part of Section 3.1.3.

#### Pre-analysis and baseline models for Danish inertia

Figure 4.12 displays the Danish inertia time series from 2018 to 2019, compared to the total level of Danish production. Compared to the Norwegian data shown in Figure 4.5, the inertia and production curves look far less related in Denmark. In general, the inertia level in Denmark appears to have very defined steps. It is also clear from the figure that the inertia level during summer 2018 is different to 2019. While the inertia stays at a level of around 4 GWs in the summer of 2018, it remains at a level of around 8 GWs in 2019. 2019, in general, appears to have a much higher base value and less variation during the year.

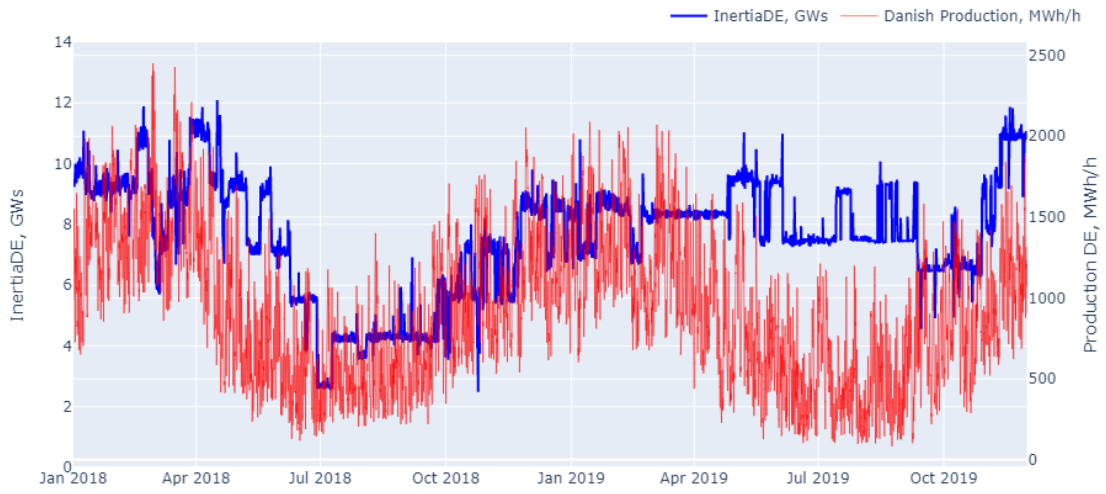



**Figure 4.12:** *Comparing the Danish inertia level to the total production in Denmark.*

Figure 4.13 displays the inertia level in the test set, July and August 2019. This figure further exemplifies the step-like nature of the inertia level in Denmark. In the test period, the inertia level has a flat base value of around 7.5 GWs, with sudden steps raising the inertia level to around 9 GWs.

Baseline models were applied to the Danish data, and in addition to the naive baseline models introduced in section 3.4.1, another baseline model was proposed - estimating the inertia value in the test set as the mean value of the inertia level during the training set. The metric results are listed in Table 4.6 along with the results of the linear models tested on the Danish data. These linear models are presented in the following.

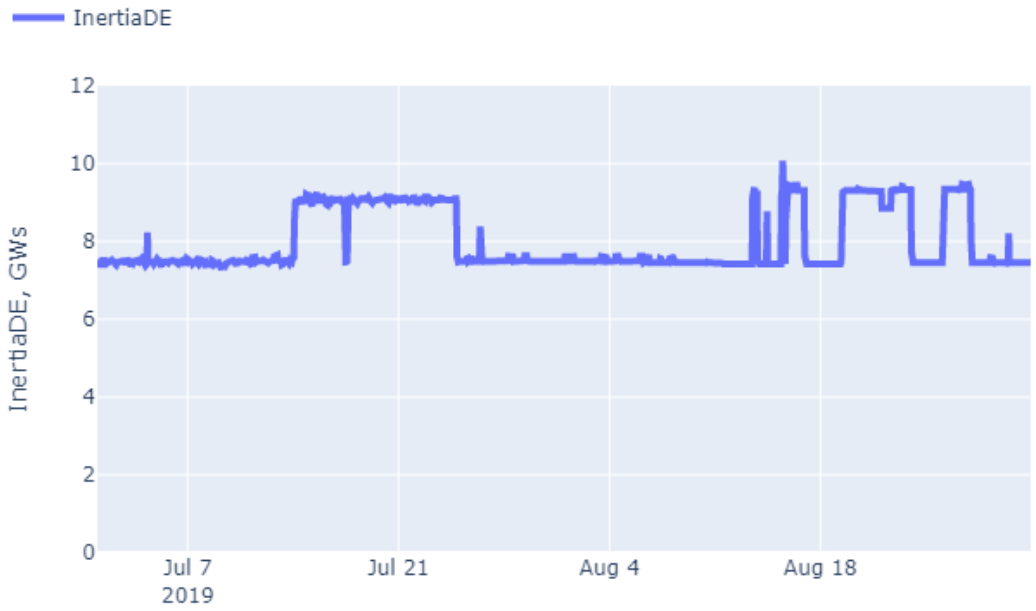



**Figure 4.13:** *The Danish inertia level in the test data, July and August 2019.*

#### 4.2.3.1 Linear models based on production level in Denmark

Figure 4.14 displays the results of the model **lr_only_total_prod** on the Danish dataset. This model had a single column as an explanatory variable, the total production in Denmark (price area DK2). The metric results on the training and test data are listed in Table 4.6. The intercept term of the model was high, 5.38 GWs compared to the average inertia level in the training period which was 7.49 GWs. The coefficient given to Danish production was 2.15. The model performance on both the training data and test data is poor; the MAPE was 19.958 % on the test set, which is much worse than all baseline models.

Three more models based on production columns were tested on Danish data. **lr_prod** and **lr_prod_forced0intercept** (trained without an intercept term) used all available columns on production in Denmark, five columns in total. The model **lr_thrm** used the data from Denmark's two thermal production columns as input. The model coefficients can be found in Appendix C, Table C.9.

**Table 4.6:** *Results of models trained and tested on the Danish data.*

| data: InertiaDE | | TEST DATA (2019-07 to 2019-08) | | | | TRAIN DATA (2018-01 to 2019-06) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Modelname** | n_features | **MSE** | **MAE (GWs)** | **MAX_E (GWs)** | **MAPE** | MSE | MAE (GWs) | MAX_E (GWs) | MAPE | intercept (GWs) |
| Naive_1h | 1 | 0.030 | 0.040 | 1.856 | 0.486% | | | | | - |
| Naive_24h | 1 | 0.483 | 0.304 | 2.651 | 3.687% | | | | | - |
| Naive_week | 1 | 0.987 | 0.618 | 2.616 | 7.440% | | | | | - |
| mean_value_predict | 1 | 0.863 | 0.540 | 2.579 | 5.951% | | | | | - |
| lr_only_total_prod | 1 | 3.576 | 1.651 | 3.639 | 19.958% | 3.087 | 1.446 | 5.238 | 23.910 % | 5.38 |
| lr_prod | 5 | 4.620 | 1.994 | 3.836 | 24.358 % | 2.489 | 1.307 | 6.493 | 21.203 % | 4.65 |
| lr_prod_forced0intercept | 5 | 15.067 | 3.537 | 6.979 | 43.699 % | 4.797 | 1.733 | 9.699 | 23.991 % | 0 |
| lr_thrm | 2 | 4.599 | 2.006 | 3.96 | 24.536 % | 2.528 | 1.323 | 6.557 | 21.488 % | 4.94 |

Evident from the results in Table 4.6 these models do not precisely estimate the inertia level in Denmark. The performance metrics on both test and train data are poor compared to baseline models. The estimations of the three models applied to the test set are shown in Figure 4.15. All three models underestimate the inertia level in the test set. The two models trained with an intercept term, **lr_prod** and **lr_thrm**, appear to give an estimation of around 6 GWs the entire test period, with no tendency to pick up the steplike increase in inertia level. **lr_prod_forced0intercept**, which had no intercept term, appears highly volatile and nowhere close to the actual inertia level.

#### 4.2.3.2 Discussion Denmark

When applying linear models to problems, the underlying assumption is that linear relationship exists between input and target variables. The results of linear models applied to Danish data suggests that such a linear relationship does not exist in this data.

The inertia level in Denmark varies in steps. A possible explanation for this is that the number of generators in Denmark, price area DK2, is relatively low compared to the other Nordic countries. 25 generators are monitored by CB status, and 22 generators by MW output, according to *Future system inertia 2* [15]. With such a limited number of connected generators, it is easy to identify in Figure 4.12 the coupling or decoupling of a unit as either a sharp rise or drop in the *InertiaDE* value.

Looking for possible relationships between production and inertia level, wind and solar production in Denmark appeared to have no relevance at all. This is as expected, as these generating units are not connected synchronously to the power system, therefore not contributing with inertia. A closer look at the test data revealed a clear relation between thermal production columns and inertia, as visualized in Figure 4.16. The two thermal production columns *DK2_Thrm_Di* (Local power production) and *DK2_Thrm_In* (Central production) are shown as a stacked area plot. The inertia level in price area DK2 is displayed on the secondary axis. During the period, the power production from local power production units is averaging 100 MWh/h. The power production from central producing units is close to zero for much of the summer, with some exceptions. From July 14, to around July 25, the contribution from central power units increases to around 70 or 80 MWh/h, and coincides perfectly with the increase in inertia value from 7.5 to 9 GWs. This relation is also visible in 4 periods in August. This indicates that some useful information regarding inertia level can be extracted from the thermal production data, but linear models do not capture it. It is possible that other regression models, such as tree-based regression models or nearest neighbour models can be applied to this data, but those are topics for further work and out of scope in this thesis.

The reason that the inertia level was much higher in 2019 than in 2018 is that Energinet kept Amager power plant, one of the larger combined heat and power plants, online throughout summer 2019. Because of ongoing line work on the power grid around Copenhagen, the power plant remained in standby to ensure N-1 security in the power grid, according to a representative from Energinet [58].

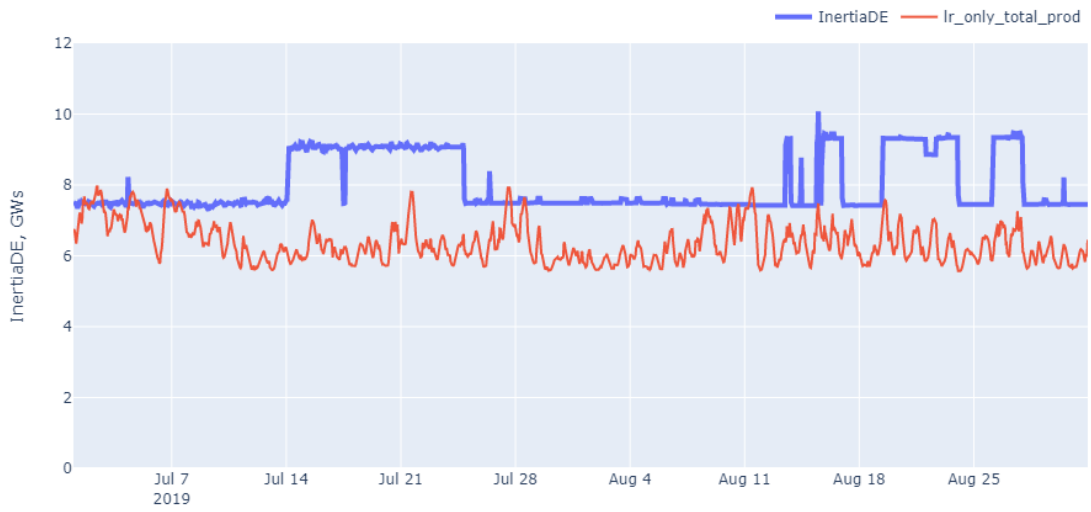



**Figure 4.14:** *Results of a linear regression model using the total Danish production as explanatory variable.*

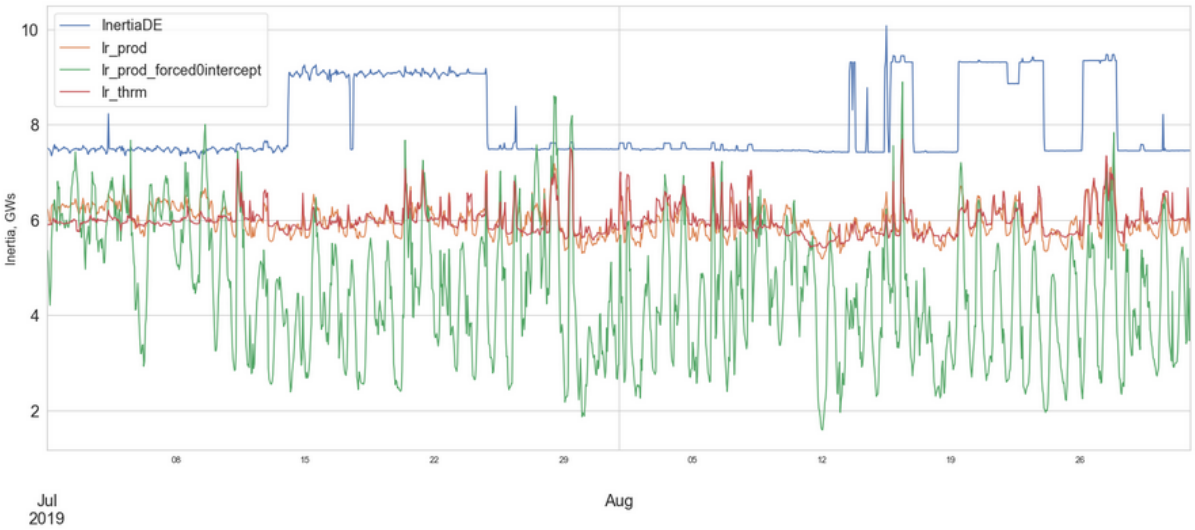



**Figure 4.15:** *Results of more linear models run on Danish data*

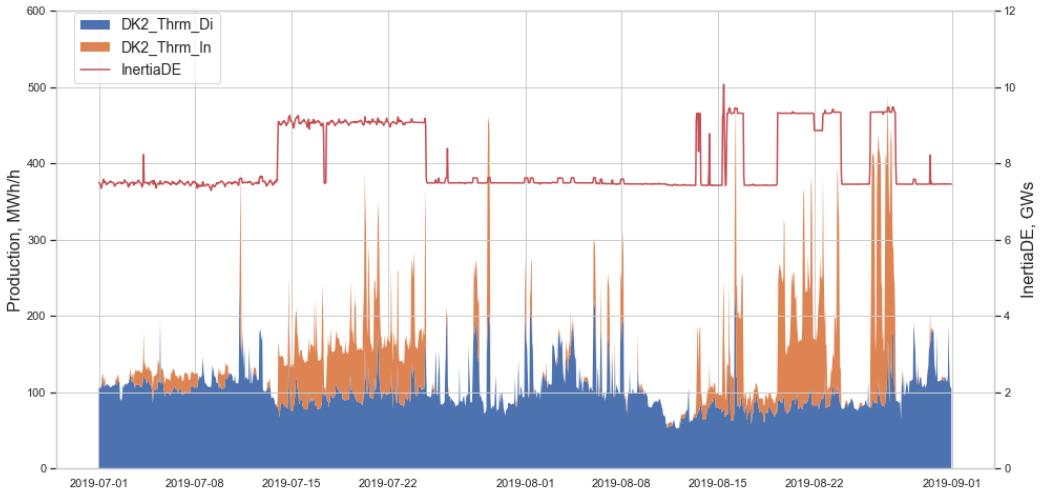



**Figure 4.16:** *Stacked area plot of thermal production compared to the Danish inertia level during two months in summer 2019.*

### 4.2.4 Finland

From 2018 through 2019, the average Finnish contribution to the kinetic energy level in the NPS was 19 %. The mean value in the period was 38 GWs, with a minimum value of 22.8 GWs and a maximum value of 50.6 GWs. Finnish inertia is displayed in blue along with the total level of Finnish production shown in red in Figure 4.17. The inertia level in Finland appears to stay on a level of about 45 GWs in the winter period, peaking at 50 GWs. In the summer period of 2018 and 2019, the inertia level was typically 30-35 GWs.

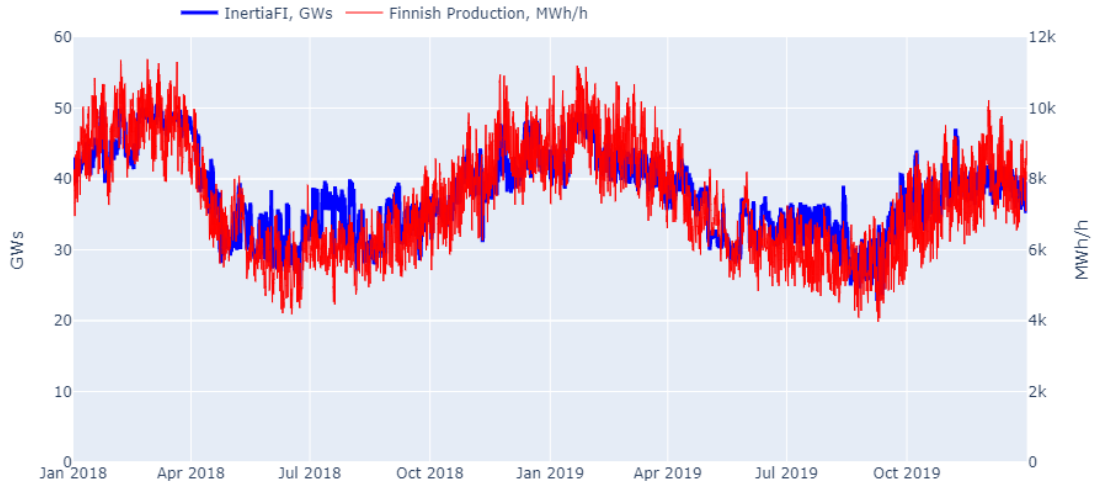



**Figure 4.17:** *Comparing the Finnish inertia level to the total production in Finland.*

#### 4.2.4.1 Baseline models - Finland

As described in section 3.4.1, baseline models were applied to the test set, and these models are shown in Figures 4.18 and 4.19. Figure 4.18 displays the Finnish inertia level *InertiaFI* shown in blue, along with **Naive_1h** (red) and **Naive_24h** (green). July 1, 2019 was a Monday. The typical inertia weekly and daily seasonal patterns are now well established, although the daily inertia level variations often appear more dampened in Finland. Weekends do also differ less from normal work days.

Figure 4.19 displays the inertia level (blue) in August 2019, the second month in the test data. In this figure, the **Naive_24h** model (red) and **Naive_week** model (green) are also shown. This figure illustrates the weaknesses of naive models, clearly exemplified August 13, and later, where the inertia level rises suddenly for a couple of days before entering a week with lower inertia. As indicated by the model names, the 24h model can only react to these changes a full day after the inertia changes, while the week model first underestimates inertia for a whole week, before completely overestimating the inertia level the next week, starting August 18th. The result of baseline models applied to the Finnish test data can be found in Table 4.7, along with the results of the linear models described below.

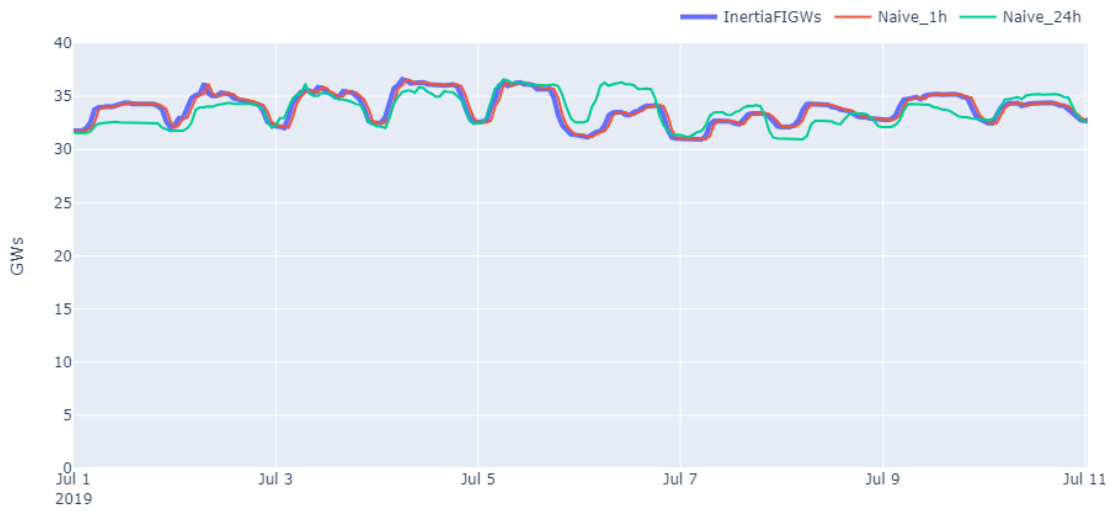



**Figure 4.18:** *Baseline models applied to the Finnish test data. Naive 1h and 24h models are displayed along with the kinetic energy level in Finland InertiaFI the first nine days in the test set. July 1, 2019 was a Monday.*

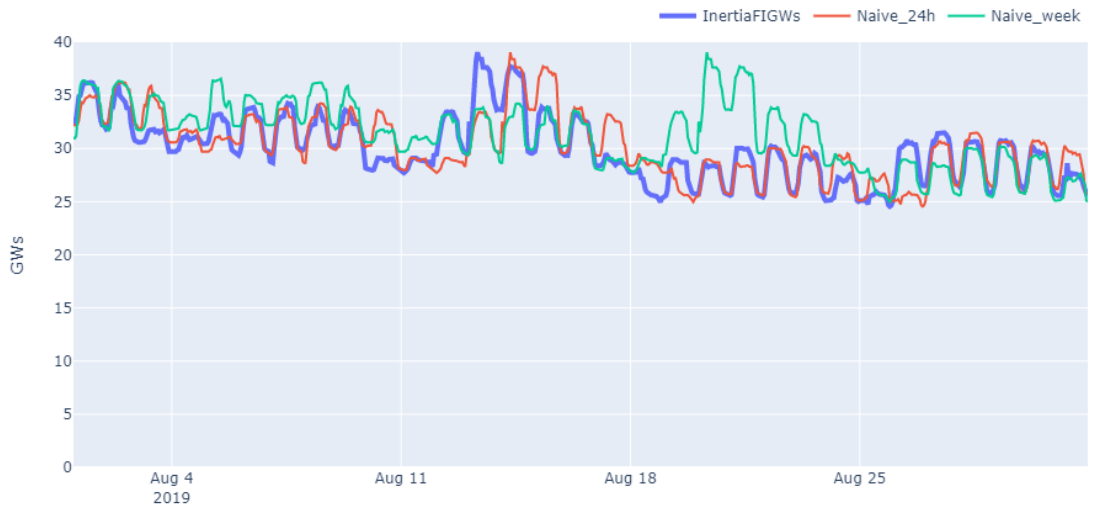



**Figure 4.19:** *Baseline models applied to the Finnish test data. The* ***Naive_24h*** *and* ***Naive_week*** *models are displayed along with the kinetic energy level in Finland InertiaFI in August 2019 (the second month in the test set). August 4, 2019 was a Sunday.*

**Table 4.7:** *Results of models trained and tested using Finnish level of inertia as target variable.*

| data: InertiaFI | | TEST DATA (2019-07 to 2019-08) | | | | TRAIN DATA (2018-01 to 2019-06) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Modelname** | **n_features** | **MSE** | **MAE (GWs)** | **MAX_E (GWs)** | **MAPE** | MSE | MAE (GWs) | MAX_E (GWs) | MAPE | intercept (GWs) |
| Naive_1h | 1 | 0.282 | 0.320 | 3.199 | 1.011 % | | | | | |
| Naive_2h | 1 | 0.901 | 0.611 | 5.303 | 1.928 % | | | | | |
| Naive_24h | 1 | 2.900 | 1.210 | 5.858 | 3.821 % | | | | | |
| Naive_week | 1 | 4.712 | 1.463 | 10.716 | 4.801 % | | | | | |
| lr_only_total_prod | 1 | 4.892 | 1.755 | 6.38 | 5.747 % | 7.207 | 2.182 | 9.903 | 5.804% | 11.58 |
| lr_prod | 6 | 3.423 | 1.469 | 4.629 | 4.970 % | 1.818 | 1.073 | 9.622 | 2.834 % | 9.85 |
| lr_prod_forced0intercept | 6 | 3.515 | 1.687 | 3.714 | 5.401 % | 2.516 | 1.259 | 11.513 | 3.363 % | - |
| lr_prod_no_unkn | 5 | 3.308 | 1.519 | 4.092 | 5.053 % | 3.986 | 1.558 | 10.990 | 4.052 % | 10.07 |
| lr_prod_no_wind | 5 | 3.494 | 1.495 | 4.545 | 5.050 % | 1.830 | 1.076 | 9.524 | 2.841 % | 10.19 |

#### 4.2.4.2 Linear models Finland

**Using total production as an explanatory variable**

The first model tested out in Finland was **lr_only_total_prod**, which used only 1 column as an explanatory variable - total production in Finland. In terms of MAPE, listed in Table 4.7, this model achieved similar results on the training and test set, scoring 5.804 % and 5.747 % respectively. However, this result is worse than all baseline models applied to the test data in Finland. The intercept for this model was 11.58 GWs, and the coefficient given to the explanatory column was 3.54. Figure 4.20 displays the model estimate as an orange line compared to the target value *InertiaFI* shown in blue. The model estimate often exaggerates peaks and crests, and the estimates are in periods constantly underestimating inertia, such as the period between July 15 and 29. The model does not estimate the significant increase in inertia August 13, but stays at the same level. The model also estimates inertia at the same level when *InertiaFI* drops to a lower level the last two weeks in August. This drop in *InertiaFI* is assumed to be caused by the disconnection of the nuclear power plant block Loviisa 2 for planned maintenance on August 19th (Table 4.5).

**Using all available production columns as explanatory variable**

The production data in Finland is split into six columns (production types), listed in Table 3.6. These columns were explanatory variables in the **lr_prod** and **lr_prod_forced0intercept** models. Compared to the model using total production, both models scored a better estimation of inertia in the test set, achieving MAPE of 4.970 % and 5.401 %, respectively. The intercept of **lr_prod** was 9.85 GWs. The coefficients for these models can be found in Appendix C, Table C.10, and for both **lr_prod** and **lr_prod_forced0intercept**, the production column unknown is given a high coefficient. The results of these models on the test data is displayed in Figure 4.21, where the target value *InertiaFI* is displayed in blue, **lr_prod** in orange, and **lr_prod_forced0intercept** is the green line. The general trend throughout the whole period is that both models overestimate inertia. Both models capture the increase in inertia on August 13th, but overestimate inertia for the rest of the test set.

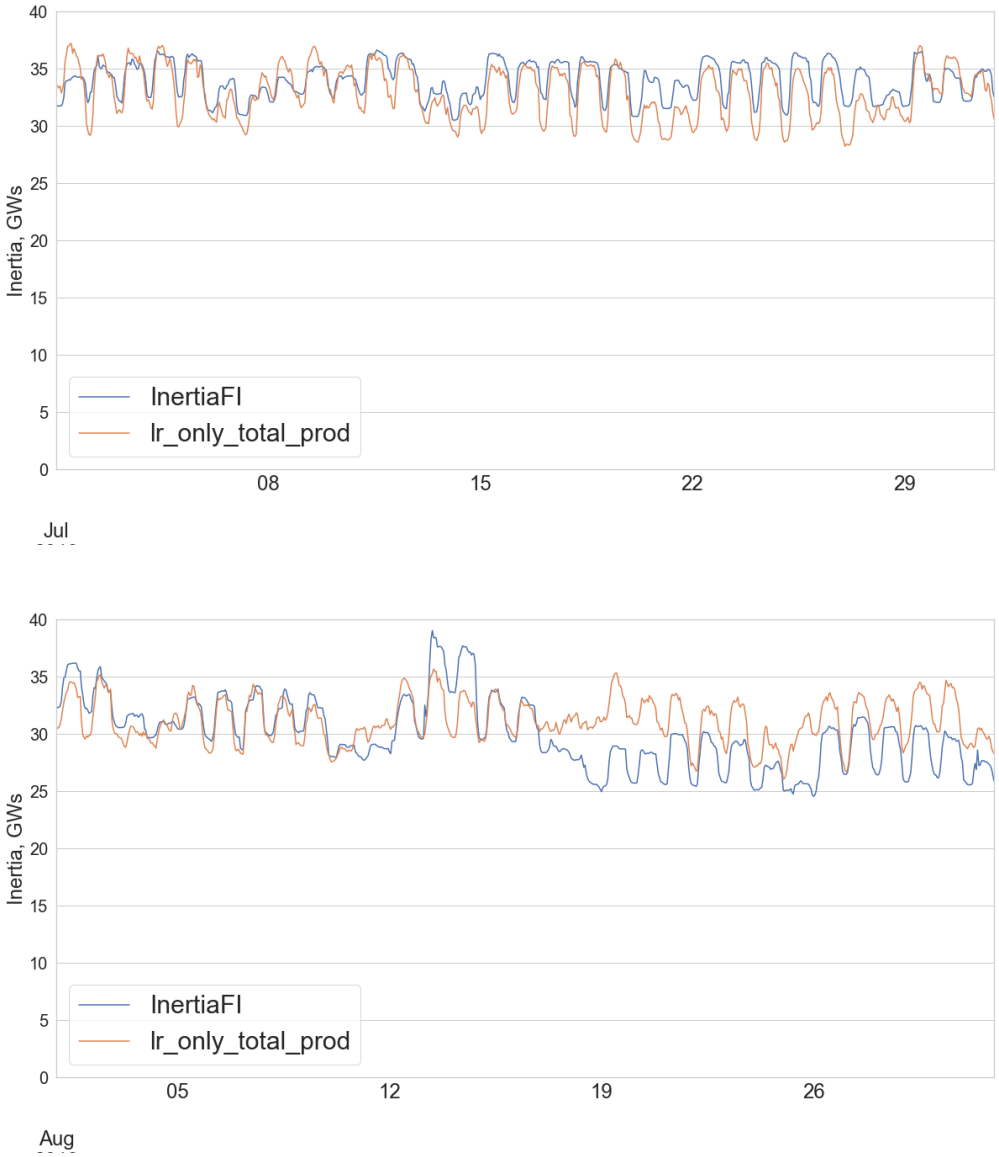



**Figure 4.20:** *Result of applying a linear regression model using the total production in Finland as an explanatory variable to estimate Finnish inertia.*

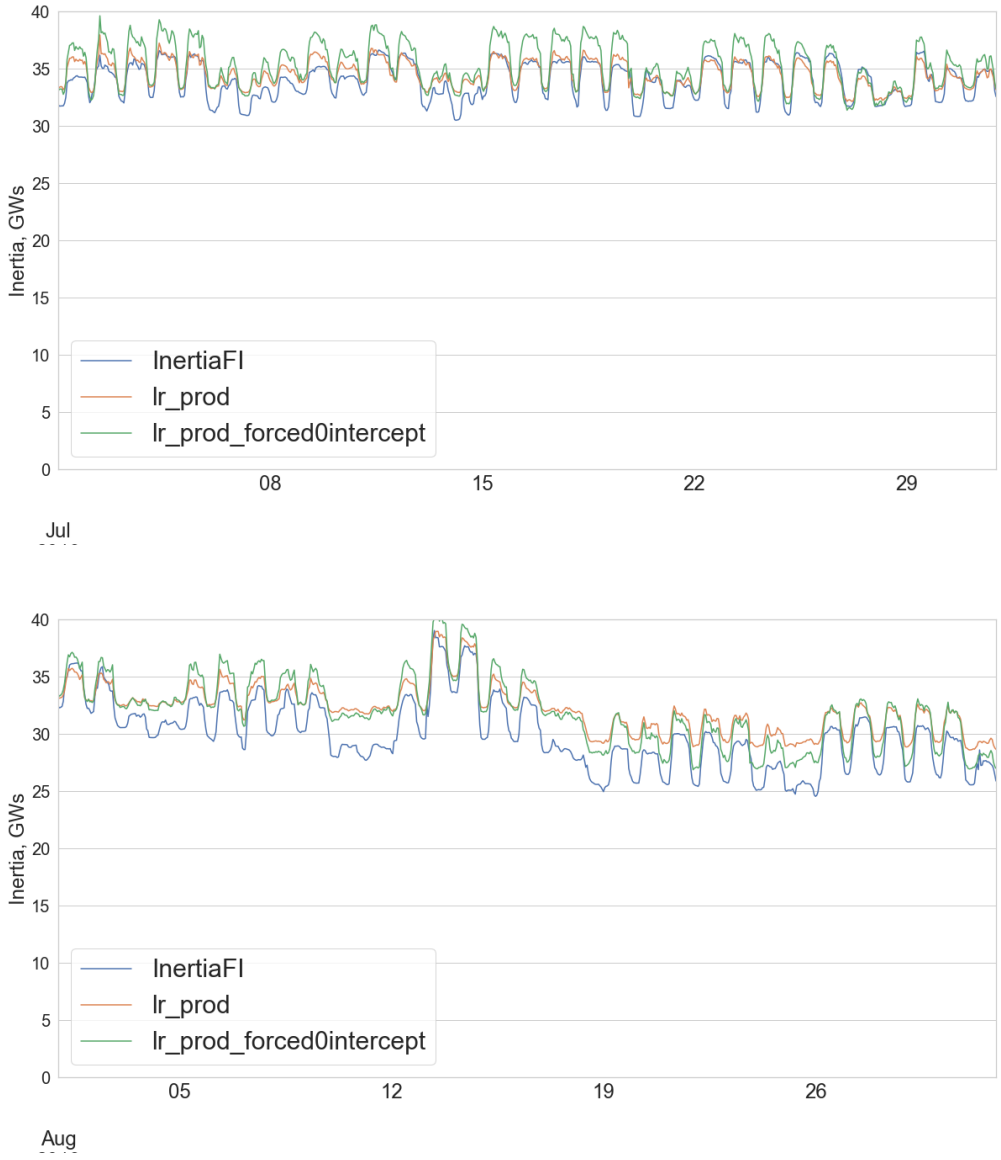



**Figure 4.21:** *Result of applying linear regression models using all available production columns in Finland as explanatory variables.* ***lr_prod*** *is trained with, and* ***lr_prod_forced0intercept*** *is trained without an intercept term.*

**Leaving out production columns in Finland**

Two more models were tested in Finland, where production columns were left out. The first model excluded unknown production, **lr_prod_no_unkn**. This models performance on test data was a MAPE of 5.053 % (Table 4.7), while the fit on training data had a MAPE of 4.052 %. Compared to **lr_prod**, the metric results on the training data dropped significantly, indicating that Finland's unknown column does include production that contribute with inertia.

The second model excluded the wind production column, **lr_prod_no_wind**. For this model, the metric results on the training data were close to unaltered compared to **lr_prod**, both models getting the same results across all metrics. For the test set, this model got a MAPE of 5.050 %.

Model coefficients for these models are listed in Appendix C, Table C.10. Thermal columns got higher coefficients as a result of leaving out the unknown production. When leaving out wind production, the other production coefficients were mostly unchanged.

The results of these models applied to the test set are displayed in Figure 4.22. The target value *InertiaFI* is shown in blue. **lr_prod** is shown as a reference in orange, **lr_prod_no_unkn** is shown in green and **lr_prod_no_wind** is shown in red. There are two key observations: The orange and red lines are hard to distinguish, meaning the two models produce similar inertia estimates. Reviewing production data, there were considerable changes in wind production, also in the test set. Wind production, therefore, does not seem to contribute additional information to the top-down inertia estimation in Finland. The other key observation from the figure is that **lr_prod_no_unkn** shown in green fail to capture the peak in inertia that occurs on the 13th and 14th of August. **lr_prod** and **lr_prod_no_wind**, which include unknown production as an explanatory variable are estimating this peak.

#### 4.2.4.3 Discussion Finland

In Finland, no linear models trained on production data were able to produce a better estimation than naive persistence baseline models on the chosen test set. The best performing model on the test set was **lr_prod** using all six production columns, but removing the wind column as an explanatory variable did not lead to significant differences in the estimated inertia.

Listed in Table 4.5, the nuclear reactor Loviisa 2 went into maintenance on August 19. After this date, *InertiaFI* fell to a daily level of 25-30 GWs. This period of low-level inertia was not picked up by any linear model, which is a potential problem as it is exactly situations with low-level inertia that are necessary to predict, when TSOs are planning for future operation.

The results on Finnish data indicated that the production column "Unknown" was important for the linear models' inertia estimation. When investigating this, it was discovered that the naming of the column was misleading - in the raw data, the naming of this production column is "*Electricity production, reserve power plants and small-scale production - real-time data*". While some small scale production is estimated and included in this production data, which was the reason for naming the column "Unknown", it also includes real-time measurement of data from the Finnish reserve power

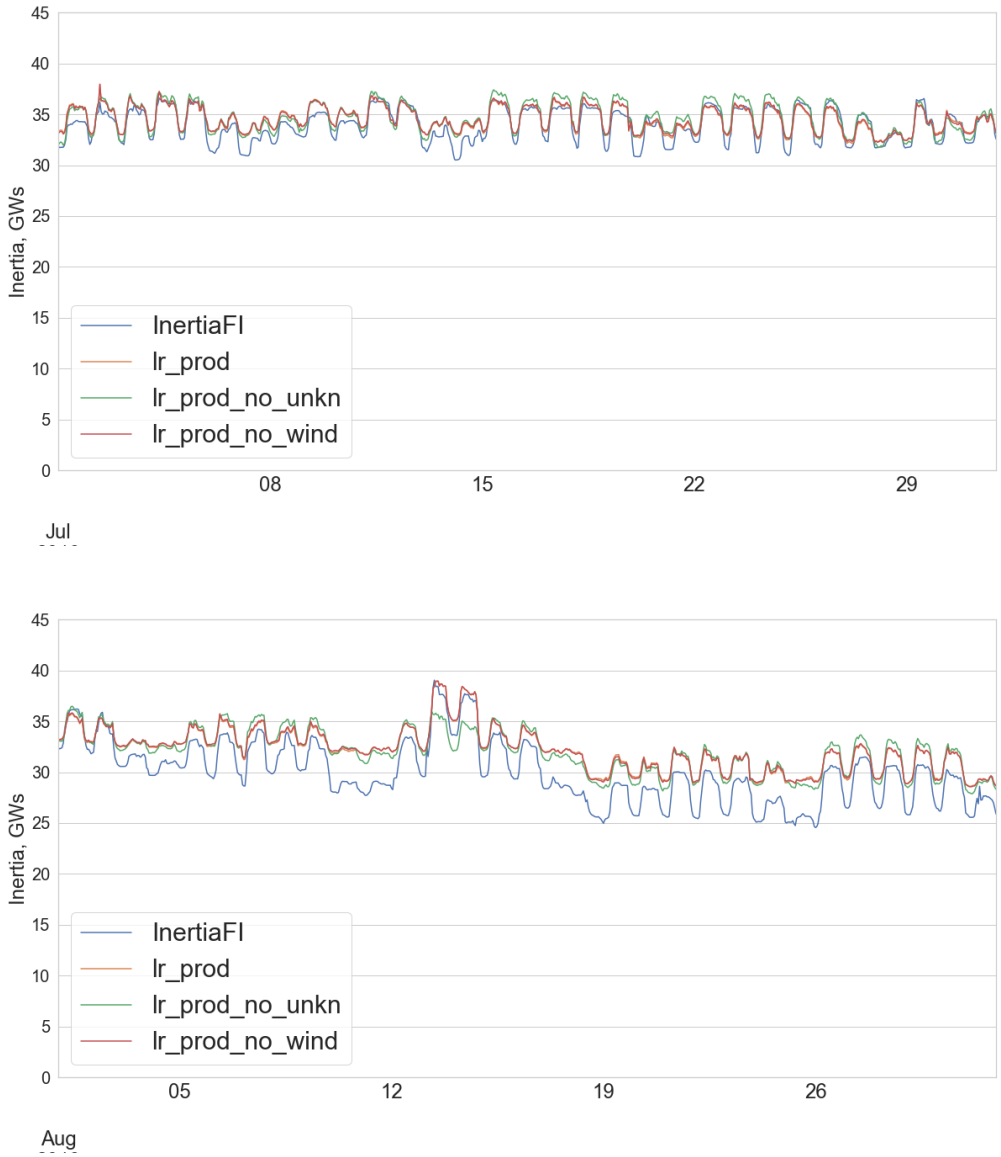
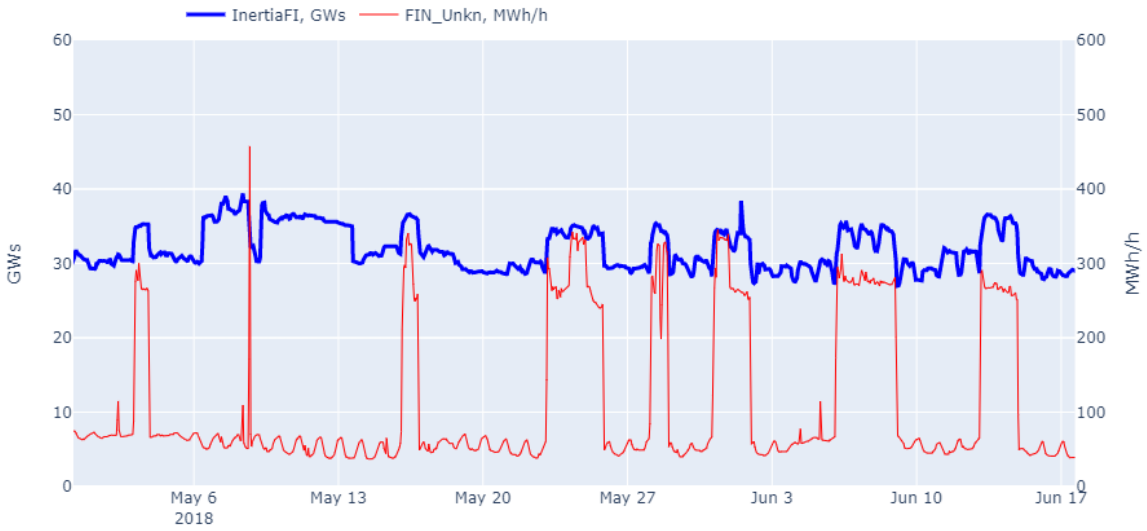


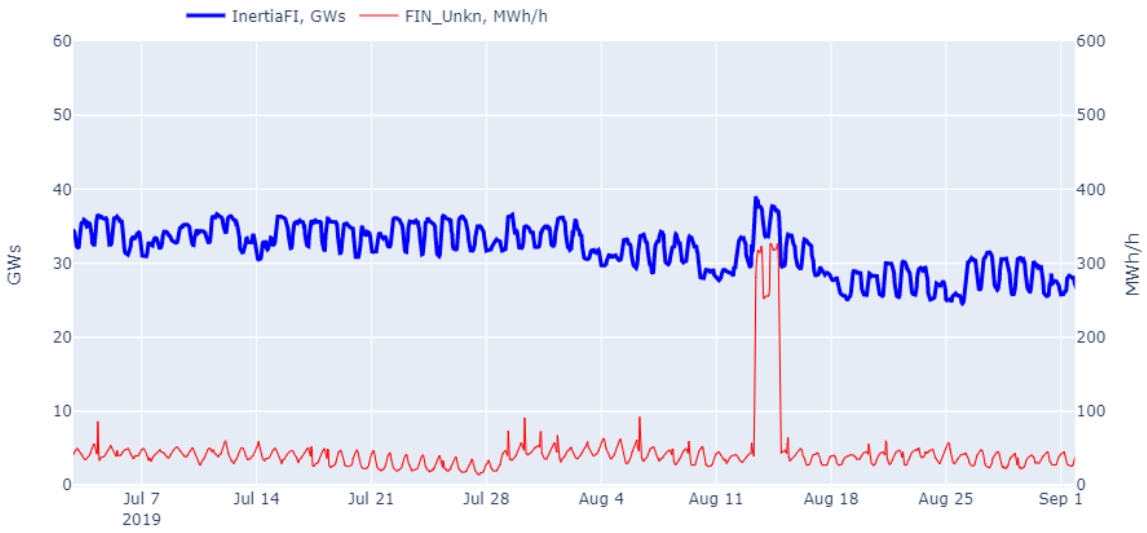

**Figure 4.22:** *Result of applying linear regression models where production columns from unknown and wind production were held out.*

plants. Fingrid owns power plants with a total of 927 MW capacity and leases reserve power plants with 291 MW capacity [59]. To investigate why linear models gave the column such a high coefficient, "Unknown" production and inertia were compared in a dual-axis line plot shown in Figure 4.23. In **(a)** and **(b)**, *InertiaFI* is shown in blue, while *FIN_Unkn* is shown in red. In **(a)**, a period in the training set chosen arbitrarily, *FIN_Unkn* has a base value of around 50 MWh/h but also spikes to values of 300 MWh/h in several periods. In those same periods, the inertia level becomes significantly higher.[3] The same can be observed in the test period, August 13 and 14, shown in **(b)**.

[3]The period with high inertia between May 6-14 shown in **(a)** does, however, not seem related to *FIN_Unkn*.




**(a)** *Finnish production column FIN_Unkn and Finnish inertia in May and June 2018.*




**(b)** *Finnish production column FIN_Unkn and Finnish inertia in the test data, July and August 2019.*

**Figure 4.23:** *Comparing the Finnish inertia level to the production column FIN_Unkn.*

### 4.2.5 Discussion of results per country

After running models per country, this subsection discusses some general observations.

#### Norwegian inertia estimation

The results on the Norwegian data differ significantly compared to the other countries, and there appears to be a stronger correlation between production and inertia level in Norway than in the other countries. Norway was the only country where using only total production as an explanatory variable could estimate inertia better than the baseline models.

The results indicate that the estimation of kinetic energy implemented in Norway was based on a different model than in the other countries. According to *Future System Inertia 2* "*the online kinetic energy estimation is based on the total production level in each of eight consumption areas, scaled with an average inertia constant, 3.44 s, factor for inertia, operation point, and power factor.*" [15]. The coefficients of some of the Norwegian models in this thesis are listed in Table 4.3, and the model coefficients are comparable to the average inertia constant (3.44), that [15] reports.

The inertia estimations of the other countries in the NPS are all based on circuit breaker status and inertia values per generator to estimate inertia, which is previously described as a bottom-up method in section 3.3 and in "*Ensuring future frequency stability in the Nordic synchronous area*" [18]. The Norwegian estimation appears to be a top-down approximation already. Because of this, there is no reason to do further investigation on the Norwegian data unless the Norwegian inertia estimation is changed to a bottom-up method. If the current Norwegian estimation of inertia is accurate enough, then forecasting of inertia based on forecasted production values appears to be ready for implementation in Norway.

Because the Norwegian inertia estimation is based on total production, the model coefficients given to individual production columns in Norway are meaningless to discuss, and therefore, the following discussion will be based on the results from Sweden, Denmark and Finland.

#### Swedish models

Sweden accounts for around 50 % of the Nordic inertia, and the results of models in Sweden implies that top-down estimations using production data can estimate inertia more accurately than day-ahead forecasting with persistence models. It is also possible that the Swedish model can be further improved, as thermal production columns appeared to not contribute in the NNLS models in Sweden that both had a MAPE of below 2 % on training data and 3 % on test data. The results from Denmark and Finland indicate that thermal columns do contribute with inertia, and thermal production plants use generators that are synchronously connected to the grid. Based on this, it is likely that the quality of data on Swedish production, especially data on thermal production, is of poor quality.

### Danish models

The Danish results indicate that the production level alone includes too little information to estimate the inertia level in Denmark, as linear models had poor performance both on the training data and the test data. There appears to be little to no value in applying this method in future work of creating a forecasting tool of Danish inertia, as the naive persistence models far outperformed linear models results.

### Finnish models

The results of models in Finland are somewhere in between the results from Sweden and Denmark. These models got a MAPE of roughly 3 % on training data, but this metric fell to 5 % on the test data, which is worse than all the tested baseline persistence models. In Finland, linear models also performed poorly on periods in test data with the lowest inertia level. In these periods, the models were overestimating the inertia level. In future work, this issue should be addressed - the low inertia hours are the essential periods from a forecasting perspective, as it is in these periods that FFR need is calculated. If forecasts of inertia are too high, it can lead to TSOs not procuring enough FFR, which threatens system stability in case of an underfrequency event.

### General remarks

As a general trend, the tested models were more accurate when the underlying data included more production. This could be due to the assumption of linearity - more production units lead to a closer linear relation between production and inertia level.

Sweden and Finland both include nuclear power units. Nuclear production columns are given stable high coefficients (around 8 in Sweden, and 5-6 in Finland) in all models where they are part of the explanatory variables. However, linear models seem unable to immediately pick up variations in inertia caused by nuclear units being decoupled for maintenance in the test data. With the number of nuclear units in the NPS being limited, the inertia contributions from these units can possibly be estimated and forecasted stand alone, as these units typically run at full capacity except for planned maintenance periods, usually with several weeks notice. To see how the models are affected, it could be interesting to re-run the models tested in this thesis with the inertia contribution from nuclear units excluded from the data.

Regarding other types of production, the results from Sweden and Finland indicate that production from hydro power plants are included in the models, usually getting coefficient values of around 2-4. While some thermal production columns in Sweden were given negative coefficients, and later were left out by NNLS models, results from Finland indicate that this type of production is also contributing in the models, with coefficients also in the range 2-4.

Wind production columns were in Sweden given coefficients close to 0 in models which allowed negative coefficients, and was set to 0 in nnls models. Leaving this production type out in Finland and Denmark made little to no difference to model performance. This is as expected, wind turbines are not connected synchronously to the grid, and are presumably not included in the Nordic TSOs estimation of inertia. Solar production columns were interestingly included in Swedish nnls models, indicating a positive

correlation with inertia. For future modelling, however, it should be safe to leave solar production out, as the contribution in the production mix is minor.

Unknown production was left out of Swedish NNLS models, which indicates that those columns were unimportant. This type of production is often not forecasted. In Finland, the results indicated that linear models were using the Finnish unknown production in the estimation. As discussed in section 4.2.4.3 this column includes estimated small scale production and electricity production from Finland's reserve power plants. Thus, the naming of the unknown column in Finland was not appropriate.

## 4.3 General Discussion

Creating a good forecasting model of inertia will enable a more secure calculation of the amount of FFR needed, which will enable TSOs to more accurately secure enough fast reserve capacity in the grid. From an economic point of view, it is important that the correct amount of reserves is procured. Buying too little reserves increases the risk of frequency falling out of system limits and causing blackouts that can have high economic and social consequences, but at the same time, securing sufficient reserves is also costly - Statnett bought FFR at the cost of 4.6 MNOK in the demonstration project in 2020 [30], and 22.5 MNOK[4] in 2021 [31].

This work has been a step towards creating a model to forecast inertia in the NPS. So far, historical values on production have been found to be able to estimate inertia with some precision through linear models. The same models can be used to forecast inertia. This could be done by simply substituting the input data to the created models with forecasted production values. Provided the forecasted production columns are comparable regarding production type and area, the models will generate forecasts of inertia. The evaluations of such forecasts are topics for further work.

### Strengths and weaknesses of method

#### Strengths

Using linear methods are straightforward, and it is transparent what goes into the model and what comes out. Further, the computations are not resource demanding. When training and testing, the coefficients are calculated quickly and applying the model on previously unseen data to create estimates are also done in seconds. Because data going into the models are not transformed or scaled in any way, expanding the models to be forecasting models only requires a good data source that provides forecasted values of production data. It could be the case that forecasts are being made using different geographical ranges, or generation types. In this case, the models must be tuned or retrained with data of the same geographical area as the forecasts.

#### Weaknesses

Because inertia is not a value easily measured, but a calculated value done by each TSO, it is not necessarily the case that the estimated inertia values that are used to train the linear models are the actual inertia values in the power system at all times.

[4]The actual cost is not stated. This value is calculated with the given details in [31].

The implementation of the inertia estimation is described briefly in Section 2.7, based on the reports *Future System Inertia 1* [13] and *2* [15]. The proposed method of forecasting inertia requires the inertia estimation by the TSOs to be good. Changes in how inertia is estimated might significantly impact the results of the work performed in this thesis.

A specific weakness with the method in this thesis is regarding the test set chosen as July and August 2019. While this period was chosen because it is typically a period of low inertia, testing the models on a single limited test period is only giving an indication of the model quality. The results of the models will likely be different if another test set is used. To further assess model quality, it is recommended to expand both the training and test set. Notable events in the power grid during summer 2019 might have had a big impact on the test set results. Therefore, the results in this thesis are exploratory and preliminary, and will require further verification and testing on more data before conclusions can be made regarding model accuracy.

Linear models assume that a linear relation exists between production power output and inertia level. This assumption might hold for a system with many generators that operate near total capacity when connected to the grid. However, if many generators operate with reduced capacity or are running idle, they still contribute with inertia. In this case, the assumption of linearity might not hold. E.g. the same observed power output can be associated with different inertia level estimations in hours one and two, if a power plant running idle in hour one is disconnected in hour two.

### Ideas of improvement

It is likely that system condition changes throughout the year. For this work, the inertia level during the summer, especially the periods with the lowest amount of inertia, is of interest. The models tested so far have been trained on data throughout an entire year, that includes both high and low values of production and inertia. A possible modification could be to train models specifically on low inertia hours and address if this improves the estimations in the summer periods.

# Chapter 5

# Conclusion and further work

## 5.1 Conclusion

The main goal in this thesis was to (**M**) develop linear models using production data to estimate the amount of kinetic energy in the Nordic Power System. During the work of the thesis, this goal has been addressed, and models have been applied both on a Nordic level, and later per country. These linear models were compared to naive persistence models to evaluate the performance if estimations are extended to forecasts in further development of the method.

On a Nordic level, the models got a MAPE of around 2 % on the test set in July and August 2019. This result was around the same precision as a naive persistence model using 1h lagged values to estimate inertia and much better than such persistence models with longer time horizons (24h or week) that both had a MAPE of around 5 %. The Norwegian implementation of inertia estimation was found to be already based on a top-down approach that uses production level rather than generator circuit breaker statuses to estimate inertia. In Sweden, using data on nuclear and hydro power production resulted in estimations of the inertia level with a MAPE of below 3 % on the test set. This is slightly less accurate than a 1h persistence model, which scored a MAPE below 2 %, but better than 24h and weekly persistence models. For Denmark, linear models performed poorly both on the training and test set. Forecasting inertia in Denmark will likely require other input data or models. In Finland, all generation types except wind contributed to a model that achieved a MAPE just above 5 % on test data, which means they do not estimate inertia better than persistence models applied to Finnish data. In addition, all linear models tested in Finland overestimated periods of low inertia.

Nuclear production was usually given a high stable coefficient in the various models, but in the test data, there was a tendency that variations in inertia caused by nuclear power plants being decoupled for planned maintenance was not picked up immediately. Also, hydro and thermal production were given stable positive coefficients in the models, while VRES production columns (wind and solar) were found to have little impact on model performance.

## 5.2 Further work

Because of the differences in production mix and data availability, it is recommended to proceed with models split per country when developing forecasting models. Even though the results of the models applied to the total Nordic level of inertia showed good results in terms of MAPE on the test set, domain-specific knowledge of power plants, as well as differences in the power mix within each country suggests that the best way forward is to proceed with individually estimating/forecasting inertia values per country. Further, the data used for the training of the models is calculated as the sum of four separate inertia estimations performed in Norway, Sweden, Denmark and Finland. Therefore, contributions to inertia by the generators in each country can technically only be observed in that respective country's inertia estimation.

Based on the results of this work, the continuation could focus on developing the Swedish and Finnish models further, and begin to map the performance of such models used in forecasting.

The estimation method currently in place for Norwegian data is already based on a top-down approach. Inertia forecasts can also be made in Norway, by providing forecasts of production. However, the evaluation of such forecasts will only measure how accurate the production forecasts are. Statnett is recommended to implement an estimation of inertia similar to that of the other countries in the NPS. Once a new bottom-up estimation of inertia is in place, it would be interesting to redo the methodology, and explore the results with the Norwegian power mix.

In Denmark, other models of forecasting should be tested, not based on a linear relation of production and inertia level. Other types of data, for instance the number of connected synchronous generators in Denmark, could be a better estimator than power measurements.

# Bibliography


[1] Norwegian Ministry of Petroleum and Energy. The development of norway's energy supply system. Online., February 2018. Available at https://energifaktanorge.no/en/om-energisektoren/energiforsyningens-utvikling-i-norge/, last accessed 18.feb 2020.

[2] Alexandra von. Meier. *Electric power systems, a conceptual introduction.* John Wiley & Sons, Ltd, 2006.

[3] Prabha Kundur, Neal J. Balu, and Mark G. Lauby. *Power system stability and control.* McGraw-Hill, 1994.

[4] Statnett. Nettutviklingsplan 2019, 2019. Downloaded from https://www.statnett.no/for-aktorer-i-kraftbransjen/planer-og-analyser/nettutviklingsplan-og-kraftsystemutredning/ [Downloaded 2020-03-09].

[5] Bloomberg New Energy Finance. New energy outlook 2019. Online, 2019. Available at https://about.newenergyfinance.com/new-energy-outlook/, Last accessed 2020-09-23.

[6] Prabha Kundur, John Paserba, Venkat Ajjarapu, Göran Andersson, Anjan Bose, Claudio Canizares, Nikos Hatziargyriou, David Hill, A.M. Stankovic, Carson Taylor, Thierry Van Cutsem, and V. Vittal. *Definition and Classification of Power System Stability IEEE/CIGRE Joint Task Force on Stability Terms and Definitions. Power Systems, IEEE Transactions on*, 19:1387 – 1401, 09 2004.

[7] NationalGrid. Technical report on the events of 9 august 2019. techreport, National Grid plc, September 2019.

[8] Marino Sforna and M. Delfanti. Overview of the events and causes of the 2003 italian blackout. In *2006 IEEE PES Power Systems Conference and Exposition*, pages 301 – 308, 12 2006.

[9] ENTSO-E. Requirement for minimum inertia in the nordic power system. techreport, Statnett, Svenska kraftnät, Fingrid, Energinet.dk, June 2021.

[10] ENTSO-E. Explanatory document for FRR dimensioning rules. Online, 2018. Available at https://consultations.entsoe.eu/markets/nordic-tsos-proposals-for-frr-and-frce-and-frequen/supporting_documents/4.%20Explanatory%20document%20for%20FRR%20dimensioning%20rules.pdf, last accessed 2020-03-30.

[11] Statnett. "NordLink", 2020. URL: https://www.statnett.no/en/our-projects/interconnectors/nordlink/. Accessed 2020-08-17.

[12] Statnett. "North Sea Link", 2020. URL: https://www.statnett.no/en/our-projects/interconnectors/north-sea-link/. Accessed 2020-08-17.

[13] Erik Ørum, Mikko Kuivaniemi, Minna Laasonen, Alf Ivar Bruseth, Erik Alexander Jansson, Anders Danell, Katherine Elkington, and Niklas Modig. Future system inertia. Technical report, ENTSO-E, 2015.

[14] Dimitrios Zografos. *Power system inertia estimation and frequency response assessment*. phdthesis, KTH, School of Electrical Engineering and Computer Science (EECS), 2019.

[15] Erik Ørum, Liisa Haarla, Mikko Kuivaniemi, Minna Laasonen, Anders Jerkø, Inge Stenkløv, Fredrik Wik, Katherine Elkington, Robert Eriksson, Niklas Modig, and Pieter Schavemaker. Future system inertia 2. Technical report, ENTSO-E, 2017.

[16] Päivi Leinonen. What is inertia? Online, September 2018. Available at https://www.fingridlehti.fi/en/what-is-inertia/, Last accessed 2020-03-02.

[17] Statnett. Fast Frequency Reserves, FFR, 2020. URL: https://www.statnett.no/for-aktorer-i-kraftbransjen/systemansvaret/kraftmarkedet/reservemarkeder/ffr/. Accessed 2021-11-26.

[18] Robert Eriksson, Niklas Modig, and Mikko Kuivaniemi. Ensuring future frequency stability in the Nordic synchronous area, 2019. *in 18th Wind Integration Workshop*, 2019.

[19] International Energy Agency. Electricity, 2021. Available at https://www.iea.org/fuels-and-technologies/electricity, last accessed 2021-29-11.

[20] Statnett. Langsiktig markedsanalyse - norden og europa 2020-2050, October 2020. Available at https://www.statnett.no/globalassets/for-aktorer-i-kraftsystemet/planer-og-analyser/lma/langsiktig-markedsanalyse-norden-og-europa-2020-50_revidert.pdf, last accessed 2021-11-29.

[21] Norwegian Ministry of Petroleum and Energy. The power market. Online., October 2021. Available at https://energifaktanorge.no/en/norsk-energiforsyning/kraftmarkedet/, last accessed 2021-11-30.

[22] Nordpool. Nordpool. Available at https://www.nordpoolgroup.com/. Last accessed: 2021-12-13.

[23] Nordpool. Day-ahead overview, 2019. URL: https://www.nordpoolgroup.com/maps/#/nordic. [Accessed 2020-03-16].

[24] D. P. Kothari and I. J. Nagrath. *Modern power system analysis*. Tata McGraw-Hill Education, 2003.

[25] ENTSO-E. Nordic balancing philosophy. techreport, Statnett, Svenska kraftnät, Fingrid, Energinet.dk, June 2016.

[26] Statnett. Primærreserver - FCR. URL: https://www.statnett.no/for-aktorer-i-kraftbransjen/systemansvaret/kraftmarkedet/reservemarkeder/primarreserver/. Accessed 2020-03-16.

[27] ENTSO-E. System Operation Agreement. 2019.

[28] ENTSO-E. Fast Frequency Reserve – Solution to the Nordic inertia challenge. techreport, Statnett, Svenska kraftnät, Fingrid, Energinet.dk, December 2019.

[29] Statnett. Fast Frequency Reserves 2018 - pilot for raske frekvensreserver. Online, September 2018. Available at https://www.statnett.no/contentassets/250c2da4dd564f269ac0679424fdfcfc/evaluering-av-raske-frekvensreserver.pdf, last accessed 2020-04-01.

[30] Statnett. Demonstrasjonsprosjekt for Fast Frequency Reserves (FFR) 2020, March 2020. URL: https://www.statnett.no/for-aktorer-i-kraftbransjen/systemansvaret/kraftmarkedet/reservemarkeder/ffr/ffrdemo2020/. Accessed 2021-11-30.

[31] Statnett. Demonstrasjonsprosjekt for Fast Frequency Reserves (FFR) 2021, 2021. URL: https://www.statnett.no/for-aktorer-i-kraftbransjen/systemansvaret/kraftmarkedet/reservemarkeder/ffr/ffrdemo2021/. Accessed 2021-11-30.

[32] Minna Laasonen and Mikko Kuivaniemi. Inertia of the Nordic power system. Available at https://www.fingrid.fi/en/electricity-market/electricity-market-information/InertiaofNordicpowersystem/. Last accessed 2021-12-03.

[33] Sebastian Raschka and Vahid Mirjalili. *Python Machine Learning, 2nd Ed.* Packt Publishing, Birmingham, UK, 2 edition, 2017.

[34] Seema Singh. Understanding the Bias-Variance Tradeoff. *Towards Data Science*, May 2018. Available at https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229. Last accessed 2021-12-14.

[35] T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2009.

[36] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009.

[37] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, and Carol Willing. Jupyter notebooks – a publishing format for reproducible computational workflows. In F. Loizides and B. Schmidt, editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87 – 90. IOS Press, 2016.

[38] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010.

[39] The pandas development team. pandas-dev/pandas: Pandas, February 2020.

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos,

D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[41] Plotly Technologies Inc. Collaborative data science, 2015.

[42] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[43] Michael L. Waskom. Seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.

[44] Mikko Kuivaniemi. Kinetic energy of the nordic power system - real time data. `https://data.fingrid.fi/en/dataset/kinetic-energy-nordic-realtime`. Last accessed: 2021-12-02.

[45] Energinet. Inertia, nordic synchronous area. `https://www.energidataservice.dk/tso-electricity/inertianordicsyncharea`. Last accessed: 2021-12-02.

[46] Mikko Kuivaniemi. Personal communication, feb 2020.

[47] International Organization for Standardization. ISO 8601 - Date and Time Format. Online, 2019.

[48] Svenska kraftnät. Elstatistik. `https://www.svk.se/om-kraftsystemet/kraftsystemdata/elstatistik/`. Last accessed: 2021-12-02.

[49] Fingrid. Fingrid open data. `https://data.fingrid.fi/en/`. Last accessed: 2021-12-02.

[50] Energinet. Electricity balance. `https://www.energidataservice.dk/dataset/electricitybalance/`. Last accessed: 2021-12-06.

[51] Niklas Modig and Birger Fält. Personal communication, March 2020.

[52] Energinet. Energi data service. `https://www.energidataservice.dk/`. Last accessed: 2021-12-02.

[53] Fingrid. Fingrid open data - api. `https://data.fingrid.fi/en/pages/api`. Last accessed: 2021-12-08.

[54] Nordpool. Historical Market Data. `https://www.nordpoolgroup.com/historical-market-data/`. Last accessed: 2021-12-17.

[55] Espen Aronsveen. Personal communication, jan 2020.

[56] Volue. Online. Available at `https://volue.com/`, last accessed: 2022-01-12.

[57] Nordic Unavailability Collection System. Online. Available at `https://www.nucs.net/`, Last accessed 2020-04-28.

[58] Erik Ørum. Personal communication, apr 2020.

[59] Fingrid. Reserve power plants, January 2022. Available online at `https://www.fingrid.fi/en/electricity-market/reserves_and_balancing/reserve-power-plants/`, last accessed 2022-01-18.

# Appendix A

# Collected data on production and inertia

The collected data used for modelling can be downloaded at the github repository `https://github.com/larspand/data_master_thesis`. Table A.1 displays the top 10 rows in the dataset *NordicInertiaWithNaN.csv.*

**Table A.1:** *Top 10 rows in the dataset NordicInertiaWithNaN.csv*

| | InertiaNordicGWs | InertiaDE | InertiaFI | InertiaNO | InertiaSE |
|---|---|---|---|---|---|
| 2017-12-31 23:00:00+00:00 | 196.12 | 9.50 | 40.55 | 46.21 | 99.86 |
| 2018-01-01 00:00:00+00:00 | 194.02 | 9.50 | 40.47 | 47.19 | 96.86 |
| 2018-01-01 01:00:00+00:00 | 189.51 | 9.46 | 40.30 | 44.15 | 95.59 |
| 2018-01-01 02:00:00+00:00 | 184.99 | 9.48 | 40.32 | 39.78 | 95.42 |
| 2018-01-01 03:00:00+00:00 | 184.62 | 9.37 | 40.26 | 39.53 | 95.46 |
| 2018-01-01 04:00:00+00:00 | 185.92 | 9.40 | 40.40 | 39.86 | 96.26 |
| 2018-01-01 05:00:00+00:00 | 189.15 | 9.39 | 40.68 | 41.43 | 97.65 |
| 2018-01-01 06:00:00+00:00 | 192.85 | 9.30 | 40.71 | 43.93 | 98.91 |
| 2018-01-01 07:00:00+00:00 | 192.88 | 9.47 | 40.72 | 42.30 | 100.39 |
| 2018-01-01 08:00:00+00:00 | 193.61 | 9.39 | 40.63 | 42.71 | 100.87 |
| 2018-01-01 09:00:00+00:00 | 196.12 | 9.50 | 40.55 | 46.21 | 99.86 |

# Appendix B

# Script to download inertia and production data from Fingrid open data

## B.1 Collecting data from Fingrid

The following script was used to download data from Fingrid open data [49].

In order to run the script you must edit the script inserting a personal API code you recieve when registering as a user.[1] The recieved API code should be put inside the quotation marks in line 29 in the script below.

**Code B.1:** *The script that was used to download Finnish data.*

```
1  # -*- coding: utf-8 -*-
2  """
3  Created on Wed Jan 15 08:40:41 2020
4
5  @author: larsan1
6  """
7  import pandas as pd
8  import requests
9
10 timeseries_prodFI =   [(188, 'FIN_Nucl'),
11                       (181, 'FIN_Wind'),
12                       (191, 'FIN_Hydr'),
13                       (202, 'FIN_Thrm_In'), #Industrial cogen
14                       (201, 'FIN_Thrm_Di'), #District cogen
15                       (205, 'FIN_Unkn'),
16                       (192, 'FIN_Prod_tot'),
17                       (193, 'Fin_Cons_tot')]
18
19 timeseries_inertiaFI = [(260, 'InertiaNordicGWs')]
20
21 def get_timeseries(start_time, end_time, variable_id, timeseries_name=None):
22     """
23     Collects all events from a time series with variable_id <variable_id>
24     from start_time to end_time from Fingrids open data platform.
25     """
```

[1] The form for registering a user is available at `https://data.fingrid.fi/open-data-forms/registration/`.

```
26      url = "https://api.fingrid.fi/v1/variable/" + str(variable_id) + "/events/json"
27      headers = {
28      'Accept': 'application/json',
29      'x-api-key':'' #<-- personal API code from fingrid goes here
30      }
31
32      params = (
33      ('start_time', start_time),
34      ('end_time', end_time),
35      )
36
37      q = requests.get(url, headers=headers, params=params) #query text
38
39      df = pd.DataFrame(q.json())
40      del df['end_time']
41      df['UTCtime'] = pd.to_datetime(df['start_time'])
42      df = df.set_index('UTCtime')
43      df.drop(['start_time'], axis=1, inplace=True)
44      if timeseries_name:
45          df = df.rename(columns={"value": timeseries_name})
46      return df
47
48
49  def get_month_timeseries(year, month, variable_id, timeseries_name=None,resample_hourly↩
        =False):
50      date = str(year) + '/' + str(month) + '/1'
51      date = pd.to_datetime(date)
52      start = str(date).replace(' ','T') + 'Z'
53
54      end_date = date + pd.DateOffset(months=1) - pd.DateOffset(seconds=1)
55      end = str(end_date).replace(' ','T') + 'Z'
56
57      df = get_timeseries(start_time=start,
58                          end_time=end,
59                          variable_id=variable_id,
60                          timeseries_name=timeseries_name)
61      return df
62
63
64  def collect_timeseries_from_tuple_per_month(timeseries_tuple, year, month):
65      df = pd.DataFrame()
66      resample_hourly = None
67      for varID, name in timeseries_tuple:
68          df = pd.concat([df,
69                          get_month_timeseries(year,
70                                               month,
71                                               varID,
72                                               name,
73                                               resample_hourly)],
74                                               axis=1)
75      return df
76
77  def collect_timeseries_year(timeseries_tuple, year):
78      df = pd.DataFrame()
79      for month in range(1,13):
80          print('Now getting month: '+str(month))
81          df = pd.concat([df,
82                          collect_timeseries_from_tuple_per_month(timeseries_tuple,
83                                                                  year,
84                                                                  month)],
85                          axis=0)
86      return df
87
88  if __name__ == "__main__":
89      """
90      If script is run directly - downloads timeseries on production listed in
91      timeseries_prodFI aswell as inertia timeseries from 2018 and 2019 and
92      stores both high resolution and hourly resolution data as csv files.
93      """
94      df2018 = collect_timeseries_year(timeseries_prodFI, year=2018)
95      df2019 = collect_timeseries_year(timeseries_prodFI, year=2019)
96
97      df_combined = pd.concat([df2018, df2019],axis=0)
```

```
    df_combined.to_csv('finnish_prod_con_data_2018_2019_minute_from_API.csv',
                       sep=';',
                       decimal=',')
    df_hourly = df_combined.resample('H').mean()
    df_hourly.to_csv('finnish_prod_con_data2018_2019_hourly_from_API.csv',
                     sep=';',
                     decimal=',')

    inertiadf2018 = collect_timeseries_year(timeseries_inertiaFI, year=2018)
    inertiadf2019 = collect_timeseries_year(timeseries_inertiaFI, year=2019)

    inertiadf_combined = pd.concat([inertiadf2018, inertiadf2019],axis=0)
    inertiadf_combined.to_csv('fingrid_inertia_data_2018_2019_minute_from_API.csv',
                              sep=';',
                              decimal=',')
    inertiadf_hourly = inertiadf_combined.resample('H').mean()
    inertiadf_hourly.to_csv('fingrid_inertia_data_2018_2019_hourly_from_API.csv',
                            sep=';',
                            decimal=',')
```

# Appendix C

# Model Coefficients

## C.1 Coefficients from LR models on the Nordic dataset

**Table C.1:** *Model coefficients for the model* ***lr_prod***

| Name | W_lr_prod |
|---|---|
| NO1_Unkn | 2170.826809 |
| NO3_Unkn | 210.788814 |
| NO4_Unkn | 134.785628 |
| SE1_Solr | 128.436154 |
| FIN_Unkn | 17.387636 |
| SE4_Solr | 16.588191 |
| SE1_Unkn | 16.352101 |
| SE4_Hydr | 12.977983 |
| NO4_Thrm | 10.577922 |
| NO1_Wind | 10.492891 |
| SE3_Solr | 8.707684 |
| SE3_Nucl | 8.466378 |
| SE4_Thrm | 6.836189 |
| SE2_Hydr | 4.401406 |
| FIN_Nucl | 4.389493 |
| SE1_Hydr | 4.109513 |
| NO4_Hydr | 3.488099 |
| NO5_Hydr | 3.397831 |
| NO2_Hydr | 3.391949 |
| NO4_Wind | 2.841869 |
| NO1_Hydr | 2.709625 |
| NO2_Wind | 2.596140 |
| DK2_Thrm_In | 2.530729 |
| FIN_Thrm_In | 2.468408 |
| NO3_Wind | 2.325450 |

| Name | W_lr_prod |
|---|---|
| SE3_Hydr | 2.205373 |
| SE3_Thrm | 1.896247 |
| DK2_Wind_Onshore | 1.430291 |
| SE2_Thrm | 1.333187 |
| NO3_Hydr | 1.260350 |
| FIN_Thrm_Di | 0.952575 |
| SE1_Wind | 0.531602 |
| FIN_Hydr | 0.459401 |
| SE3_Wind | 0.425990 |
| FIN_Wind | 0.003075 |
| SE2_Wind | -0.150424 |
| SE4_Wind | -0.371938 |
| NO1_Thrm | -0.431886 |
| DK2_Wind_Offshore | -0.529245 |
| DK2_Thrm_Di | -1.439189 |
| NO5_Thrm | -1.581035 |
| DK2_Solr | -2.476222 |
| NO3_Thrm | -15.312572 |
| SE1_Thrm | -18.841135 |
| SE3_Unkn | -21.878663 |
| SE4_Unkn | -26.661903 |
| SE2_Solr | -36.046248 |
| NO2_Thrm | -47.786668 |
| NO2_Unkn | -302.305897 |
| NO5_Unkn | -770.778465 |
| SE2_Unkn | -3465.955926 |

**Table C.2:** *Model coefficients for the model* ***lr__prod__forced0intercept***

| Name | W_lr_prod_forced0intercept | Name | W_lr_prod_forced0intercept |
|---|---|---|---|
| SE1_Unkn | 1196.435912 | SE2_Thrm | 2.158346 |
| NO1_Unkn | 781.829691 | NO3_Hydr | 2.142138 |
| NO3_Unkn | 424.144232 | NO3_Wind | 2.090699 |
| NO5_Thrm | 37.315173 | FIN_Hydr | 2.002674 |
| NO4_Thrm | 35.533917 | DK2_Solr | 1.960619 |
| FIN_Unkn | 19.849941 | DK2_Wind_Offshore | 1.904792 |
| NO1_Wind | 14.069990 | SE1_Wind | 1.401954 |
| SE3_Unkn | 13.816712 | DK2_Wind_Onshore | 1.196697 |
| NO1_Thrm | 10.428567 | DK2_Thrm_Di | 0.827916 |
| SE3_Solr | 10.338456 | SE3_Wind | 0.724870 |
| SE3_Nucl | 9.726472 | DK2_Thrm_In | 0.624295 |
| SE4_Hydr | 7.999746 | FIN_Wind | 0.234377 |
| FIN_Thrm_In | 7.809646 | SE2_Wind | -0.199899 |
| SE4_Thrm | 7.355023 | SE4_Wind | -0.333768 |
| FIN_Nucl | 6.592076 | FIN_Thrm_Di | -0.853234 |
| NO4_Wind | 5.076860 | SE4_Solr | -3.670514 |
| NO4_Hydr | 4.346296 | SE1_Thrm | -11.330300 |
| SE1_Hydr | 4.323283 | NO3_Thrm | -11.763139 |
| NO1_Hydr | 4.024130 | SE2_Solr | -16.029367 |
| SE2_Hydr | 3.837076 | SE1_Solr | -62.039024 |
| SE3_Thrm | 3.422844 | SE4_Unkn | -66.816841 |
| NO2_Hydr | 3.360811 | NO2_Thrm | -118.741023 |
| SE3_Hydr | 3.025198 | NO4_Unkn | -163.821644 |
| NO5_Hydr | 2.828166 | NO5_Unkn | -207.567192 |
| NO2_Wind | 2.655602 | NO2_Unkn | -273.710691 |
| | | SE2_Unkn | -491.157103 |

**Table C.3:** *Model coefficients for the model* ***lr__prod__nnls***

| Name | W_lr_prod_nnls | Name | W_lr_prod_nnls |
|---|---|---|---|
| NO1_Unkn | 1350.906085 | SE3_Thrm | 1.322033 |
| SE1_Solr | 200.159464 | SE3_Hydr | 1.202922 |
| SE1_Unkn | 117.011184 | NO3_Hydr | 0.752162 |
| NO1_Thrm | 62.437187 | DK2_Thrm_In | 0.718775 |
| NO3_Unkn | 25.239176 | FIN_Thrm_Di | 0.655877 |
| NO1_Wind | 20.367010 | DK2_Wind_Offshore | 0.649786 |
| FIN_Unkn | 18.038351 | FIN_Hydr | 0.503773 |
| NO4_Thrm | 10.640387 | DK2_Wind_Onshore | 0.247296 |
| SE4_Hydr | 10.333320 | SE1_Wind | 0.151596 |
| SE4_Solr | 9.518437 | SE1_Thrm | 0.000000 |
| SE3_Nucl | 8.854628 | SE3_Wind | 0.000000 |
| SE4_Thrm | 7.730238 | FIN_Wind | 0.000000 |
| SE3_Solr | 6.866338 | SE2_Solr | 0.000000 |
| SE2_Thrm | 5.910801 | SE2_Wind | 0.000000 |
| SE2_Hydr | 4.310281 | SE4_Unkn | 0.000000 |
| NO4_Hydr | 4.303810 | SE3_Unkn | 0.000000 |
| SE1_Hydr | 4.289897 | SE2_Unkn | 0.000000 |
| FIN_Nucl | 4.073588 | NO5_Unkn | 0.000000 |
| NO4_Wind | 3.883131 | NO5_Thrm | 0.000000 |
| NO5_Hydr | 3.774353 | NO4_Unkn | 0.000000 |
| NO2_Hydr | 3.104191 | SE4_Wind | 0.000000 |
| FIN_Thrm_In | 3.073447 | NO2_Unkn | 0.000000 |
| NO2_Wind | 2.977296 | NO2_Thrm | 0.000000 |
| NO3_Wind | 2.160669 | NO3_Thrm | 0.000000 |
| NO1_Hydr | 2.033181 | DK2_Thrm_Di | 0.000000 |
| | | DK2_Solr | 0.000000 |

**Table C.4:** *Model coefficients for the model* ***lr__prod__nnls__forced0intercept***

| Name | W_lr_prod_nnls_forced0intercept |
|---|---|
| SE1_Unkn | 1186.920705 |
| NO1_Unkn | 483.656512 |
| SE1_Solr | 234.326588 |
| NO3_Unkn | 140.762317 |
| NO1_Thrm | 76.949759 |
| NO5_Thrm | 38.963962 |
| NO4_Thrm | 37.998291 |
| SE2_Solr | 26.898738 |
| NO1_Wind | 22.960735 |
| FIN_Unkn | 19.111024 |
| SE3_Nucl | 9.782134 |
| SE3_Solr | 9.321491 |
| SE3_Unkn | 8.669664 |
| FIN_Thrm_In | 6.851378 |
| NO4_Wind | 6.240437 |
| FIN_Nucl | 6.073588 |
| NO4_Hydr | 4.623491 |
| SE1_Hydr | 4.592161 |
| SE4_Hydr | 4.288743 |
| SE2_Hydr | 3.779990 |
| SE4_Thrm | 3.480157 |
| NO1_Hydr | 3.371399 |
| NO2_Hydr | 3.095645 |
| NO5_Hydr | 3.081512 |
| NO2_Wind | 2.896021 |

| Name | W_lr_prod_nnls_forced0intercept |
|---|---|
| SE3_Hydr | 2.444456 |
| DK2_Wind_Offshore | 2.027919 |
| FIN_Hydr | 1.979710 |
| NO3_Hydr | 1.704253 |
| NO3_Wind | 1.616643 |
| SE3_Thrm | 1.264430 |
| SE1_Wind | 1.262513 |
| SE2_Thrm | 1.020036 |
| DK2_Solr | 0.722366 |
| DK2_Thrm_Di | 0.542383 |
| SE4_Solr | 0.466514 |
| DK2_Wind_Onshore | 0.399351 |
| SE3_Wind | 0.238599 |
| FIN_Wind | 0.100877 |
| NO2_Unkn | 0.000000 |
| NO2_Thrm | 0.000000 |
| NO4_Unkn | 0.000000 |
| NO5_Unkn | 0.000000 |
| SE2_Wind | 0.000000 |
| SE2_Unkn | 0.000000 |
| SE4_Wind | 0.000000 |
| SE4_Unkn | 0.000000 |
| DK2_Thrm_In | 0.000000 |
| FIN_Thrm_Di | 0.000000 |
| SE1_Thrm | 0.000000 |
| NO3_Thrm | 0.000000 |

## C.2 Coefficients from LR models on Norwegian data

**Table C.5:** *Model coefficients for the model* ***lr_prod*** *on the Norwegian dataset.*

| Name | W_lr_prod |
|---|---|
| NO1_Unkn | 1345.687868 |
| NO1_Thrm | 22.596486 |
| NO2_Thrm | 9.176743 |
| NO4_Thrm | 8.535394 |
| NO1_Wind | 5.740622 |
| NO2_Wind | 3.592537 |
| NO4_Hydr | 3.533113 |
| NO5_Hydr | 3.465941 |
| NO2_Hydr | 3.439252 |
| NO3_Wind | 3.415414 |
| NO1_Hydr | 3.358781 |
| NO3_Hydr | 3.144954 |
| NO4_Wind | 2.222477 |
| NO5_Thrm | -0.584282 |
| NO3_Thrm | -2.365427 |
| NO4_Unkn | -39.274812 |
| NO2_Unkn | -67.516239 |
| NO3_Unkn | -105.476122 |
| NO5_Unkn | -280.387067 |

## C.3 Coefficients from LR models on Swedish data

**Table C.6:** *Model coefficients for* ***lr_prod and lr_prod_forced0intercept*** *on the Swedish dataset.*

| Name | W_lr_prod |
|---|---|
| SE1_Solr | 157.917476 |
| SE3_Nucl | 8.310900 |
| SE3_Solr | 5.633400 |
| SE4_Solr | 5.265822 |
| SE1_Hydr | 4.859401 |
| SE2_Hydr | 2.797125 |
| SE3_Thrm | 1.421118 |
| SE1_Wind | 1.288550 |
| SE3_Hydr | 0.551859 |
| SE4_Wind | 0.038168 |
| SE3_Wind | -0.151125 |
| SE2_Wind | -0.346637 |
| SE4_Hydr | -1.136663 |
| SE4_Thrm | -1.534753 |
| SE2_Thrm | -2.840839 |
| SE4_Unkn | -32.300566 |
| SE3_Unkn | -35.198325 |
| SE1_Thrm | -41.524265 |
| SE2_Solr | -76.280118 |
| SE1_Unkn | -485.100383 |
| SE2_Unkn | -4073.530351 |

| Name | W_lr_prod_forced0intercept |
|---|---|
| SE2_Unkn | 2230.972402 |
| SE1_Unkn | 897.822368 |
| SE1_Solr | 317.361596 |
| SE2_Solr | 12.758666 |
| SE3_Nucl | 10.206737 |
| SE3_Unkn | 8.136398 |
| SE1_Hydr | 5.483317 |
| SE3_Solr | 4.618291 |
| SE2_Thrm | 3.481137 |
| SE2_Hydr | 3.025292 |
| SE1_Wind | 2.755600 |
| SE3_Hydr | 0.854742 |
| SE3_Wind | 0.344438 |
| SE4_Wind | 0.286690 |
| SE2_Wind | -0.363340 |
| SE3_Thrm | -0.556107 |
| SE4_Thrm | -2.508364 |
| SE4_Solr | -3.296387 |
| SE4_Hydr | -7.537615 |
| SE1_Thrm | -17.958141 |
| SE4_Unkn | -71.723326 |

**Table C.7:** *Model coefficients for* ***lr__prod__nnls*** *on the Swedish dataset.*

| Name | W_lr_prod_nnls |
|---|---|
| SE1_Solr | 378.265865 |
| SE4_Solr | 10.662888 |
| SE3_Nucl | 8.061258 |
| SE1_Hydr | 4.757882 |
| SE3_Solr | 4.220443 |
| SE2_Hydr | 2.681092 |
| SE3_Hydr | 0.354688 |
| SE4_Hydr | 0.000000 |
| SE1_Unkn | 0.000000 |
| SE2_Solr | 0.000000 |
| SE4_Unkn | 0.000000 |
| SE3_Unkn | 0.000000 |
| SE2_Unkn | 0.000000 |
| SE4_Thrm | 0.000000 |
| SE1_Wind | 0.000000 |
| SE3_Thrm | 0.000000 |
| SE1_Thrm | 0.000000 |
| SE4_Wind | 0.000000 |
| SE3_Wind | 0.000000 |
| SE2_Wind | 0.000000 |
| SE2_Thrm | 0.000000 |

**Table C.8:** *Model coefficients for* ***rr__prod__nnls*** *on the Swedish dataset.*

| Name | W_rr_prod_nnls |
|---|---|
| SE3_Nucl | 7.970225 |
| SE1_Hydr | 4.710231 |
| SE2_Hydr | 2.711258 |
| SE3_Hydr | 0.459705 |
| SE3_Solr | 0.175861 |
| SE4_Solr | 0.074638 |
| SE2_Solr | 0.009807 |
| SE1_Solr | 0.001810 |
| SE2_Wind | 0.000000 |
| SE3_Wind | 0.000000 |
| SE4_Wind | 0.000000 |
| SE1_Thrm | 0.000000 |
| SE1_Wind | 0.000000 |
| SE3_Thrm | 0.000000 |
| SE4_Thrm | 0.000000 |
| SE1_Unkn | 0.000000 |
| SE2_Unkn | 0.000000 |
| SE3_Unkn | 0.000000 |
| SE4_Unkn | 0.000000 |
| SE4_Hydr | 0.000000 |
| SE2_Thrm | 0.000000 |

## C.4 Coefficients from LR models on Danish data

**Table C.9:** *Model coefficients for models on the Danish dataset.*

| W_lr_prod | |
|---|---|
| Name | |
| DK2_Thrm_Di | 8.856472 |
| DK2_Wind_Offshore | 3.533570 |
| DK2_Thrm_In | 2.838322 |
| DK2_Solr | 1.541590 |
| DK2_Wind_Onshore | -0.945953 |

| W_lr_prod_forced0intercept | |
|---|---|
| Name | |
| DK2_Thrm_Di | 25.177350 |
| DK2_Wind_Offshore | 12.614552 |
| DK2_Solr | 11.048126 |
| DK2_Thrm_In | 3.830418 |
| DK2_Wind_Onshore | -1.903883 |

| W_lr_thrm | |
|---|---|
| Name | |
| DK2_Thrm_Di | 9.210428 |
| DK2_Thrm_In | 2.695061 |

## C.5 Coefficients from LR models on Finnish data

**Table C.10:** *Model coefficients for models on the Finnish dataset.*

| W_lr_prod | |
|---|---|
| Name | |
| FIN_Unkn | 13.558345 |
| FIN_Nucl | 6.562157 |
| FIN_Thrm_Di | 2.892768 |
| FIN_Thrm_In | 2.061910 |
| FIN_Hydr | 1.987700 |
| FIN_Wind | 0.256099 |

| W_lr_prod_forced0intercept | |
|---|---|
| Name | |
| FIN_Unkn | 13.653046 |
| FIN_Nucl | 8.368340 |
| FIN_Thrm_In | 5.506961 |
| FIN_Hydr | 3.253513 |
| FIN_Thrm_Di | 1.641587 |
| FIN_Wind | 0.743239 |

| W_lr_prod_no_unkn | |
|---|---|
| Name | |
| FIN_Nucl | 5.928483 |
| FIN_Thrm_Di | 3.910237 |
| FIN_Thrm_In | 3.721950 |
| FIN_Hydr | 2.067639 |
| FIN_Wind | -0.606562 |

| W_lr_prod_no_wind | |
|---|---|
| Name | |
| FIN_Unkn | 13.293666 |
| FIN_Nucl | 6.537443 |
| FIN_Thrm_Di | 2.945612 |
| FIN_Thrm_In | 2.025998 |
| FIN_Hydr | 1.920660 |

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

Postboks 5003
NO-1432 Ås
Norway