Norwegian
University of
Life Sciences

Master's Thesis 2024 30 ECTS
Faculty of Science and Technology

Integration of stochastic modeling and machine learning for estimating inertia in the Nordic power system

Thomas Drageset
Environmental Physics and Renewable Energy

Abstract

This thesis proposes methods for estimating the stored kinetic energy within the Nordic power system using both a stochastic modeling of frequency to estimate a regional inertial constant and a Long Short-Term Memory (LSTM) model that directly estimates stored kinetic energy within the power systems. The stochastic model uses a linear model to estimate kinetic energy stored within the power system from regional inertial constant and aggregate power produced in the region. This approach relies on modeling frequency behavior as an Ornstein-Uhlenbeck process and using the diffusion coefficient to estimate a regional inertia constant.

The attempted estimation using the diffusion and LSTM models to estimate kinetic energy based on steady-state frequency measurements has shown promising results. The model's accuracy was evaluated using various metrics, including the Mean Absolute Percentage Error (MAPE) and the R2 score. For the year 2020, the diffusion model achieved a MAPE of 4.2% and an R2 score of 0.837, and the LSTM model achieved a MAPE of 3.1% and an R2 score of 0.917. Both results indicate a high ability to replicate the transmission system operator's(TSO's) estimation of kinetic energy stored in the power system, which was used as ground truth.

These findings highlight the potential of using stochastic models or LSTM networks for the estimation of kinetic energy in power systems. The two models display different tendencies during periods of low kinetic energy stored in the power system, with the LSTM model favoring underestimations and the diffusion model favoring overestimations. While the study suggests that the models are effective for the Nordic power system, their applicability might vary across different regions, necessitating further validations within these regions to ensure their generalizability and practical relevance on a global scale. Similar models could be utilized to improve the operational efficiency of the power grid, allowing for more rapid electrification of the Nordic power system and aiding in reaching the goal of net zero emissions.

Sammendrag

Denne master oppgaven foreslår metoder for å estimere lagret kinetisk energi i det Nordiske kraftsystemet ved bruk av både stokastisk modellering av frekvens for å estimere en regional treghetskonstant, og en Long Short-Term Memory (LSTM) modell som direkte estimerer lagret kinetisk energi i det Nordiske kraftsystemet. Den stokastiske modellen bruker en lineær modell for å estimere lagret kinetisk energi i kraftsystemet fra en regional treghetskonstant og samlet kraftproduksjon per produksjon type i regionen. Den stokastiske estimeringen er basert på å modellere frekvensoppførsel som en Ornstein-Uhlenbeck-prosess og bruk av diffusjonskoeffisienten for denne prosessen til å estimere en regional treghetskonstant.

Bruken av diffusjons- og LSTM-modellene for å estimere kinetisk energi basert på målinger av steady-state frekvens har vist lovende resultater. Modellens nøyaktighet ble evaluert ved hjelp av ulike målemetoder, inkludert Gjennomsnittlig Absolutt Prosentvis Feil (MAPE) og R2-score. For året 2020 oppnådde diffusjonsmodellen en MAPE på 4,2% og en R2-score på 0,837, og LSTM-modellen oppnådde en MAPE på 3,1% og en R2-score på 0,917. Begge resultatene indikerer en høy evne til å reprodusere TSOs estimering av lagret kinetisk energi i kraftsystemet, som ble brukt som fasit.

Disse funnene fremhever potensialet for bruken av stokastiske modeller eller LSTM-modeller for estimering av kinetisk energi i det Nordiske kraftsystemet. Estimatene fra de to modellene viser ulike tendenser for lav inertia situasjoner i det Nordiske kraftsystemet. LSTM modellen har en tendens til å underestimere inertia for disse periodene, mens diffusjons modellen har en tendens til å overestimere inertia for de samme periodene. Selv om oppgaven antyder at modellene er effektive for det nordiske kraftsystemet, kan deres anvendelighet variere på tvers av forskjellige regioner, noe som nødvendiggjør videre forskning innen disse regionene for å sikre at funnene generaliserer til andre kraftsystemer med ulike oppbygging. Lignende modeller kan potensielt benyttes for å forbedre driftseffektiviteten til det Nordiske kraftnettet, noe som kan tillate raskere elektrifisering av det Nordiske kraftsystemet og bidrar til å nå målet om nullutslipp raskere.

Acknowledgements

This thesis marks the end of my studies at the Norwegian University of Life Sciences (NMBU). I want to thank my family and friends for their support throughout my studies at NMBU and during the writing of my master's thesis. A special thanks goes out to my supervisor Leonardo Rydin Gorjão. Leo has been an excellent supervisor during the entire process. Our discussions around the research objectives, theory, methods, and results of my thesis have given me a much greater understanding of the topics I cover in the thesis. You have shown a constant willingness to proofread the different sections that you already have read four times over, and you have been supporting and helping from day one to the last days of writing. Thank you for being a great supervisor.

Thomas Drageset Ås, May 2024

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background and scope

Industrialization was the driving force when the Nordic power system and power generation in the Nordic region were built. From the 1960s and onwards, it became a collaboration between several countries [1] with the introduction of cross-border transmission lines. The cross-border cooperation has expanded and given rise to a Nordic synchronous region, as shown in the figure 1.

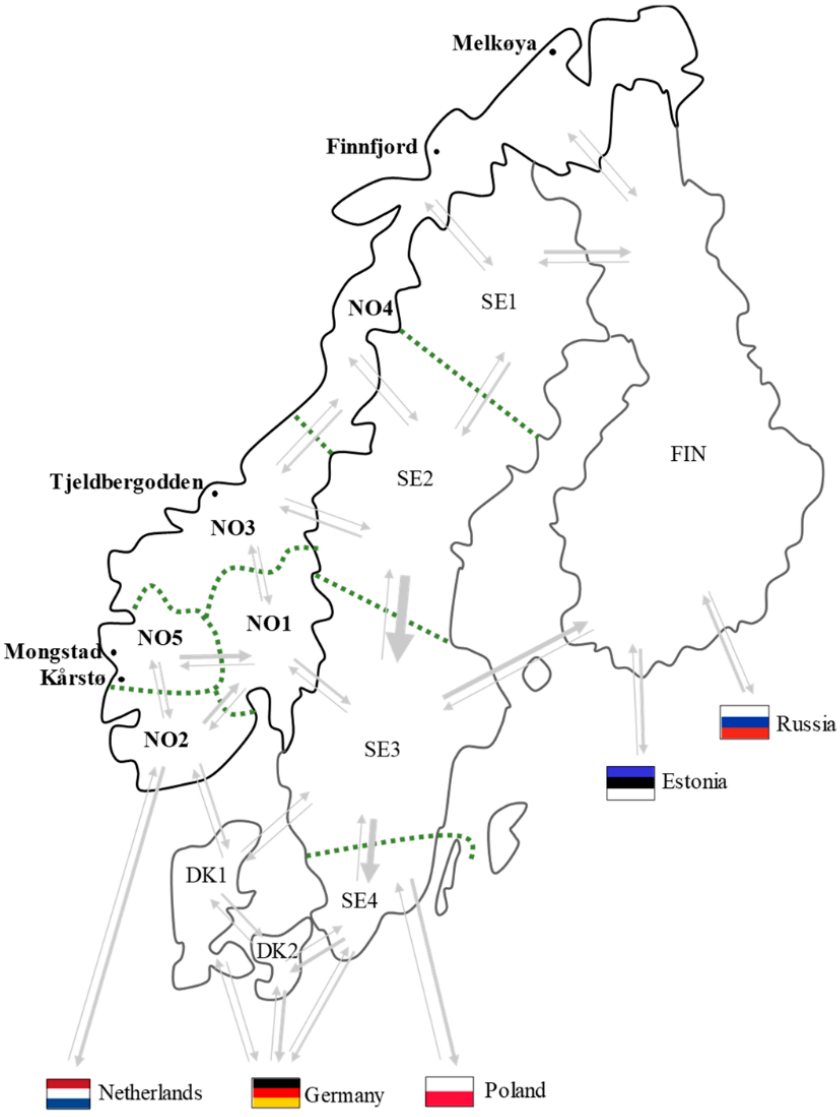

Figure 1: A map showing the Nordic synchronous region. It comprises the Norwegian, Swedish, Finnish, and a part of the Danish power systems, the DK2 price area. These regions share a common grid frequency of 50 Hz to facilitate the easy exchange of power within the region. Taken with permission from [2].

Within this region, the countries operate with a shared grid frequency of 50 Hz to easily facilitate the power exchange within the region. However, the challenges related to the environmental crisis we face to-

day demand a change to the power system. Norway's power system operator Statnett's long-term market analysis (LMA) states that the installed effect of inverter-based renewable energy (IBRE) in Norway will increase from around 5 GW to 25 GW by 2050 [3]. Furthermore, LMA indicates that the other Nordic countries will develop similarly towards 2050. Despite the positive impact IBRE has on the environment, the intermittent and disrupted nature of this production type poses new problems for transmission system operators (TSO) of power systems. The TSO is responsible for ensuring the operational security of the power system. One critical aspect is maintaining the power grid frequency within a predefined range of 48–52Hz [4]. If there is a trip of a single large power generator, i.e., a nuclear power plant or a large cross-border transmission line, large frequency deviations can occur due to the sudden mismatch between produced and consumed power. If not corrected, the frequency deviations can trigger safety features that lead to the disconnection of other production units and loads, which can lead to a cascading effect and possibly result in power outages. The kinetic energy stored in large synchronously connected generators causes the power system to resist these sudden changes in frequency [5]. However, IBRE is starting to displace the synchronous generation in the Nordic region. Most modern IBRE systems interface to the power grid using power electronics in such a way that their generation frequency is independent of the power-grid frequency, causing this type of production not to contribute kinetic energy to the power system. The expected increase in IBRE production would change the relative contribution from IBRE to the power produced in the Nordic power system [3], resulting in a larger share of power coming from these production types. Additionally, the focus on renewable energy production in the Nordic region [3] also causes thermal power plants to spend less time synchronized to the power system, further reducing the kinetic energy available in the power system.

The Nordic region also faces rapid electrification of many sectors in an attempt to reach the sustainability goals proposed by United Nations [6] causing an increasing need for electrical power [3, 7]. To meet the growing demand for electrical power, power system operators must operate the power system closer to their operational limits [8]. Utilizing more of the power system's capacity will, together with faster frequency dynamics resulting from increased production from IBRE [9], make power system operation more challenging. Historically, when the Nordic power system mainly consisted of synchronous generation, an assumption of adequate and almost constant inertia in the power system was satisfactory for frequency stability. This assumption is not necessarily valid for a power system with sufficient penetration of IBRE [9]. To operate the Nordic power system closer to operating limits, accurately forecasting inertia in real-time becomes essential [8]. The current methods employed by the Nordic transmission operator for real-time inertia estimation are described and outlined in the report Future System Inertia by Ørum *et al.* [5]. This method uses information about what power plants are connected to the power grid, from now on referred to as circuit breaker position closed, together with the following equation,

$$E_{k,sys} = \sum_{i=1}^{K} S_{ni} H_i, \tag{1.1}$$

where $E_{k,sys}$ is the stored kinetic energy in the power system, K is the total number of generators, S_{ni} is the rated apparent power of generator i, and H_i is the inertial constant of generator i. When the circuit breaker is closed, the generator's stored kinetic energy is available to the power system. Information regarding $S_{n,i}$ and H_i is available to the individual TSOs for each of the four countries, and they calculate the inertial contribution of their respective regions. These estimations also use information on production units and rotating condensers [5].

The increased volatility from more IBRE is expected to reduce the accuracy of the current inertia estimation method used by the Nordic TSOs [8] because the increase in IBRE changes the amount of controllable production capacity. Historically, the production capacity of the Nordic power system was located in large

hydro and nuclear power plants; these types of power plants give a high degree of control over production capacity. With the increase of IBRE, more production capacity is located in smaller, less controllable power generation in the form of wind turbines and photovoltaic cells. The accessibility of energy resources extracted from wind turbines and photovoltaic cells has significant variability and needs to be utilized when available. It is anticipated that this will decrease the precision of the projected production capacity of other production types, leading to less accurate circuit breaker information for synchronous generation.

Previous works have proposed a top-down approach for inertia estimation [10, 5] using production data per production type, i.e., wind, hydro, nuclear, etc., and linear regression machine learning models to estimate power-system inertia for the Nordic power system. A notable limitation of this approach is the lack of accurate ground truth for the kinetic energy stored within the power system, and this method used estimations from the TSO as ground truth [10]. Also, this estimation method relies on the linear relation between kinetic energy stored in the power system and exogenous variables in the model. A method to estimate regional inertia constant outlined in the report Future system Inertia [5] is the estimation of regional inertia constants using disturbances. This method can accurately find the regional inertia constant within a short period using the swing equation (2.7) together with the frequency response right after the disturbance. However, this method cannot accurately estimate regional inertia constants for the power system for moments outside of the disturbance. Previous work in creating time series data for simulation by Rydin Gorjão *et al.* [11] has shown that the characteristics of power-grid frequency can be recreated using stochastic modeling. The approach is a data-driven stochastic model, where the model's parameters are estimated using data-driven analysis tools and an approach based on the Fokker-Planck equation.

To improve inertia estimations in future power systems, it could be beneficial to examine nonlinear relationships between the production and consumption of electrical power in the power grid and the kinetic energy stored in the synchronous generators within the power grid. Memory cell-based Long Short-Term Memory (LSTM) [12] machine learning algorithms have been used to analyze other dynamical systems, like spot price, through multivariate time series analysis [13]. These models can find complex nonlinear relationships in data [13], which can be hard or impossible to find with conventional models.

1.2 Motivation

The increasing need for accurate estimations and predictions of kinetic energy stored within the power system demands new approaches for analyzing available data to have more reliable results. The report Future System Inertia [5] by the European Network of Transmission System Operators (ENTSO-E) describes a disturbance-based method for estimating the regional inertia constant. However, this method suffers from the fact that large-scale disturbances needed for the estimations are rare within stable power systems, such as the Nordic power system. A more reliable method for estimating regional inertia constants could rely on information measured during a steady-state operating period within the power system, which is more common for stable power systems.

The current state-of-the-art models for kinetic energy estimation, also outlined in ENTSO-E report [5], only have the ability to capture linear relationships within the data. A model capable of capturing nonlinear relations could utilize more of the currently available information in the data to estimate the stored kinetic energy in a power system.

1.3 Research objectives

Objective 1 *Propose a method of estimating a regional inertial constant for the Nordic power system using steady-state power-grid frequency measurements together with a stochastic model of frequency behavior in the power grid.*

Objective 2 *Examine the ability of a simple LSTM neural network to perform multivariable analysis of production, load, and power-grid frequency data to predict the stored kinetic energy within the Nordic power system.*

2 Theory

The theory section is structured around three topics: power systems, stochastic processes, and machine learning. The segment on power systems begins with an overview of the fundamentals of the Nordic power system. Then, it explains how the swing equation can be applied to model frequency. Finally, it explains the concept of inertia within a power system. The segment on stochastic processes covers the basics of the Fokker-Planck equation and introduces a special case of the Fokker-Planck equation, called the Ornstein-Uhlenbeck equation. Then, it covers the Nadaraya-Watson estimator and how the estimator can be used to estimate the drift and diffusion coefficients for processes described by a Fokker-Planck equation. The segment on machine learning covers the fundamental goal of machine learning and methods to introduce non-linearity to the learning process. It continues explaining how recurrent neural networks can be used for sequential analysis and explains a modified recurrent neural network cell called a long-short-term memory cell. Finally, the pipeline from unprocessed input data to the trained model is outlined.

2.1 The Nordic power system

The power system in Norway is a part of the Nordic power system, which includes Norway, Sweden, Finland, and the eastern part of Denmark, corresponding to the DK2 price region. Figure 1 shows a map of the price regions that make up the Nordic synchronous region. Within the shared Nordic synchronous region, the Nordic TSOs chose the frequency of the AC power out of a synchronous generator connected to the power system to be 50 Hz [1]. The power grid frequency is dictated by synchronous generators connected to the power grid. Therefore, the frequency of the Nordic power system becomes 50 Hz. The shared grid frequency facilitates easy exchange of power within the region because the synchronous power producers/consumers can exchange power directly by changing the relative rotor angle between synchronous machines. Figure 2 depicts the conceptual reasoning behind a shared power grid frequency. In the figure, the power grid is represented as a collection of synchronous machines, the red circles, with a common power angle, δ, often chosen to be zero. When the synchronous machines operate with a common frequency, any change to the power angle relative to the common power angle induces an exchange of power. In figure 2, this is represented with a change of the power angle, $(\delta+\theta)$, of the leftmost synchronous machine, which leads to the power flow P.

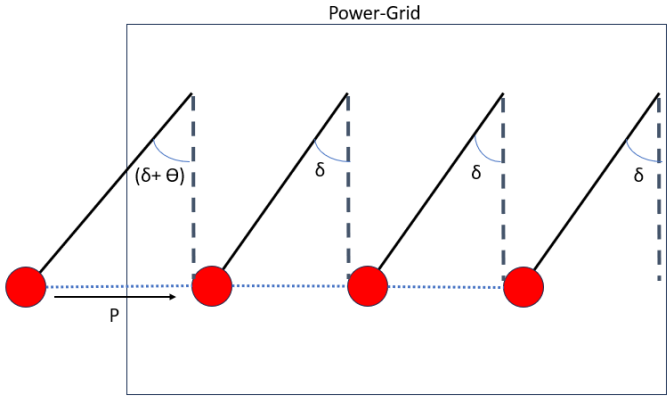

Figure 2: Physical interpretation of the power grid as a collection of pendulums connected by springs. Because the leftmost pendulum oscillates with a different initial angle, a force from the pendulum goes into the system. This force from the pendulum into the system is analogous to the power flow from a synchronous machine into the power grid with a rotor angle greater than the average rotor angle of the power grid.

2.2 The swing equation

A natural place to start when creating a model of frequency behavior is the swing equation. The swing equation is derived from Newton's second law of rotation and relates the acceleration of a synchronous generator or turbine to the imbalance of mechanical- and electrical torque. The following derivation follows closely from the book [14]. For a single synchronous generator, Newton's second law gives

$$J\frac{\partial \omega_m}{\partial t} = \tau_m - \tau_e, \tag{2.1}$$

where J is the moment of inertia for the turbine, shaft and generator combined, given in [kgm^2], ω_m is the mechanical angular frequency of the generator measured in [rad/s], τ_m is the mechanical torque given in [Nm], and τ_e is the electrical torque given in [Nm]. For applications in power system analysis, it is useful to express the swing equation (2.1) in terms of power and not torque. At the same time, it is beneficial to convert the equation into a per-unit system. Expressing the equation in terms of per-unit power is done by multiplying with the mechanical angular frequency, ω_m, and divided by rated apparent power, S_{rated}

$$\frac{J\omega_m}{S_{rated}}\frac{\partial \omega_m}{\partial t} = P_{m,pu} - P_{e,pu}, \tag{2.2}$$

where $S_{rated} = P_{max}/pf$ is the rated apparent power of the generator with pf denoting the power factor of the generator, $P_m = \omega_m \tau_m$ is the mechanical power from turbine into generator and $P_e = \omega_e \tau_e$ is the electrical power from generator to grid. Both P_m and P_e are given in [W].

Generators can vary considerably in rated apparent power and moment of inertia, therefore a normalized inertia constant is used. This inertia constant is defined as

$$H = \frac{1}{2}\frac{J\omega_m^2}{S_{rated}}, \tag{2.3}$$

Equation (2.3) is a function of the mechanical angular frequency, which entails that the H constant does change with the angular frequency of the generator. However, a synchronous generator needs to generate electrical power with a frequency matching the frequency of the power grid. So, for all practical purposes, one uses a fixed ω_m such that the frequency of the generated power matches the grid frequency. Since the H-constant is normalized by rated apparent power, it typically falls within a narrow range of 1-10s [14]. This gives an advantage over using the moment of inertia because one can easily compare different-sized generators.

Using equation (2.3) and equation (2.2) the swing equation can be written as,

$$2H\frac{\omega_m}{\omega_{m,syn}^2}\frac{\partial \omega_m}{\partial t} = P_{m,pu} - P_{e,pu}, \tag{2.4}$$

where $\omega_{m,syn}$ is the synchronous mechanical angular frequency, $P_{m,pu}$ is the mechanical power into the generator, and $P_{e,pu}$ is the electrical power out of the generator. Equation (2.4) is given with mechanical angular frequency. When using the swing equation to model grid frequency response, the electrical angular frequency is of more interest than the mechanical angular frequency. Converting from mechanical to electrical angular frequency follows the following relation:

$$\omega = \frac{\mu}{2}\omega_m, \tag{2.5}$$

where μ is the number of poles in the generator. To further simplify the swing equation, the per-unit angular frequency, ω_{pu}, is defined in the following manner.

$$\omega_{pu} = \frac{\omega_m}{\omega_{m,syn}}. \tag{2.6}$$

Using equations (2.6) and (2.5), equation (2.4) can be rewritten using the electrical angular frequency instead of the mechanical angular frequency,

$$\frac{2H}{\omega_{syn}}\omega_{pu}(t)\frac{\partial\omega(t)}{\partial t} = P_{m,pu}(t) - P_{e,pu}(t), \tag{2.7}$$

where ω_{pu} is the per-unit angular frequency, ω is the electrical angular frequency. The equation above represents an idealized frequency response of a generator. To better represent a real-world generator, it is often added a term to represent the dampening torque applied on the generator when it deviates from synchronous speed, and a second term can also be added to represent frequency deviations resulting from noise. A more realistic swing equation, taking both noise and dampening into account, can be described as

$$\frac{2H}{\omega_{syn}}\omega_{pu}(t)\frac{\partial\omega(t)}{\partial t} = P_{m,pu}(t) - P_{e,pu}(t) - D\frac{\omega(t)-\omega_{syn}}{\omega_{syn}} + \beta\epsilon(t). \tag{2.8}$$

The equation now contains a dampening factor, D, a noise factor, ϵ, and the strength of the noise β. The dampening term represents the natural dampening of oscillations in the power grid, and the noise term represents the natural fluctuations in power grid frequency. For smaller-scale power generation, voltage dynamics are often included in the swing equation to give a more realistic representation of frequency dynamics. At the transmission grid level, voltage is approximately constant, given that the TSO regulates voltages in an attempt to make it constant. Since the frequency dynamics being studied are at the transmission grid level, the voltage dynamics are not considered; therefore, the swing equation as described in equation (2.8) is the most complex representation of frequency dynamics addressed in the thesis.

2.3 Inertia and kinetic energy in power grids

The inertial constant of a power grid can be expressed as a sum of the inertial contributions of all the synchronous machines in the system. Considering N synchronous machines, the total inertial constant, H_{sys}, is given as

$$H_{sys} = \frac{\sum_{i=1}^{N} S_{i,n}H_i}{\sum_{i=1}^{N} S_{i,n}}, \tag{2.9}$$

where $|S_{i,n}|$ and H_i are the rated apparent power and inertial constant of generator number i. Synchronous machines refer to synchronously connected turbine generators and motors and should be included similarly. The TSOs in the Nordic power system do not measure a total system inertial constant but the kinetic energy stored in the synchronously rotating generators connected to the power system. The kinetic energy stored in the system is closely connected to the inertia constant of the power system. Rewriting equation (2.9), the system kinetic energy $E_{k,sys}$, measured in [Ws], can be expressed as

$$E_{k,sys} = S_{sys}H_{sys} = \sum_{i=1}^{N} S_{i,n}H_i, \tag{2.10}$$

where S_{sys} is the total rated apparent power and H_{sys} is the total inertia constant for the power system. This equation is only valid for synchronous machines that deliver power to the grid. A modified version of equation (2.10) takes the state of the circuit-breakers into account,

$$E_{k,sys} = S_{sys}H_{sys} = \sum_{i=1}^{N} S_{i,n}H_iK_i, \tag{2.11}$$

where K_i is the state of the circuit breaker. The circuit breaker state can be expressed as

$$K_i = \begin{cases} 1, & \text{grid connected,} \\ 0, & \text{not grid connected.} \end{cases} \tag{2.12}$$

When $K_i = 1$, the synchronous machine i is connected to the power grid and can supply kinetic energy, and when $K_i = 0$, the synchronous machine i is not connected to the power grid, and the kinetic energy is no longer accessible to the power grid. Note that from a power system perspective, a synchronous generator might not be actively providing/withdrawing power but still provides inertia to the system. In this thesis, the terms inertia and kinetic energy are used interchangeably to refer to the stored kinetic energy within the power system. When referring to the inertial constant, denoted H, the term inertia constant is used.

2.4 Fokker-Planck equation

The evolution in time for many dynamic systems in power engineering, such as the swing equation, is described deterministically by differential equations. Then, a stochastic term is added to represent random fluctuations of a value, as in equation (2.8). The Fokker-Planck (FP) equation is a partial differential equation that describes the evolution of a probability distribution function over time and space and has been utilized to model other dynamical systems containing a stochastical component, notably within the fields of physics and finance. The general FP equation for one variable x is given as [15].

$$\frac{\partial W(x,t)}{\partial t} = \left[-\frac{\partial}{\partial x} D^{(1)}(x,t) + \frac{\partial^2}{\partial x^2} D^{(2)}(x,t) \right] W(x,t), \tag{2.13}$$

where W is the probability distribution function, $D^{(1)}(x)$ is the drift coefficient and $D^{(2)}(x)$ is the diffusion coefficient. Solving the FP equation yields a distribution function as a function of space x and time t. The drift $D^{(1)}(x)$ and diffusion $D^{(2)}(x)$ are the first and second-order Kramers-Moyal coefficients. The j-th order Kramers-Moyal coefficients can be estimated using the forward derivative definition of Kramers-Moyal coefficients, which is defined as [16]

$$D^{(j)} = \frac{1}{j!} \lim_{\tau \to 0} \frac{1}{\tau} K^{(j)}(x,t,\tau) = \frac{1}{j!} \lim_{\tau \to 0} \frac{1}{\tau} \langle [x(t+\tau) - x(t)]^j \rangle \Big|_{x(t)=x}, \tag{2.14}$$

where $K^{(j)}(x,t,\tau)$ is the conditional moment and τ represents a small increment in time. Using the general Kramers-Moyal equation (2.14), the drift, $D^{(1)}(x)$, and the diffusion, $D^{(2)}(x)$, are defined as

$$D^{(1)}(x,t) = \lim_{\tau \to 0} \frac{1}{\tau} K^{(1)}(x,t,\tau), \tag{2.15a}$$

$$D^{(2)}(x,t) = \frac{1}{2} \lim_{\tau \to 0} \frac{1}{\tau} K^{(2)}(x,t,\tau). \tag{2.15b}$$

There are several special cases of the general FP equation where drift and diffusion have special properties. One of these is the Ornstein-Uhlenbeck process. In an Ornstein-Uhlenbeck process the drift coefficient, $D^{(1)}(x) = -\theta x$, is a linear function of x, and the diffusion coefficient, $D^{(2)}(x) = D$, is a constant. The FP equation for an Ornstein-Uhlenbeck process can then be expressed as

$$\frac{\partial W(x,t)}{\partial t} = \left[\theta \frac{\partial}{\partial x} x + D \frac{\partial^2}{\partial x^2} \right] W(x,t). \tag{2.16}$$

The Ornstein-Uhlenbeck process can also be described as a Langevin equation [15]. The Langevin representation of the Ornstein-Uhlenbeck process is

$$\frac{dx(t)}{dt} = -\theta x(t) + \alpha \Gamma(t), \tag{2.17}$$

where x is the variable that behaves according to the Ornstein-Uhlenbeck process, θ is a constant denoting the strength of the change related to $x(t)$, $\Gamma(t)$ is assumed to be a Gaussian random variable with zero mean and δ correlation, i.e., $\langle \Gamma(t+t')\Gamma(t) \rangle = \delta(t,t')$ [15], and α is a constant denoting the strength of $\Gamma(t)$. Using the definition of the Kramers-Moyal expansion coefficients [15] the drift coefficient and the diffusion coefficient can be related to θ and α in the following way

$$D^{(1)}(x,t) = -\theta\, x, \tag{2.18}$$

$$D^{(2)}(x,t) = \alpha^2. \tag{2.19}$$

Importantly, equation (2.17) only works on processes that have a mean of zero. For processes with a mean different from zero, subtracting the mean value from the process allows the equation to be applied.

2.5 Nadaraya-Watson Estimator

To estimate the drift and diffusion coefficients empirically from a non-stationary time series [16], the Nadaraya-Watson (NW) estimator can be utilized. For a non-stationary time series, the Kramers-Moyal (KM) coefficients have the potential to be time-dependent and KM coefficients for the entire time series can be hard to attain. The NW estimator allows the estimating of 'local' KM coefficients in each time step [16]. The method is based on estimating a conditional averaging, $\langle Y|X=x \rangle$, for the KM coefficients using a kernel. The conditional averaging, $\langle Y|X=x \rangle$, of the joint probability distribution functions, $p(x,y)$, can be defined as [16]

$$\langle Y|X=x \rangle = \frac{\int y p(x,y) dy}{\int p(x,y) dy} = \frac{\sum_{i=1}^{n} k_x(x-x_i) y_i}{\sum_{i=1}^{n} k_x(x-x_i)}, \tag{2.20}$$

where $k_x(x)$ is a kernel, $x \in X$ is value from set X at the current time step, $x_i \in X$ and $y_i \in Y$ represents the value of set X, and Y at time step i. The KM conditional moments can be expressed on the form $\langle Y|X=x \rangle$ where $Y \equiv (x(t+\tau)-x(t))^n \forall n \in \{1,2,...\}$. Using this expression for KM conditional moments together with equation (2.20), the j-th order KM conditional moments can be expressed as [16],

$$K^{(j)} = \frac{1}{\tau} \langle (x_{(i+1)\tau} - x_{i\tau})^j |_{x_i=x} \rangle = \frac{\sum_{i=1}^{n} k\left(\frac{x_{i\tau}-x}{h}\right) \left(x_{(i+1)\tau} - x_{i\tau}\right)^j}{\sum_{i=1}^{n} k\left(\frac{x_{i\tau}-x}{h}\right) \tau}, \tag{2.21}$$

where h is the bandwidth of the kernel $k(x)$. The kernel can be any smooth function satisfying,

$$\int x^2 k(x) dx < \infty, \quad k(x) \geq 0, \quad \int k(x) dx = 1, \quad \int x k(x) dx = 0. \tag{2.22}$$

The most common kernels include the Gaussian kernel $k(x) = \frac{1}{\sqrt{2}} exp(-x^2/2)$ and the Epanechnikov kernel $k(x) = \frac{3}{4}(1-x^2), x \in [-1,1]$. Using equation (2.21) and (2.14), the j-th order KM coefficient can be estimated using the following equation,

$$D^{(j)}(x,t,\tau) = \frac{1}{j!} \lim_{\tau \to 0} \frac{1}{\tau} K^{(j)}(x,t,\tau) = \frac{1}{j!} \lim_{\tau \to 0} \frac{1}{\tau} \left[\frac{\sum_{i=1}^{n} k\left(\frac{x_{i\tau}-x}{h}\right) \left(x_{(i+1)\tau} - x_{i\tau}\right)^j}{\sum_{i=1}^{n} k\left(\frac{x_{i\tau}-x}{h}\right)} \right]. \tag{2.23}$$

2.6 Machine learning fundamentals

Machine learning methods comprise three main parts: data, model, and loss function. Data is a collection of data points. A data point is a unit of information, as an example in a time series a timestep is a data

point and represents a unit of information. The data can be described with features and labels. Features are often denoted by x and represent measurable quantities such as temperature, pixel values, text, etc. Labels are often denoted with the letter y and represent more complex facts or quantities associated with a given data point, for example, categorization such as dog or cat, the sentiment of a text, quantities resulting from a combination of features, etc. Considering an ML model with a feature space $\mathcal{X}$ and a label space $\mathcal{Y}$ the goal of the model is to learn a mapping $h : \mathcal{X} \to \mathcal{Y}$ such that

$$h(x) = \hat{y} \approx y \; \forall \; x \in \mathcal{X}, \tag{2.24}$$

where h is a transformation from the feature space $\mathcal{X}$ to the label space $\mathcal{Y}$ [17]. $\hat{y}$ is the label estimated by the ML model for data point x, and y is the true label of data point x. The entire space of possible $h(x)$ maps is called a hypothesis space and is denoted by $\mathcal{H}$. For the machine learning model to learn the best $h(x)$ mapping for a given problem, a loss function is used to find the discrepancy between the estimated $\hat{y}$ and target y. The loss function $L(h, x, y)$ can be generalized as a map

$$L : \mathcal{X} \times \mathcal{Y} \times \mathcal{H} \to R : ((x, y), h) \to L((x, y), h). \tag{2.25}$$

with the h mapping, x features, and y labels as inputs and the numerical value $L((x, y), h)$ as output. The objective of a machine learning algorithm is to minimize the loss function, thereby having the smallest difference between the predicted values and the actual values and thereby finding the best h map from $\mathcal{X}$ to $\mathcal{Y}$. There are many different approaches to designing the model architecture, the one used in this thesis is a version of a recurrent neural network feeding into a fully connected feed-forward network also called a dense layer.

2.7 Activation functions

In machine learning, activation functions can be used to allow neural networks to learn nonlinear relationships in data by transforming the input data from linear space into nonlinear space. Two activation functions used in this thesis are the sigmoid and tanh functions. The sigmoid function is defined as

$$f(x) = \frac{1}{1 + e^{-x}}, \tag{2.26}$$

where x is the input variable and e is Euler number. The tanh function is expressed as

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \tag{2.27}$$

As shown in figure 3, the sigmoid and tanh functions output nonlinear values between $[0, 1]$ and $[-1, 1]$, respectively. In machine learning algorithms, the activation functions can go from linear input to nonlinear output. Allowing the models to learn nonlinear relations in data.

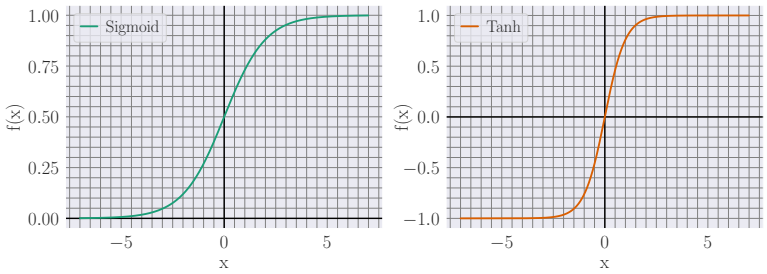

Figure 3: Two activation functions commonly used in machine learning: sigmoid and tanh. They both take an input vector,$x = [x_1, x_2, ..., x_n]$, and create a non-linear output, $f(x)$, for the sigmoid function the output is between $[0, 1]$, and for tanh it is between $[-1, 1]$. The activation functions allow a neural network to learn nonlinear relations in data.

2.8 Fully connected feed-forward network

A fully connected feed-forward network, referred to as a multilayered perceptron (MLP), is a neural network consisting of nodes and weights. How the nodes and weights are connected is shown in figure 4. The nodes are the colored circles, and the weights are the lines connecting them.

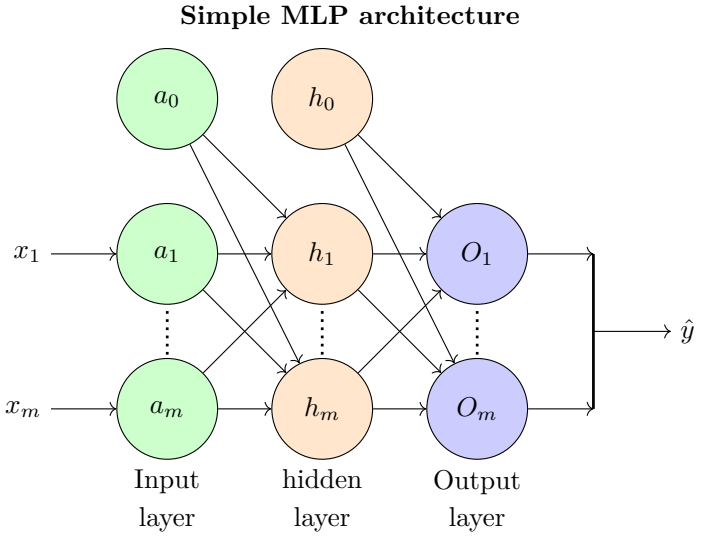

Figure 4: A simple graph representation of an MLP. The nodes are the colored circles, and the weights are the lines connecting them. Thea_0 and h_0 nodes represent the bias of the input and the hidden layer, respectively. The other node represents the activated information from the previous layer. x is the input features and $\hat{y}$ is the estimated label for the model. Figure inspired by [18]

Mechanisms must be introduced to facilitate information flow from input data to the estimation of labels. To go from input data to input layer the basis a_o is added to the feature vector, $X = [x_1, \ ... \ , x_m]$, in order to create the $A = [a_0, x_1, \ ... \ , x_m]$. The flow of information from the input layer to the hidden layer is can be computed using

$$h = \phi_h \left(AW_{ah} \right). \tag{2.28}$$

Where h is the activated hidden layer, A is the input layer, W_{ah} is a matrix containing all the weight between the input and the output layer, ϕ_h is the activation function of the hidden layer. In the same manner, the activated output layer, O, can be computed using

$$O = \phi_o \left(hW_{ho} \right), \tag{2.29}$$

where ϕ_o is the activation function for the output layer, and W_{ho} is the matrix containing all the weights from the hidden layer to the output layer. From the activated output layer the label estimation $\hat{y}$ can be calculated and used to evaluate the loss function against the true label. To make the MLP capable of learning, the weight matrices must be updated based on the loss function. This is achieved by using backpropagation [18]. In backpropagation, the gradient of the loss function is calculated with respect to the different weights in the MLP. The weight connecting the output and hidden layers can be calculated using

$$\frac{\partial L}{\partial W_{ho}} = h^T \delta_{out}, \tag{2.30}$$

where L is the loss function used to evaluate the predictions of the MLP and δ_{out} is the error outputted after evaluating the prediction using the loss function. For the weights connecting the hidden layer and

the input layer, the gradient is calculated using

$$\frac{\partial L}{\partial W_{ah}} = A^T \left(\delta_{out} (W_{ho})^T \odot \frac{\partial \phi_h z_h}{\partial z_h} \right), \tag{2.31}$$

where $z_h = AW_{ah}$ is the unactivated hidden layer and $\odot$ represents a elemt-wise multiplication.

2.9 Recurrent neural networks and Long Short-Term Memory models

When analyzing time series data, the order of the features matters as they are not independent of each other. Therefore, when applying machine learning models to time series, it is important that the model can retain the sequential information in the time series. Recurrent neural networks (RNNs) are a type of machine learning architecture useful for analyzing sequential data because they can extract and retain the sequential information found within the time series. A simplified visualization of the basic RNN architecture can be seen in figure 5.

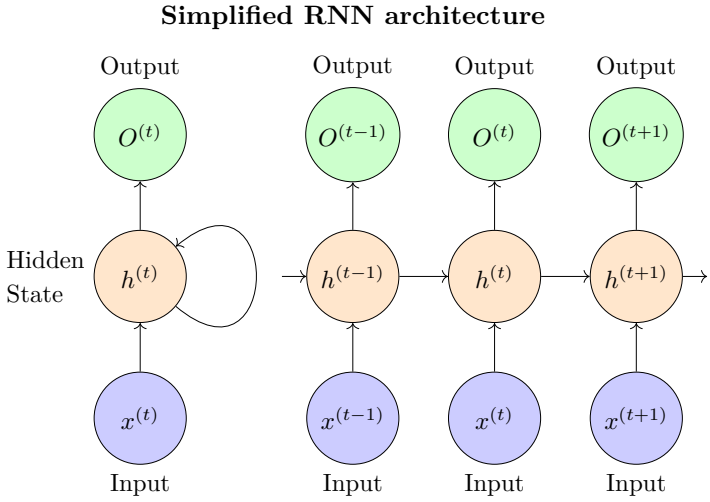

Figure 5: Representation of a simple RNN. The figure to the left is a compacted representation showing the flow of information from the input layer, $x(t)$, to the hidden layer, $h(t)$, to the output layer, $O(t)$. The loop from $h(t)$ to $h(t)$ has been expanded in the figure to the right and represent a movement of information in time. The flow of information from $h^{(t-1)}$ to $h^{(t)}$ is the part of the RNN that allows for the retention of sequential information. Figure inspired by [18].

Remembering the goal of machine learning is to learn a mapping $h : \mathcal{X} \to \mathcal{Y}$ from the input data to the output data such that equation (2.24) is satisfied. To do this, there needs to be a way for the information to propagate through the system. For neural networks, this is achieved using linear algebra and activation functions. Looking at figure 5 the flow of information from the input layer, $x^{(t)}$, to the hidden layer, $h^{(t)}$, can be described with

$$h^{(t)} = \Phi_h \left(w_{xh} x^{(t)} + w_{hh} h^{(t-1)} + b_h \right), \tag{2.32}$$

where Φ_h is the activation of the hidden layer, w_{xh} is the weight matrix connecting the input and the hidden layer, w_{hh} is the weight matrix connecting the precious hidden layer with the current hidden layer, and b_h is the bias vector for the hidden units. From this the output layer, $O^{(t)}$, can be computed using

$$O^{(t)} = \Phi_o \left(w_{ho} h^{(t)} + b_o \right), \tag{2.33}$$

where Φ_o is the activation function of the output layer. The loss function can now be evaluated and used to adjust the weight matrices. The learning process used to adjust the weight matrices for the hidden layers in an RNN is called backpropagation through time. In backpropagation through time, the gradient of the total loss for the model is used to update the weight matrices. The total loss is the sum of all the losses at all the timesteps and can be expressed as

$$L = \sum_{t=1}^{T} L^{(t)}. \tag{2.34}$$

L is the total loss for the RNN, T is the total number of time steps, and $L^{(t)}$ is the loss at time step t. To update the weight matrix, the gradient of the loss function with respect to the hidden weights is calculated. The gradient can be calculated using the following equation [18],

$$\frac{\partial L^{(t)}}{\partial w_{hh}} = \frac{\partial L^{(t)}}{\partial O^{(t)}} \frac{\partial O^{(t)}}{\partial h^{(t)}} \left(\sum_{k=1}^{t} \frac{\partial h^{(t)}}{\partial h^{(k)}} \frac{\partial h^{(k)}}{\partial w_{hh}} \right), \tag{2.35}$$

where $O^{(t)}$ is the activated output layer, $h^{(t)}$ is the activated hidden layer, $h^{(k)}$ represents the previous hidden layers, and $L^{(t)}$ it the loss at hidden layer number t. The gradient of the hidden layer being evaluated, $h^{(t)}$, with respect to the previous hidden layers, $h^{(k)}$, can be expressed as

$$\frac{\partial h^{(t)}}{\partial h^{(k)}} = \prod_{i=k+1}^{t} \frac{\partial h^{(i)}}{\partial h^{(i-1)}}. \tag{2.36}$$

This multiplication factor introduces one of the challenges of learning time dependencies in long time series using RNN, namely the vanishing and exploding gradient problems. It arises from the fact that the partial derivative of the current hidden layer $h^{(t)}$ with respect to the previous hidden layer $h^{(k)}$ in equation (2.35) can be expressed using equation (2.36). This shows that $\frac{\partial h^{(t)}}{\partial h^{(k)}}$ has $t-k$ number of multiplications. The weights between the hidden layers are denoted w_{hh}. If $w_{hh} < 1$, this factor becomes very small when $t-k$ is large, leading to the vanishing gradient issue. If the $w_{hh} > 1$ this factor becomes very large when $t-k$ is large, leading to the exploding gradient issue. Ideally, $w_{hh} \approx 1$. In a Long short-term memory (LSTM) cell, a $w_{hh} \approx 1$ is accomplished by creating a cell state with this property and controlling the flow of information into and out of the cell state with three gated cells. A representation of an LSTM cell is shown in figure 6. There are three types of gates in the LSTM layer: the forget-, input- and output- gates.

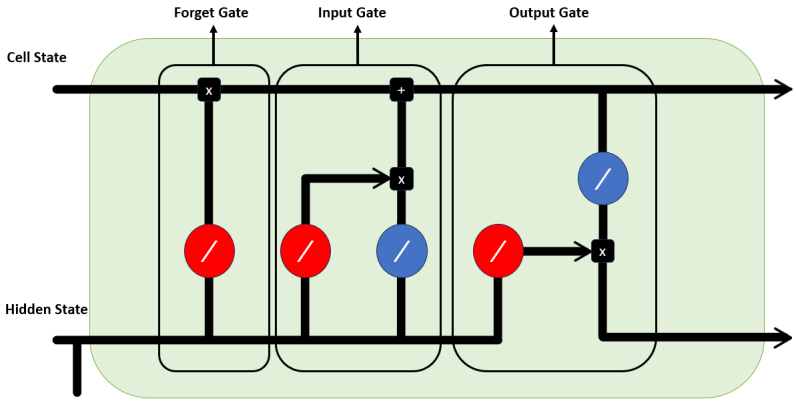

Figure 6: An LSTM layer is based on the gated units: forget-, input-, and output gates. The goal of the forget gate is to take the combined hidden and input states and determine what information in the cell state should be forgotten. The input gate determines what information should be added to the cell state. The output gate creates a new hidden state using data from the cell state and the current hidden state. The + represents piecewise addition, and × represents piecewise multiplication of matrices. The red and blue circles represent the application of the sigmoid and the tanh activation functions, respectively. Figure inspired by [13].

Forget gate

The purpose of the forget Gate is to determine what information can be removed from the cell state. To achieve this, the importance of the information from the previous hidden state is evaluated using the sigmoid activation function, where information with a value closer to zero is deemed less important, and a value closer to one is deemed more important. The resulting information matrix is then piecewise multiplied together with the cell state. Thus, less important information for long-term dependencies is removed from the cell state.

Input gate

The input gate determines what information should be added to the cell state. First, the importance of the input information is decided using a sigmoid function. Then, the input information is normalized using the tanh activation function. The importance of the information and the normalized information are added together with piecewise multiplication. After this, the information is added to the cell state with piecewise addition.

Output gate

The output gate creates a new hidden state for the next iteration. To create a new hidden state, the sigmoid function is applied to the previous hidden state to determine what information is important. Information from the cell state is normalized between [-1,1] using the tanh function and piecewise multiplied with the matrix resulting from applying the sigmoid function to the hidden state. This creates a new hidden state from the cell state while emphasizing the important features found in the hidden state.

2.10 Overfitting and underfitting of models

Some common problems in machine learning are overfitting or underfitting of models. Assuming that Y is the target label and $\hat{f}$ is the model, the expected square error for the point X, $Err(x)$, can be expressed as [19]

$$Err(x) = E\left[\left(Y - \hat{f}(x)\right)^2\right], \tag{2.37}$$

where E is the expected value operator. The error can be further broken down into three parts: bias, variance, and fixed error. The error can then be expressed as

$$Err(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right] + \sigma_e^2, \tag{2.38}$$

$$Err(x) = Bias^2 + Variance + Fixed\ \ error, \tag{2.39}$$

where $f(x)$ is the correct value at point x and σ_e^2 is a measure of the noise in the data. The bias contains information about the difference between the average prediction of the model and the target value it is trying to predict [19]. Variance tells something about the variability of predictions for a given data point. Overfitting and underfitting occur when a model fails to strike the right balance between bias and variance, resulting in increased error. For supervised learning, underfitting occurs when a model can not capture the underlying relations in the data. Typically, these models have high bias and low variance. Often, this can be a result of model complexity being too low for the given task. Overfitting occurs when a model pays too much attention to the training data, capturing the underlying noise in the data together with the relations. Typically, these models have low bias and high variance. Often, this can be a result of model complexity being too high for the given task. In order for the model to have the best performance possible, it is important to find the right balance between bias and variance. The model needs to have enough complexity to learn the patterns in the data without including the noise, giving the best performance on unseen data. Figure 7 shows the error of a machine learning model as a function of the model complexity.

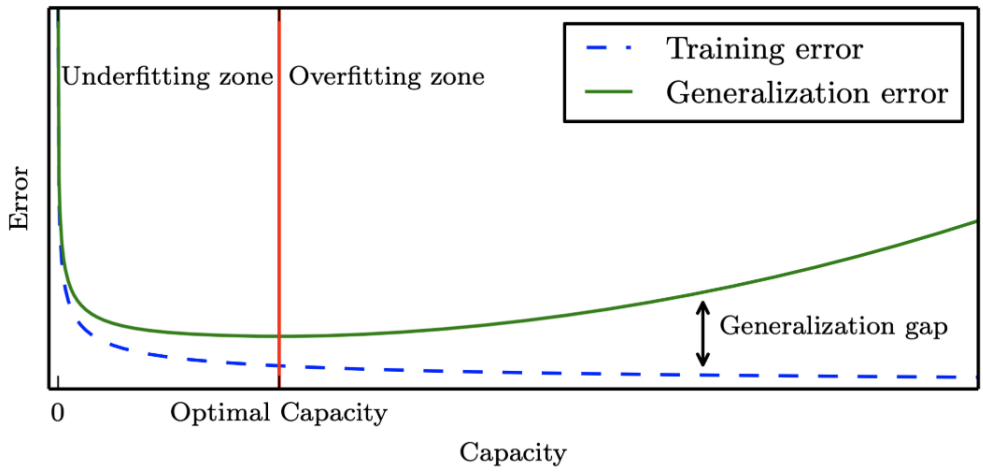

Figure 7: Trying to create the correct model for a given task is important to avoid over or underfitting the model. An attempt at visualizing when over/underfitting occurs is depicted by plotting error as a function of learning capacity, with learning capacity on the horizontal axis and error on the vertical axis. The training error is the blue dotted line, and the test error is the green whole line. The underfitting zone contains low complexity models that are incapable of learning the relations present in the dataset and, therefore, have high errors on both test and train data. The overfitting zone contains models with two high levels of complexity for the given task. These models can fit all data points in the training data perfectly but cannot generalize this information to unseen data. This leads to a large discrepancy between the training and generalization errors, shown in the figure as the generalization gap. The model with the optimal complexity minimized the overall generalization error. Taken with permission from [20]

Increasing model complexity is a good way to combat underfitting. There are several methods to combat or safeguard a model against overfitting; some of them are covered below.

2.10.1 Test-train split

A common way to check if a model is overfitting is the test-train split. The data containing features is split into two datasets: one for training and validating the machine learning model and one for testing the model's ability to generalize on unseen data. The model's performance on the test data gives a good indication of how well it is expected to perform on new unseen data.

2.10.2 Cross validation

Cross-validation is a method for ensuring that the validation data selected is representative of the training data. Cross-validation is done by separating the data into different sections and picking a random part of the section to be the validation data. If the validation data was just picked as a single part of the data, it could be the case that the validation data is not representative of the data set. This would cause the model to be wrongly adjusted during the training process, leading to overfitting toward data points in the validation data.

2.11 Supervised learning

Supervised learning can be applied to create the mapping described by equation (2.24). The process from raw data to prediction can be described using a four-step process.

Step 1: Preprocessing

Data needs to be preprocessed before being used as input features in a machine-learning model. Preprocessing can include extraction, scaling, selection of features, dimensional reduction, sampling, and elimination of nan values. This ensures that the model has no prior bias regarding feature importance, meaning all features start as equally important. After the preprocessing, the data is usually split into training and test data.

Step 2: Training

The next step is using the training data to train the model. During training, the model learns the underlying relations in the data by applying the model to the input data, evaluating the given prediction against the actual target using the loss function, and updating the model weights using backpropagation. Techniques including model selection, cross-validation, and hyperparameter tuning can be used during training to ensure that models are well suited to the given task.

Step 3: Evaluation

After the model is trained, its performance is evaluated on the test data created in Step 1 by applying a chosen performance metric. The goal is for the model to achieve the best possible result on the test data. If the performance on the test data is much worse than on the training data, the usual procedure is to go back to step 2 and develop the model further.

Step 4: Prediction

When steps 2 and 3 have been repeated many times, and the results in step 3 are adequate, the model can be applied to new data. The predictions can now be used in practical applications, such as forecasts.

2.12 Evaluation metrics

Four evaluation metrics were used to evaluate the performance of models in this thesis. These are root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and R2 score. RMSE and MAE can be mathematically described as [18]

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \ , \tag{2.40}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \ , \tag{2.41}$$

where y_i is the actual target, $\hat{y}_i$ is the predicted target, and n is the number of predictions. Both RMSE and MAE measure the residuals between the predicted and actual targets. They are scale-dependent errors, meaning that the error calculated from the metrics needs to be seen in relation to the scale of the targets. MAPE and R2 scores are, on the other hand, scale-independent errors and can be mathematically described as

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \times 100\%, \tag{2.42}$$

$$R2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}. \tag{2.43}$$

MAPE is a scale-independent version of MAE and is given as the average percentage error of residual, $|y_i - \hat{y}_i|$, with respect to the targets. The R2 score measures the proportion of the variance in the dependent variable, label, that is predictable from the independent variables, features. When R2$\approx$ 1, the regression model almost perfectly explains the variability in the dependent variable, and the model fits the data almost perfectly. If R2= 0, the regression model captures none of the variability around the mean and performs as well as using the mean of the dependent variable as the model prediction for all observations. R2 can also take negative values. When R2 < 0, this indicates a model that fits the data worse than a model that simply predicts the mean value for all of the observations, regardless of the input features [21].

3 Data and use of artificial intelligence

This section covers the use of artificial intelligence and the different datasets used in the thesis.

3.1 Use of generative artificial intelligence within the thesis work and writing

The generative model Chat.GPT 4.0 was used to assist in improving code for the data processing and displaying figures. Chat.GPT 4.0 and Grammarly were used to assist with spellchecking the text in the thesis.

3.2 Data Description

This section covers the data used in the thesis. Table 1 shows from what organization the data was obtained for different datasets, the units in which the data is recorded, and the time resolution of the data are also presented.

Table 1: Data and sources in this thesis, with their respective resolutions, units, and governmental agencies that provide them.

Data type	Organization	Unit	Time resolution	Years	Reference
Production	ENTSO-E	MW	1 Hour	2020-2023	[22]
Load	ENTSO-E	MW	1 Hour	2020-2023	[22]
Frequency	Fingrid	Hz	0.1s	2020-2023	[23]
Inertia	Energinet	GWs	1 Hour	2020-2023	[24]

3.2.1 Production and load data

The production and load data used to estimate the kinetic energy in the power grid was gathered from ENTSO-E, using the ENTSO-E API [22] in combination with the python library entsoe-py [25]. The dataset gathered is time series data with a duration of one hour. The measured values are the active power produced by different production types in the Nordic regions of Norway, Sweden, and Finland. Production from Denmark denoted DK2 for the equivalent electricity market bidding zone, was omitted as the quantity is not comparable to the other regions [24]. This data was combined into a single data frame showing the active power produced by different production types for the whole Nordic region. Data was gathered for 2020–2023 and contained some variations in the production types, but the production types with the highest power contributions in the Nordic regions remain available. The production types contributing around 90% of the Nordic region's total power are present in all the 2020–2023 datasets. Norwegian and Swedish hydropower, together with Finnish and Swedish nuclear power, compose around 70% of the power production in a given year. Figures showing the contributions of all the generation types for the years 2020–2023 can be found in Appendix B. The production types seen as synchronous and used to estimate kinetic energy with the linear models are thermal, nuclear, hydro, and wind power. The hydropower production type consists of five columns from the production data and is summarized in table 2

Table 2: Columns in the production data corresponding to hydroelectric production in the Nordic power system, what country they come from, and their unit of measure (SE for Sweden, NO for Norway, FI for Finland).

Country	Name of column	Unit
Norway	Hydro Pumped Storage, Hydro Run-of-river and poundage	MW
Sweden	Hydro Water Reservoir_SE	MW
Finland	Hydro Run-of-river and poundage_FI	MW

The nuclear power generation data consists of two columns, with country, name, and unit of measure in table 3.

Table 3: The columns in the production data corresponding to nuclear production in the Nordic power system, what country they come from, and their unit of measure.

Country	Name of column	Unit
Sweden	Nuclear	MW
Finland	Nuclear_FI	MW

The data concerning wind power generation consists of three columns in the generation data, shown in table 4.

Table 4: The columns in the production data corresponding to wind production in the Nordic power system, what country they come from, and their unit of measure.

Country	Name of column	Unit
Norway	Wind Onshore_NO	MW
Sweden	Wind Onshore_SE	MW
Finland	Wind Onshore	MW

Thermal generation in the Nordic power system is represented using 13 columns, shown in table 5 with country, unit of measure, and column name.

Table 5: The columns in the production data corresponding to thermal production in the Nordic power system, what country they come from, and their unit of measure.

Country	Name of column	Unit
Norway	Fossil Gas, Other_NO	MW
Sweden	Other_SE	MW
Finland	Gas_FI, Fossil Hard coal, Fossil Oil, Fossil Peat, Waste_FI, Biomass, Other renewable, Other	MW

The load data for all years gathered contain three features: load in Norway, Sweden, and Finland. Load refers to the total active power, in MW, consumed within the region, and the load data is a time series with a time resolution of one hour. A modified version of the production and load data is utilized for the application in the LSTM model. This dataset contains the columns present in all the production and load data from 2020 to 2023; the first five rows of this dataset can be found in the apendix B

3.2.2 Frequency data

This dataset contains the local frequency measured in Finland by Fingrid, the Finish TSO, for 2020–2023. The data was gathered from Fingrid's open data platform [23]. The dataset is a time series with

a temporal resolution of 0.1 seconds. It has only one feature, the power-grid voltage frequency, given in Hz. It is also assumed that the single frequency value from Fingrid represents the frequency for the entire Nordic power system.

3.2.3 Inertial data

To validate how well the different methods described in this section work, an existing data set containing information about power system inertia, measured as kinetic energy in GW·s, was used as a best guess. Data was available for the years 2020–2023. The data was gathered from Energy Data Service API [24]. The data set is a time series with a temporal resolution of one hour. It contains six features: the total kinetic energy in the Nordic power system, kinetic energy in Norway, kinetic energy in Sweden, kinetic energy in Finland, kinetic energy in the Danish price region DK2, and time of the recordings. All the values for kinetic energy are given in GW·s.

4 Methods of kinetic energy estimation

The following section proposes four models for estimating the kinetic energy stored in the Nordic power system. The first three models are based on the same linear regression models for estimating kinetic energy, with different methods of finding the Nordic power system's regional inertial constant. The last model is a non-linear machine-learning model used to estimate the kinetic energy stored in the power system directly from data containing production, load, drift, diffusion, and oscillation point data.

4.1 Inertia estimation using generic H constants

The goal of the model using generic H constants for estimation is to see what type of behavior should be expected when modeling kinetic energy in the power system from aggregate production data together with inertia constants. The results of this model are then further used to modify the diffusion model to account for some of the expected errors resulting from the choice of model converting inertia constants into kinetic energy estimations. The inertia constants used in the generic H model are gathered from the ENTSO-E report [5] and can be found in the table (6).

Table 6: Inertia constants used for estimating kinetic energy in the ENTSO-E report [5]. The Inertia constants represent the average inertia constant of a generator within a production type.

Production Type	H[s]
Nuclear	6.3
Other thermal	4
Hydro conventional	3
Hydro small scale	1
Wind	0

The values in table 6 represent average values for inertia constants for generators by their production type [5]. The method by which these average values are calculated is not explicitly stated in the report. Still, it is assumed for this thesis that the values of the H constant in table 6 are the average value of the inertia constants for all generators within the respective production types that the TSOs have information about. A modified version of equation (2.11) is used to estimate the kinetic energy in the power system from aggregate production and average H constants. The equation used for estimation is given as

$$E_{k,sys} = \sum_{m=1}^{N} S_m H_m = \sum_{i=1}^{N} \frac{P_m}{pf} H_m, \tag{4.1}$$

where P_m is the total active power generated and S_m is the total apparent power generated by production type m. H_m is the average inertial constant and pf is the power factor. The production that is used as synchronous for this estimation can be found in table 2, 3, 4. Since aggregated values for production were used, a single value for hydro was deemed sufficient, and only the conventional hydro constant was used for estimation.

The modified equation was chosen because information about the circuit breaker state, K_i, and rated apparent power, S_i, used in equation (2.11) are not publicly available and proved challenging to acquire. The inertial estimation used a power factor of 0.9 for all production types, following the report by ENTSO-E [5]. The use of aggregated active power produced is expected to lead to underestimations of the kinetic energy in the power system. It allows the inertial contribution per production type, production type m,

to take any value between $E_{k,m} \propto [0, P_{0,max}]$ in a continuous manner, not taking into account the binary nature of inertia. Meanwhile, the TSO estimation only allows discrete increases or decreases in the inertia contribution from production type m, which is related to the circuit breaker state of the generators in production type m.

4.2 Estimation of inertia from Kramers-Moyal coefficients

This section explains the estimation of a regional inertial constant using the drift and diffusion coefficients estimated from Fingrid's frequency data. The goal is to estimate the regional inertial constant of the Nordic power system with a time resolution of one hour. From table 1, the time resolution of the frequency data is 0.1 s. For an hourly estimation of the inertial constant, a single drift and diffusion coefficient must accurately describe the frequency behavior within an hour. It is therefore assumed that within an hour of steady-state operation of the power system, where $\Delta P \approx 0$, the frequency behavior can be described using an Ornstein-Uhlenbeck process (2.17). If this is the case, the Langevin expression of the Ornstein-Uhlenbeck process (2.17) is used to evaluate the frequency change. Removing the mean value from the frequency data, the equation governing the frequency behavior.

$$\frac{d\omega(t)}{dt} = \theta(\omega_{syn} - \omega(t)) + \alpha\Gamma(t). \tag{4.2}$$

Evaluating the equation in relation to the realistic swing equation (2.8), with $\omega_{pu} = 1$, yields the equality,

$$\theta(\omega_{syn} - \omega(t)) + \alpha\Gamma(t) = -\frac{D}{2H}(\omega(t) - \omega_{syn}) + \omega_{syn}\frac{\beta\epsilon(t)}{2H}. \tag{4.3}$$

It stands to reason that if the Ornstein-Uhlenbeck process and swing equation can be used as the governing equation for frequency dynamics, they should describe the same behavior. This would indicate that the two dampening terms describe the same dampening behavior, and the two noise terms describe noise with similar properties. Together with the estimated drift and diffusion coefficients, this is used in the following section to propose estimation processes for the regional inertial constant. The estimations of local conditional moments described by equation (2.23) are done using the function `kmc()` from the python package kramersmoyal [26]. The `kmc()` function takes one hour of frequency data, a chosen bandwidth, a chosen number of bins, and how many powers of local conditional moments should be in the outputs. For the estimations of local conditional moments, the bandwidth was set to the resolution of the time series 0.1 s, the number of bins was chosen to be 5000, and the function was set to output the first and second local conditional moments for every point in the times. To go from local conditional moments to drift and diffusion coefficients, the relations

$$D^{(1)} = \frac{K^{(1)}}{\tau}, \tag{4.4}$$

$$D^{(2)} = \frac{K^{(2)}}{2\tau}, \tag{4.5}$$

are used on the first and second-order local conditional moments calculated by the `kmc()` function. With τ begin equal to the time resolution of the time series, 0.1 s.

4.2.1 Estimation using drift

The hourly estimations of an inertial constant are based on the assumption that the Ornstein-Uhlenbeck process (2.17) governs frequency dynamics for an hour of steady-state operation. As a result of this, only hours were the drift coefficient, $D^{(1)} = -\theta\omega(t)$, is a linear function of angular frequency can be used for the estimation. To evaluate this for an hour, the `kmc()` function was used to estimate the first-order

local conditional moment, described in equation (2.21), for every timestep in the frequency data. Then, a first-order polynomial fit was done on the first-order local conditional moments using the `np.polyfit()` function from the python package `numpy` [27], this resulted in a single global value representing the first-order conditional moment for the hour being analyzed. The single value for the first-order conditional moment was evaluated against the first-order local conditional moments using the $R2$ score. The $R2$ score was created by evaluating a line where the slope is equal to the global value against the time series of first-order local conditional moments. For all hours having an R2 score > 0.8, the first-order global conditional moment was deemed sufficient to describe the behavior of the first-order local conditional moments for the whole hour. Using the relations from equation (2.15a), (2.18) together with equality (4.3) the inertia constant for the Nordic power system can be related to the drift coefficient using the following relation

$$\theta = \frac{D}{2H}. \tag{4.6}$$

Solving equation 4.6 with respect to the inertial constant, H, gives an expression for the inertial constant for one hour of steady-state operation for the Nordic power system

$$H = \frac{D}{2\theta}. \tag{4.7}$$

The TSOs do not have publicly available information regarding the regional inertial constant. However, they have information regarding the stored kinetic energy in the power system. Considering this, a simple linear model was used to create estimations for stored kinetic energy in the Nordic power system from the regional inertial constants. These estimations were then evaluated against the TSO's estimations of stored kinetic energy in the Nordic power system. The equation used to estimate the kinetic energy is

$$E_{k,drift} = \frac{P}{pf}\frac{D}{2\theta}, \tag{4.8}$$

where P is the total active power produced by synchronous generators. It is important to note that $P = \sum P_m$ where P_m is a specific production type. pf is a universal power factor set to 0.9 following the report [5]. The constant D was found by finding the best fit of equation (4.8) for the evaluated year. To find the best fit, the function `curve_fit` from the `scipy.optimize` [28] package was used with the initial guesses of $D_0 = 0.05$. The initial value for D_0 was found in the thesis [29]. Equation (4.8) is a modification of the one used by the TSO (1.1) because the ENTSO-E datasets contain information about active power produced per production unit, not the rated apparent power of every generator. One of the drawbacks of this method is that the inertial contributions per production type can take any value to take any value between $E_{k,m} \propto [0, P_{0,max}]$. When this is seen together with the TSO estimation, that only allows a single value for the possible contribution from generator i, $E_{k,i} = S_{n,i}H_i$, equation (4.8) is expected to underestimate the kinetic energy stored in the power system. The production used as synchronous is the same for this estimation as for the generic H estimation.

4.2.2 Estimation using diffusion

The hourly estimations of an inertial constant are based on the assumption that the Ornstein-Uhlenbeck process (2.17) governs frequency dynamics for an hour of steady-state operation. As a result of this, only hours with a diffusion coefficient, $D^{(2)} = \alpha$, is approximately constant can be used in this estimation. To evaluate this for an hour, the `kmc()` function was used to find the second-order local conditional moment, described in equation (2.21), for every timestep in the frequency data. Then, a zero-degree polynomial fit was done on the second-order local conditional moments using the `np.polyfit()` function from the python package numpy [27], this resulted in a single global value representing the conditional moment for the hour being analyzed. This global second-order conditional moment was evaluated against the

second-order local conditional moments using the MSE score. For the hours assessed using the R2 score mentioned in the previous section, the MSE from evaluating the constant line defined by α against the time series of the second-order local conditional moment never exceeded $3.969 \times 10^{(-12)}$. This value was deemed sufficiently small for the assumption of constant diffusion to be valid for the hours in the time series.

Assuming that $\Gamma(t)$ and $\epsilon(t)$ in equation (4.3) have the same properties, it is reasonable that α can related to the inertia in a similar fashion to the noise term in equation (2.8). α can then be related to the diffusion coefficient and the inertial constant using the relations from equation (2.15b),(2.19) together with equality (4.3). This results in the following relation

$$\alpha = \sqrt{D^{(2)}} = \frac{C}{H}, \tag{4.9}$$

where C is a constant. Solving equation 4.9 with respect to the inertial constant, H gives an expression for the inertial constant for one hour of steady-state operation for the Nordic power system as

$$H = \frac{C}{\alpha}. \tag{4.10}$$

Following the same reasoning as above for the drift estimation, a simple linear model was used to evaluate the diffusion-based estimated regional inertial constant by using them to estimate the stored kinetic energy in the Nordic power system. These estimations were then evaluated against the TSO's estimations of stored kinetic energy in the Nordic power system. The equation used to estimate the kinetic energy is

$$E_{k,diffution} = \frac{P}{pf}\frac{C}{\alpha} + \Upsilon, \tag{4.11}$$

where P is the total active power produced by synchronous generators, pf is a universal power-factor, set to 0.9 following the report [5]. C is a constant, and Υ represents an offset. They were estimated by finding the best fit via least squares for the evaluated year. The offset parameter was added to address modeling errors found for the generic H model. A proposed cause for the error being corrected can be found in section 4.1. To find the best fit, the function `curve_fit` from the `scipy.optimize` [28] package was used with the initial guesses of $C_0 = 1.4/3 \times 10e - 3$, $\Upsilon_0 = 100$. The initial values for Υ_0 and C_0 were found by a manual fitting of equation (4.11) to the Energinet inertia target for the year 2020. Equation 4.11 is a modification of the one used by the TSO (1.1). This model suffers from the same issues as the drift model discussed above. The same production data is used as synchronous production for the diffusion model as for the previous models.

4.3 Estimation using LSTM

The inertia of the power system was also estimated using a machine-learning regression model based on an LSTM layer. The model was created using the package `tensorflow.keras` [30] in Python. It is a sequential model with two layers: one LSTM and one Dense layer. Within the LSTM layer, there is also defined an input layer, in table 7 this layer is extracted to show the input dimensions of the training data. As shown in table 7, the model is rather simple, with 40 LSTM nodes that learn the relations between the features in the data and the target and one fully conceded layer, the Dense layer that outputs one value as a final estimation.

Table 7: A summary of the machine learning model used to estimate inertia. All layers in the model are shown with their layer type, number of nodes, activation function, output shape, and the number of trainable parameters. The model described is a regression model with total trainable parameters equal to 9801, and it outputs a single floating point value as its prediction.

Layer	Number of nodes	Activation function	Output shape	Trainable parameters
Input	20	linear	(none,24,23)	0
LSTM	40	tanh	(none, 40)	9760
Dense	1	linear	(none,1)	41

The training process for the models follows the outline in section 2.11 and is described in the sections below.

Step 1: Preprocessing

A combination of the three datasets mentioned earlier was used to create the dataset that was used to train the LST. As previously mentioned, the ENTSO-E data does not have consistent features for the years 2020-2023. To be used in the LSTM, the data feature space must be consistent for all the years. Because of this, 20 features were retained for the production data from 2020-2023. The features were filtered based on their contribution to the overall power production and the number of missing values. Figures showing the relative production from production types for all the years can be found in appendix C. These 20 chosen features were combined with the diffusion, $D^{(2)}$, R2 score, and oscillation point extracted from the frequency data. The table below shows the first five rows of the data frame. As the target for supervised learning, the best guess inertial data from the Danish TSO Energinet was used. The features were normalized using the `MinMaxScaler()` from the sklearn.preprocessing package [31]. The normalized input features have a shape of $(26163, 23)$, and for applications in an LSTM model, the data needs to have three dimensions. To achieve this, every step of the data was given knowledge of the previous 24 hours of feature information, giving the data the dimensions of $(26163, 24, 23)$. The reason behind this is to give the LSTM knowledge about the 24-hour cycle that exists in power consumption/production. The data was then split into training and testing data, which were used during the training and evaluation part of supervised learning.

Step 2: Traning

The model was trained and cross-validated at the same time. The cross-validation was done using the `KFold()` function from the `sklearn.model_selection package` [31]. This function creates five random splits of the testing data, called folds. The folds are then split into training and validation parts at an 80-20 ratio. The model's training is then done over the five folds, keeping the best-performing models at the end of each fold. This is done because it ensures that the validation data is representative of the dataset as a whole. Each fold was trained for 30 epochs with a batch size of 32. Early stopping was used to combat overfitting and retain the best-performing model for each fold.

Step 3: Evaluation

During the evaluation stage, the model was tested on the test data to evaluate its performance on unseen data. The model's performance was evaluated using the R2, RMSE, MAE, and MAPE scores.

Step 4: Prediction

The model can now be trained on all the available training data, and the fully trained model can predict new data points.

4.4 Residual analysis of diffusion and LSTM models

A residual analysis was performed to evaluate the diffusion and LSTM model's ability to capture relevant information from the available data. Additionally, the analysis is used to examine if the models have tendencies towards either overestimating or underestimating the target value. Kurtosis and skewness are used to evaluate the distributions, along with mean, median, and standard deviation. For large datasets, Kurtosis and Skewness can be defined as[21]

$$\text{Kurtosis} \approx \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 - 3, \tag{4.12}$$

$$\text{Skewness} \approx \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3. \tag{4.13}$$

Where $\bar{x}$ is the mean value and s is the standard divation of the time series $x = [x_1, x_2, ... \ x_i \ .., x_n]$. Kurtosis is a number quantifying the shape of a frequency distribution. If the kurtosis is zero, the frequency plot of the residuals will have a Gaussian distribution. For a residual plot, a positive kurtosis results in a pointier frequency plot than a Gaussian distribution, with more values around the mean and a faster fall-off in the tails. Negative kurtosis indicates a flatter frequency plot with fewer values around the mean and a slower fall-off in the tails. Skewness is, on the other hand, a measure of the symmetry of a distribution. A skewness of zero indicates a distribution symmetric around the mean value. If skewness is positive, most values fall on the left of the mean value, and the mean and median are typically larger than the mode. Negative skewness indicates that most values fall to the right of the mean value, and the mean and median are typically less than the mode. The residual used in the analysis was created using equation (4.14)

$$Residuals \ = estimations - targets. \tag{4.14}$$

The skewness of the residuals was found using the `skew()` function, and the kurtosis of the residuals was found using the `kurtosis()` function both from the `scipy.stats` package [28]. Finally, the residual distribution was plotted using the `plt.hist()` function from the `matplotlib` package [32].

The residuals were examined further to see if they had constant variance concerning the target value and if the residuals were independent with respect to themselves and the target value. A scatter plot was created to examine variance using the `plt.scatter()` function from the `matplotlib` package [32]. The plot shows the residuals plotted against the target values and is used to investigate the variance of the residuals with respect to the estimated values. If the linear model is a good representation of the data, the variance should be constant and have no relation to the target variable. To examine whether residuals are independent or dependent of each other, they were plotted over time using the `plt.plot()` function from the matplotlib package [32]. The plot shows the time series containing the residual values and is used to examine the existence of patterns between the residuals in the time series.

5 Results

The following sections show the results of the four methods of inertia estimation described in the method sections, section 4. They are presented in increasing order of performance, with the worst-performing model first and the best-performing model last. The two first models get a less in-depth analysis, while the last two models get a more in-depth analysis. They are all applied to the year 2020, with the third model also applied to the year 2022. The figures also show more detailed figures from January, May, and August in an attempt to show the seasonal variations for each model.

5.1 Estimation of inertia using drift

Applying the method of inertial estimation using a drift coefficient calculated from power-grid frequency, outlined in section 4.2, on the year 2020, yielded the results shown in figure 8. The fitting of equation (4.8) resulted in the change of the constant power-system dampening from the initial value of $D = 0.05$ to the value $D = 4.73{\times}10^{-5}$. Utilization of the drift coefficient for estimating the kinetic energy stored in the power system yielded the plots in figure 8 and failed to capture the relations expected to be present within the data. The estimated values, in green, fluctuate around the target values and seem to capture a minimal amount of the relations in the data. The estimations also spike to over two times the largest target value several times; on some occasions, the spikes exceed almost three times the largest target values.

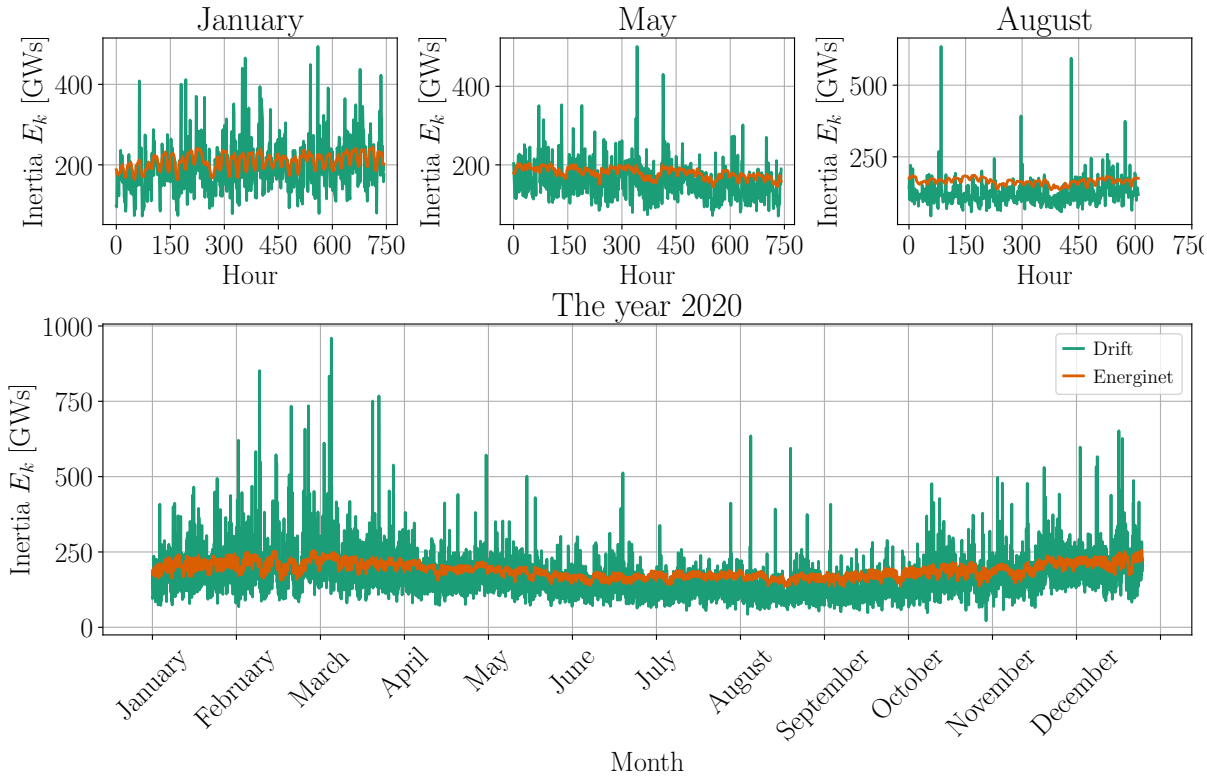

Figure 8: Estimation of inertia constants using drift-based estimation method outlined in section 4.2. The estimated values are shown in green, and the target values are shown in orange. There is no discernable relation between the estimated and target kinetic energy. For the monthly plots, the horizontal axis depicts hours. For the yearly plot, the horizontal axis depicts months. The vertical axis shows the kinetic energy, E_k, stored in the power system measured in GW·s.

Because of the generally underwhelming performance of this model, only the results for the whole year will be described; the reasoning behind this can be seen in figure 8 where the monthly variation does not contribute much meaningful insight. Still, the evaluation metrics for the months shown in figure 8 can be found in table 8. For the year 2020, the model had a MAPE of 24.3%. The estimation ranges from $[22, 960]$ GW·s with an MAE of 46 GW·s, and the target values ranges from $[135, 256]$ GW·s. The maximum error of the predictions is 723 GW·s, almost three times the largest target value.

Table 8: Evaluation metrics for the 2020 estimation using drift coefficient. The metrics used to evaluate the estimations against the targets are R2 score, RMSE, MAE, and MAPE. The RMSE and MAE are scale dependent and have the unit GW·s. The R2 Score and MAPE are scale-independent and have no units.

	January	May	August	Year
R2	-7.485	-12.115	-28.754	-4.929
RMSE	56.085	47.852	60.673	60.293
MAE	42.047	38.093	49.932	45.790
MAPE	20.20%	21.30%	30.50%	24.30%

5.2 Estimation of inertia using generic H constants

The inertial estimation with the generic H constant, outlined in section 4.1, yielded the results shown in figure 9. The estimations show a similar trend to inertia data from Energinet with an offset. The evaluation metrics from table 9 show that on the months tested, the model has the best accuracy for January with a MAPE of 14.2%. While the actual targets range from January are from $[163, 245]$ GW·s, the values of estimations range from $[131, 218]$ GW·s with a MAE with an average absolute error of 29 GW·s, as seen in table 9. The accuracy of the 2020 model reduces from January to May. For May, the

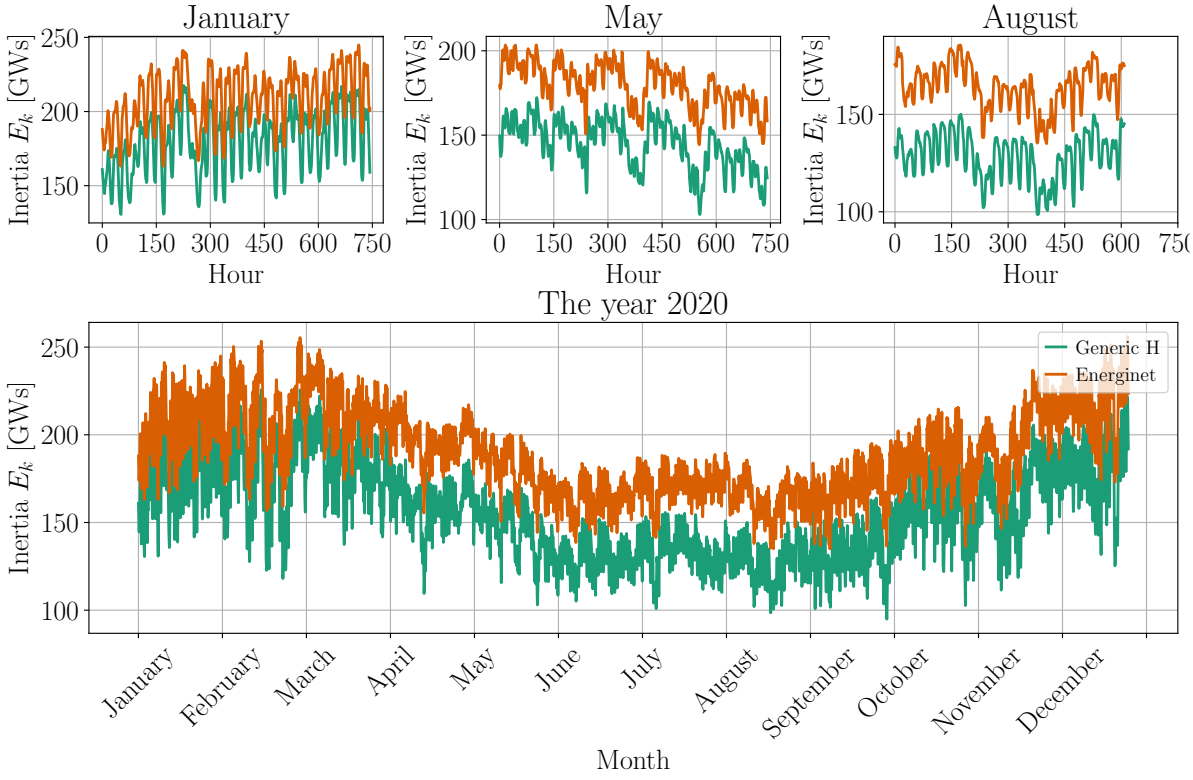

Figure 9: Inertia estimation via the generic H constant method described in section 4.1 applied to the year 2020. Estimations are shown in green, and target values are in orange. Analyzing the trends in the four figures, the estimated inertia seems to follow a similar trend to the ground truth inertia but with an offset around 35 GW·s. The model is most accurate during the winter months and the least accurate in the summer months. For the monthly plots, the horizontal axis shows hours. For the yearly plot, the horizontal axis shows months. The vertical axis does, in all cases, show kinetic energy stored in the power system measured in GW·s.

model had a MAPE of 19.6%. Figure 9 shows that the overlap of estimation and target values in January is no longer present for May. However, it seems the estimations can still reproduce the general trend of the target data because the peaks and valleys shown coincide for estimation and target inertia. For May, the estimations range from $[103, 172]$ GW·s, with an average absolute error of 35 GW·s. The actual targets have a range of $[144, 203]$ GW·s. From May to August, the MAPE rises to 21.8%. For August, the model still shows an ability to recreate the general trends in the data. The values of the estimations are in the range $[99, 150]$ GW·s, with an average absolute error of 36 GW·s. The target values in this month are in the range $[135, 186]$ GW·s. A summary of the different performance metrics evaluated on the different months can be found in table 9.

Table 9: Evaluation metrics for the 2020 estimation using generic H constants. The metrics used to evaluate the estimations against the targets are R2 score, RMSE, MAE, and MAPE. The RMSE and MAE are scale dependent and have the unit GW·s. The R2 Score and MAPE are scale-independent and have no units.

	January	May	August	Year
R2	-1.383	-6.170	-9.430	-0.983
RMSE	29.739	35.520	36.117	34.850
MAE	29.272	34.970	35.617	34.093
MAPE	14.20%	19.60%	21.80%	18.20%

Model performance for the year 2020 is slightly better than in May and August. For the year as a whole, the MAPE is 18.2%. The inertial estimations have the range of $[95, 227]$ GW·s with an average absolute error of 34 GW·s. The range of the inertia targets is $[135, 256]$ GW·s. The bottom plot in figure 9 shows how the estimated and target inertia changes over the year. The generic H model gives values closest to the target inertia for the winter months of January, February, and December. Values furthest from the target inertia occur in the summer months of June, July, and August.

5.3 Estimation of inertia using diffusion

Applying the method for estimating kinetic energy using the diffusion coefficient calculated from the power-grid frequency, outlined in section 4.2, for the year 2020 yielded the results shown in figure 10. The best fit of equation (4.11) to the inertial data for 2020 resulted in the change of the constant $C = 0.00457\,\text{s}$ from its initial value of $C_0 = 0.00466\,\text{s}$. The offset value was also changed from its initial value of $\Upsilon_0 = 100\,\text{GW·s}$ to the fittest value $\Upsilon = 93.81\,\text{GW·s}$. Looking at figure 10 and the evaluation metrics in table 10, the model performed best in January with a MAPE of 3.6% and a R2 score of 0.73. For January the estimations ranged from $[152, 270]$ GW·s with an MAE of 7.70 GW·s. The actual targets for this period were within the range of $[163, 245]$ GW·s, which is contained within the estimation range for the period. The MAPE increases slightly in May to 4.7% with an R2 score of 0.38. The estimation are in the range $[153, 227]$ GW·s with an MAE of 8.28 GW·s. The actual targets are in the range of $[144, 203]$ GW·s. Comparing this range to the estimation range, it is evident that some of the target values are not within the estimation range. MAPE rises again for August, where it is 5.1% with an R2 score of 0.13. The estimations are within the range $[142, 207]$ GW·s with an absolute error of 8.2 GW·s. The actual targets for August are contained in the range $[135, 186]$ GW·s; comparing this range to the estimation range, it becomes clear that some of the target values for August are not within the range of estimation.

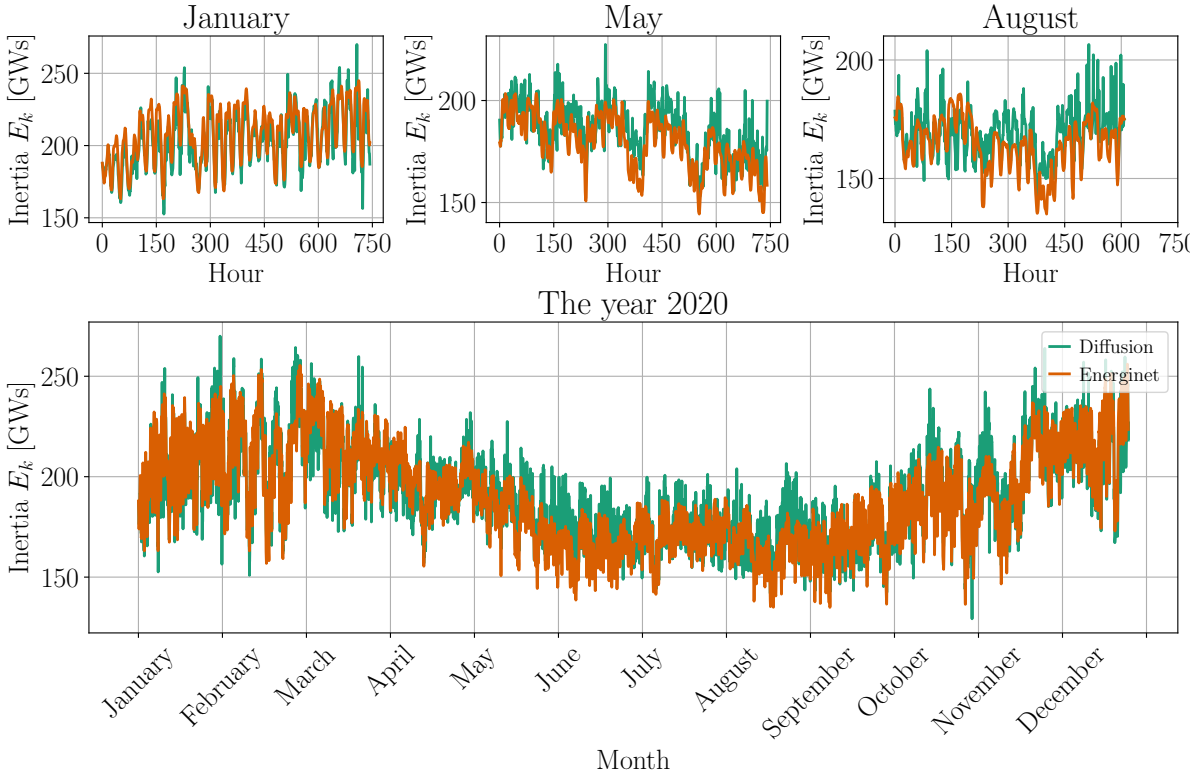

Figure 10: Inertia estimation via the diffusion coefficient outlined in section 4.2 applied on the year 2020. The estimated kinetic energy values are shown in green, and the targets are shown in orange. The model is most accurate for the winter months and least accurate for the summer months. For the monthly plots, the horizontal axis shows hours. For the yearly plot, the horizontal axis shows months. For all plots, the vertical axis shows kinetic energy in the power system measured in GW·s.

For the year 2020 as a whole, the model had a MAPE of 4.2% and an R2 score of 0.84. The yearly MAPE is better than the MAPE for May and August and slightly worse than the MAPE for January. Moreover, the R2 score for the whole year is better than for the individual months. The estimations are within the range of $[129, 270]$ GW·s with an MAE of 7.83 GW·s. The target values are within the range $[135, 256]$ GW·s notably, all targets are within the range for the estimations. The complete table of evaluation metrics for the diffusion estimation can be found in table 10.

Table 10: Evaluation metrics for the 2020 estimation using diffusion coefficient. The metrics used to evaluate the estimations against the targets are R2 score, RMSE, MAE, and MAPE. The RMSE and MAE are scale dependent and have the unit GW·s. The R2 Score and MAPE are scale-independent and have no units.

	January	May	August	Year
R2	0.725	0.375	0.133	0.838
RMSE	10.102	10.444	10.355	9.975
MAE	7.698	8.281	8.186	7.832
MAPE	3.60%	4.70%	5.10%	4.20%

To further examine if the model tends to over- or under-predict the target values, the histogram-based residual analysis, outlined in section 4.4, was used. When applied to the year 2020, it yielded the residual

distribution shown in figure 11.

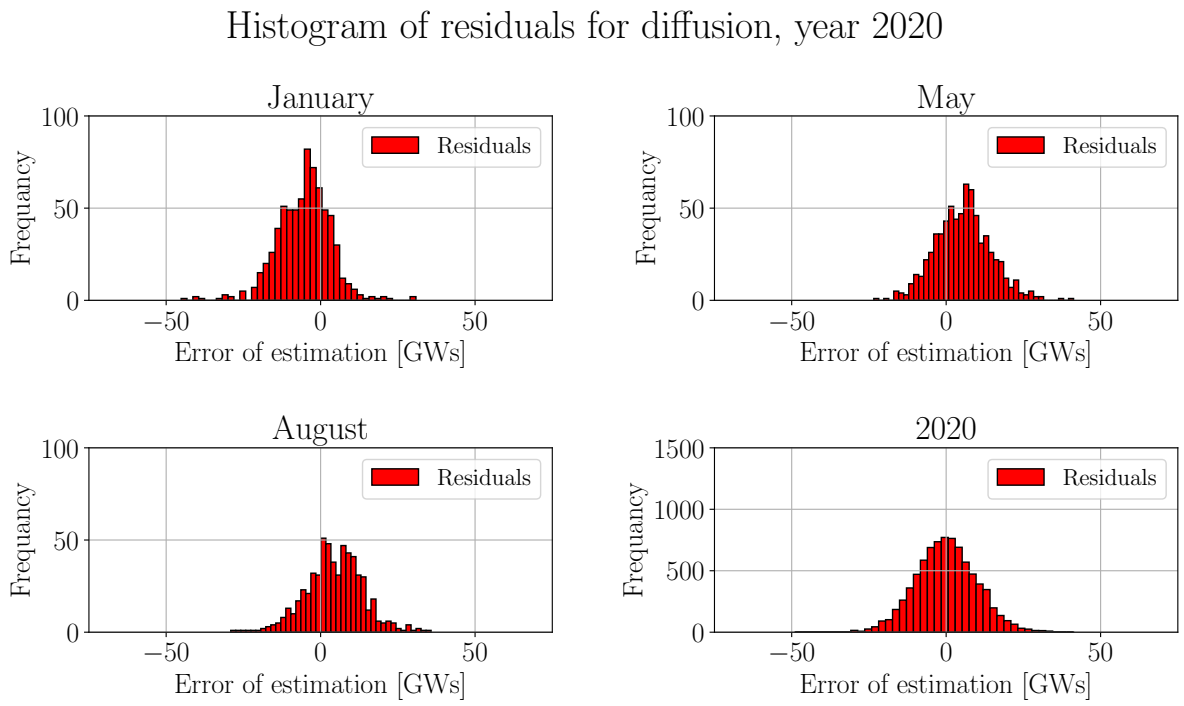

Figure 11: Residual distribution for inertia estimation based on diffusion coefficient, 2020. With the magnitude of the residual on the horizontal axis, given in GW·s and the frequency of the residual given on the vertical axis.

The distribution in figure 11, in combination with table 11, shows that the diffusion model for 2020 tends to under-estimate the kinetic energy for January with a mean error of $-0.12\,\text{GW·s}$. The model tends to overestimate the kinetic energy for May and August with a mean error of $5.3\,\text{GW·s}$ and $1.8\,\text{GW·s}$, respectively. Seeing the year as a whole, the model shows no tendency to over or under-predict, having a mean error of $0\,\text{GW·s}$.

Table 11: Statistical information about the residuals from the diffusion estimation for the year 2020. Skewness measures the discrepancy between the given distribution and a symmetrical distribution around the same mean. Kurtosis measures the discrepancy between a normal distribution and the given distribution.

	January	May	August	2020
Skewness	−0.29	0.19	−0.07	0.01
Kurtosis	2.19	0.43	0.71	0.54
Mean	−5.33	5.44	4.44	0.00
Median	−4.56	5.63	4.39	−0.13
Std	8.58	8.91	9.35	9.97

The method of estimating kinetic energy from the diffusion coefficient outlined in section 4.2 showed more promising results than the methods using the drift coefficient. Thus, the fitting of equation (4.11) was done for the other years individually, and the combined data of all the years 2020–2023. A residual

analysis similar to the analysis to the one done for 2020 was performed for the yearly fitted models and for the model fitted to all the data. The model fitting yielded coefficients for C and Υ; the coefficients can be found with their corresponding year in table 12.

Table 12: Fitting of the parameters C and Υ in equation (4.11). The fitting was done for the years individually and the combined 2020–2023 data. The variance and covariance are also shown for C and Υ parameters.

	2020	2021	2022	2023	All years
C [s]	0.0046	0.0045	0.0031	0.0034	0.0032
Υ [GW·s]	93	100	140	140	130
variance C [s^2]	5.08×10^{-10}	4.00×10^{-10}	8.01×10^{-10}	6.23×10^{-10}	1.94×10^{-10}
variance Υ [$(GW\cdot s)^2$]	0.24	0.19	0.24	0.19	0.078
covariance [$GW\cdot s^2$]	-1.08×10^{-5}	-8.41×10^{-6}	-1.30×10^{-5}	-1.02×10^{-5}	-3.69×10^{-6}

To further investigate if the fitting should be carried out yearly or globally, the residual analysis method described in section 4.4 was applied for all the years and the combined data for all years. Figure 12 shows the resulting frequency plots for the yearly fits, and figure 13 shows the residual plot for the entire dataset, 2020-2023. The fit for 2020 and 2023 has a skewness of -0.01 and -0.81 and a kurtosis of 0.53 and 0.22, respectively, indicating normally distributed residuals for both years. The year 2022 has, on the other hand, a skewness of -0.81 and a kurtosis of 1.63. The kurtosis of this distribution indicates a distribution that is starting to resemble a leptokurtic distribution, with a much larger peak around the mean value and more outliers. The resemblance to a leptokurtic distribution is even more evident for 2021, with a kurtosis of 6.78 and a skewness of -1.05. The residual distributions are overall similar from year to year, with skewness in the range of $[-1.05, -0.01]$ and kurtosis in the range of $[0.22, 6.78]$. The distributions also have a similar mean and standard deviation, with the mean values being in the range $[0, 0.07]$ GW·s and the standard deviation within the range $[9.97, 14.91]$ GW·s. The distribution showing the residuals for the global fit has a kurtosis of 49.76 and has the characteristics of a leptokurtic distribution. Therefore, this distribution should not have the same properties as the yearly ones, even though the mean of 0.0 GW·s, a median of 0.22 GW·s, and standard deviation of 16.44 GW·s are reasonable compared to the yearly values.

Table 13: Statistical information about the residual distribution for the yearly and globally fitted estimations.

	2020	2021	2022	2023	All years
Skewness	-0.01	-1.05	-0.81	-0.08	-1.64
Kurtosis	0.53	6.78	1.63	0.22	49.76
Mean [GW·s]	0.00	0.00	0.07	0.00	0.00
Median [GW·s]	-0.13	0.34	1.77	0.22	0.02
Standard deviation [GW·s]	9.97	11.58	14.91	14.02	16.44

Due to the differences between the yearly and global residual distributions, the analysis is proceeding with only the yearly fittings. The scatter plot in figure 14 can be used to check if the residuals are independent of target values. When plotting the residuals against the target values, no significant correlation between the magnitude of the error and the amount of kinetic energy stored in the power system can be seen.

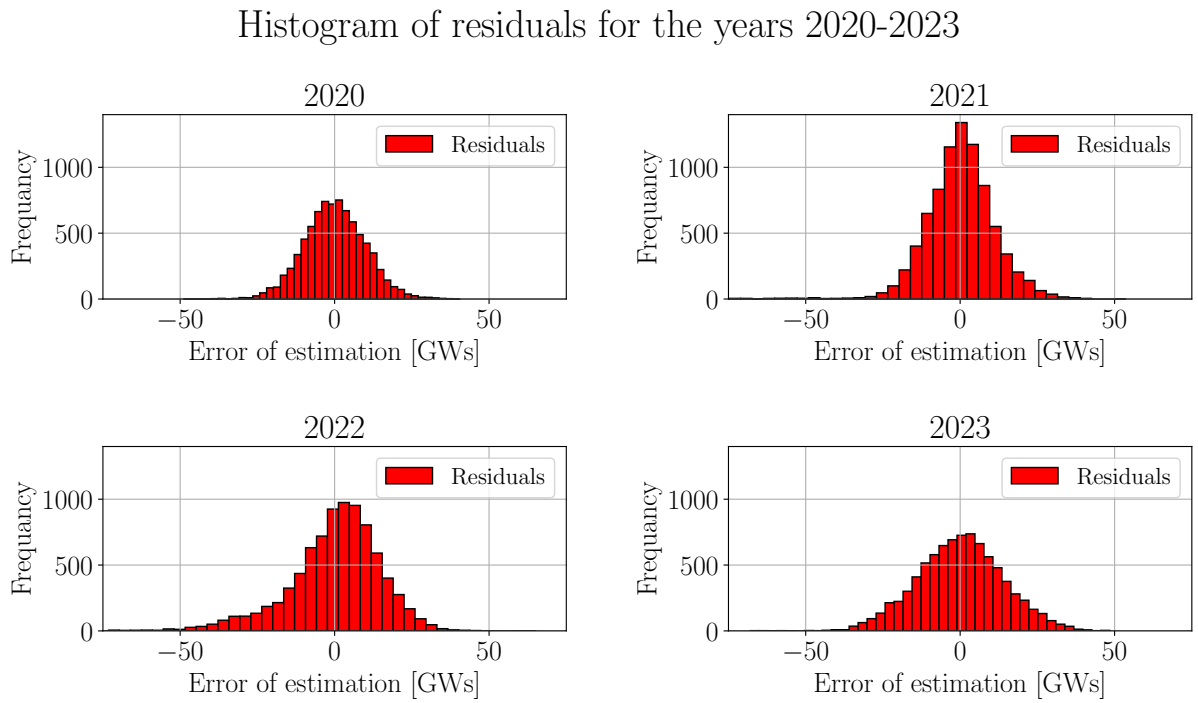

Figure 12: Residual distribution for kinetic energy estimation based on diffusion coefficient, for the years 2020–2023. With the magnitude of the residual on the horizontal axis, given in GW·s and the frequency of the residual given on the vertical axis. Statistical information about the distributions can be found in table 13.

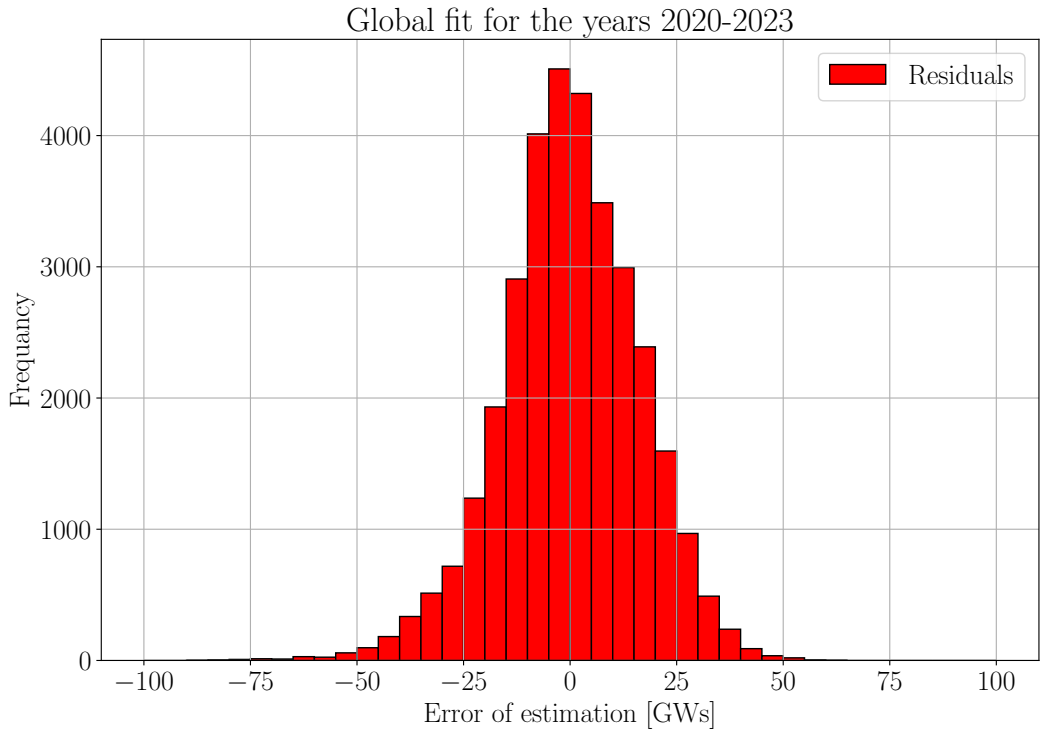

Figure 13: Residual distribution of the global dataset. Statistical information about the distribution can be found in table 13 and shows that this distribution is leptokurtic. The figure has the magnitude of the error in GW·s on the horizontal axis and the error frequency on the vertical axis.

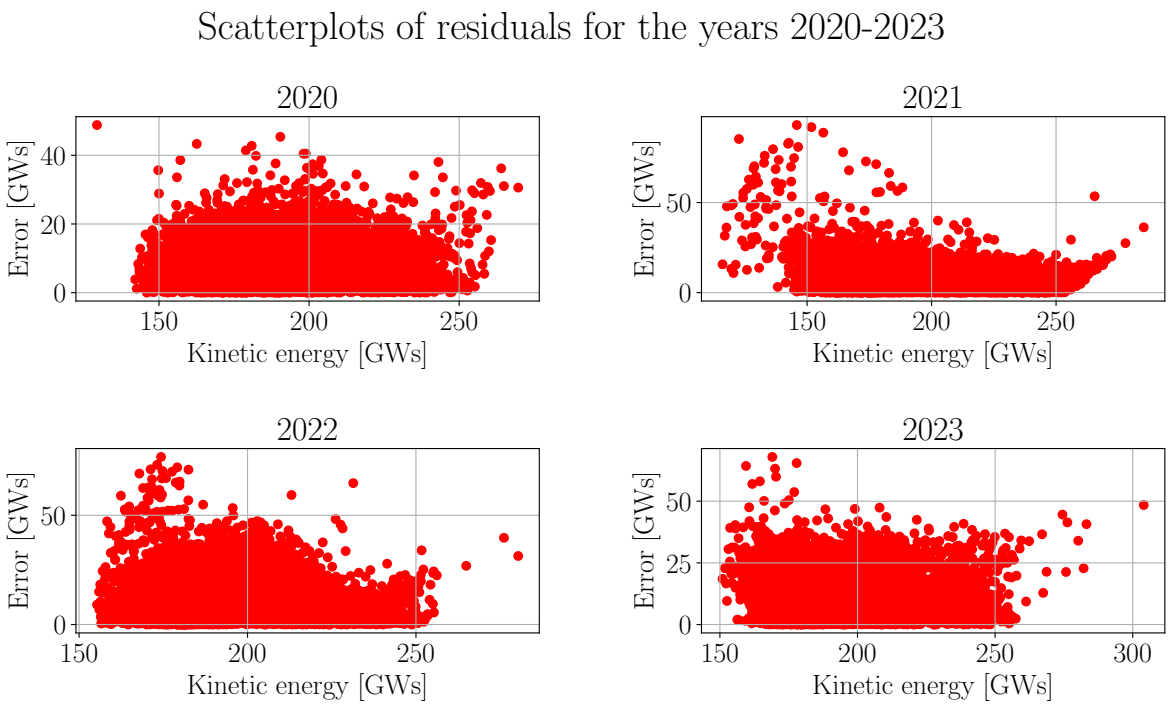

Figure 14: Scatter plot of the absolute magnitude of the residuals given in GW·s against the kinetic energy stored in the power system, also given in GW·s.

Plotting the residuals over time can be used to investigate if the residuals depend on each other. Figure 15 shows the magnitude of the residuals plotted over time, with no discernable pattern between the residuals. Indicating that the residuals are independent of each other.

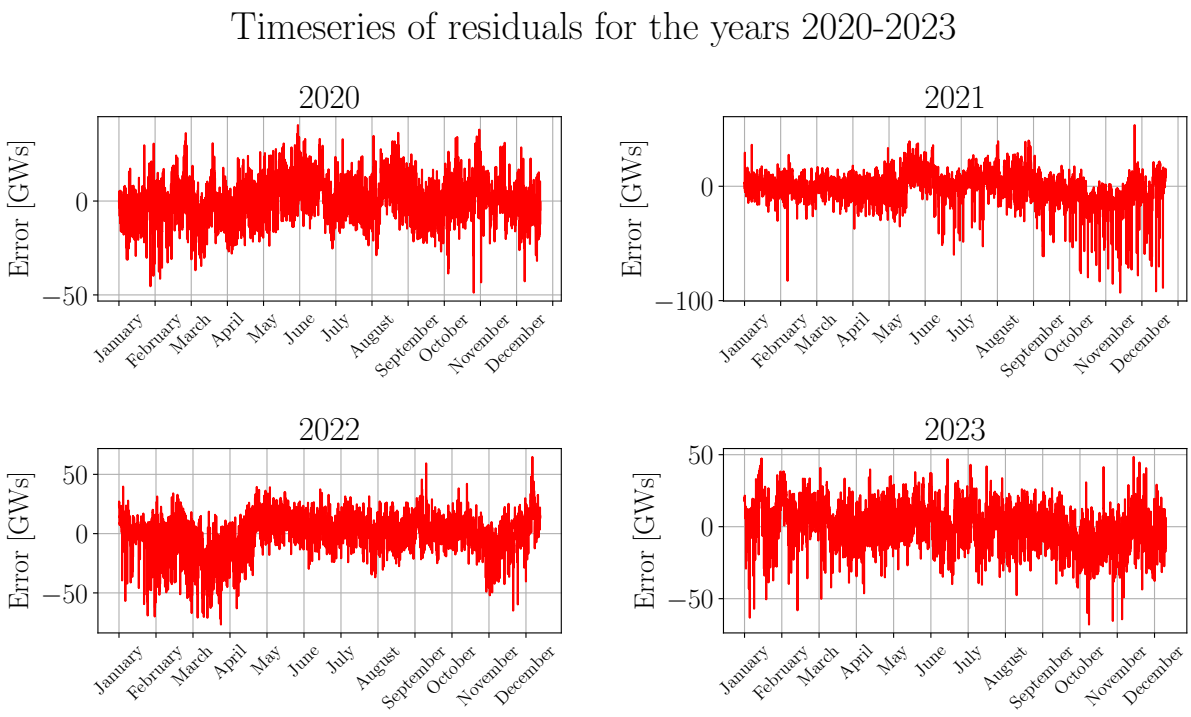

Figure 15: Time series of the magnitude of the error, residuals, between the estimated stored kinetic energy and actual kinetic energy in the power system given in GW·s.

Given the different distributions of the residuals, seen in figure 12, and the different Υ and C values, seen in table 12 another year was evaluated using the evaluation metrics. The year 2022 was chosen because of the values of the parameters $C = 0.0031\,\text{s}$ and $\Upsilon = 1.4{\times}10^2\,\text{GW·s}$, and residual distribution with a kurtosis of 1.63. These values are the furthest from the C and Υ of the testing year 2020, and the residual distribution is more leptokurtic than the residual distribution for 2020. Applying the diffusion-based estimation method outlined in section 4.2 on the year 2022 yielded the results displayed in figure 17.

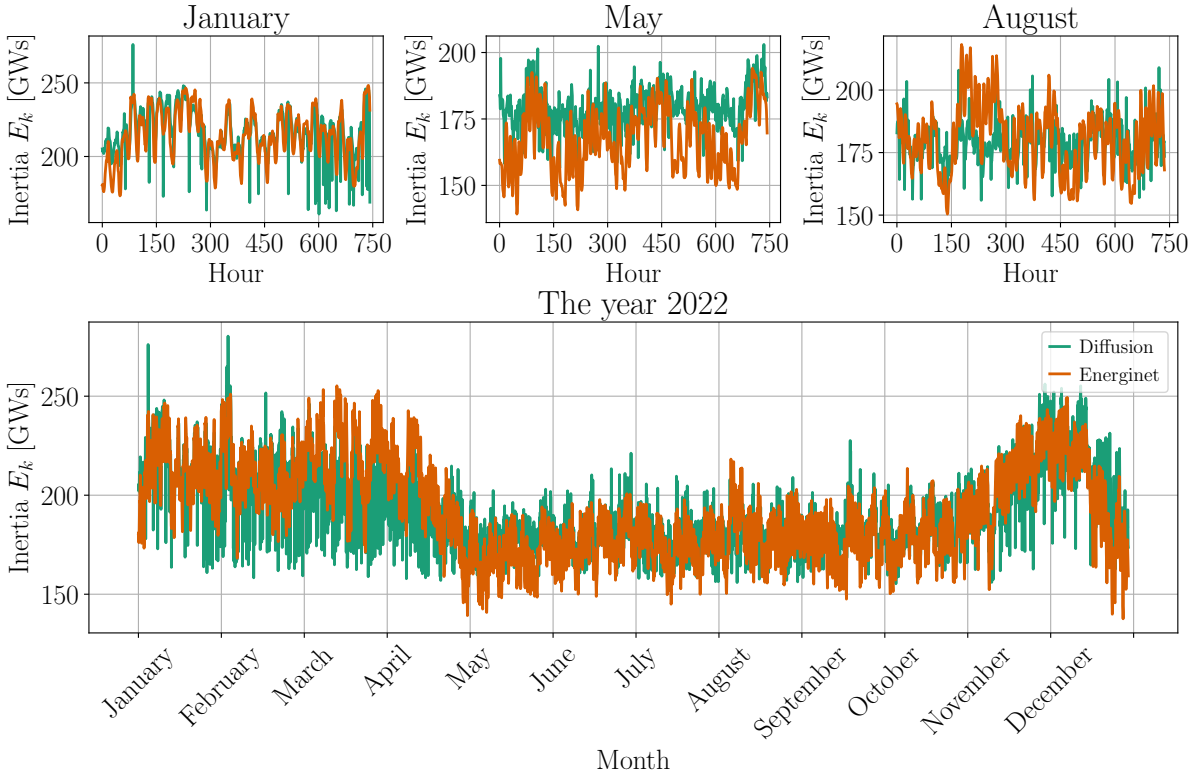

Figure 16: Inertia estimation via the diffusion coefficient, $D^{(2)}$, outlined in section 4.2, applied on the year 2022. The estimated kinetic energy values are shown in green, and the targets are shown in orange. The model is most accurate for the winter months and least accurate for the summer months. Notably, for the year 2022, there is a large discrepancy between estimated and target values from late January to April. For the monthly plots, the horizontal axis shows hours. For the yearly plot, the horizontal axis shows months. For all plots, the vertical axis shows kinetic energy in the power system measured in GW·s.

Looking at the evaluation metrics in table 14, the estimation for 2022 presents the best performance for January with a MAPE of 4.0% and R2 score of 0.46. This month's estimation has the range $[161, 276]$ GW·s with an MAE of 8.3 GW·s, and a target range of $[173, 248]$ GW·s. Notably, all observed target values fall within the estimated range. Model accuracy for May is lower when comparing 2022 to 2020, with a MAPE of 8.1% and an R2 score of -0.66. The R2 score indicates that the model loses its ability to capture the general relations of the data for the month of May. The estimations are within the range of $[156, 203]$ GW·s with an MAE of 13.1 GW·s, while the target range for this month is within the range $[139, 194]$ GW·s. Notably, all the target values for this month are not within the range of estimation. For August, the model has a MAPE of 5.4% and an R2 score of 0.14, which is close to the MAPE and R2 score seen for August 2020. The estimations have a range of $[156, 209]$ GW·s with a MAE of 11.2 GW·s, and the actual target range were within $[150, 218]$ GW·s.

Table 14: Evaluation metrics for the 2022 estimation of stored kinetic energy, using diffusion coefficient. The metrics used to evaluate the estimations against the targets are R2 score, RMSE, MAE, and MAPE. The RMSE and MAE are scale dependent and have the unit GW·s. The R2 Score and MAPE are scale-independent and have no units.

	January	May	August	Year
R2	0.464	−0.659	0.136	0.619
RMSE	12.583	15.595	12.443	14.908
MAE	8.295	13.092	9.797	11.211
MAPE	4.00%	8.10%	5.40%	5.90%

For the year as a whole, the diffusion estimation for 2022 has a MAPE of 5.9% and an R2 score of 0.62. The estimations are contained within the range $[155, 280]$ GW·s with a MAE of 7.85 GW·s. The target values are within $[138, 255]$ GW·s. Notably, not all target values are within the estimation range. Following the same method as for the year 2020, a histogram-based analysis was used to examine if the model tends to over- or under-estimate the kinetic energy in the power system. The method yielded the distributions seen in figure 17.

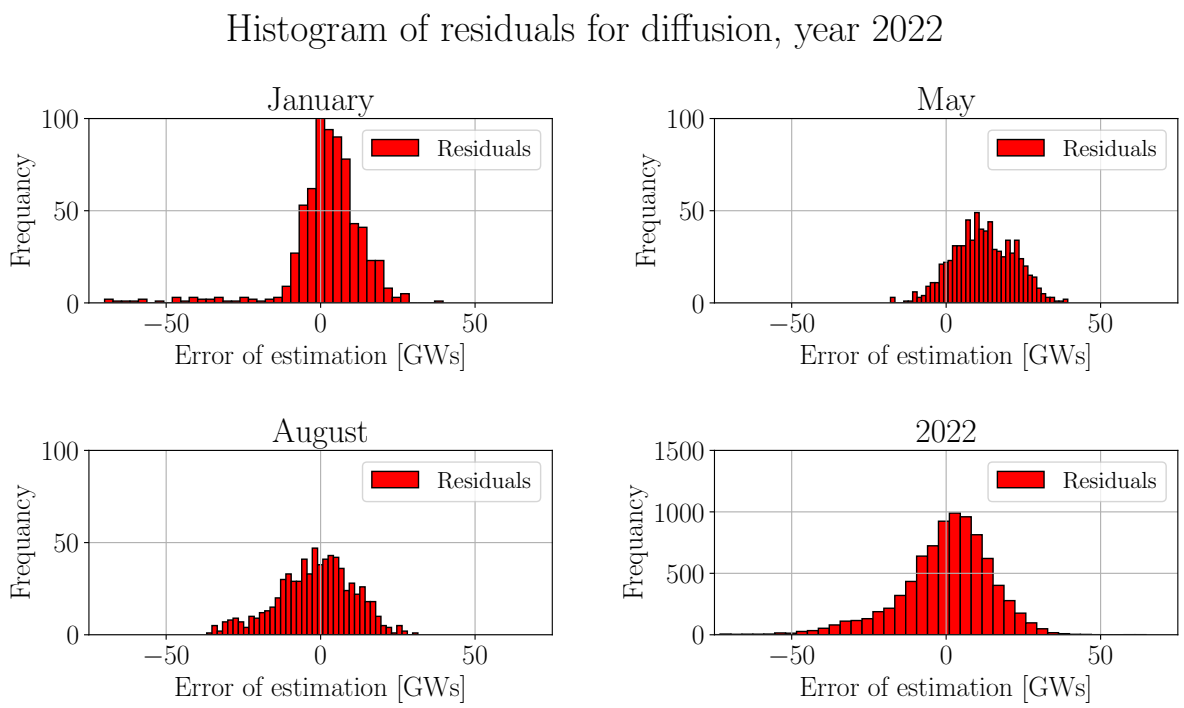

Figure 17: Residual distribution for kinetic energy estimation based on diffusion coefficient, 2022. With the magnitude of the residual on the horizontal axis, given in GW·s and the frequency of the residual given on the vertical axis.

The statistical information given in table 15 shows that the model for 2022 has a trend of overestimating the kinetic energy in the power system for January, May, and the year as a whole with a mean error of 0.94 GW·s, 13.92 GW·s, and 1.57 GW·s respectively. For August, the model tends to underestimate the

kinetic energy in the power system with a mean error of $-0.91\,\text{GW}\cdot\text{s}$.

Table 15: Statistical information about the residuals from the diffusion estimation for the year 2022. Skewness measures the discrepancy between the given distribution and a symmetrical distribution around the same mean. Kurtosis measures the discrepancy between a normal distribution and the given distribution.

	January	May	August	2020
Skewness	-2.24	-0.03	-0.33	-0.81
Kurtosis	9.30	-0.28	-0.08	1.62
Mean	2.18	12.11	-1.74	0.00
Median	3.00	11.83	-0.99	1.67
Std	12.39	9.83	12.32	14.91

5.4 Estimation of inertia using LSTM

Applying the LSTM model outlined in section 4.3 to the year 2020 yielded the results shown in figure 18. From looking at both the estimation in green and the targets in orange, the general trend is well captured by the model. Table 16 provides the evaluation metrics for the LSTM estimation kinetic energy stored in the power system in 2020; the model performed best in January with MAPE 1.6% and an R2 score of 0.96. The estimations are within the range of $[165, 242]\,\text{GW}\cdot\text{s}$ with an MAE of $3.29\,\text{GW}\cdot\text{s}$. The targets for January are contained within the range of $[163, 245]\,\text{GW}\cdot\text{s}$, notably some of the target values are not within the estimation range. For May, the model performed worse with a MAPE of 4.3% and an R2 score of 0.52. The estimations range from $[129, 203]\,\text{GW}\cdot\text{s}$ with MAE of $7.69\,\text{GW}\cdot\text{s}$. The targets for May were in the range of $[144, 203]\,\text{GW}\cdot\text{s}$; notably, all the target values are within the estimation range. For August the LSTM model had an MAPE of 4.4% and an R2 score of 0.44. The estimations for this month have the range of $[127, 184]\,\text{GW}\cdot\text{s}$ with an MAE of $7.20\,\text{GW}\cdot\text{s}$. The targets are contained within the range of $[135, 186]\,\text{GW}\cdot\text{s}$, notably this means some of the target values are not within the estimation range. For 2020, the LSTM model had a MAPE of 3.1% with an R2 score of 0.92. The estimations across the year are within the range $[121, 248]\,\text{GW}\cdot\text{s}$ with a MAE of $5.63\,\text{GW}\cdot\text{s}$. The targets are within the range $[135, 256]\,\text{GW}\cdot\text{s}$; notably, not all the target values are within the estimation range. The full table showing the LSTM evaluation metrics can be found in table 16.

Table 16: Evaluation metrics for the 2020 estimation using the LSTM model. The metrics used to evaluate the estimations against the targets are R2 score, RMSE, MAE, and MAPE. The RMSE and MAE are scale dependent and have the unit GW·s. The R2 Score and MAPE are scale-independent and have no units.

	January	May	August	2020
R2	0.955	0.515	0.443	0.917
RMSE	4.051	9.156	8.216	7.125
MAE	3.291	7.685	7.197	5.626
MAPE	1.60%	4.30%	4.40%	3.10%

As for the previous model, the distribution of residuals is interesting in examining tendencies to overestimate or underestimate the kinetic energy stored in the power system. A histogram showing the distribution of the residuals can be seen in figure 19. Looking at where the distributions in figure 19 seem to have their mean values indicates the model's tendency to over or under-estimate kinetic energy in the power system. Comparing the figure with the statistical information regarding mean in table 17, the LSTM model seems to overestimate the stored kinetic energy in January, with a mean error of $2.69\,\text{GW}\cdot\text{s}$,

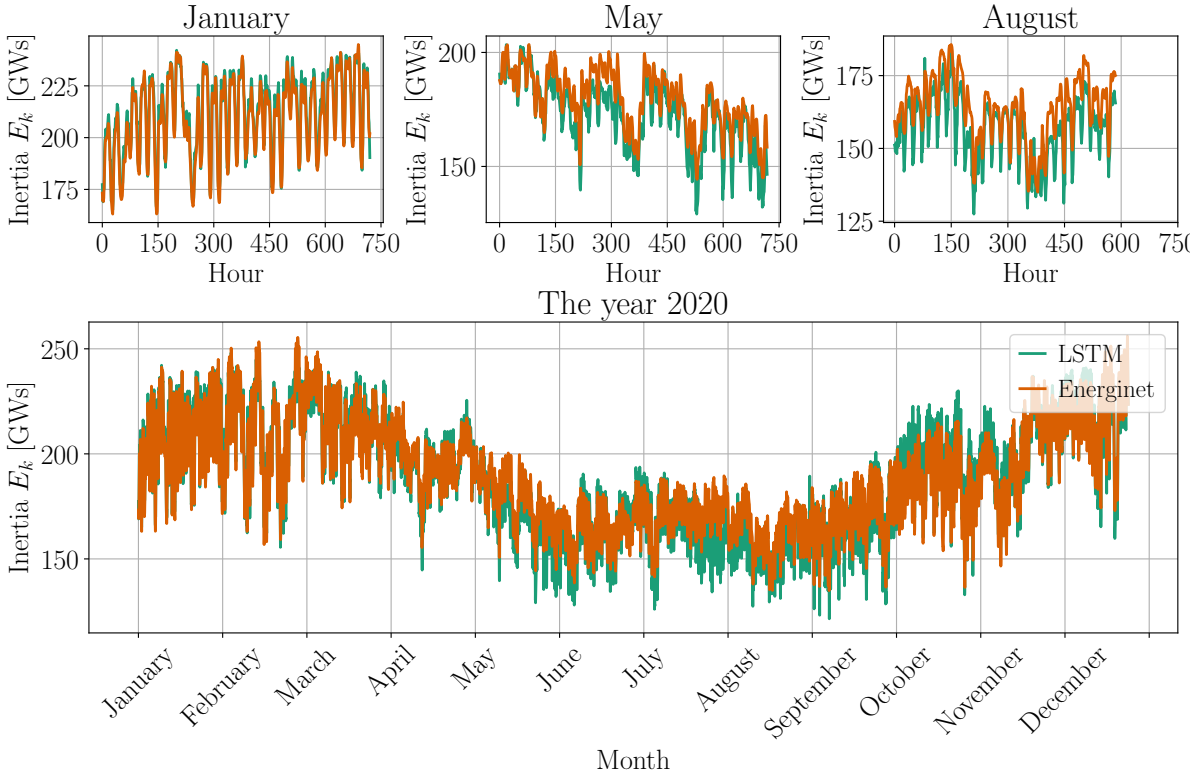

Figure 18: Inertia estimation using an LSTM machine learning model outlined in section 4.3 for the year 2020. The estimated values are given in the green figure, and the target values are shown in the orange figure. The model seems to capture the general trend in the data, and the estimations follow the target values closely. The model seems to have the best performance during the winter months and slightly worse performance during the summer months. For the monthly plots, the horizontal axis shows the timestep in hours. For the yearly plot, the horizontal axis shows months. For all plots, the vertical axis shows Stored kinetic energy in the power system given in GW·s.

and underestimate the stored kinetic energy for the rest of the months tested. May has a mean error of $-7.17\,\mathrm{GW{\cdot}s}$ and August has a mean error of $-7.30\,\mathrm{GW{\cdot}s}$. For the whole year, the model has a mean error of $-0.98\,\mathrm{GW{\cdot}s}$, showing that for 2020, the model underestimates the stored kinetic energy more often than it overestimates it.

Table 17: Statistical information about the residual distribution displayed in figure 19. Skewness measures how symmetrical the distribution is around its mean, and kurtosis measures how similar the distribution is to a normal distribution. Skewness and kurtosis of zero indicate a symmetrical and normal distribution, respectively.

	January	May	August	2020
Skewness	-0.07	-0.08	0.06	-0.12
Kurtosis	0.57	0.03	-0.04	0.16
Mean[GW·s]	2.69	-7.17	-7.05	-1.32
Median[GW·s]	2.61	-7.28	-7.30	-0.98
Standard deviation[GW·s]	3.03	5.69	4.22	7.00

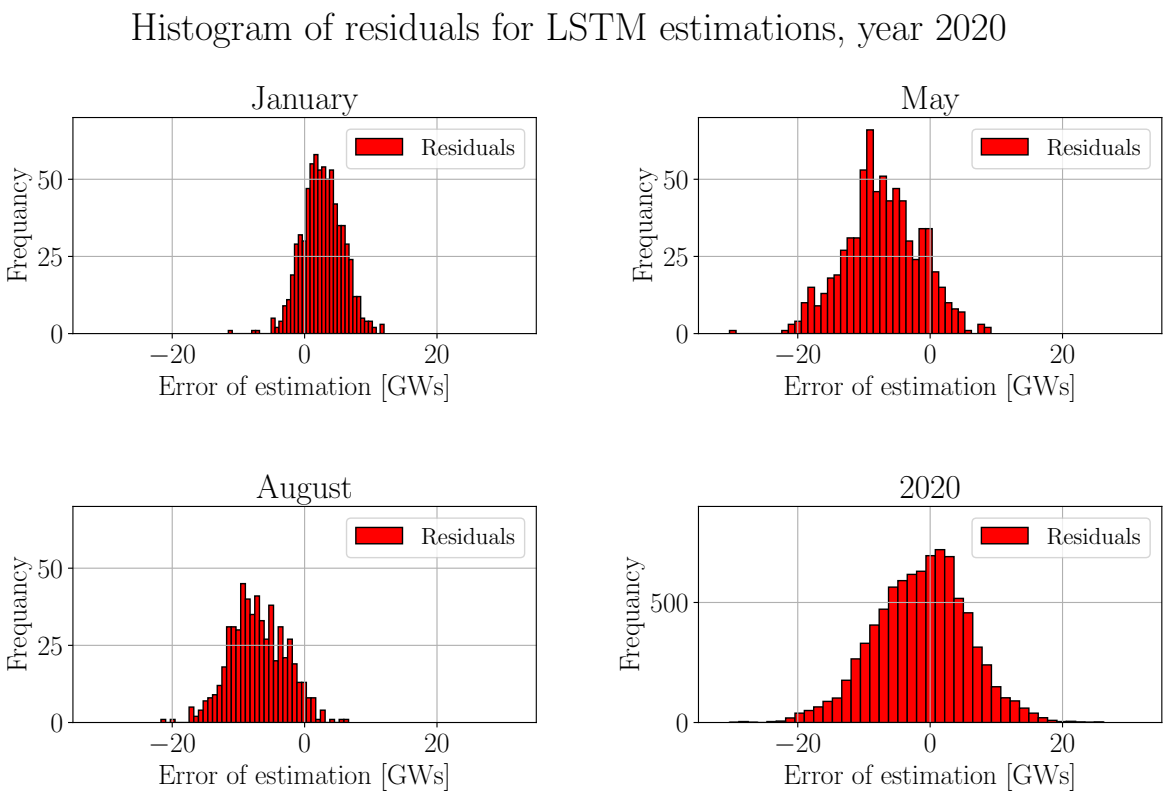

Figure 19: Residual distribution for kinetic energy estimation by LSTM machine learning model, 2020. With the magnitude of the residual on the horizontal axis, given in GW·s and the frequency of the residual given on the vertical axis.

6 Discussion

The discussion of the results presented in this thesis has three main paragraphs. *Key findings and interpretation* outline the key findings from the results presented in section 5 and the interpretations of these results. *Limitations* presents some limitations regarding the choice of the stochastic model and the equation used to estimate kinetic energy. This section also covers some generalizations that are not necessarily true for other power systems, given the results in this thesis. *Implication and recommendations* cover the implication of the key findings and limitations with respect to current practices and give suggestions on further work based on what has been explored in this thesis.

6.1 Key findings and interpretation

Four models were created to estimate inertia in the Nordic power system, outlined in section 4, the generic H-, the drift-, the diffusion- and the LSTM models. The generic H, drift, and diffusion model uses production data together with H constants for their estimations. The generic H model uses average values for the inertial constant H to estimate inertia. The drift and diffusion models were based on stochastic modeling of power-system frequency using an Ornstein-Uhlenbeck process. The drift model uses the drift coefficient, $D^{(1)}$, and the diffusion model uses the diffusion coefficient, $D^{(2)}$, to estimate regional inertial constants hourly for the year 2020. To evaluate the generic H, drift, and diffusion models, the inertia constants were combined with data of synchronous production, also for 2020, to create inertia estimation so that the inertia estimation could be compared to the inertia estimation of a power system operator. Finally, an LSTM model was trained on production, load, and frequency data from 2021–2023 and then used to estimate the inertia in the power system directly from the input features from 2020. The first five rows of the LSTM data are available in appendix B.

The main reason for creating the generic H model was to examine the behavior of the linear model, outlined in section 4.1, to get a better understanding of how the linear model, used to create estimations of kinetic energy from the different inertial constants, behaves compared to the TSO estimations used as target values. The generic H constants in table 6, gathered from the ENTSO-E report [5], represent the average inertial constant of a generator by production type. The most important finding from the generic H model is the existence of an offset of around 35 GW·s, as seen in figure 9. The presence of the offset is likely due to the use of aggregated active power output rather than rated separate power and not because the H constants fail to represent true values accurately. To ensure that the metrics evaluating the performance of the diffusion model reflect the precision of the regional inertial constant rather than the accuracy of active power measurement, the offset parameter Υ is introduced to equation (6.1). This parameter is designed to correct for the offset caused by the use of aggregate active power. The generic H model displays an ability to capture much of the underlying patterns and dynamics of the data, as seen by how the trajectory of the estimation plot closely mirrors the trajectory of the target plot in figure 9. This could indicate that much of the estimation error can be attributed to a constant offset error, resulting in the estimations consistently being lower than the target values by a fixed amount. The metric table 9 shows a seasonal variation in model performance. During the test months, the model performed better in January with a MAPE, MAE and RMSE lower than the annual average, while the months of August and May had higher MAPEs. Because of the offset that exists between the estimated inertia and the target inertia for the generic H model, the R2 score does not indicate anything useful and is not discussed further. Figure 9 indicates that the amount of stored kinetic energy in May and August is significantly lower compared to January. It could be the case that the model performed better for periods with high amounts of stored kinetic energy within the power system than for periods with low kinetic energy stored within the power system. The periods with high and low inertia also coincide with periods of high and

low production and consumption of electrical power, as the winter months in the Nordic region create a larger need for power than the summer months.

With a MAPE of 4.2%, R2 score of 0.837, MAE of 7.8 GW·s, and an RMSE of 9.98 GW·s for the year 2020, the diffusion model showed an ability to estimate kinetic energy stored within the power system with a surprising degree of accuracy. This shows that the diffusion-estimated regional inertia constant discussed in section 4.2 captures relevant information from the data-driven analysis of frequency measurements for the Nordic power system. The model's ability to capture relevant data from frequency strengthens the assumption that frequency behavior can be estimated by an Ornstein-Uhlenbeck process on an hour-to-hour basis in the Nordic power system and that a stochastic representation of frequency allows the extraction of useful information from noise-based deviations in power grid frequency. The equation used for the diffusion estimation (4.11), and equation (1.1) proposed in the report [5] is rewritten below for convenience,

$$E_{k,\text{diffusion}} = \frac{P}{pf}\frac{C}{\alpha} + \Upsilon, \tag{6.1}$$

$$E_{k,\text{TSO}} = \sum_{i=1}^{K} S_{ni} H_i. \tag{6.2}$$

Equation (6.1) uses an offset parameter, Υ, in an attempt to correct the offset error that can be seen in figure 9. The offset parameter compensates for the average difference between the total rated apparent power and the active power produced for a specific type of production. However, in periods where generators operate closer to maximum capacity, it is expected that the produced active power of a production type P_m starts to approximate the aggregate rate apparent power, $S_{n,m}$, for production type m. If the two power terms become more similar, it would cause a smaller discrepancy between the active power generated and the rated apparent power used in equation (6.1) and equation (6.2), respectively. This could result in equation (6.1) starting to approximate equation (6.2) for periods where the power system uses more of the available capacity, resulting in a smaller discrepancy than the annual average. A potential result of this is a fitting of equation (6.1) favoring the periods with high amounts of stored kinetic energy, leading to better performance for these periods.

To ensure that 2020 was not a random case of success, the model was also tested for the year 2022. The model performed a bit worse for the year 2022. This includes the loss of its ability to capture much of the relevant variance of the label for May, indicated by the negative R2 score; however, given the model's simplicity, the performance did not drop significantly enough for the rest of the year to indicate that the estimation for 2020 is an isolated case of success. This suggests that stochastic modeling can be helpful for extracting information regarding regional/nodal inertial constants from steady-state frequency data. However, the loss of ability to accurately capture the variance in power system inertia, together with the odd behavior that can be seen in figure 10, could indicate some relations between frequency behavior and regional inertial constants that the Ornstein-Uhlenbeck process is not capable of capturing. The odd behavior in figure 10 is a large increase in the discrepancy between the estimated and target values, starting in late January and persisting to April; this was not investigated further due to time constraints.

To examine the diffusion ability to capture relevant information from the features, a residual analysis was done. The analysis involved examining the distribution, variance, and independence of the residuals in relation to themselves and to the target variable. Looking at figure 14, the diffusion model displays variance that seems constant with respect to the target variable. The residuals also display no pattern against the targets, indicating that the residuals are independent of the target values. Figure 15 shows

that the residuals have no pattern that is dependent on previous or following residuals; this indicates that the residuals are independent of other residuals. Reisudals that are both independent with constant variance indicate that the linear model with regional inertia constants estimated from steady-state frequency, outlined in section 4.2, captures relevant relations between production, inertia constants, and stored kinetic energy within the power system available in the data. Furthermore, the analysis showed that the diffusion model tended to favor underestimations for periods of high kinetic energy, like January, and overestimation for periods with high amounts of kinetic energy, like May and August. A possible conjecture for this behavior is the reverting to the mean nature of the Ornstein-Uhlenbeck process, as the process describes a system that deviates randomly from a mean value and has a linear drift back the mean value (2.17), possibly resulting in the model leaning towards estimating the mean value. When measuring this against the need for better estimations of stored kinetic energy for more optimal operation of the power system [8], this behavior is not favorable. This is because an overestimation for a period with low stored kinetic energy can result in an operational state where the power system is not working at optimal levels but outside safe operational boundaries.

To determine whether the model should be fitted yearly or a single fit over all the data, 2020–2023, a residual analysis for 2020–2023 was done. Given that the residuals for the yearly fit display close to normally distributed residuals for the years, shown in figure 12, and displayed fairly similar properties, shown in table 13, a yearly fitting of equation (4.11) was chosen over a global fitting for over all the years. The residuals of the global fitting had a vastly different distribution to the yearly fit, shown in figure 13. However, a more interesting finding from the analysis regarding the choice of fitting approach can be seen in table 12. Namely, the value of C is similar for two consecutive years, 2020 and 2021. Then, it changes to a new value for 2022 and remains similar for the consecutive year 2023. The fact that the parameter is similar in pairs of years but changes significantly from 2021 to 2022 could indicate that the 'constant' is affected by an underlying change in a dynamic process within the power system.

Moving to the LSTM model, it was trained on data from 2021–2023 and then tested on data from 2020. A MAPE, MAE, RMSE, and R2 score of 3.1%, 5.6 GW·s, 7.1 GW·s and 0.92 for the year 2020 indicating that the model could extract the relationships between the input features and the output label, the stored kinetic energy in the power system. The model performed best for the month of January and somewhat worse in May and August. Relationships between features and labels may be more straightforward for January compared to relationships in May and August. By the same reasoning as for the diffusion model, it is most important to capture the relations in the periods of low kinetic energy stored in the power system, like August and May. The residual analysis for the LSTM shows that the model tends to underestimate the stored kinetic energy for the periods where the actual stored kinetic energy is low. Underestimations during low inertia periods are beneficial for a method created to estimate stored kinetic energy in a power system. Suppose the estimated value is less than the actual value. In that case, any actions taken based on the estimation will still result in the safe and efficient operation of the power system, given its actual state. This would allow for a more optimal operation of the power system. The LSTM model, like other machine learning models, can easily add additional features to the model without requiring significant changes in the model architecture. This makes it possible to add different data that could contain relations to the amount of inertia in the power system, but where the relations in the data are hard to find with classical methods, such as weather patterns. Compared to the diffusion model's results for 2020, the LSTM model outperforms the diffusion model in all the metrics tested. This also remained the case for all the individual moths tested. The LSTM model also showed a tendency to underestimate inertia for periods where the inertia in the power system is low; this is favorable to the diffusion model's tendency to overestimate the inertia in the power system for the same periods.

The drift model does not seem to capture much of the relations available in the data. Examining figure 8 and the evaluation metrics in table 8, the drift models prove no ability to predict the kinetic energy stored in the data from the available parameters. The most probable cause of this is the assumption that power dampening, D, in equation (2.17) is not valid for the Nordic power system.

6.2 Limitation

One of the significant limitations of inertia estimations, in general, is the lack of a proper ground truth. All the metrics used to evaluate the models show how well the model can estimate the TSO's estimation of stored kinetic energy. However, the TSO estimation does not necessarily depict the actual amount of kinetic energy stored within the power system. For instance, the TSO model does not consider synchronous consumption of electrical power [5], which also contributes to the overall stored kinetic energy in a power system. When using the TSO's estimation as the ground truth, any model that depicts kinetic energy stored in the power system more realistically would perform worse according to the metrics. Creating ambiguity around the errors shown by the evaluation metrics; the errors may result from an inability to capture relevant information, or the error could result from the model capturing more relevant information than is done in the TSO estimation, causing it to better represent the true state of inertia in the power system.

The use of the generic constant H to examine the behaviour of the linear equation (see equation (4.1)) introduces a limitation regarding the accuracy of the H constants shown in table 6 for the years analyzed. The constants in table 6 represent the average H constant of the generators that make up the different production types shown in table 6 and are gathered from the ENTSO-E report [5] from 2015. Given that the Nordic TSOs chose the average H values from the data available to them, it is assumed that they are representative of the production types for the year they were created. However, changes to the power system after 2015 could impact the true average value of the H constant for the different production types.

Further limitations arise from the different time resolutions that can be seen in table 1, where the frequency data has a much finer time resolution than the other datasets. The reduction in time resolution from 0.1 s to 1 hour does entail the loss of information that could contain information about the regional inertia constant. A possible remedy for this would be to increase the time resolution for the production data. An increased time resolution on production data would allow for a smaller reduction in the time resolution of the frequency data, retaining more of the information available in the dataset. Another limitation of the method lies within the stochastic model used when estimating the hourly regional inertial constants. This equation relies on several assumptions outlined in section 4.2. Based on the general performance of the model on the two years tested and the residual distribution of the modeling errors for the other years, shown in the result section, these assumptions seem reasonable for the operational years 2020–2023. However, this does not necessarily generalize outside the years being analyzed. The same is true for the scale of the system being analyzed. The methods used for the analysis are applied at the transmission grid level, and the results might not generalize to smaller-scale power systems, such as the regional power grid. Additionally, any conclusions around applying stochastic modeling in extracting inertial information from frequency and the further estimations of kinetic energy stored in the power system are only valid for the Nordic power system. The Nordic power system may be ideal for modeling frequency dynamics in the power system with stochastic processes, and similar behavior might not be the case for other power systems.

There are also limitations regarding the change in the value of the C parameter from 2021 to 2022.

This could, as mentioned, point to the fact that the C parameter is influenced by a power system dynamic not captured by the Ornstein-Uhlenbeck equation used to model frequency dynamics. Since the Ornstein-Uhlenbeck equation is a stochastic differential equation in time, a model based on this equation would not capture a deterministic process impacting the parameter C. While the residual histograms of all the years, figure 12, indicate that a constant C works on the timescale of one year, there is a potential that the change in C from 2021 to 2022 is a result of a change in a deterministic process. A slowly or rarely changing deterministic process affecting frequency behavior could change the best fit for the C parameter from 2021 to 2022. This could be one of the potential causes of deviation seen in figure 17 from late January to April. This is, however, not the only limitation of using the Ornstein-Uhlenbeck equation to model frequency behavior. If the change in the parameter C results from a change in a non-time-dependent stochastic process from 2021 to 2022, the Ornstein-Uhlenbeck equation could not model this change in behavior. Since the Ornstein-Uhlenbeck is a stochastic differential equation with respect to time, the equation can only be used to model time-dependent stochastic processes, and the model will not capture any information regarding non-time-dependent stochastic processes within the frequency deviations.

A possible limitation of inertial estimation with machine learning models is the lack of training data. Energinet has only recorded the stored kinetic energy in the Nordic power system from the middle of 2019 and onwards. The low amount of available data to train the model makes it harder to ensure that the LSTM model generalizes to new unseen data. This would become even more prevalent if relations between the explanatory variables and stored kinetic energy become more complex when more IBRE is introduced in the power system. There is also the potential that the added computational complexity of using an LSTM instead of other machine learning models is not worth the LSTM model's ability to capture long-term dependencies. Even though the LSTM model shows a good ability to capture the available information from the data, it is hard to evaluate how much of this is due to the LSTM's capability of handling both non-linear relations and long-term dependencies or if any machine learning model capable of handling sequential information would perform similarly, with a lower computation cost. Although incorporating extra features into an LSTM model is straightforward in terms of architecture and computational demands, this comes at the expense of reduced explainability, primarily because each additional feature adds complexity to the model's decision-making process, making it harder to trace how inputs influence outputs. So, even though adding more features can enhance the model's performance, it becomes hard to determine whether this improvement is due to the model discovering new relationships in the data or simply overfitting the training set's peculiarities.

6.3 Implication and recommendations

The promising results regarding estimating regional inertial constants from frequency using the Ornstein-Uhlenbeck process to model frequency behavior could have several advantages for the Nordic TSOs compared to the methods mentioned in the introduction [10, 5, 8]. The diffusion model can estimate a regional inertia constant by analyzing steady-state frequency measurements. This could, in theory, allow for a real-time estimation of regional inertial constants that includes inertial contributions from synchronous loads. The new method for estimating regional inertial constants could allow the Nordic TSOs to better understand how the kinetic energy contribution is distributed and how it changes in real-time for the Nordic power system. A real-time estimation of the regional stored kinetic energy within the power system would be a valuable tool for operating the power system at higher capacities in order to meet the increasing demand for electrical power that is expected in the near future [3]. The stochastic estimation of inertia constant from frequency also requires less data when compared to the TSO method, as described in report [5]. This method utilizes all available information on the inertial constants of the synchronously

connected generators. With this method, estimating regional inertia constants for large power systems requires extensive knowledge of the inertial constants of all synchronous generators, leading to the issue of having up-to-date knowledge of as many generators as possible. The diffusion model only needs frequency data to create a regional inertial constant for the region, reducing the amount of accurate data required for the estimation. This method also allows for a finer time resolution for the estimation of regional inertial constant, it is possible to estimate an inertia constant for every time step making up the frequency data. The limiting factor in time resolution for the inertia estimation results from the resolution of the production data, necessitating the use of stochastic modeling with the Ornstein-Uhlenbeck process. Some practical applications of regional estimation of inertial constants depend on whether or not the results found in this thesis are generalized to smaller-scale power systems. Therefore, further work should include applying stochastic modeling to smaller-scale power systems, such as regional distribution grids, to see if the methods yield similar results for smaller power systems as for the transmission grid.

Following the same reasoning as for the Nordic power system, the method of estimating inertia constants would be useful for other power systems outside the Nordic region if stochastic modeling yields similar results for these power systems. In order to examine whether or not the findings generalize to other power systems, it would be beneficial to do a similar analysis for more power systems in Europe. When considering such an analysis, three power systems initially stand out: the British, Irish, and European. The British power system has a similar scale as the Nordic power system but differs in terms of what production types provided the majority of electrical power [33]. The Irish power system is smaller than the Nordic power system but still generates much of its power from synchronous sources in hydro and thermal power generation [22]. This power system could be used to see if similar stochastic models provide similar results to smaller transmission grids and to investigate further if stochastic frequency analysis could be a tool for other power systems. Finally, the European power system is of a much larger scale and has much more varied production than the Nordic power system. It could be used to examine if the finding of this thesis generalizes to a larger-scale power system, where power is produced from several different countries containing vastly different production types.

Moving to the potential of the two best-performing models proposed in this thesis. A comparison of the LSTM and diffusion models indicates different strengths and weaknesses. As previously stated, the LSTM model provided estimations that, in all cases, were closer to the TSO estimation than the estimations made by the diffusion model. The residuals of the LSTM model also indicate that the LSTM model has a tendency to underestimate inertia for periods where inertia in the power system is low. Both the better performance and the more favorable direction of the error for the LSTM model give it more merit as a potential tool for Nordic TSOs to estimate the inertia in the Nordic power system than the diffusion model. However, this model does not give further insight into which features are important for the estimation process, leading to a model that gives accurate estimation but no insight that can be used to create a power system that is more resilient to frequency deviation in low inertia periods. This is where continuing the development of stochastic models like the diffusion model proposed in this thesis can be beneficial. It could be modified to increase the accuracy of the regional inertial constants estimated from the stochastic model.

The modification could involve including more time-dependent variables in the stochastic model, making it into a higher-order stochastic differential equation. This could potentially resolve periods where the diffusion model loses its ability to capture the variance of the data. Alternatively, integrating non-time-dependent processes that change with respect to another variable, making it a partial stochastic differential equation, could allow for the modeling of frequency behavior that is not time-dependent. Or a combination of the two modifications, resulting in a higher-order stochastic partial differential equation. If the modifications provide better results, it would be because they allow the stochastic model to more closely model the actual behavior of the power grid's frequency. More work is needed to understand why the model loses its ability to capture the relations in the data for May 2022 and what dynamics govern the C parameter from equation (4.11). Both the sudden change of this parameter seen in table 12 and the model's inability to capture variance for May 2022 could result from changes in a deterministic or stochastic process not accounted for by the model outlined in section 4.2, that a different stochastic model might be able to capture.

7 Conclusion

To conclude, objective (1) was explored by proposing the diffusion model outlined in section 4.2. The results yielded by the diffusion model indicate that stochastic modeling of power-grid frequency can extract useful information regarding regional inertial constants from noise-based deviations of steady-state frequency by modeling the frequency behavior as an Ornstein-Uhlenbeck process. Objective (2) of the thesis was explored with the LSTM model outlined in section 4.3. The LSTM model showed great ability to capture relevant relations between features of the training data and the label, stored kinetic energy in the Nordic power system. The LSTM model also showed some beneficial behavior as it tended to underestimate the stored kinetic energy in the Nordic power system in periods when this value was lower than average.

The analysis performed for this thesis focuses on estimating stored kinetic energy for the Nordic power system since this region already has historical estimations of stored kinetic energy using other methods. This limits the generalizability of the findings to this region, and further research on different power systems would be needed to validate the models and ensure their effectiveness across various power systems.

Based on the diffusion model's results, power systems operators should consider investigating similar methods of estimating regional inertial constants in real time for their respective power systems. If similar models based on stochastic power-grid frequency analysis yield similar results for other power systems, this method could depict the state of stored kinetic energy in a more realistic way than the methods outlined in other literature [5, 8]. Successfully implementing LSTM or stochastic models to improve the current techniques for real-time estimation of stored kinetic energy could result in new tools, allowing TSOs to operate their power systems more optimally and efficiently than what can be done with the information available today. A more efficient power system could facilitate a more rapid change to a society with a higher degree of electrification for the Nordic region and Europe as a whole. With a high potential to mitigate emissions and decarbonize energy supply chains, the electrification of Europe is one of the considerable hurdles that need to be passed to achieving the sustainability goals set by the United Nations [6] and reach net-zero carbon emission.

References

[1] Norges Vassdrags- og Energidirektorat (NVE). Overview of Norway's Electricity History. *Last Accessed: 2024-04-16*, URL: https://publikasjoner.nve.no/rapport/2017/rapport2017_15.pdf.

[2] J. Clauß, S. Stinner, C. Solli, K. B. Lindberg, H. Madsen, and L. Georges. Evaluation Method for the Hourly Average CO_{2eq}. Intensity of the Electricity Mix and Its Application to the Demand Response of Residential Heating. *Energies* **12**, 1345, 2019. DOI: 10.3390/en12071345.

[3] Stattnett. Langsiktig markedsanalyse. Last Accessed: 2024-04-16, 2022. URL: https://www.statnett.no/globalassets/for-aktorer-i-kraftsystemet/planer-og-analyser/lma/langsiktig-markedsanalyse-2022-2050.pdf.

[4] F. Trengereid, K. Brekke, and S. Parelius. Leveringskvalitet i kraftsystemet. Last Accessed: 2024-04-16, 2004. URL: https://publikasjoner.nve.no/dokument/2004/dokument2004_03.pdf.

[5] E. Ørum, M. Kuivaniemi, M. Laasonen, A. I. Bruseth, E. A. Jansson, A. Danell, K. Elkington, and N. Modig. Future System Inertia. *Last Accessed: 2024-04-16*, 2015. URL: https://eepublicdownloads.entsoe.eu/clean-documents/Publications/SOC/Nordic/Nordic_report_Future_System_Inertia.pdf.

[6] United Nations. Sustainable Development. *Last Accessed: 2024-04-16*, URL: https://sdgs.un.org/goals.

[7] International Energy Agency. Electrification. *Last Accessed: 2024-04-26*, URL: https://www.iea.org/energy-system/electricity/electrification.

[8] E. Heylen, F. Teng, and G. Strbac. Challenges and opportunities of inertia estimation and forecasting in low-inertia power systems. *Renewable and Sustainable Energy Reviews* **147**, 111176, 2021. DOI: 10.1016/j.rser.2021.111176.

[9] A. Ulbig, T. S. Borsche, and G. Andersson. Impact of Low Rotational Inertia on Power System Stability and Operation. *IFAC Proceedings Volumes* **47**, 7290–7297, 2014. DOI: 10.3182/20140824-6-ZA-1003.02615.

[10] L. P. Andersen. *Applying a Top-Down Estimation of Inertia in the Nordic Power System*. Master's thesis. Norwegian University of Life Sciences, 2022. URL: https://hdl.handle.net/11250/2993969.

[11] L. Rydin Gorjão, M. Anvari, H. Kantz, C. Beck, D. Witthaut, M. Timme, and B. Schäfer. Data-Driven Model of the Power-Grid Frequency Dynamics. *IEEE Access* **8**, 43082–43097, 2020. DOI: 10.1109/ACCESS.2020.2967834.

[12] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation* **9**, 1735–1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.

[13] W. Demloj. *Electricity Price Forecasting using Multivariate Price Time Series*. Master's thesis. Oslo Metropolitan University, 2023. URL: https://hdl.handle.net/11250/3100586.

[14] D. J. Glover, J. T. Overbye, S. M. Sarma, and B. B. Adam. Power System Analysis & Design. 7th ed. Cengage, 2022. ISBN: 978-0-357-67619-6.

[15] H. Risken. The Fokker–Planck Equation. 2nd ed. Springer, 1996. ISBN: 978-3-540-61 530-9.

[16] M. R. R. Tabar. Analysis and Data-Based Reconstruction of Complex Nonlinear Dynamical Systems Processes. 1st ed. Springer, 2019. ISBN: 978-3-030-18471-1.

[17] A. Jung. Machine Learning. 1st ed. Springer Singapore, 2022. ISBN: 978-981-16-8192-9. DOI: 10.1007/978-981-16-8193-6.

[18] S. Raschka and V. Mirjalili. Python Machine Learning. 3rd ed. Packt Publishing, 2019. ISBN: 978-1-78995-575-0.

[19] S. Singh. Understanding the Bias-Variance Tradeoff. *Last Accessed: 2024-04-16*, URL: https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229.

[20] J. Trebbien. *Explainable Artificial Intelligence and Deep Learning for Analysis and Forecasting of Complex Time Series: Applications to Electricity Prices*. Master's thesis. University of Cologne, 2023. URL: https://kups.ub.uni-koeln.de/70766/1/JTrebbien_Thesis_final.pdf.

[21] G. G. Løvås. Statistikk for universiteter og høgskoler. 4th ed., 300–325, 2018. ISBN: 978-82-15-03104-0.

[22] ENTSO-E. Transparency Platform RESTful API - user guide. *Last Accessed: 2024-04-23*, URL: https://transparency.entsoe.eu/content/static_content/Static%20content/web%20api/Guide.html.

[23] Fingrid. Frequency - historical data. *Last Accessed: 2024-04-26*, URL: https://beta-data.fingrid.fi/en/datasets/339.

[24] Energinet. Inertia, Nordic Synchronous Area. *Last Accessed: 2024-04-23*, URL: https://www.energidataservice.dk/tso-electricity/InertiaNordicSyncharea.

[25] J. Pecinovsky. entsoe-py. *Last Accessed: 2024-04-16*, URL: https://github.com/EnergieID/entsoe-py.

[26] L. Rydin Gorjão and F. Meirinhos. kramersmoyal: Kramers–Moyal coefficients for stochastic processes. *Journal of Open Source Software* **4**, 1693, 2019. DOI: 10.21105/joss.01693.

[27] C. R. Harris et al. Array programming with NumPy. *Nature* **585**, 357–362, 2020. DOI: 10.1038/s41586-020-2649-2.

[28] P. Virtanen et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**(3), 261–272, 2020. DOI: 10.1038/s41592-019-0686-2.

[29] T. M. Sandvik. *Area Based Frequency Control in the Nordic Power System*. Master's thesis. Norwegian University of Science and Technology, 2016. URL: http://hdl.handle.net/11250/2616077.

[30] M. Abadi et al. TensorFlow: A system for large-scale machine learning. *2016, Last Accessed: 2024-04-16*, URL: https://arxiv.org/abs/1605.08695.

[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. Scikit-learn: Machine Learning in Python. *2012, Last Accessed: 2024-04-16*, URL: https://arxiv.org/abs/1201.0490.

[32] J. D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90–95, 2007. DOI: 10.1109/MCSE.2007.55.

[33] ENTSO-E. Actual Generation per Production Type. *Last Accessed: 2024-04-26*, URL: https://transparency.entsoe.eu/generation/r2/actualGenerationPerProductionType/show.

A Appendix A: Gitlab

The data and code used to produce the results for this thesis can be found on:
https://gitlab.com/tomazzo/master-oppgave

B Appendix B: LSTM Data

Table B.1: The first five rows of the input data for the LSTM model, transposed.

	Hour 1	Hour 2	Hour 3	Hour 4	Hour 5
Actual Load	9548	9237	9141	9149	9206
Actual Load_NO	16151	15787	15532	15427	15321
Actual Load_SE	14957	14860	14595	14204	14263
Biomass	549	546	547	542	542
D2	0.000002	0.000002	0.000002	0.000002	0.000002
Fossil Hard coal	551	519	515	514	515
Fossil Peat	472	466	466	468	483
Hydro Pumped Storage	309	309	309	309	309
Hydro Run-of-river and poundage	1368	1372	1372	1373	1371
Hydro Run-of-river and poundage_FI	1166	980	929	889	912
Hydro Water Reservoir_NO	13716	13597	13027	12005	11099
Hydro Water Reservoir_SE	4284	3731	3663	3620	3725
Nuclear	7767	7767	7767	7770	7768
Nuclear_FI	2798	2798	2798	2797	2797
Other renewable_FI	25	25	26	26	26
Other_NO	54	54	54	53	54
Other_SE	1128	1128	1116	1103	1094
R2_score	0.972649	0.972649	0.985341	0.956299	0.984712
Waste_FI	7	7	7	7	7
Wind Onshore	1219	1197	1190	1050	916
Wind Onshore_NO	1422	1314	1276	1274	1250
Wind Onshore_SE	4863	4878	5109	5230	5500
oscillation point	49.775360	50.010848	49.994559	50.039888	50.042532

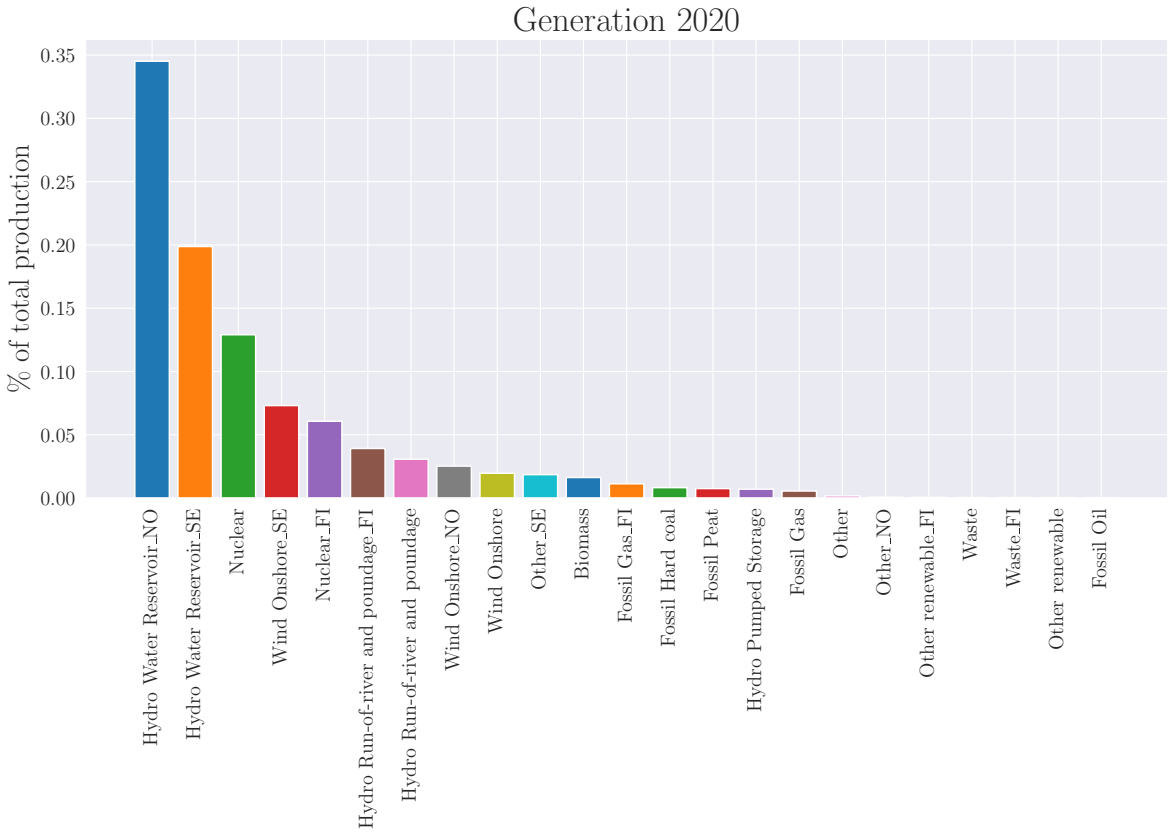

Figure C.1: The relative production of all production types found in the production data for the Nordic Power system for 2020.

C Appendix C: Production data

The Figures C.1–C.4 show the relative production of all production types available at the ENTSO-E transparency platform [33] for the Finish, Norweigan and Swedish power systems.

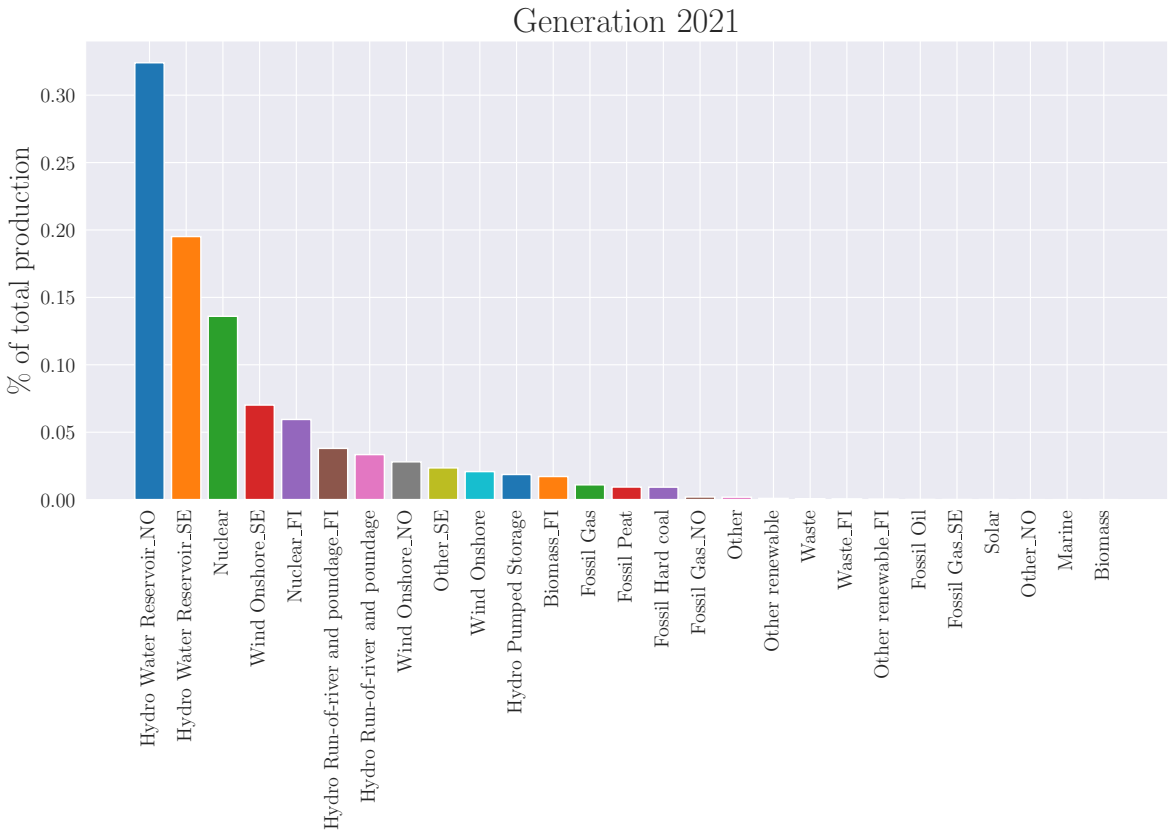

Figure C.2: The relative production of all production types found in the production data for the Nordic Power system for 2021.

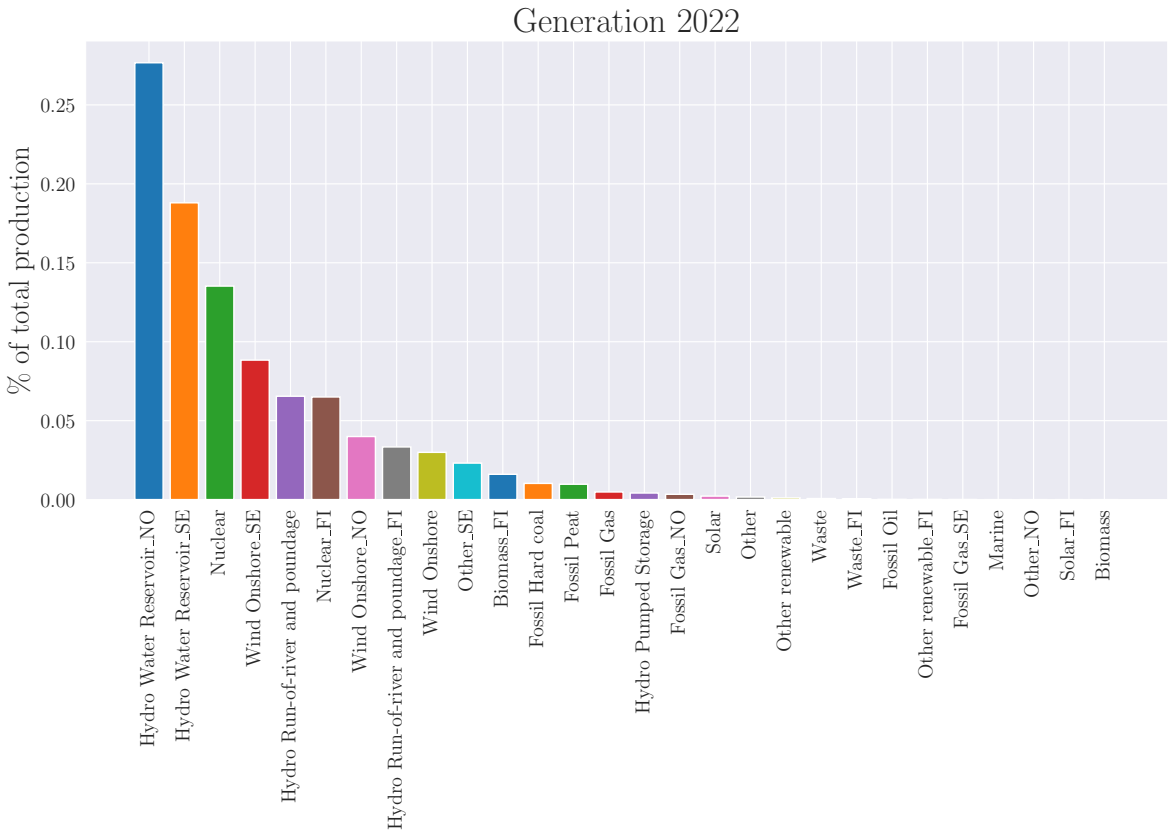

Figure C.3: The relative production of all production types found in the production data for the Nordic Power system for 2022.

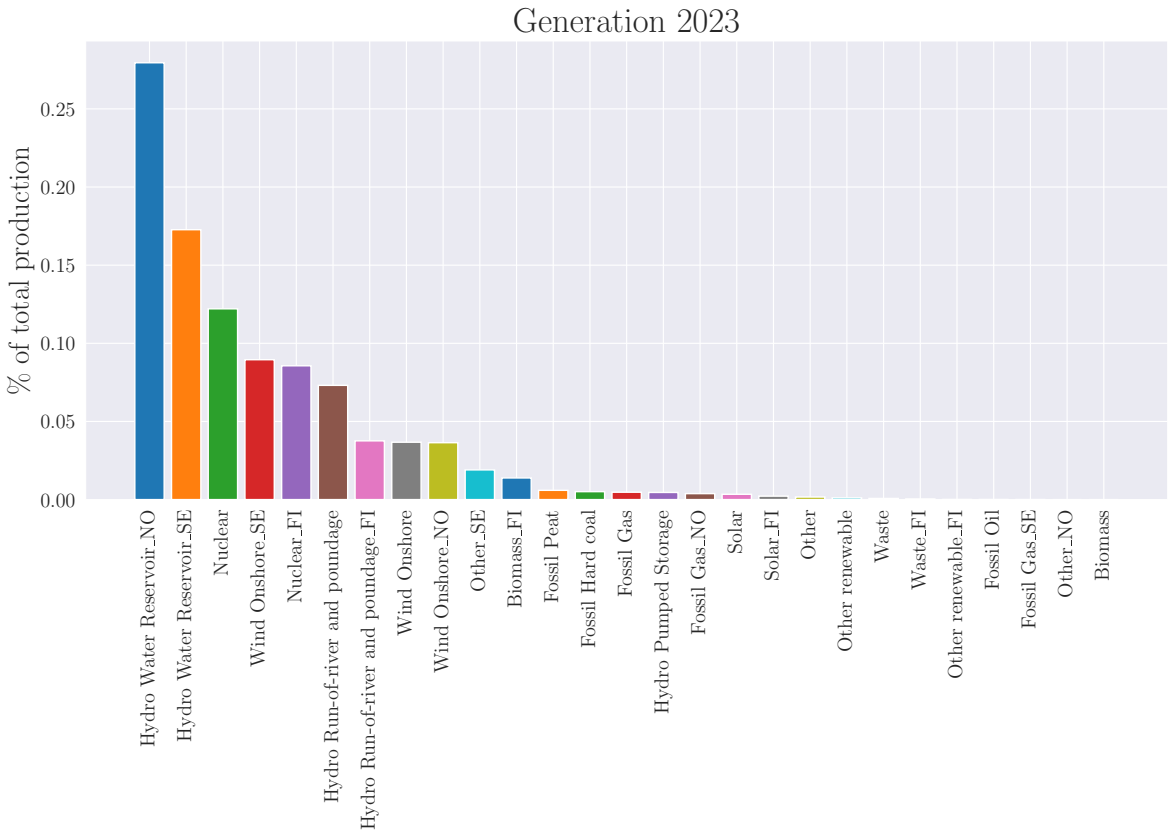

Figure C.4: The relative production of all production types found in the production data for the Nordic Power system for 2023.

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

Postboks 5003
NO-1432 Ås
Norway