Forecasting regulation market balancing volumes from market data and weather data using Deep Learning and Transfer Learning

David Andreas Bordvik

Supervised by
Prof. Kai Olav Ellefsen (UiO)

Dr. Signe Riemer-Sørensen (SINTEF)

Thesis submitted for the degree of
Master in Informatics: Robotics and intelligent systems
60 credits

Department Department of Informatics
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

Spring 2022

Forecasting regulation market balancing volumes from market data and weather data using Deep Learning and Transfer Learning

David Andreas Bordvik

Supervised by
Prof. Kai Olav Ellefsen (UiO)
Dr. Signe Riemer-Sørensen (SINTEF)

Supervised by
Prof. Kai Olav Ellefsen (UiO)
Dr. Signe Riemer-Sørensen (SINTEF)

Forecasting regulation market balancing volumes from market data and weather data using Deep Learning and Transfer Learning

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Abstract

The energy and power sector is a major value contributor to our society and our high living standards. In recent times the power sector has gained increased complexity while undergoing significant changes, with the increased share of renewable production being one of the contributors. An increased portion of renewable contributors in the power mix from, e.g., wind power, results in more volatile power production, increasing the need for grid balancing, making the regulating power market more challenging for power producers to participate in. The purpose of the regulating power market is to compensate the gap between the planned production that has been settled in the day-ahead market and the actual production and demand. The ability to forecast the regulating power volumes and prices some hours in advance of the hour when it is actually traded would enable power producers to balance their positions in the market more optimally. This project exploits historical regulation data together with different market data and weather data to train deep learning models to forecast future regulation volumes. A thorough time-series analysis of regulating power volumes revealed some predictive potential. Furthermore, Bidirectional LSTM showed satisfying results when forecasting up to four hours into the future using data from 2016-to 2021. No previous research was found that uses more than two years of data, no previous research uses recent data, and no previous work has utilized deep learning to forecast the Norwegian regulation market volumes. Additionally, this project did a deep analysis of topographical weather images and transfer learning to evaluate the potential of predicting regulating power volumes using weather images. Different weather forecasts, actual weather, and weather uncertainties were all utilized. The weather data was generally not found to have a considerable direct influence on regulation volumes. However, the weather is considered to have an increasing influence in the future as more volatile renewable power production is expected in the power markets. No previous research has been found to investigate weather images in the context of the regulation market.

Acknowledgement

I would like to thank Prof. Kai Olav Ellefsen and Dr. Signe Riemer-Sørensen for supervising me on this project, I really appreciate the accessibility, and guidance from the both of you. A special thanks to Dr. Signe Riemer-Sørensen for making this project possible and for introducing me to the KoBas project. I would also like to thank the people I have been working with in the KoBas project: Dr. Stian Backe and Dr. Christian Andre Andresen. Thanks for all the meetings and interesting talks.

I am grateful for my girlfriend Elisabeth Olsen. You are the family rock holding things together. I am grateful for my son Leander Elias Olsen-Bordvik. You are my best friend.

This master has been very challenging to complete with corona difficulties, lockdowns, and closed kindergartens. Last but not least I would like to send my thoughts to my sister Amanda, who decided to end her own life during the pandemic. I will always remember you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and motivation

The power markets in Norway and Europe are undergoing changes concerning complexity, technology, and market structures. One of the reasons the electrical power systems and markets are considered complex is because power must be consumed or transported in the same instant as it is produced. Another main reason behind the complexity is that power supply and demand must be instantaneously balanced. Adding to the existing complexity, a shift toward a higher fraction of renewable energy in power production poses challenges due to their volatile nature (in particular solar and wind energy). Furthermore, the power market complexity is increasing due to infrastructure integration among the Nordic countries and Europe[4], where some of the most recent interconnections are between Norway and Germany and Norway and England. Therefore, to tackle the changes happening to the power markets, new data-driven technology is needed.

The Nordic power sector consists of 12 bidding zones, as shown in Figure 1.1, of which the five Norwegian zones are named NO1, NO2, NO3, NO4, and NO5. The zones are designed to take congestion in the transmission grid into account in the market. The price for power within each zone is determined by the balance between demand and supply relative to the transmission constraints[5]. Each zone constitutes several markets; day-ahead (spot), intra-day, and the regulation market.

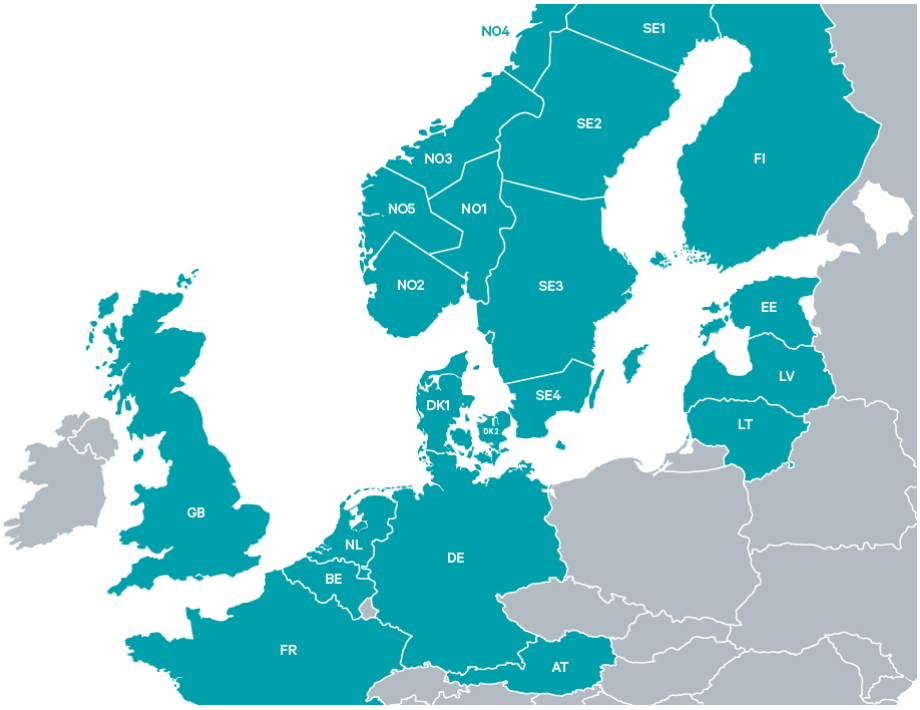


Figure 1.1: Bidding zones for the Nordic energy markets. Image from Nord Pool[1].

In the deregulated Nordic power market, the electricity volumes for all producers and consumers are settled 12-36 hours in advance of the operational hour in the day-ahead market. However, neither consumption nor production can be predicted with 100% certainty. The producers can either balance their position through trade in a continuous intraday market (such as XBID[2]) or settle them in the balancing market [6]. The day-ahead, intra-day, and regulation market are presented in more detail in Subsection 2.1.1, 2.1.2, and 2.1.3. Figure 1.2 shows the Nordic and Norwegian market structure with emphasis on the interaction between grid operation (balancing) and the market operation side.

The ability to forecast the manually activated tertiary reserves (regulation volume) shown in Figure 1.2 some hours ahead of their realisation will provide the power producers with decision support for improved production planning and strategic decisions regarding in which market to settle potential imbalances. With the increase of non-dispatchable renewable sources such as solar and wind energy both in Norway and Europe, as well as increased electrification of society, the weather significantly affects both electricity production and consumption. Continuously updated weather forecasts may contain information that will allow for predicting the regulation volumes some hours ahead but after the settlement of the day-ahead market. This would give the producers a chance of optimising the production to the actual situation.

[2] https://www.epexspot.com/

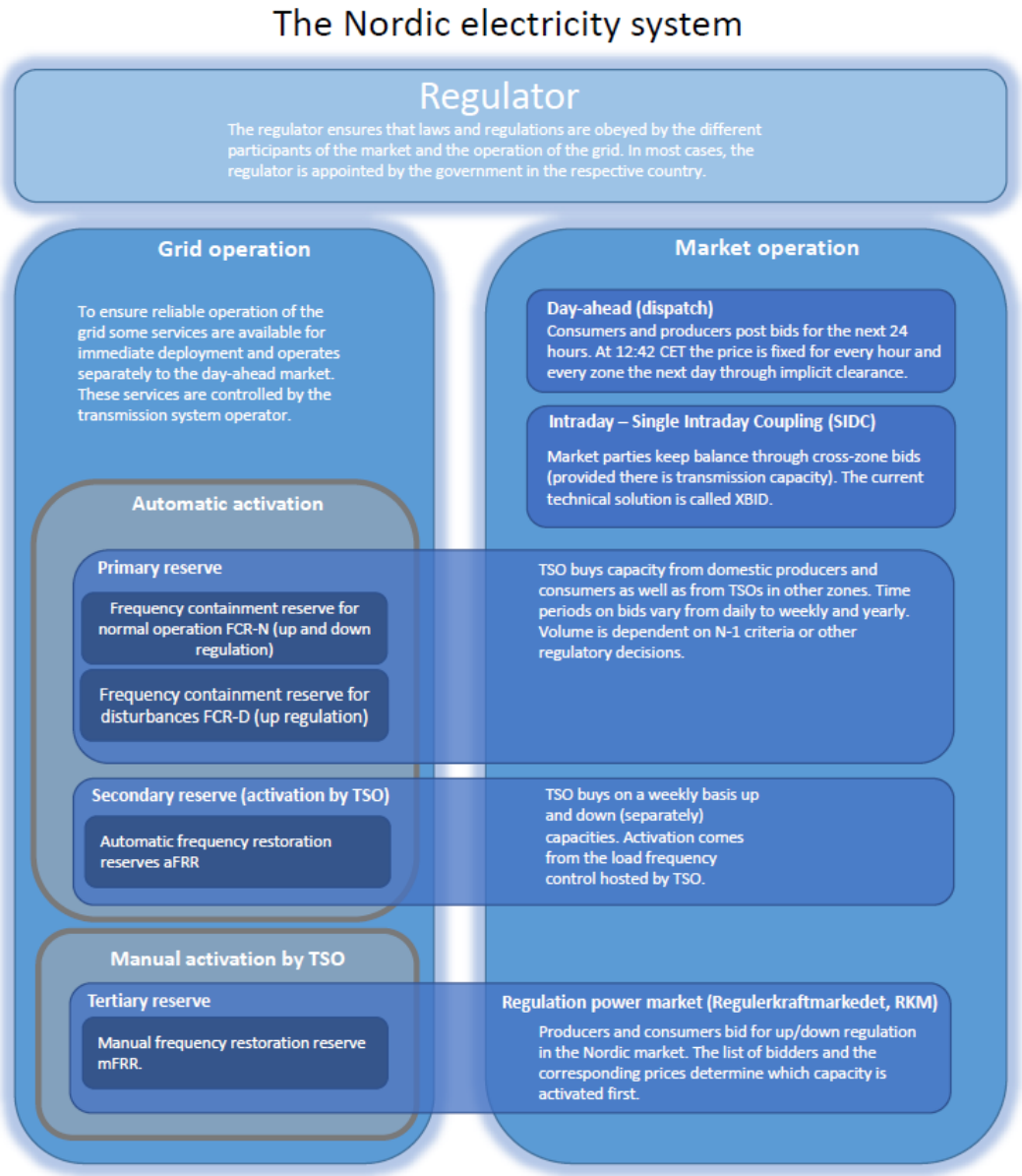


Figure 1.2: Different power markets and interactions with the transmission grid operator. Drawing made by Pål Forr Austnes in the KoBas project 1.1. Permission to reuse the drawing was granted.

The thesis is coupled to the KoBas project described in Section 1.1, and do not cover the entire scope of the project. KoBas aims to develop predictive models for the power regulation market that can aid power producers in better planning their production and trading strategies in an increasingly competitive and challenging regulation market. In general, one can say that detailed power production planning is like guessing the actions of an "invisible hand," with many complex relationships and fields to consider at a micro and macro level. Thus, deviations from plans made in the day-ahead market and the actual hourly condition of the power market are inevitable, which leads to the need for adjustments through grid balancing. Besides complexity in modeling the regulation market volumes, another overall motivation is that very little research have been done on regulation volume prediction, and the power producers who are partners in KoBas do

at the moment not use any models for predicting regulation in their daily operations.

The drivers of the power markets and the volatile nature of weather are significant topics for the thesis. Therefore, both market data and very rich weather data are investigated together with the use of neural networks. Neural networks are powerful functional approximators that can be trained to generalise over complex relationships in the data, making the use of suitable neural network a strong candidate for forecasting future power regulation volumes (timescale of hours). To fully grasp the scope of the problem and challenges faced for the work, it is imperative to investigate and understand the available data describing the energy markets.

The power sector constitutes a multidisciplinary field, and this thesis fall into a intersection between data science, image analysis and computer vision, and energy informatics and power markets shown in the figure 1.3.

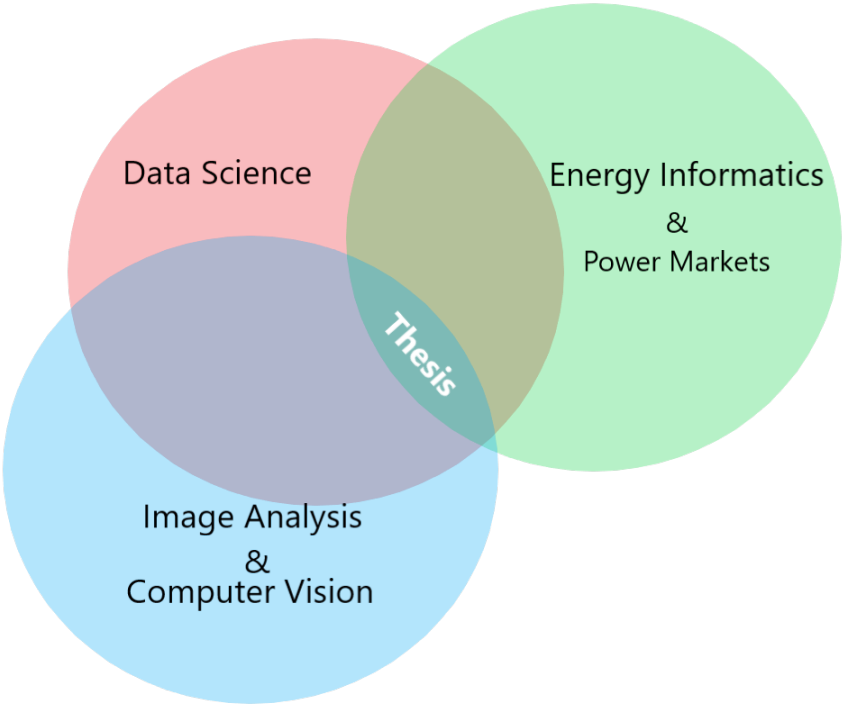


Figure 1.3: Thesis intersection between data science, computer vision and energy informatics

1.1 The KoBas project

The thesis is coupled to the KoBas project[3] funded through the Energy-X programme from the Research Council of Norway [7]. The KoBas project aims to create solutions for forecasting the regulation power market volume and price for the project partners to position themselves better in the regulating power market r[4]. The project is an innovation project lead by Skagerak Kraft AS together with the partners Eviny AS and Equinor

[3] grant number 309315
[4] https://www.sintef.no/en/projects/2020/collective-for-balance-services-kobas/

Energy AS. Several students have been involved in the project together with researchers from SINTEF, jointly working on individual models that can be combined into a real-time decision support system with better predictive powers than single models can achieve.

Some of the ideas in this thesis were motivated by the project and related work done by other project members, but everything presented here is my own implementations unless explicitly otherwise stated. Pål Forr Austnes was an extended summer student primarily working with Urgent Market Messages (UMM). Figure 1.2 was made by Pål and is reused in the thesis with permission. The case study that I performed on the North Sea Link failure in Section 4.37 is influenced by a similar study performed by Pål but fitted into the context of the thesis. Pål did a superficial study of regulation and unprocessed temperature and wind data, whereas I use the full images with both spatial and temporal imformation.

1.2 Research questions

Along with changes happening in the power markets, the power producers have an increasing need for insight into how they can better position themselves in the markets. Unfortunately, the regulating power market is not much researched and is a very stochastic market. The research questions investigated in the thesis are:

- How does the regulation market relate to other markets and are there any significant drivers of activation of tertiary reserves volume?
- Is regulation volume predictable or just stochastic white noise?

- Weather influence on the regulating power markets in Norway
  - Does weather influence the regulating power markets in Norway, and do certain weather conditions drive the need for down or up-regulation?
  - The hourly net or grid balancing is a effect of the current market situation not going as scheduled from the day-ahead market. The plans made during the day-ahead market is based on different types of forecasts. Therefore, it is hypothesised that weather forecast uncertainties influence the need for grid balancing trough the activation of tertiary volumes (regulation volumes).
  - Can information from a domain with images of animals, humans, etc, be transferred to better forecast the power regulation volumes from weather data?

Contributions can be found in Section 3.3.

Chapter 2

Background

This chapter introduces background theory and concepts used in thesis. The chapter also introduces high level background information relating to the Nordic and Norwegian power markets. The machine learning community is a vast and rapidly developing community. Therefore, a considerable amount of information and methodology is communicated through blog posts while other is communicated through more traditional published articles and books. I am aware that published material is preferred due to it being scientifically rooted. However, blog posts and GitHub repositories were frequently used in addition to published material in order to maximise inspiration and knowledge on how to tackle the challenges.

2.1 The Nordic and Norwegian power markets

The wording regulation market and balancing market are used interchangeably. Likewise the regulation volume has many names that are used interchangeably; regulation volume, balancing volume, regulating power, balancing power, and activation of tertiary reserves. Therefore, the thesis adopts the interchangeably usage of the mentioned wordings.

2.1.1 Day-ahead (spot) power market

In todays primary power market, the day-ahead market, the Nordic power producers must deliver bids for how much power they plan to produce hour by hour the following day. This bid must be delivered to the system operator within 12:00 PM the current day, meaning that the bids are based on estimates incorporating uncertainties arising from looking 12-36 hours into the future. Power production from certain renewable energy sources is volatile and can vary significantly. Therefore, uncertainties stemming from a 12-36 hour timespan can be particularly challenging for non-adjustable green and renewable energy producers such as wind and solar. Under and over production from non-adjustable volatile energy sources will affect the entire power marked, and deviation from bids and expected supply of power leads to either financial losses, fluctuations in the power markets, or power blackouts due to power shortage. Considering the financial losses, not delivering on the supply of power placed during the bidding process in the day-ahead market triggers certain penalties for the participant under contract from the bidding process. These risks propose a significant threat to the society as a whole

including; power operators, the power market, and the consumers of power.

2.1.2 Intra-day energy market

When the day-ahead market is closed at 12:00, the intra-day market (e.g. XBID) opens (at 14h) in the Nordic energy marked. In contrast to day-ahead, the intra-day market allows for continuously trading and for trading closer to the period of delivery [8]. Looking specifically at the Nordic power sector, if a trader wants to buy 10 MWh for a 30-minute interval from 18:00-18:30 PM, it needs to find a seller willing to supply the power, and the trade must be closed at 17:00 PM, 60 minutes prior to the delivery[9]. Another difference comparing the intra-day and the day-ahead market is the pricing. Whereas pricing in the day-ahead market is determined by the market-clearing price setting a fixed price for the following day for all transactions within each hour, the intra-day trading[1] of power is based on continuous bilateral agreements between producers. Market participants foremost utilize trading in the intraday market to balance their positions[2] since intra-day trading serves the purpose of minimizing power shortfalls and surpluses that may occur after the day-ahead market is closed. Therefore, the ability to buy power from the market to meet the contract set in the day-ahead market can be financially vital to the participant. The added trading flexibility from intra-day trading is especially suitable for market participants that mainly rely on energy production from renewable sources since it enables them to adjust to sudden changes in weather conditions or other changing circumstances.

2.1.3 Balancing power market

Plans made in the day-ahead market and the activity within the intra-day market both intend to create a balance between consumption and production of power. However, there will undoubtedly be events that upset the balance within a particular operating time, Footnote 2. The Transmission System Operator (TSO) administrates the balancing market in order to maintain a balance between demand and response (and ensure proper frequencies). The balancing market is also market-based in the sense that market participants can offer tertiary reserve power at a given price. However, the only trading partner for a producer is the TSO.

As shown in Figure 1.2 grid balancing consists of three different types; primary, secondary, and tertiary reserves. The primary and secondary reserves relate to balancing the frequency, while the tertiary reserves relate to manually activated power volumes in MegaWatt-hour. The tertiary reserves are activated when the TSO physically calls a power producer participating in the regulation market and asks if they can produce more power than planned with a 15 minutes maximum response time, Footnote 2. On the other side, the producers can receive a call from the TSO asking them not to produce according to schedule but still get paid for not producing. The balancing conducted in the regulation market (a part of the balancing markets) is to some degree to counter unexpected events after the day-ahead market clearance. The events leading to balancing are often short lived and balanced out by the TSO. Therefore, the nature of the regulation market is "short lived", stochastic and can be highly volatile in terms of activated tertiary power volumes.

[1] https://www.next-kraftwerke.com/knowledge/intraday-trading
[2] https://energifaktanorge.no/en/norsk-energiforsyning/kraftmarkedet/

2.2 Machine Learning (ML) - Basic concepts

Artificial Intelligence (AI) has become increasingly popular, and today it plays a significant role in our daily lives whether we notice its present or not. However, AI is a rather vast and abstract term, and a subset of AI that gives it a concrete realization is called machine learning. To get an intuition about what machine learning is and understand what differentiates machine learning from traditional computer-based systems, it is sensible to look at what separates the two at the highest level possible. A computer-based systems brains are constructed by rules to tackle specific situations or problems in traditional programming. Such a systems output consists of answers, targets or solutions given the input data relative to the constructed rules. Therefore, a traditional system cannot autonomously adapt to changing circumstances unless every conceivable circumstance is already incorporated within the rule-based mechanisms. Comparing a traditional computer-based system with Machine Learning, Machine Learning is a computer-based system that can learn rules rather than applying rules to produce results or answers. The primary idea of Machine Learning is to construct a system or a model that can learn a function that maps inputs to outputs. The learning process serves the purpose of extracting and memorize important patterns within the data. Furthermore, the extracted information can be utilized in the decision-making process considering new input data not previously seen by the system. Figure 2.1 below visualize the main differences comparing a traditional computer-based system with a supervised machine learning based system. The illustration above and figure 2.1 draws inspiration from the *datalya.com*[3] data science blog and from the well known American online publishing platform *medium.com*[4].

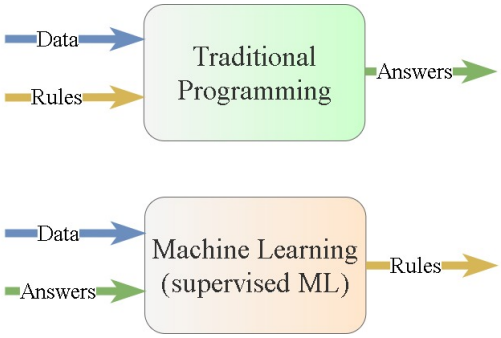


Figure 2.1: Traditional Programming vs Supervised Machine Learning.

For knowing when to benefit from using Machine Learning and utilize advances in data-driven development, two deciding factors should be considered; The complexity of the problem and the need for adaptivity[10]. Furthermore, access to data is crucial, and the quality of training data is the most essential element to consider in machine learning for the data-driven learning process. One of the main challenges in machine learning relates to the quality of the data Geron[2], e.g., if information or phenomenons affecting the outcome(answers) is not incorporated within the input data, the model will not learn the cause or pattern that influences the outcome(answers). The lack of relationships between input data and answers leads the model to learn a random set of rules resulting in arbitrary outcomes.

[3] https://datalya.com/blog/machine-learning/machine-learning-vs-traditional-programming-paradigm
[4] https://medium.com/@amjad.baba913/traditional-programming-vs-machine-learning-11c9abd51928

There are three commonly known subcategories within the machine learning paradigm named; Supervised Learning, Unsupervised Learning, and Reinforcement Learning, Geron[2]. Each sub-category relate to a unique learning process for different types of problems, data, and environments.

2.3 Supervised Learning

Out of the three mentioned sub-categories, supervised learning is the most mature and explored learning method today. The term supervision fits nicely to an analogy where a student(learner) learns while being guided such that the student has access to all answers during learning. This process is often referred to as Inductive Learning where observed truths, examples, solutions, or answers leads to knowledge behind rules, patterns, or functions that can generalize to predict new reasonable answers given a new set of previously unseen data[11]. In other words, examples in Inductive Learning comprise of both the data input x and known targets or answers y. The learning process learns the relationship between x and y and constructs a model that is able to predict y denoted as $\hat{y}$ given a function such that $\hat{y} = f(x)$. However, the input data x and its relationship to target data y is rarely a perfect fit because of noise. Therefore, a more accurately expression of the function including the noise ϵ can be as follows [12]:

$$\hat{y} = f(x) + \epsilon \tag{2.1}$$

Looking further, Supervised Learning is commonly formulated either as a classification or a regression problem. Furthermore, machine learning problems can also be tuned for time-series or time-dependent data to preserve the adjacency of data within different time intervals.

2.3.1 Classification

The problem of classifying or place a name tag for specific situations or conditions occurs relatively often in our daily lives. An example of such a situation is; will the weather be good or bad tomorrow? A similar but more complex classification problem concerning the weather condition for tomorrow could be the following; Will it rain, snow, be cold, be warm, cloudy, or sunny. Another essential application of classification is to predict a medical condition of a patient. All these conditions can be viewed as different classes because they are categorized and separated from each other. The role of a classifier in this regard is to process new information and identify which category it belongs to. The possible categorical outcomes of such a classifier can be either binary, the one or the other class, or it can be of many possible classes[12].

2.3.2 Regression

The idea of regression has been around for a long time, and it has long been used in statistical modelling for estimating quantitative measures such as size, weight, and so forth. Regression is all about numerical values, not categorical as in classification. This results in regression having numbers for the target data (y), and not categories[12].

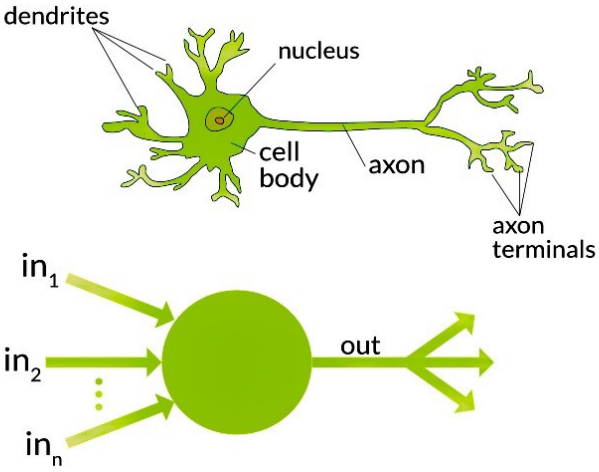


Figure 2.2: A neuron and its connection to deep learning[1]

2.4 Deep Learning

Deep learning is a rapidly evolving field which is a subset of the machine learning field. It shares the same overall idea of learning by analyzing vast quantities of data. The main idea behind deep learning is to mimic the functionality of the human brain. Deep learning is inspired and built on the same concepts as neurons in the human brain. Neurons collect and process electrical signals, and each neuron fires an electrical signal based on how strong its input signals are. The vast network of neurons in the brain work together, forming higher-level actions enabling humans to feel, sense the environment, and control the body. Humans are not born with knowledge, and knowledge must be learned. Therefore, the behavior of the majority of neurons, when it fires and how they relate to each other, must be trained from birth. Luckily, some basic functionality is incorporated within our DNA. In deep learning the neurons are ordered in networks called layers, which is why deep learning is often referred to as a Artificial Neural Network[13][1]. Figure 2.2 is taken from a medium article [1], and it shows the connection between Real-life learning and deep learning. Figure 2.3 illustrates a neural network having two input neurons $\{x_1,x_2\}$ in the first input layer, 5 neurons in the hidden intermediate layer, and 1 neuron $\{y_1\}$ at the output layer. As the signals flow from the input layer to the output layer, they change. A transition from one neuron to another changes the signals strength or signal value according to weights. The weights between neurons represent how strong the bond is between the two, and these weight are trained during training of the network. The resulting signal at the output neuron represents the deep learning models predicted value.

2.4.1 Activation function

Activation functions in neural network mimics the activation of a neuron in the human brain to some degree. The neuron calculates a weighted sum of inputs, adds a bias and decides if it should fire on the output or not[5]. The activation function in a neural network is essential since it enables the information to flow from the input to the output depending on which neurons are fired from the activation functions. This thesis mainly

[5] https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0

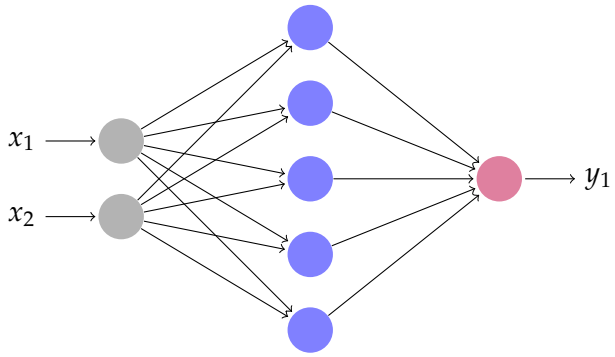


Figure 2.3: Illustration of a artificial neural network with one hidden layer

uses the Rectified Linear Units (ReLu) activation function. This function gives x on the output if x is positive. Otherwise, it gives out zero.

2.4.2 Dropout

Dropout is a commonly used regularization technique in neural networks that have been successfully used to gain better generalization and increased performance [2]. The main idea of dropout is that it gives every neuron a probability of temporally being dropped out in the training step or epoch.

2.5 Convolutional neural networks (CNN) and Transfer Learning

The word convolution is specified by Nielsen [14] as follows: *Convolution means applying a kernel (a matrix) to a larger matrix by sliding it across the larger matrix, forming a new one. Each element of the new matrix is the sum of element-wise multiplication of the kernel and a subsection of the larger matrix*. In the context of convolutional neural networks, they consist of a network of convolutional layers which incorporate several kernels sometimes called filters or neurons. *Convolutional neural networks (CNNs) emerged from the study of the brain's visual cortex, and they have been used in image recognition since the 1980s* [2]. The CNN are know to be good at at extracting features from images that are passed through the convolutional layers. The convolutional operations are extracting features such as edges and other patterns from images[6].

The pre-trained models used with transfer learning in this thesis are CNN models initially trained on a subset from the ImageNet database introduced by Deng et al. [15]. The database incorporates millions of images and 1000 different labeled classes within the 12 following categories: mammal, bird, fish, reptile, amphibian, vehicle, furniture, musical instrument, geological formation, tool, flower, and fruit. Transfer learning works by transferring knowledge (e.g., trained model weights) from one or several source domains to a new target domain, where the knowledge needed in the target domain is different but might be related to the source, Pan et al. [16]. There are several motivations for utilizing transfer learning, and one major benefit relates to how much training data is available. Training, e.g., deep CNN models, generally require a large number of data. However,

[6] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

when common knowledge is transferred from a source to the target domain, the target domain model can focus more on fine-tuning the target problem even with smaller datasets, Pan et al. [16]. This is particularly relevant to the weather forecast, and weather forecast uncertainty images used in this thesis since only data from 2021 were created due to the time-consuming process, Section 4.14 and Subsection 5.4.1. Additionally, the Norwegian Meteorological Institute also deletes old weather forecasts to free up storage space after some time, limiting the available data for download. Another benefit of using transfer learning is that it often leads to faster training convergence since the model is relived from learning parts of the knowledge needed for the domain target. Additionally, knowledge transfer often leads to a better starting point for modeling which again may improve the performance of the final model, Pan et al. [16]. Gao et al. [17] summarize two common strategies that are usually considered for transfer learning; a feature extractor and fine-tuning. For this thesis, only the fine-tuning strategy is considered. Fine-tuning generally involves freezing some pre-trained network parts and training the remaining layers through backpropagation and gradient descent.

2.6 Time series

Time series data is sequenced and time-dependent data. Analysis of the same type of data observed at different timestamps is a major and important field. Data that incorporate historical events which relate across times open up unique possibilities. Time-series data can help us understanding how and why thing have changed. Furthermore, it enable us to forecast the future when knowing how the data will evolve or behave based on deep historical insight [18].

2.6.1 Crucial steps for working with time-series

In the book by Auffarth [19] seven crucial steps for working with time-series are defined: importing the dataset, data cleaning, understanding variables, uncovering relationships between variables, identifying trend and seasonality, prepossessing (including feature engineering), training a machine learning model. All these steps are considered part of a cyclic loop, which the book refers to as *The time-series machine learning flywheel*. The book emphasizes the cycle being the iterative nature of working with time-series, where steps are revisited after new discoveries are made. In the defined process, machine learning modeling count for $\frac{1}{7}$ of the cycle, and $\frac{3}{7}$ of the cycle relate to time-series data analysis as shown in Figure 2.4. Additionally, for this thesis, data download and organization are added prior to the step named *importing the data* resulting in 8 steps in total. These eight steps form the backbone in answering the research questions raised and apply to both the market data and the image data later discussed in this thesis.

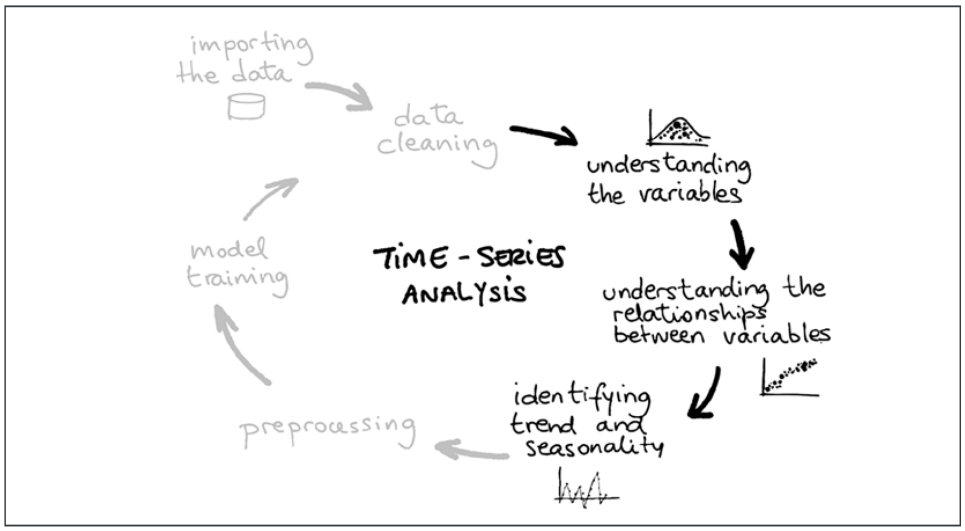


Figure 2.1: The time-series machine learning flywheel

Figure 2.4: The cycle of time-series defined by Auffarth [19] published by Packt[7]

2.6.2 Linear regression and Ridge regression

Both Linear regression and Ridge regression are designed to make predictions by computing a weighted sum of the input features with the inclusion of a bias [2]. A simple linear regression formula is as follows:

$$\hat{y} = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n$$

Where $\hat{y}$ is the predicted value, n is the number of features, x is the input data, and β is the coefficients. The coefficients of the model linear regression model are calculated by the Ordinary Least Squares procedure, which minimizes the sum of the squared residuals. The squared residuals are the difference between predicted results, and the actual targets squared relative to the regression line going through the data. The sum of all squared errors (residuals) together is what the least-squares method minimizes [2]. The Ridge regression is a linear regression with an additional regularization term added to the cost function. The purpose of the regularization is to punish the model weights by forcing them to be as small as possible. This cancels out features that do not influence the outcome of the model to prevent them from introducing more noise [2].

2.6.3 Support Vector Regression (SVR)

The Support Vector Regression model is basically the Support Vector Machine used with discrete values. The SVR model tries to find the best fit hyperplane among the data with the maximum number of points within a given band. This hyperplane reflects the relationship between data [20].

2.6.4 eXtreme Gradient Boosting (XGBOOST)

XGBOOST is a gradient boosting algorithm based on decision trees, famous for its speed and performance. It sequentially adds predictors to an ensemble. The added ensemble performs predictions and stores the residual error. Then a new ensemble is added, correcting the errors of the existing predecessor [2].

2.6.5 Recurrent Neural Network and Long Short-Term Memory

The Recurrent Neural Network (RNN) is known for its ability to model temporal dependencies and behavior in data using hidden internal states. The hidden states act as memory and pass information from previous timesteps forward in time. Through shared weights across all times, the memory becomes universal and the RNN is therefore known as a feedback network. The RNN belongs to a class of artificial neural networks where connections between units form a directed cycle [21]. One significant disadvantage of the vanilla RNN is its inability to capture long-term dependencies for long sequences of data [2]. The Long Short-Term Memory (LSTM) model, which has been frequently used for electricity consumption forecasting [19], extends the memory of the RNN by forget, input and output gates. These gates are optimized during model training, and the model learns the most significant dependencies to memorize. At the same time, the model can forget the lesser important dependencies freeing up memory for the most significant patterns.

Regular RNN and LSTM are only looking at the past and present input data during training. However, in some cases it might be beneficial to describe relationships from a reverse order. This has for example proven beneficial in natural language processing domain where the last part of a sentence can carry a lot of context for the first part. The bidirectional recurrent layer makes it possible to encode a given word with information of the next word ahead coming from also traversing the sentence in revers order [2]. Figure 2.5 shows a high level visualization of a bidirectional recurrent layer. The arrows represents the direction the sequence is processed during training. The hidden states from traversing the input data twice are concatenated.

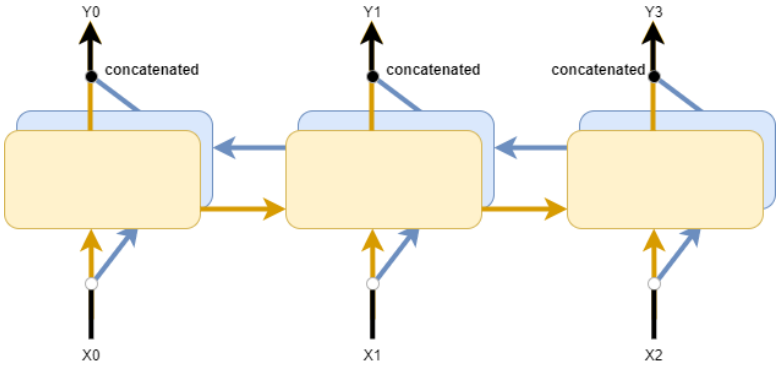


Figure 2.5: Bidirectional recurrent layer inspired by Geron [2] published by O'REILLY.

The bidirectional recurrent neural network was introduced by Schuster et al. [22], and processing sequences in both ways has proven successful for time-series by Siami et al. [23]. They showed that Bidirectional LSTM (BiLSTM) provide better results than regular unidirectional LSTM and ARIMA models on financial time-series data.

2.6.6 ConvLSTM

Convolutional LSTM (ConvLSTM) combines convolution with LSTM for spatial-temporal modeling. The ConvLSTM model consists of recurrent layers just as the regular RNN design. However, the internal weight updates are performed using convolution instead of matrix multiplications. This enables the use of time-series images as input data

over the 1-dimensional feature vector required by the regular LSTM model[8].

2.7 Evaluation metric and loss function

The Mean Squared Error (MSE) is commonly used evaluation metric for regression problems. It measures the *average of the squared difference between the original and predicted values in the data set*[9]. This represents the variance of the residuals and it is formulated as follows:

$$MSE = \frac{1}{n}\sum_{n=1}^{n}\left(y - \hat{y}\right)^2 \tag{2.2}$$

Where n is the number of instances in the dataset, $\hat{y}$ the predicted value, and y is the target. In this project, the MSE is used both as a evluation metric and also as a loss function during model training.

2.8 Image statistics

The theory for this section is collected from the book by Gonzales et at.[24] published at Pearson[10].

2.8.1 Gray Level Co-ocurrence Matrices (GLCM)

First-order statistics are often used in statistical analysis. However, first-order statistics do not describe geometry or context in a data grid or an image. Statistics relative to geometry and context that characterize the texture in an image can be computed as second-order statistics derived from Gray Level Co-occurrence Matrices (GLCM), an abstract image representation. GLCM incorporates the relationship between pixel pairs within a given image. As such, statistics derived from GLCM consider how the gray level pixel at position i relate to the gray level pixel at position j when including the distance d and angle θ between the pixel pairs. The pixel pairs i and j can be similar or different, triggering different statistical properties. Additionally, the magnitude of the pixel difference and their direction also influence the statistics. Extracting the second-order statistics, also referred to as image features, are done by applying different weights on the Gray Level Co-occurrence Matrix, which results in an aggregated weighted sum. More details on how the GLCM is constructed for this project can be found in Section 7.2.3 in Appendix.

2.8.2 Second order image statistics

This subsection lists the different second order image statistics used in this thesis. Detailed information and formulas for the different statistics can be found in Gonzales et at.[24].

[8] https://medium.com/neuronio/an-introduction-to-convlstm-55c9025563a7

[9] https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd

[10] https://www.pearson.com/us/higher-education/program/Gonzalez-Digital-Image-Processing-4th-Edition/PGM241219.html

2.8.2.1 Entropy from GLCM

Entropy is based on the value of the GLCM and it measure of how uniform the texture in a image is. In other words, it measures the randomness of the image textures, Yang et al. [25].

2.8.2.2 Angular Second Moment (ASM) from GLCM

The Angular Second Moment (ASM) feature measures homogeneity from the values in the GLCM. ASM is also knows as energy, Yang et al. [25].

2.8.2.3 Inverse Difference Moment (IDM) from GLCM

The Inverse Difference Moment feature measures local homogeneity of an image from the positions in the GLCM, Yang et al. [25].

2.8.2.4 Contrast / Inertia (INR) from GLCM

Inertia measuring the contrast (local variances in an image) from pixels-pairs from positions in the GLCM, Yang et al. [25].

2.8.2.5 Dissimilarity (DIS) from GLCM

Dissimilarity is a feature that measures the distance between pixel-pairs using positions in the GLCM.

2.8.2.6 Cluster Shade (SHD) from GLCM

Cluster Shade measures the skewness of the GLCM, by utilizing pixel positions in the GLCM. A high value means the image textures are asymmetric, Yang et al. [25].

2.9 Outlier

A threshold for determining outliers can be data points above or below 1.5 times the Interquartile range (IQR). The IQR can be seen as the length of the box in a boxplot. The IQR can also be referred to as the middle 50%. Besides being a less sensitive measure of the spread of a data set, the interquartile range has another important use. Due to its resistance to outliers, the interquartile range helps identify when a value is an outlier. The interquartile range rule is what informs us whether we have a mild or strong outlier[11].

[11] https://towardsdatascience.com/why-1-5-in-iqr-method-of-outlier-detection-5d07fdc82097

Chapter 3

Previous work and Research questions

Research related to modeling the regulation volume in the Norwegian and Nordic region regulating power markets is scarce, with only a handful of papers found. Research on regulation volume outside the Nordic region in Europe is also limited. Research outside the Nordic region is not considered relevant since balancing is rooted in local circumstances. Other European countries have different power sources and demands than the Nordic region and base their plans on different criteria. Additionally, while the Nordic countries are similar, what defines a bidding zone in other European countries differs more. However, five previous research publications are considered related and used as inspiration for the thesis work, of which four target the Norwegian markets and one the Swedish markets. Despite being considered related, their results are not directly comparable with this thesis work as discussed in Section 3.1 to 3.2. In contrast, research topics rooted in the day-ahead and partially the intra-day markets have received more attention [26, 27]. However, *compared to the wholesale electricity market, the balancing power is a considered niche market* [28]. When looking at the regulation markets, the existing research mainly dealt with forecasting using traditional or conventional statistical methods (without considering weather data) and not state-of-the-art deep learning methods. The aforementioned indicate the novelty level in forecasting regulation volumes for the Norwegian regulating power markets, especially when considering usage of advanced deep learning methods.

Previous work is grouped into two sections; Section 3.1 is considered related work, as mentioned above. Section 3.2 considers work that is not directly coupled to forecasting the regulation volumes for the Norwegian or Nordic regulating markets.

3.1 Related work - Norwegian regulating power markets

3.1.1 Skytte 1999

Skytte [6] found patterns in the regulating power market from an economical and business perspective on data from 1996. He established a relationship (correlation) between the day-ahead price and the regulation price, named premium. He also linked the premium to the amount of regulation for down- and up-regulation using a linear

model. This thesis uses the words premium and delta price, Δpr, interchangeably.

3.1.2 Jaehnert et al. 2009

Jaehnert et al. [29] built upon the work of Skytte [6] and proposed a linear statistical model (SARIMA) for modeling the regulation volume thereby extending the linear model by error terms. He modeled the effect of regulating volumes on electricity prices and social welfare using 2007 data. Jaehnert et al. utilized the price difference (delta price) between the regulation price and the day-ahead as follows:

$$\Delta pr = pr_{reg} - pr_{spot}$$

Where pr_{reg} is the regulation price and pr_{spot} is the day-ahead (spot) price.

Jaehnert et al. found a price delta correlation in NO1 with the regulation volume presented as:

$$\varrho(vol_{reg}, \Delta pr) = 0.7811$$

Where ϱ is the correlation and vol_{reg} is the regulation volume. Jaehnert et al. found a high linear correlation between regulation volume and the Δpr, and got relatively good results. However, the experiment was conducted on data from 2007 on bidding zone NO1, which was later split into two zones, NO1 and NO2. Therefore, his results are not directly comparable to this thesis work.

3.1.3 Klæboe et al. 2015

The paper by Klæboe et al. [27] is the most extensive research found covering both regulation price and volume forecasting in the Norwegian regulation market. The research benchmark several time-series based forecasting models for regulation price and volumes in the balancing market such as; different types of autoregressive models, Markov models, and arrival rate models for predicting the balancing state. The main motivation was to create and benchmark models that make forecasts related to market pricing and volume, since producers need good forecasts for balancing market prices to a achieve qualitative decision support when participating in the regulation market. Good bidding strategies rely on forecasts, and for the producers it influences capacity allocations between the day-ahead and balancing market as highlighted in the paper. The cost and need for utilizing the balancing market become even greater given the increased fraction of power from volatile renewable sources[30], which seems to enhance the motivation of studying the balancing market. The data used in the paper are collected from the bidding zone NO2 for the period of 19.07.2010–23.12.2012 and it consists of; balancing volumes (power regulated), overall production volumes(power produced), balancing states, balancing prices, and day-ahead prices. Weather data was not included in their research, and weather data was not publically accessible before 2017.

Klæboe et al. discovered that the correlation found by Jaehnert et al. [29] between delta price and regulation volume had declined from 0.7811 on 2007 data down to 0.47 on the 2010-2012 data. The decreasing correlation is discussed further in the exploratory data analysis chapter 4. None of the model frameworks used in Klæboe et al. leads to high predictive power. Instead they conclude that its is hard to foresee the balancing market before the day-ahead market is closed which was their motivation behind the work. They specifically state the following: *"The balancing market is designed to handle unforeseen*

events and fluctuation, and therefore we are not surprised by concluding that the volume and the premium in the balancing market are random. In fact, it could be interpreted as a sign of an efficient electricity market that it is not possible to predict the balancing market price."[27]. This conclusion leads to the research question 3 on whether the regulation volume is predictable using recent publicly available data or if it just resembles random stochastic white noise .

Lastly, it is not practical to directly compare the results from Klæboe et al. with the results obtain in this thesis. The research takes many statistical and conventional models into account. Due to data availability it is not possible to test the models used in this thesis on the same data range as Klæboe, and as Klæboe et al do not provide their final models, their models cannot be tested on the same data available for this thesis work.

3.1.4 Dimoulkas et al. 2016

The work of Dimoulkas et al. [31] uses a Hidden Markov Model on the Swedish bidding zone SE2 on 2014 market data to forecast both regulation price and volume. They got good results on one hour ahead forecast. However, they did not get good results when forecasting several hours ahead, and they concluded the poor results on the data being quite random without having seasonal patterns. They did not utilize wind, solar, power production, and consumption load but propose them for future work. The results are not directly comparable to this thesis since it is not the Norwegian regulation market. However, the Nordic markets are quite similar, and the way they analyze the data serve as an inspiration for this thesis. They investigated the partial autocorrelation for regulation volume and found the first and second lag significant, as well as significant correlation between regulation volumes and price premiums. The project is conducted on 2014 data and hence outside the data public available.

3.1.5 Salem et al. 2019

The work done by Salem et al. [32] is relevant in general but not directly comparable since the work was conducted with the TSO (Statnett) as a partner that provided the project with data that are not publicly available. They used data with a 5 minute resolution, which is not comparable to the 1 hour resolution used in this thesis. Moreover, the project was conducted using 2015-2016 data which is outside the range of public data used here. The project serves as motivation for this thesis work, and it is the most recent project focusing on forecasting regulation volume imbalances for the Norwegian markets.

The model built, a quantile forest regression model, was performing well, at least on a very short-term horizon. The model performance was good enough to be deployed and used as a prototype model in the Statnett control room, where only Statnett has access to the tool. They trained the model using market prices, planned transmission flows, production plans, and historical imbalances (lags) as features. They excluded weather information because the features were found to not contribute to improving the results. However, they hypothesize that weather may be influential if a model could incorporate consumption forecasts together with temperature forecasts or if the power systems, in general, had a greater share of renewable wind or solar production. Comparing the effect of having a 5 min resolution versus an hour of aggregated values would have been an interesting topic. However, the data and project remains internal to Statnett. In comparison, the KoBas project and this thesis utilize publicly available data with lower

quality to propose a method for forecasting the imbalances for those on "the other side," namely those participating in the regulation markets that the TSO governs.

3.2 Partially relevant previous work - European power markets

3.2.1 Garcia and Kirschen 2006

The work of Garcia and Kirschen [33] is the only found project that uses artificial neural network techniques to forecast the regulation volume. The project used data for England and Wales from 2001-2004. Both the data and its time span make the project not so relevant for a direct comparison. However, they discovered better results using neural networks than conventional forecasting methods when forecasting the regulation volume due to the neural nets ability to capture non-linear relationships between variables and influencing contributors. They used the following models: linear networks (LNs), multilayer perceptron (MLP), radial basis function (RBF), probabilistic neural networks (PNNs), and generalized regression networks (GRNNs). This work motivates the use of deep learning in general when modeling the regulation volume, which makes the paper interesting for this thesis.

3.2.1.1 Hirth et al. 2015

The work of Hirth et al. [28] looks at how variable renewable energy sources (VRE) in Germany, such as wind and solar, and the impact of forecast errors on balancing reserve requirements; the supply of balancing services by VRE generators; and the incentives to improve forecasting provided by imbalance charges. The essence of their finding is that the balancing reserves depend on many factors and that wind and solar power forecast errors are two of several possible drivers. They also mention that other factors are possibly more quantitatively important than VRE in influencing the balancing.

3.2.2 Aggregation of numerical market data

While some existing research focus, e.g., on one particular power plant or wind farm, this thesis intends to utilize spatially and temporally aggregated numerical market data for each bidding zone. Working with aggregated data has been proven beneficial in other problems. In Miettinen and Holttinen [34], aggregation of data is shown to have a significant benefit on minimizing day-ahead forecasting errors when working with large geographical areas. The paper concludes that data aggregation for forecasting power production in the Nordic region significantly reduces large forecasting errors. However, the paper does not address aggregation benefits targeting balancing markets and regards this as future work. Nevertheless, it is regarded as relevant for this thesis regarding data utilization. On the other side, aggregating data may lead to a weaker relationships between different data and features. Losing resolution in the data by aggregation may filter out noise, but it can also make the data less descriptive for solving the problem. The market data relevant for this thesis is mainly available at an hourly resolution. Thus, it is unknown how the temporal aggregation has influenced the potential for forecasting the activation of tertiary volumes.

3.3 Contributions

In this thesis, a thorough data analysis is performed on data from 2016 to 2022. Previous studies focus on 1-2 years of data, this is the first public study analysing and modelling from a longer time period. Other works state that correlations between delta price (price premiums) and regulation volume was decreasing from 2007 to 2010-2012 Klæboe et al. [27]. It is found that the correlation is still present but it is shifted toward a more non-linear relationship in Section 4.12.1. All the related work in Section 3.1 considered and included day-ahead and regulation price in their modeling of regulation volume. In Subsection 4.9.1 it was found that both day-ahead and regulation price for NO1, NO2, and NO5 has become non-stationary from 2020-2022 which reflects resent changes happening in the power markets. In addition to price, the relationship between regulation volumes and the delta between power production and production prognosis (delta production) was studied. No previous work is found analysing this relationship and it was found to be significantly correlated in Section 4.12.1.

We are not aware of any use of deep learning in the literature for Norwegian regulation markets. The observed level of noise in regulation volumes motivates use of deep learning. The observed non-stationary trends of price and the non-linear relationship between delta price and regulation volumes enhance the motivation of using deep learning. Additionally, the observed relationship between regulation volumes and production deltas motivates the use of power production and production prognosis as features in for multivariate modeling. Despite not having groundbreaking predictive powers, the models derived in Section 5.3 are better than the current non-data-driven approach used by the partners of the KoBas project, which base their decisions simply on personal experience.

We did not find any existing study of relevance targeting the relationship between weather and regulation volumes. This motivated a thorough investigation both from an analysis perspective and machine learning to substitute for missing research on the topic. Extreme weather was mainly found impact the regulation volumes in a secondary manner, e.g., from power blackouts and not directly from the weather in Subsection 4.13.1. Good results were achieved when modeling using random shuffled training, validation, and test data. However, the model was not found to generalize when the training, validation, and test data.

Chapter 4

Data and exploratory analysis

4.1 Motivation behind the data and exploratory analysis

The lack of previous research addressing tertiary reserves activation in the Norwegian regulation market motivates a thorough data analysis of regulation volume to understand its nature and relationships with other market data such as; day-ahead price, day-ahead volumes, regulation price, and power production, production prognosis, and weather. Furthermore, the exploratory data analysis addresses some of the research questions raised while also serving as input for further problem-solving by suggesting appropriate data pre-processing and Machine Learning methods. This data and exploratory analysis chapter is written linearly prior to the machine learning experiments in Chapter 5. However, the data analysis and experiments have been a continuous process going back and forth between analysis and machine learning experiments following the flywheel cycle described in the background subsection 2.6.1. The overall guide to working with time-series problems shown in Figure 2.4 ensures a proper time-series problem-solving approach with overall crucial steps intact.

A summary section named **Data analysis and takeaways** (Section 4.15) highlights why machine learning is relevant to utilize on the analyzed data as well as points out relevant experiments and machine learning approaches. Besides an observed high level of noise for regulation volumes, the data is known to be dominated by zero values (no regulation) for all zones. For example, 67.55% of the regulation volume for bidding zone NO1 is 0.0, and a significant portion of the rest is close to zero. Thus, predicting 0.0 for all situations results in a relatively good baseline score (Mean Squared Error score) concerning all zones. This adds to the motivation of reserving significant space in the thesis for analyzing the data. If not stated otherwise, absolute values of correlation coefficients are presented for an easier representation of relationships between variables.

4.2 Numerical data

The dataset with all numerical market data consists of data between 2016-01-01 and 2022-02-06, 53 488 hours of data in total. Around 160 different numerical data types (features) were investigated and tested during experiments. All the different data types are sequenced time-series data. All the data except the data from The Norwegian Energy Regulatory Authority (NVE) is equally spaced with a frequency of one hour between

each observation. The data from NVE (water level in Hydro reservoir) comes sampled at weekly intervals, and the hourly resolution matching the other data was calculated using the number of hours within a week and the slope calculated from the difference between two consecutive weeks. Some of the data in the dataset investigated fits into the following categories: Regulation volume, regulation price, transmission flow and capacity, day-ahead volume, day-ahead price, delta price, production, production prognosis, delta production, consumption, consumption prognosis, delta consumption, the water level in Hydro reservoir, and seasonal features. Not all data were investigated in-depth, and the data exploratory analysis highlights which data the thesis focus on.

Multiple sources were utilized to download numerical power market data; Entsoe[35], NVE[36] and Nord pool[37]. All data is public historical data, and the data was downloaded using an Application Programming Interface (API). A student license was acquired for using the API to download from Nord pool and Entsoe. The NVE API does not require any license. Several Python scripts were made to be able to download data properly. Some of the data was provided by the KoBas project. However, python scripts and API were locally made and maintained throughout the thesis work to be self-sufficient in gathering the data needed at all times.

4.3 Weather forecasts and actual weather data

The Norwegian Meteorological Institute (METNO) provides 2-dimensional data grids of actual weather and different types of weather forecasts. The weather data covering Scandinavia and the Nordic region is categorized into two types of products; forecasts from the MetCoOp Ensemble Prediction System (MEPS) and post-processed products.

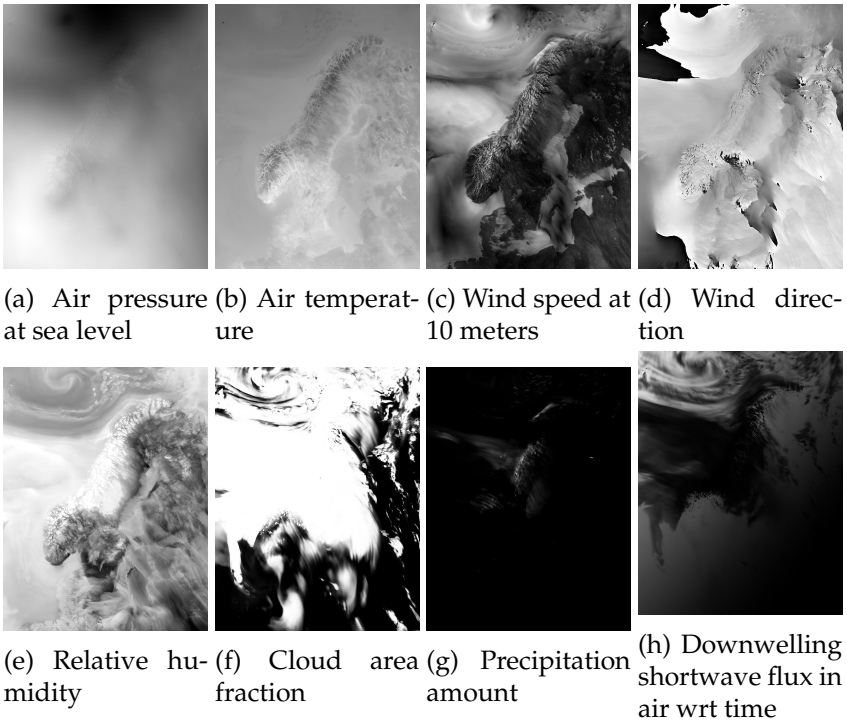
MEPS is a convection-permitting atmosphere ensemble model [1], and the model provide unprocessed forecasts up to 60 hours into the future. The 60-hour forecast is conducted four times every day at 00:00, 06:00, 12:00, and 18:00. The MEPS system commences several forecasts for the same time interval using different forecasting parameters, and in 2020 METNO started using 30 ensemble predictions. The entire Nordic region covered by the MEPS forecast incorporates a grid spacing resolution of 2.5 km pixel-wise[2].

While the ensemble forecasts can be considered raw predictions, the best forecasts are achieved from post processed products that integrate the MEPS outputs. Additionally, the post processed product is infused with measurements from radar, weather stations, and citizen stations, forming a weather grid having a pixel-wise grid spacing resolution of 1 km. The dimensions of a post processed weather image is 2321×1796, and the post processed products use the Lambert conformal conic coordinate reference system. Forecasts from the post processed product are used by *yr.no*, and it is the post processed products that are relevant for this thesis. The post processed product constitute both the MET Nordic Analysis and the MET Nordic Forecast products. The analysis product, referred to as actual weather data in this thesis, represents the weather at a particular hour. The post processed forecast product is the most up-to-date forecast 60 hours into the future[3]. A rich selection of weather data is made publicly available by METNO, and this thesis focus on the following types; air pressure at sea level, air temperature 2 m

[1] https://drive.google.com/file/d/0B-SaEtrDE91WWEJoNkJiUm5TNzg/view?resourcekey=0-PCe4JeYwEiWs8EU2KBGe-w

[2] https://github.com/metno/NWPdocs/wiki

above ground, wind speed 10 m above ground, wind direction, relative humidity, cloud area fraction, accumulated precipitation, down-welling shortwave flux in air with respect to time. Figure 4.1 shows a snapshot of the 8 different types of weather conditions at the date 2019-06-28, 20:00 PM.



(a) Air pressure at sea level
(b) Air temperature
(c) Wind speed at 10 meters
(d) Wind direction
(e) Relative humidity
(f) Cloud area fraction
(g) Precipitation amount
(h) Downwelling shortwave flux in air wrt time

Figure 4.1: 8 different actual weather conditions at 2019-06-28 20PM

Both actual weather data and weather forecasts were downloaded via an Application Programming Interface (API) to the THREDDS Data Server (TDS)[3]. The weather data, both actual weather and forecasts, were downloaded throughout the course of the thesis work and the KoBas project work. Some years were accessible through the *metpp_rerun_v2* server, and others were available through the *metpp_operational* server.

The final dataset of actual weather data consists of hourly data for all aforementioned weather types between 2016-01-01 up to 2022-01-29; 433 643 files and 82GB of data in total.

In contrast to downloading actual weather, the data for weather forecasts are more compound since it does not naturally fit an hour-by-hour download. As mentioned, weather forecasts are performed four times every day. Each downloaded forecast constitutes chunks of 60 hours of future predictions for every weather type, 48 images in total, meaning that there are many overlapping forecasting hours since every forecast stretches far into the future. Every downloaded forecast takes up 2.7GB of storage, which take up 10.8GB of storage per downloaded day and 3942GB for an entire year. It is not feasible to work with that much data, so different scripts were made to structure and save only the images of interest that support specific Machine Learning and data analysis experiments. Therefore, many different datasets of weather forecast images were created,

[3] api.met.no/product/THREDDS

and while subsets of data forecasts were downloaded across several years, the majority of forecasts images were downloaded from the most recent 2021 year.

4.4 Main sources for power production

Before studying price and aggregated volumes within the imbalance or regulation market, it is imperative first to examine what the aggregated power production quantities consolidate in general. While hydro-power has historically been dominating the power production in Norway, new sources and ways of producing power emerge within each bidding zone. The changes to the mix of power production from 2016-2022 can be observed in figure 4.2. The magnitude of hydro-power, including both Hydro Water Reservoir and Run-of-river and poundage, is revealed in both figures 4.2 and 4.3 considering all bidding zones (NO1-NO5). All values at the y-axis are normalized to reflect their portion of total power production. However, both plots squeeze in 53488 hours of data spanning 2016-01-01 up to 2022-02-06, and the plots do not perfectly visualize values fluctuation at a high frequency. Regardless, all values sum up to 1 at each hour, even though the plots can be visually misleading.
It can be observed in figures 4.2 several newcomers from 2021-2022 within power production from; Waste, Hydro Pumped Storage, and other renewables. Another recent change to the energy mix is the dropout of power from Fossil Gas from 2021 for all zones except in NO5. Wind power is showing an increase from 2016-2022 for NO2, NO3, and NO4. NO1 got wind power production in 2019, which seems to be increasing. NO5 does not yet have wind power in its energy mix. Figure 4.3 shows the similar data as in figure 4.2 with the different contributors stacked on top of each other. All bidding zones indicate a reduction in the Hydro Water Reservoir portion from 2016-2022. It can be seen in figure 4.3 that NO1 almost completely replaced Hydro Water Reservoir with the combination of Waste (<1%), Wind Onshore and Hydro Run-of-river and poundage in 2022. Moreover, a spike in Wind Onshore is observed for N04 in early 2022, and both NO2 and NO3 indicate significant wind power contribution to the grid in early 2022. This sudden increase can result from the windy weather conditions observed in early 2022, with storms hitting the country, e.g., the Gyda storm that hit Norway mid to late January. Furthermore, the Hydro Reservoirs levels are historically low. Therefore, the low reservoirs levels and several new energy sources and wind farms may significantly explain the recent rapid changes in the energy mix.

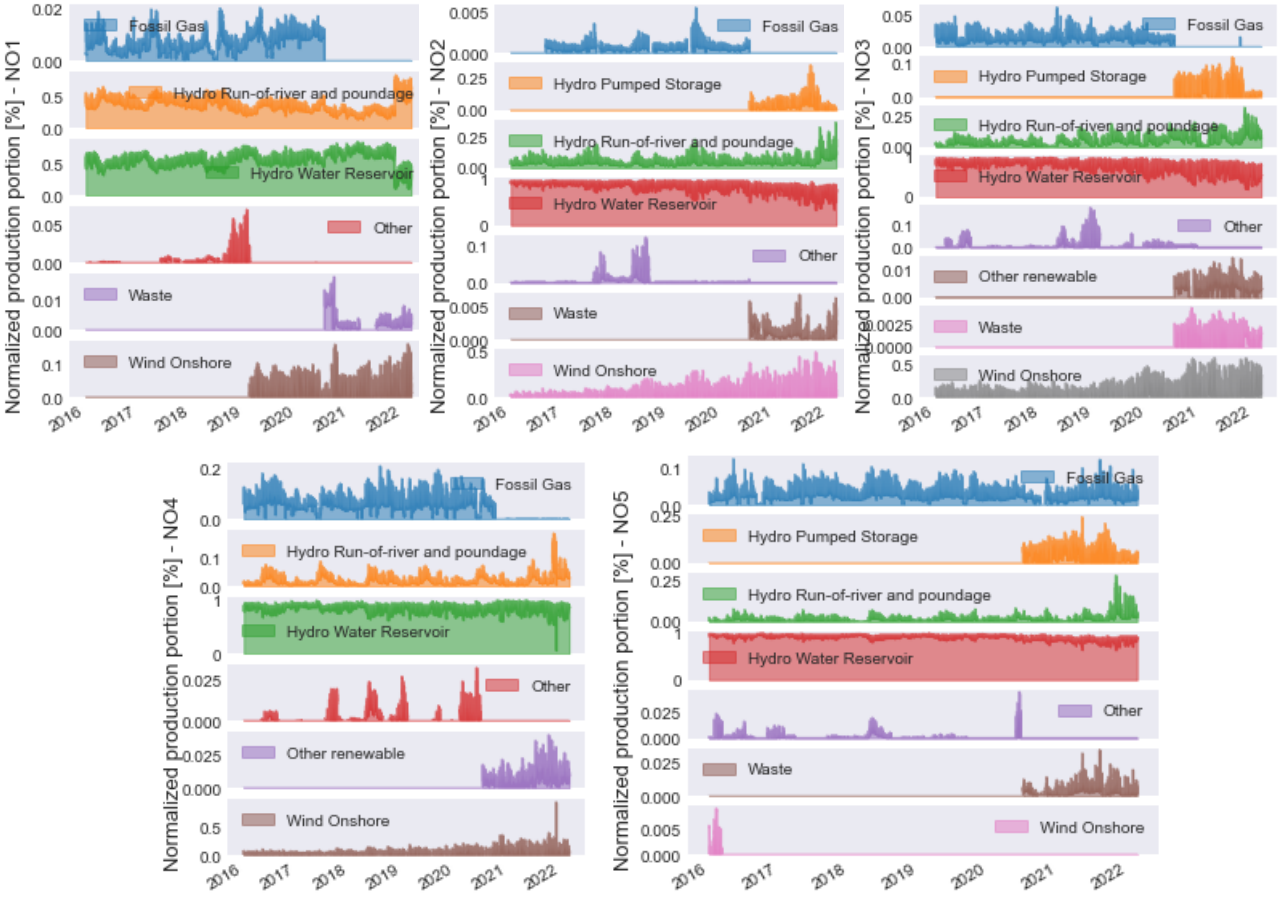


Figure 4.2: Different sources behind power production for NO1-NO5

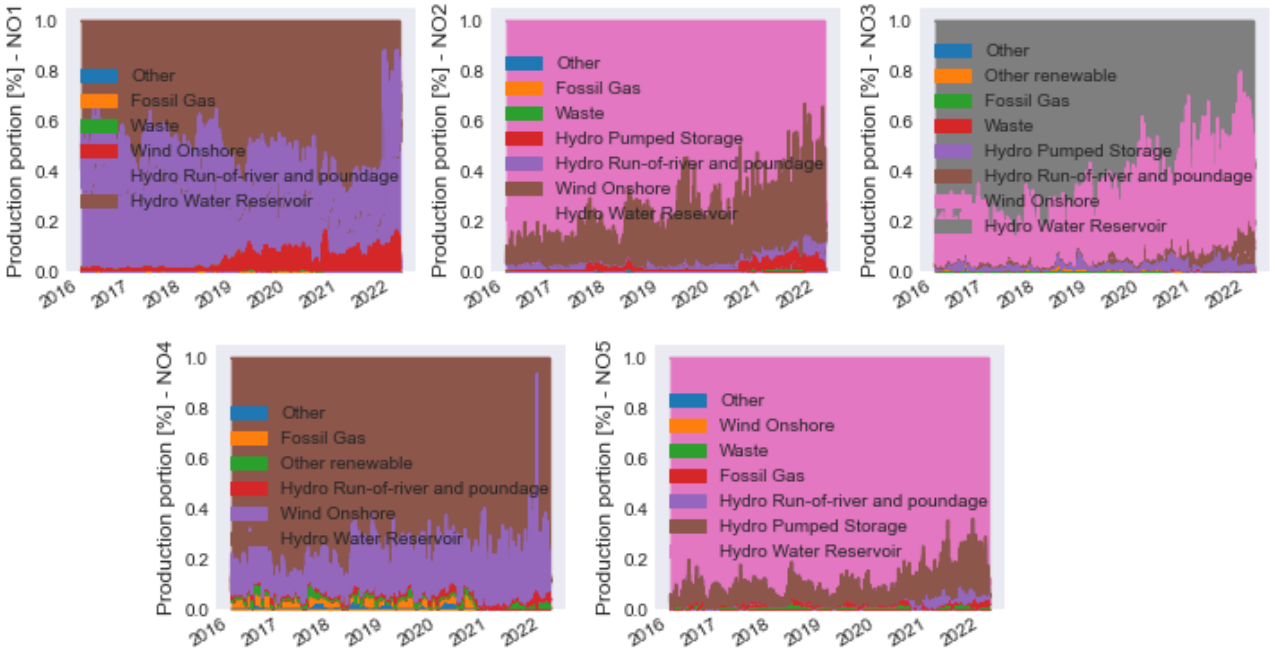


Figure 4.3: Stacked production share for NO1-NO5

4.5 Day-ahead and regulation markets comparisons

4.5.1 Volumes comparison

The day-ahead markets in the Nordic region are the largest in the Nordic region in terms of power volumes. The day-ahead markets constitute the majority of available power traded. Since the day-ahead market is the primary provider of power and electricity for society, it follows the daily needs and trends in society to a much greater extent than the intra-day and regulation markets. The overall nature of power volumes traded and consumed in the day-ahead market is show in figure 4.5. The figure shows the day-ahead volumes, the power bought, in NO5 for 2019[4]. The day-ahead volumes are the aggregated and overall hourly planned and bought power for the entire next day, published to all market participants at the closing of the day-ahead market. Therefore, the day-ahead volume is a different quantity than the aggregated actually produced power from hour-to-hour. The hour-to-hour actual power production is more affiliated to the hourly intra-day and regulation markets. Figure 4.5 clearly shows monthly fluctuations and trends, having more traded and consumed power during winter than summer. The boxplot inside figure 4.5 shows a five[5] number summary of the data relative to days. Potentially outliers are visually excluded from the plot for simplicity. The lower and upper whisker represents the minimum and maximum, while the lower and upper ends of the box represent the lower and upper quartiles. The median is the black horizontal line, and the mean value is plotted additionally as the red dot. The boxplots for each day reveal a slight difference between business days and weekends for traded volumes. The volumes do not seem to differ between business days, but the boxplot indicates slightly less traded volumes for weekends. The hourly day-ahead plot inside figure 4.5 shows less volumes traded for late evening and early morning. The volumes increase from 00:00 and first peek at 06:00 when most people start their day and commencing normal household activities prior to work. The volumes show a slight decrease during normal working hours. A new peek is reached around 18:00, which matches the end of a normal workday and the hour when most people make dinner at home and start to consume power for evening household activities. Both the hourly and the daily trends look to be aligned with our daily routines and lives.

The regulation volume is shown for NO5 the same period as for the day-ahead volume in figure 4.4. The first noticeable difference between the two is that regulation volumes can be negative, whereas day-ahead volumes must be positive at all times. This reflects the urgent balancing purpose of the regulation market, where power can be balanced up if it is below a certain threshold, or it can be balanced down. Down regulation means that the participants in the region are selling reserve power to the TSO (Statnett), which is policing the balancing by using the accessed reserves volumes to balance the grid in another region and market. An aspect of the relationship between the day-ahead and regulation volumes is that the regulation volume reserves are tightly connected to the hourly actual produced power for urgent grid balancing while also being rooted and related to the planned volumes originating from the day-ahead market. In contrast to the day-ahead volumes, the regulation volumes do not reveal similar clear patterns or relationships to society. Visual inspection do not reveal a monthly seasonal pattern, and the data looks like stochastic noise. The hourly regulation volume fluctuations are neither revealing

[4]2019 is considered a normal year for power consumption due to the corona pandemic striking early 2020
[5]The minimum, the first quartile, median, the third quartile, and the maximum

any clear patterns. The boxplot inside figure 4.4 shows that the data has a median at zero and a mean value close to zero. However, the boxplot may indicate a slight difference between business days and weekends with a slightly more positive up-regulation. Possible outliers are also here not included in the boxplot for simplicity. However, comparing the relatively small minimum and maximum values (boxplot whiskers) to the y-axis on the main plot indicates that regulation volumes may be subjected to a large portion of outliers relative to the interquartile range (IQR)[6]. The other regulation and day-ahead sibling markets in NO1, NO2, NO3 and NO4 shares the same characteristics as seen for NO5 in figure 4.4 and 4.5.

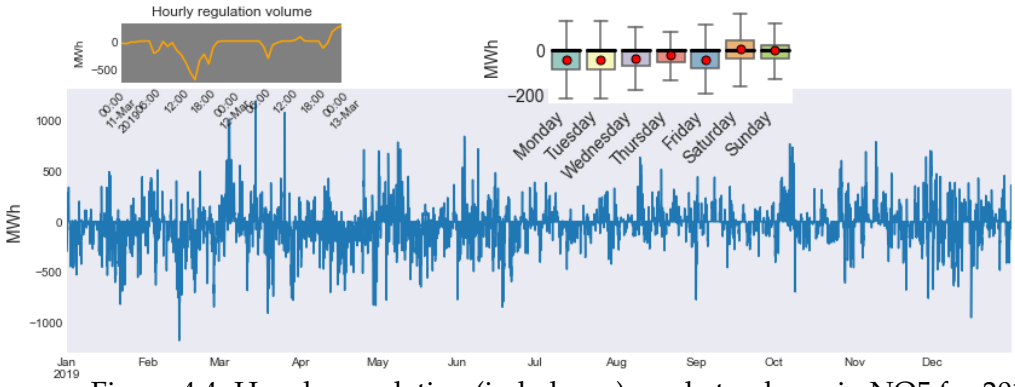


Figure 4.4: Hourly regulation (imbalance) market volume in NO5 for 2019

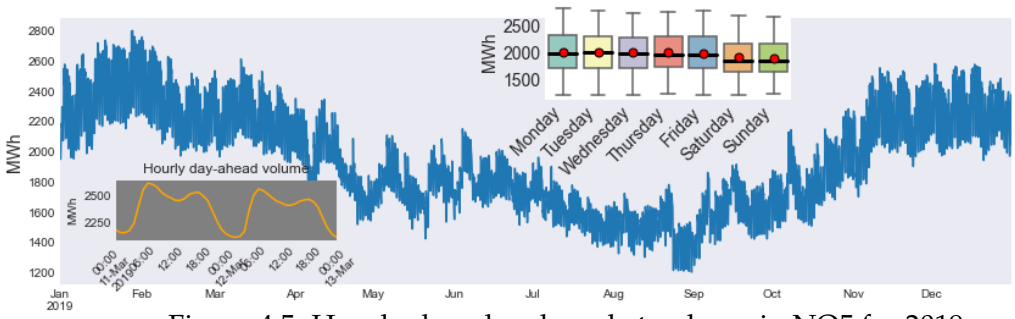


Figure 4.5: Hourly day-ahead market volume in NO5 for 2019

4.5.2 Price comparison and business aspects

A more common signature between the day-ahead and regulation market is revealed when comparing the power price in figure 4.6 and 4.7. Figure 4.6 shows that the price development in the day-ahead market price is fluctuating similar to the day-ahead volumes. The day-ahead price is lower during the summer, and the boxplot indicates a slightly higher power price during business days, and the price peaks at around 06:00 and 18:00 within a day. When looking at the price fluctuation signature for the regulation price in 4.7, it closely resembles the signature of the day-ahead price. However, the regulation price is more spiky, and the spiky characteristics is inherent from the activation of regulation volume reserves. When there is zero regulation, the regulation price is set equal to the day-ahead price. When the need for balancing occurs, the TSO choose which market participants to buy regulation reserves from. The TSO buy regulation reserves based on a ascending price order, first choosing the tertiary reserves having the

[6]Measure of the spread of the data and statistical dispersion. IQR = Q3 – Q1. Outlier if data is outside Q3 + (1.5 × IQR) or Q1–(1.5 × IQR)

cheapest price[38]. Therefor, the market participants use the day-ahead price as baseline for setting the price for regulation reserves. The business mechanisms kick in whether the participants are willing to reserve potential power production for the potential regulation reserves market, and how far from the day-ahead price they are willing to sell their power for. If the participants reserve some potential production for the regulation market and put a too high price on its volume, they may end up not selling their power and losing profit. Contrary, navigating "correctly" relative to the regulation market and other competitors can be profitable for participants in the power sector markets.

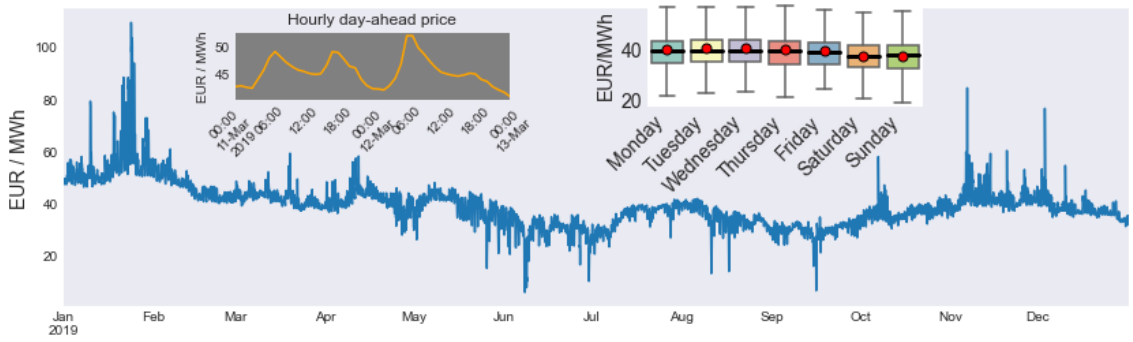


Figure 4.6: Hourly day-ahead market price in NO5 for 2019

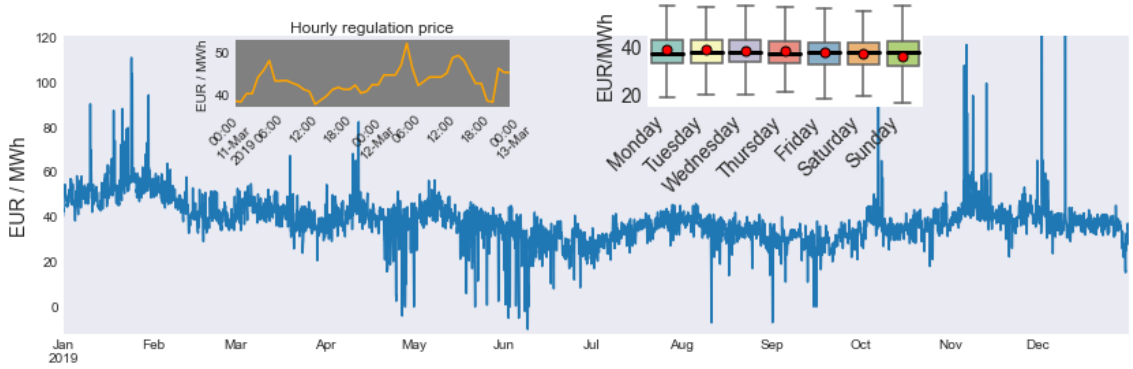


Figure 4.7: Hourly regulation (imbalance) market price in NO5 for 2019

4.6 Regulation price perspectives

The power markets and the different properties are complex, and they influence each other. Price is often the result of several different conditions and situations. Therefore, price is examined with the motivation to investigate its relationship with grid balancing (regulation volumes). The regulation price was briefly discussed in section 4.5.2, showing its close relationship with the day-ahead price. Sharing a close relationship with the day-ahead price also means that the regulation price is heavily influenced by the fluctuations in the day-ahead market in general. Power and electricity price formation for the day-ahead market is complex and a vast study topic. This thesis does not dive deeply into the underlying dynamics of price formation since it is worth a thesis solely on that topic. Therefore, this thesis regards the day-ahead price as one of the main driving forces behind the regulation price while excluding a deep dive into the detailed underlying origin of price formation. However, some significant drivers of price are looked at besides regarding the day-ahead price as the most significant one for setting the regulation price.

Its known that coupling of markets leads to a price convergence between the coupled markets[39]. Figure 4.8 shows the hourly regulation price for NO1-NO5 between 2016-01-01 and 2022-01-29. NO1, NO2, and NO5 looks equal by visual inspection, while NO3 and NO4 seems similar. This is most likely due to geographical locations and that NO4 only directly interconnect with NO3 when only considering the Norwegian bidding zones. The median, the middle quartile, is shown as the blue horizontal line in the boxplot figure 4.9. The middle 50% of the data within the inter-quartile range is represented by the middle box. The upper and lower whiskers represent upper and lower 25% of the data that falls outside the middle 50%. The mean value is plotted as the red dot, which falls within the middle box and looks equal for all zones. The inter-quartile range of the data seems equal for all the bidding zones when looking at the regulation price. The most apparent difference between box plots is the shape of possible outliers. The distant data points considered too far away are referred to as outliers, and the outlier threshold[7] is marked by the whiskers in the boxplot. Looking at the boxplots, the lower and upper tails, which are market as possible outliers, seem to be equal for NO1, NO2, and NO5, while NO3 and NO4 show a similar pattern. In general, all boxplots seem to have a relatively high number of points triggered as possible outliers. However, the price is real for buyers and sellers of power, even at extreme prices. Therefore removing or excluding possible outliers must be done with caution. Five records of a 5000 EUR/MWh price are observed, and only those data points are excluded from the boxplot to prevent them from dominating the figure. These data points are also excluded from figure 4.8. Based on the visual similarities and the data range of regulation price for NO1-NO5, it may be that investigating price for one particular bidding zone is also representative for another.

[7] Outlier threshold: $\pm 1.5 \times IQR$

4.7 The overall nature of regulation price

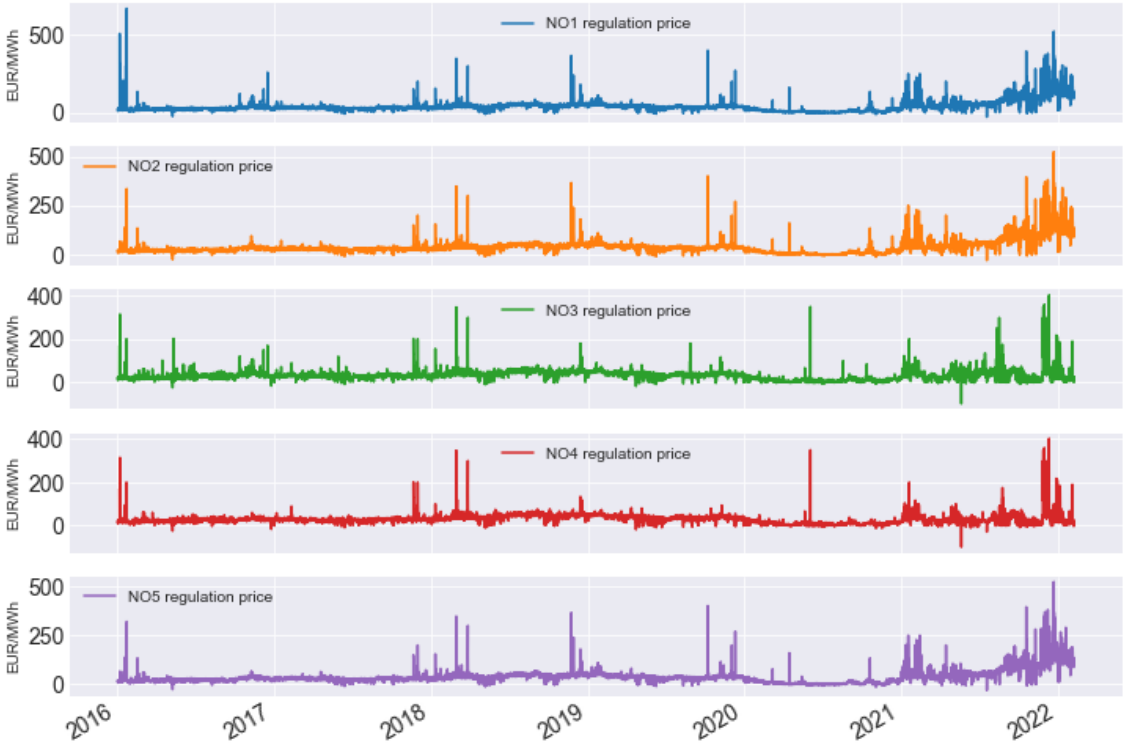


Figure 4.8: Regulation price for all bidding zones

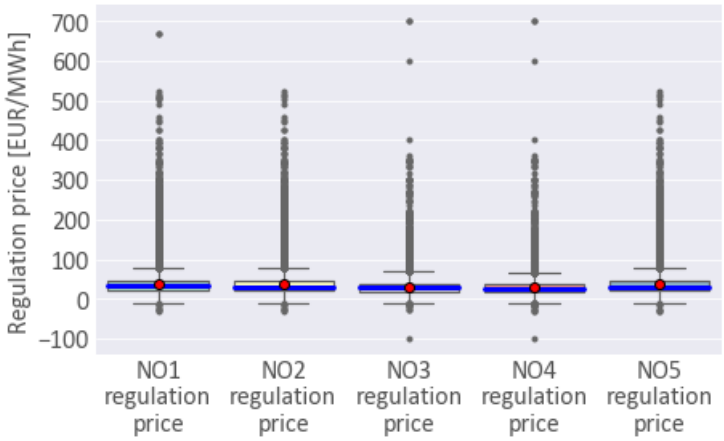


Figure 4.9: Boxplot of regulation price

4.8 Price markets relationships and drivers of price

The strong correlation between regulation price markets and how they relate to the day-ahead price is indicated in figure 4.10. The correlation coefficient number is a measures of how strong two variables are related across observations, e.g., between NO1 regulation price and NO2 regulation price, and if the data in NO1 change, the data for NO2 change

accordingly. One can say that, the higher the correlation coefficient (1.0 is max for spearman correlation), the stronger is the relationship between two compared variables of those shown in the matrix. When comparing a variable to it self, it has complete resemblance to itself yielding a correlation coefficient of 1.0. Spearman correlation is used since *"The Spearman rank correlation can handle outliers and non-linear relationships much better than the Pearson correlation coefficient"*[19].
One can see from the top-left cell in figure 4.10, representing the row for NO1 regulation price, that it has a correlation value of 0.99 when comparing NO1 regulation price with NO2 and NO5 regulation price. Moving further in the first row of the matrix, the NO1 regulation price reveal a correlation of 0.95, 0.94, and 0.94 comparing with day-ahead price for NO1, NO2 and NO5 respectively. Similar correlation coefficients can be seen in the rows representing NO2, and NO5 regulations price. The figure mainly reveal high correlations coefficients all over, but NO3 and NO4 is not equally correlated with NO1, NO2 and NO5. However, NO3 seems to be highly correlated with NO4 and visa versa. The Swedish, Danish, and Finnish markets that connect to the Norwegian markets are included to show that the Norwegian bidding zones and markets are not isolated for Norway only. Moreover, all markets are related as the power infrastructure in the whole Nordic region is connected, which facilitates the day-ahead markets, intra-day markets, and regulation markets in other countries to influence the Norwegian markets and opposite. The connection to England, the Netherlands, and Germany is not included in the matrix for simplicity.

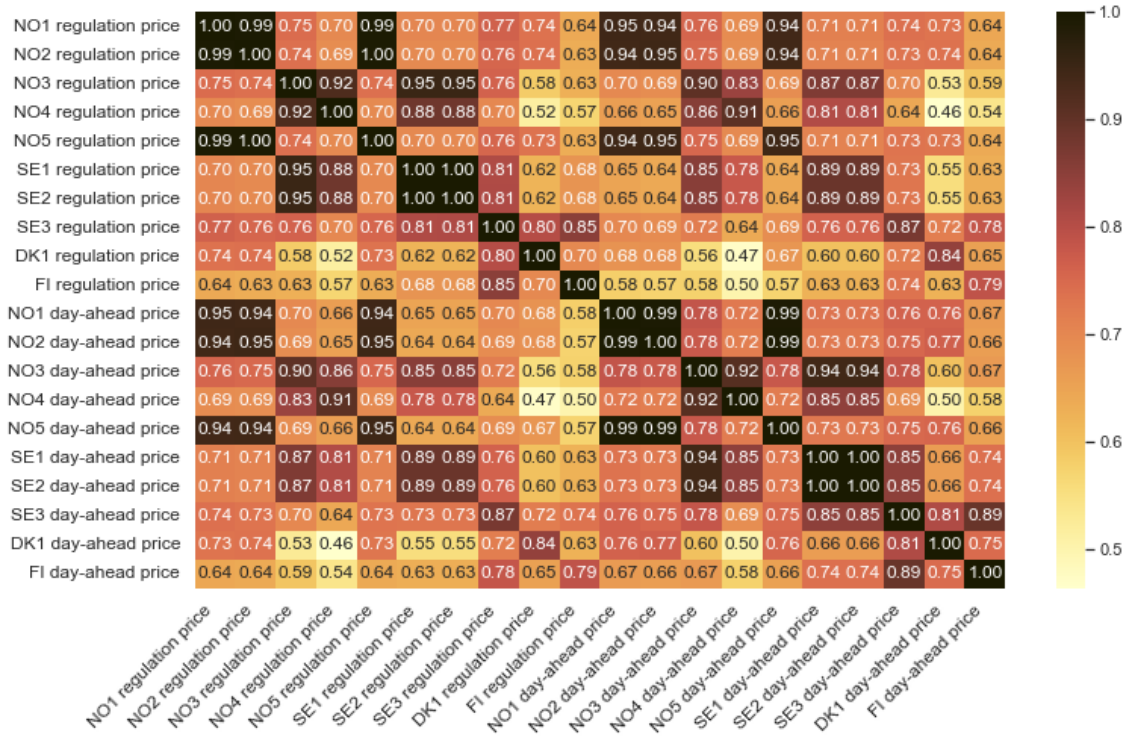


Figure 4.10: Spearman correlations matrix between regulation price and day-ahead price for all Norwegian bidding zones including other directly connected bidding zones in the Nordic region

As previously mentioned, the day-ahead price forms the base level for the regulation price, which also implies that the regulation price inherit seasonal fluctuation from the

day-ahead price and the day-ahead market in general. Other interesting drivers behind both day-ahead price and regulation price are general changes in the power sector, changes in the energy mix, new inter-connected transmission cables, and the water level in dams in Norway. The purpose of figure 4.11 is to show the influence of market coupling and the water levels in dams on the regulation price using NO2 as an example. The orange line in the plots shows the decomposed trend of the time-series data for regulation price and aggregated total water level in percentage for all dams in NO2. The day-ahead trend plot at the bottom of the figure is included only for reference. A trend is one of the essential characteristics to consider for time-series data. *"A trend is the presence of a long-term increase or decrease in the sequence"*[19]. The trend for both plots in the figure is calculated using the python package named statsmodels[8]. The process of estimating the trend can be conducted by approximating a polynomial of a certain degree for the time-series signal. A polynomial of a lower degree than the signal itself will fit the data with more smooth curvature and therefore capture a low-frequency trend which can be used to point at which direction the data moves in. Figure 4.11 shows that mid-2020 up to 2022 differentiate significantly from previous years. It is clear from the trend and the observed water levels in dams for NO2 point toward historically low water levels from mid 2021 to early 2022. For the same period, the trend for regulation price points to a rapid price increase in the middle plot in the figure. Additionally, the regulation price seems to have a denser distribution of spikes from 2021 to 2022 comparing the last 6 years. The spikes may be influenced by the change to the power mix to counter the loss of potential cheap and stable Hydro Water Reservoir power. However, the increase in the price trend and the frequency of the observed spikes are extreme from a historical perspective. It can be observed that the dynamics of the regulation price changed within the same time span as two new export transmission cables were put to use. The time at which the two export cables were put to usage is plotted as vertical lines in the plot in figure 4.11. The Nord Link cable going to Germany was in trial operation in late 2020. The plot reveals an immediate spike the day they started to test it in operation. It was in regular operation a bit later in 2021. The North Sea Link connected to England was in operation from 2021-10-01, and new electricity price records superseded old ones after it was put in operation. The effect of market coupling leading to market price convergence looks to be visible in the plot. However, one could also argue that a significant portion of the price increase stems from the low water levels in the dams. Additionally, fuel costs and coal price in Europe is said to influence the price for power and electricity, but this analysis is not included in the thesis. It may be that the newly installed export cables further enhance the influence from European fuel and coal price due to a tighter market coupling from direct connections.

[8] https://www.statsmodels.org/devel/generated/statsmodels.tsa.seasonal.seasonal_decompose.html

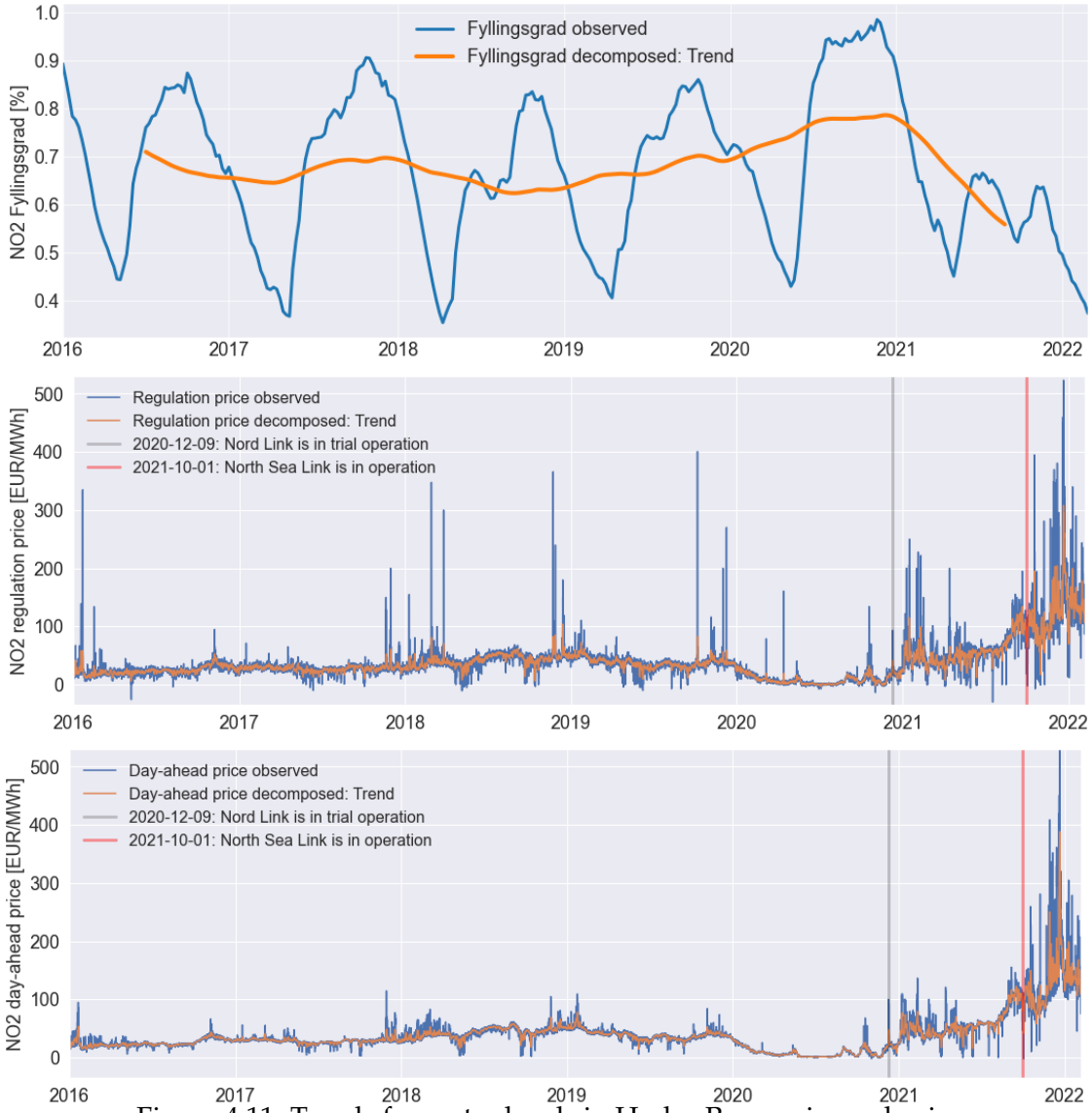


Figure 4.11: Trends for water levels in Hydro Reservoirs and price

4.9 Global and local statistics for regulation price

Global statistics of regulation price from 2016 to February 2022 is shown in table 4.1. One can see from the statistics that NO1, NO2, and NO5 has similar statistics while NO4 and NO4 is relatively equal. The mean price and standard deviation in NO3 and NO4 is lower than NO1,NO2, and NO5.

	NO1	NO2	NO3	NO4	NO5
Median	31.78	31.54	29.8	26.61	31.53
Mean $[\mu]$	37.55	37.29	30.29	28.12	37.12
Standard deviation $[\sigma]$	32.39	31.72	19.77	18.72	31.52
Variance $[\sigma^2]$	1049.14	1006.11	390.94	350.57	993.60

Table 4.1: Regulation price global statistics for NO1-NO5 in unit EUR/MWh for 2016-2022

In time-series data, global statistics are not sufficient to understand how the data develop over time. Therefore, time-series data is commonly analysed within overlapping windows of a specific window size to extract statistical properties local to several time intervals. There are different types of window functions, and figure 4.12 demonstrate the rolling mean and rolling standard deviation window functions using a one-year moving average and standard deviation. Each window stretch over one year of data and the statistics are computed using a stride of one hour for each incremental shift. These window functions create aggregated measures of the time-series data through time. Figure 4.12 reveal slight variations in the mean and standard deviation value up until 2021 while drastically change in all zones from 2021 after the commissioning of the Nord Link. Further on, it seems that an additional change is occurring between late 2021 to start 2022, which is post commissioning of the North Sea Link at 2021-10-01. This shows that the mean and standard deviation is not constant through time. The extra added number on the x-axis for NO1, NO2, NO3, and NO4 represents wind farm facilities in operation at that time. NO1, NO2, and NO4 steadily increase in wind farm facilities in operation from 2016-to 2021, while NO3 seems to increase with 5 operational wind farms between the second half in 2020 and the first half of 2021. The increase of operational wind farms does not seem to carry the change to the price development alone. It looks like the two transmission cables contribute significantly to bending the local statistical properties post-2021. NO5 does not have any wind farm facilities in operation during 2016-2022.

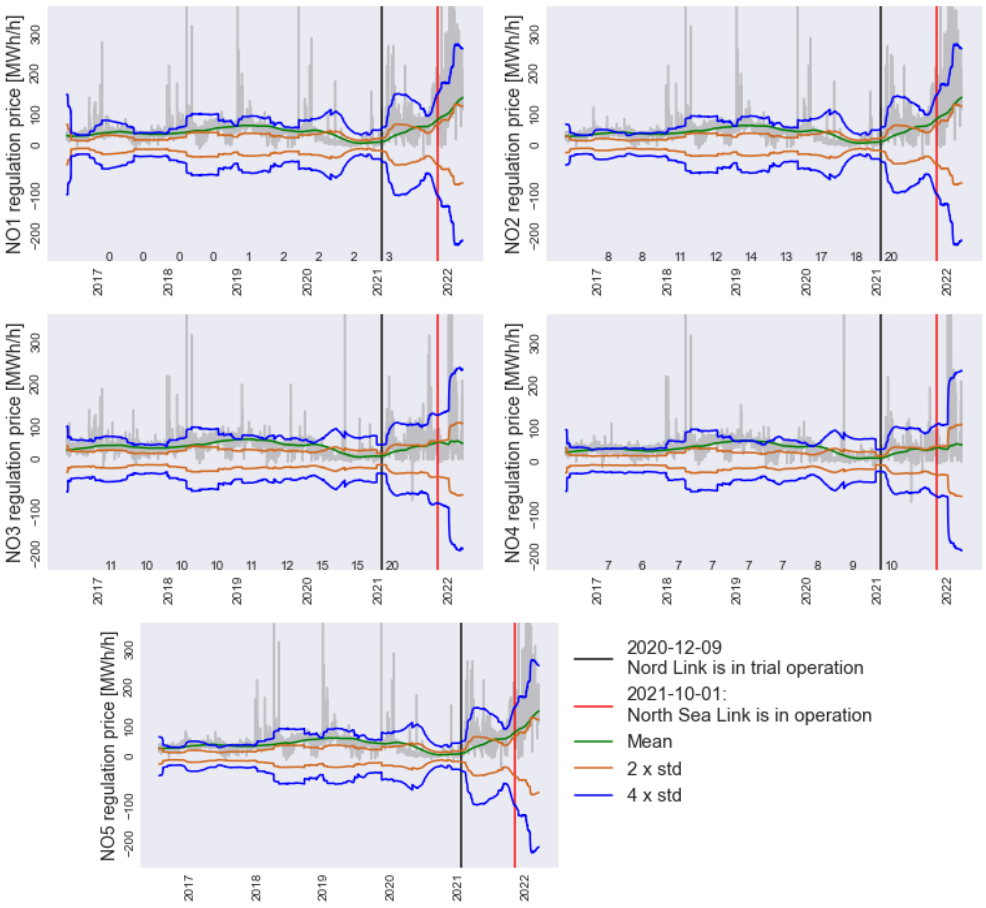


Figure 4.12: One year moving average and standard deviation for regulation price calculated using one hour incremental rolling from 2016 to February 2022.

4.9.1 Statistical stationarity

Time-series analysis involves looking at several characteristics of the data relative to time. A stationarity check is essential when analyzing the nature of time-dependent sequenced data, and some forecasting models assume stationary data to perform. Stationarity *"refers to a property of time-series not to change distribution over time, or in other words, that the process that produces the time-series doesn't change with time. Time-Series that don't change over time are called stationary (or stationary processes)"* [19]. A simple definition of stationarity is presented as follows:

The time-series is stationary if the distibution of $y_t, y_{t+1}, ..., y_{t+k}$ do not depend on t for all possible k lags [14], where t is the time-step and y is the value of the data at time-step t.

A widely used statistical method to determine if the time-series is stationary is called the Augmented Dickey-Fuller (ADF) test. The underlying process of this method is a hypothesis test that determines if the data or the process behind the data is stationary or not. The null hypothesis, H_0, states that the time-series incorporate a unit root[9]. The alternative hypothesis,H_A, states that a unit root is not present in the time series. In simpler words, the setup can be viewed as follows:

H_0**: The data sequence incorporate a non-stationary structure**

H_A**: The data sequence does not have a non-stationary structure**

If H_0 is rejected based on a given level of significance, the test recommend that the time-series has a stationary structure as stated by H_A. If the test fail to reject H_0, the test recommend that the time-series has a unique root which implies that the data has non-stationary structure [20].

To prove if the data is stationary the evidentiary level is set to 99%, resulting in a significance level of 0.01, which sets the strength of the test evidence to very high. Choosing a very high significance level is based on observed complexity in power market data having several processes influencing its characteristics. The Augmented Dickey-Fuller test is carried out using the Python package statsmodels, the same package used for the moving average calculations. Given the chosen level of significance, the test fails to reject H_0 if the P-value from the ADF test is greater than 0.01 indicating that the data is non-stationary. H_0 is rejected for P-values $\leq$ 0.01 indicating a very confident proof of the data having stationary characteristics. Table 4.2 shows the results of the ADF test for regulation price for all the Norwegian bidding zones between 2016-2022. The test shows all zones being stationary. However, drastically recent changes are observed in the power sector affecting several markets. Therefore, a ADF test on data from 2020-2022 were conducted. Table 4.3 tell a different story, it proves that the data from 2020-2022 has non-stationary components for NO1,NO2 and NO5, while NO3 and NO4 is still stationary. This proves that the regulation price is significantly trending towards a increasing prices for NO1,NO2, and NO5, while NO3 and NO4 is regarded as stationary.

[9] A unit root and the mathematics behind the process is out of scope for this thesis

Regulation price	Data range	Stationary	P-value	Significance level	Reject H0
NO1	2016-2022	Yes	1.00E-07	0.01	Yes
NO2	2016-2022	Yes	6.00E-07	0.01	Yes
NO3	2016-2022	Yes	0	0.01	Yes
NO4	2016-2022	Yes	0	0.01	Yes
NO5	2016-2022	Yes	7.00E-07	0.01	Yes

Table 4.2: Statistically stationary test for regulation price for NO1-NO5 on data between 2016-2022.

Regulation price	Data range	Stationary	P-value	Significance level	Reject H0
NO1	2020-2022	NO	0.0202856	0.01	NO
NO2	2020-2022	NO	0.0210839	0.01	NO
NO3	2020-2022	Yes	0	0.01	Yes
NO4	2020-2022	Yes	0	0.01	Yes
NO5	2020-2022	NO	0.0242998	0.01	NO

Table 4.3: Statistically stationary test for regulation price for NO1-NO5 on data between 2020-2022.

A similar test performed for the day-ahead price for comparison is shown in table 4.4 and 4.5. The day-ahead market has similar characteristics as seen for the regulation market. This proves that price trends and the non-stationary characteristics in the day-ahead market influence the regulation market price.

Day-ahead price	Data range	Stationary	P-value	Significance level	Reject H0
NO1	2016-2022	Yes	2.40E-06	0.01	Yes
NO2	2016-2022	Yes	1.14E-05	0.01	Yes
NO3	2016-2022	Yes	0	0.01	Yes
NO4	2016-2022	Yes	0	0.01	Yes
NO5	2016-2022	Yes	8.30E-06	0.01	Yes

Table 4.4: Statistically stationary test for day-ahead price for NO1-NO5 on data between 2016-2022.

Day-ahead price	Data range	Stationary	P-value	Significance level	Reject H0
NO1	2020-2022	NO	0.0215752	0.01	NO
NO2	2020-2022	NO	0.0273963	0.01	NO
NO3	2020-2022	Yes	0.000106	0.01	Yes
NO4	2020-2022	Yes	6.81E-05	0.01	Yes
NO5	2020-2022	NO	0.0219233	0.01	NO

Table 4.5: Statistically stationary test for day-ahead price for NO1-NO5 on data between 2020-2022.

4.9.2 Autocorrelation

Autocorrelation is another method commonly used in time-series data analysis. Its main purpose is to give insight into how the past influence the present. Autocorrelation, also

known as serial correlation, measures the signal correlation between two points, the signal itself and a lagged (delayed) copy of itself. Said in other words: *"Autocorrelation gives you an idea of how data points at different points in time are linearly related to one another as a function of their time difference"* [14][40]. Additionally, autocorrelation can also help in determining if the time series is mostly random stochastic data or not. If the past historical data influences the present time, the data is not random. Autocorrelation is especially important for univariate models since such modeling implies forecasting using only one variable, having the same data stream as both the input and output. Therefore, exploiting previous time-steps as a key feature in describing the present is relevant for the majority of time-series problems.

Figure 4.13 show the autocorrelation for regulation price for bidding zone NO5. It can be observed a strong correlation between the present price and historical lags up to 24 hours. The correlation seems to increase closer to a 24 hours lag which is most likely due to the daily price fluctuations. The observed strong autocorrelations proves that there is a strong linear relationship between the regulation price at time step y_t and y_{t-1} all the way up to y_t and y_{t-24}. The other zones had similar autocorrelation.

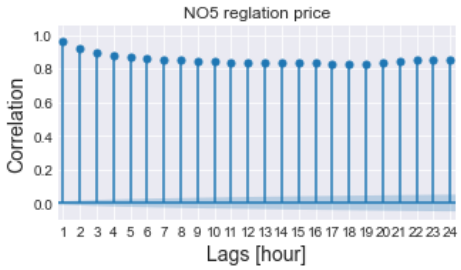


Figure 4.13: Autocorrelation of the regulation price for bidding zone NO5 over 2016-2021 data

4.10 Drivers of tertiary reserves volume activation

The regulation volume for NO2 investigated in subsection 4.5.1 did not reveal any clear patterns. In addition, it has previously been claimed that regulation volume and price premium share properties with a statistically random distribution [27] and hence will be almost impossible to predict. Therefore, it is necessary to conduct a deeper analysis in order to understand the nature of regulation volumes and to determine whether it is predictable or just random stochastic white noise.

White noise is pure random data which is not considered predictable, meaning one cant foresee the value of the next future time-steps. The criteria and data characteristics to fulfill for a time-series to be regarded as white noise is as follows [41]:

$$\begin{aligned} \mu &= 0 \text{ and constant through time} \\ \sigma &= \text{constant through time} \\ r_l &= 0 \text{ for all lags} \\ \text{stationary} &= \text{Yes} \\ \text{seasonal components} &= \text{No} \end{aligned}$$

Where r_l is the lagged correlation for lag l (autocorrelation).

Figure 4.14 shows the regulation volume as time series. The figure show the overall complex signature for of regulation volume from 2016-01-01 up to 2022-02-01 for all the Norwegian zones. The y-axis in each subplot represents regulation volume in megawatt-hour (MWh).

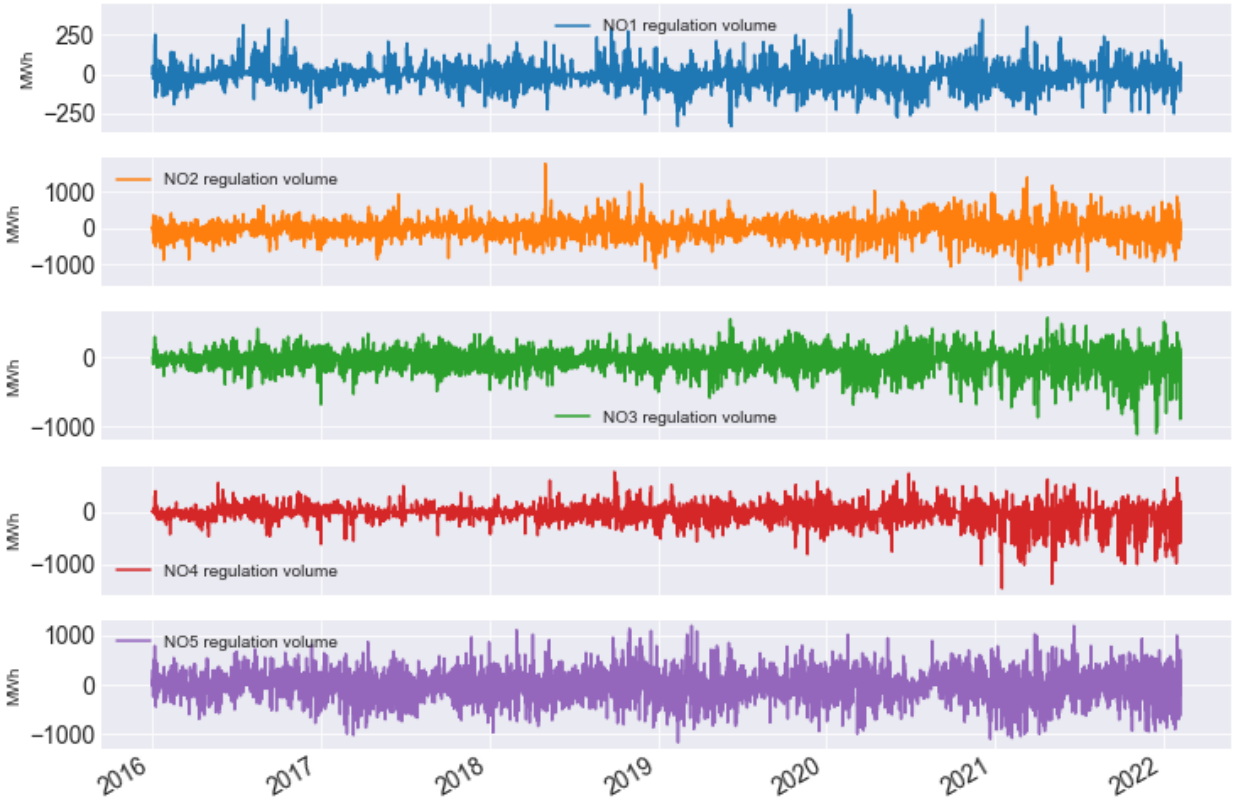


Figure 4.14: Regulation volume for all bidding zones

Investigating decomposed characteristics of the regulation volume may reveal possible trends and seasonal components. For example, the regulation price had a clear trend from 2021-2022. Figure 4.15 show the decomposed characteristics of regulation volume for NO2, NO3, and NO5. NO1 and NO4 share similar decomposition characteristics. In addition to the trend component investigated for the regulation price, the decomposed seasonal and residual components are included for regulation volume. The purpose of the seasonal component characteristic is to show seasonal patterns through time. The residual component represents the noise or the parts of the data that can not be described through signal decomposition. The green-colored line indicates the trend component. The trend plots do not reveal any clear trending patterns. However, some trends seem to be triggered in the signal decomposition. The seasonal component in the figure shows a thick horizontal pink line clamped close to zero regulation, not revealing any seasonal patterns. Orange residual lines dominate the plots and cover most of the observed data in blue. This indicates that the data is significantly dominated by random noise. Whether the Nord Link or North Sea Link cables influence the regulation volumes cant be easily determined based on the decomposition figure.

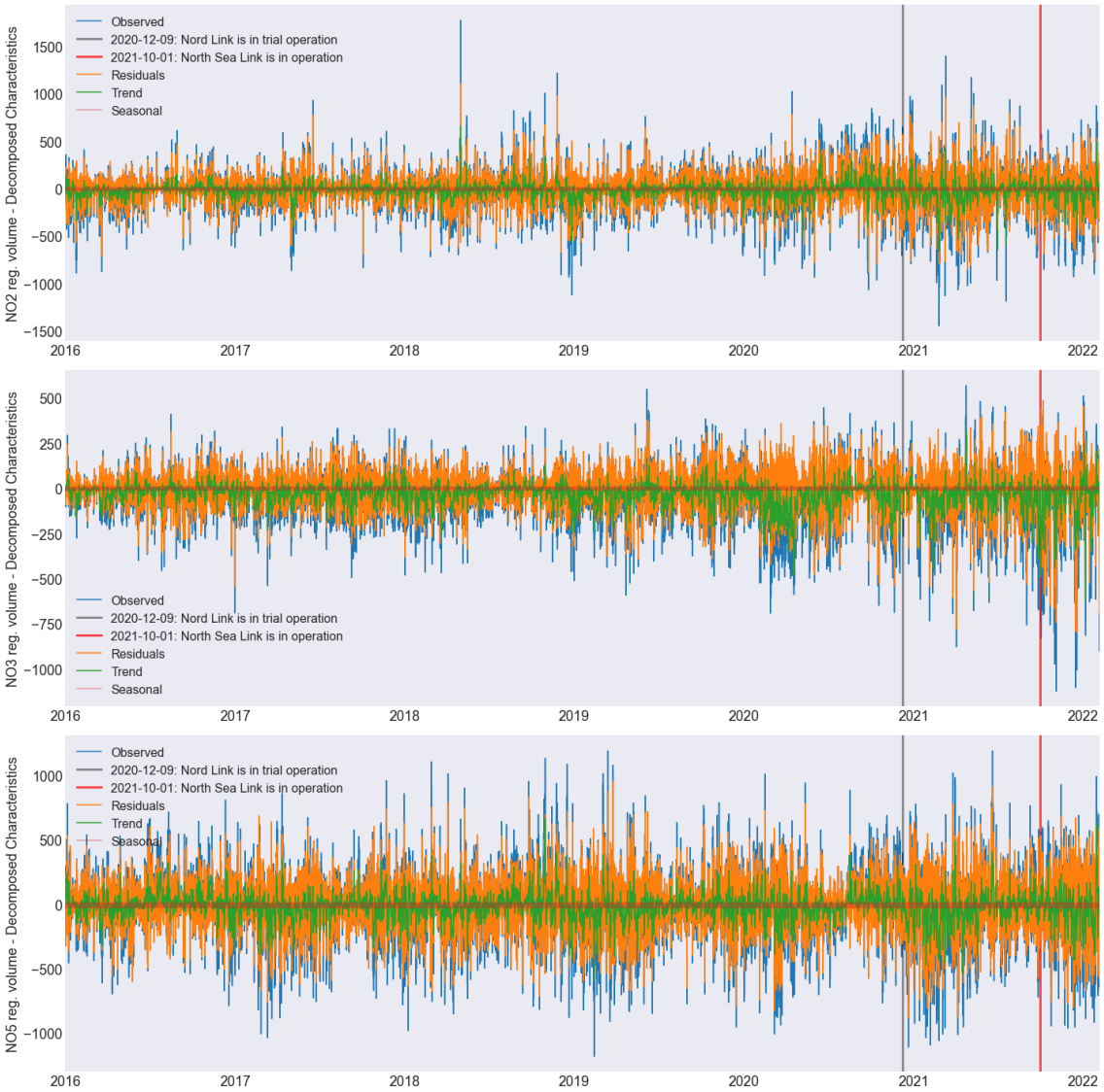


Figure 4.15: Signal decomposition of regulation volume for NO2,NO3 and NO5

Figure 4.16 show a boxplot of the regulation volume for each zone. The figure serves the purpose of highlighting overall statistics and shape of the data. Looking at the boxplot for all zones for regulation volume, many data points are located outside the thresholds negative and positive sides (the whiskers). A significant part of the power regulation is then considered an outlier using common statistical measures. It is common to deal with outliers in some way since they can, in many cases, be regarded as abnormal data points that can potentially skew the data set toward the extremes. Not dealing with outliers, especially extreme outliers, can potentially dominate the modeling process when expecting a model capable of approximating reasonable output signals given the input data. However, dealing with outliers must be done with caution, and activation of regulation reserves is considered to originate from abnormal conditions by its nature. Further on, both the mean and median values in the distribution for each zones seems to be close to zero. NO2 stands out having the longest upper and lower tails comparing all zones.

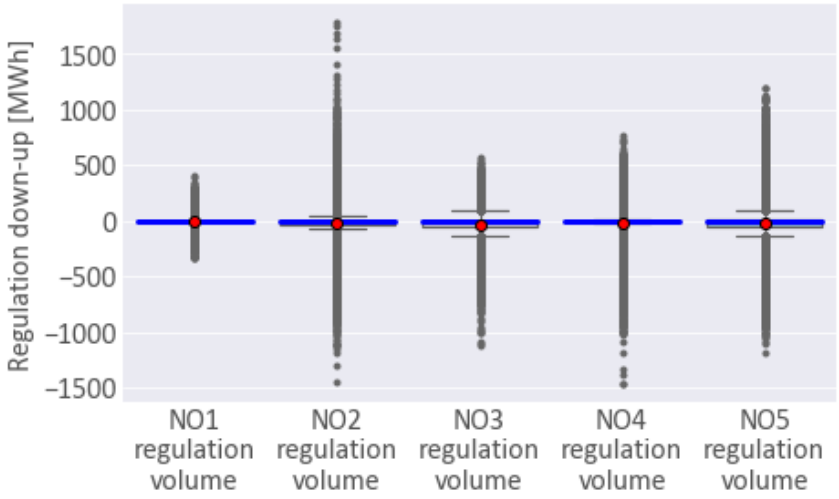


Figure 4.16: Boxplot of regulation volume for all zones

Investigating the histograms of the regulation volumes in figure 4.17 clearly show that zero regulation is dominating the distribution for each zone looking at the high frequency on the y-axis at zero regulation. The fraction of the data being zero for each zone is shown table 4.6. The high zero fraction is expected since zero regulation represents normal conditions in the power market where the power is in balance. Deviations from normal conditions resulting in power balancing are represented by each tail seen in the histograms, where negative values on the x-axis represent down regulation[10], and positive values represent up regulation[11]. The mean, μ, and standard deviation, σ, are plotted in the histogram subplots and shown in the table. For a model to approximate a function of the regulation volume, it must adapt to and predict similar characteristics of the shown histograms. Adapting to the dominating zero, while also giving attention to of both tails can prove challenging for some Machine Learning models. As previously mentioned, the tails are constructed as part of the data pre-processsing. The distribution does not have two tails in the data from Nordpool, since they represents two different scenarios; one process or tail is tied to power not being produced as originally planned (a market participant gets payed for not producing power), the other is rooted in larger demand than planned for which is corrected with additionally power reserves being produced. However, both cases serves the purpose of balancing the grid.
Figure 4.18 is showing the histograms zoomed in on the tails. Visually inspecting the zoomed histogram show a distribution resembling a Laplacian distribution rather than a Gaussian distribution considering all zones. Table 4.6 shows that all zones have a small negative mean values which indicate the distributions are barely skewed towards down regulation. NO2 and NO5 stands out having high variance comparing all zones in table 4.6.

[10]The zone is giving (selling) power to other zones
[11]The zone is getting (buying) power to other zones

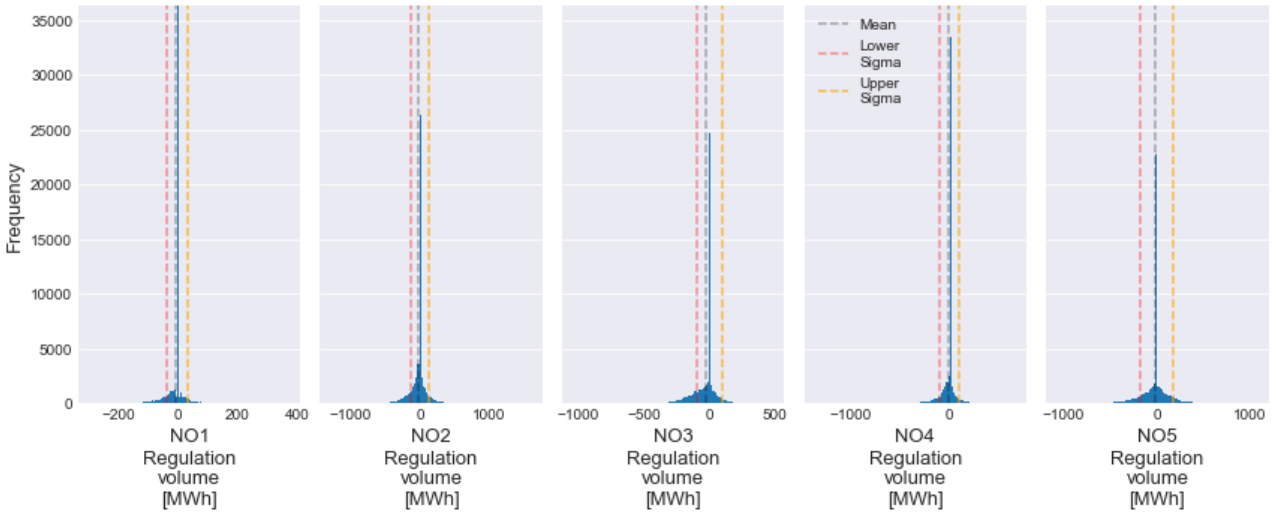


Figure 4.17: Regulation volume histogram for all bidding zones

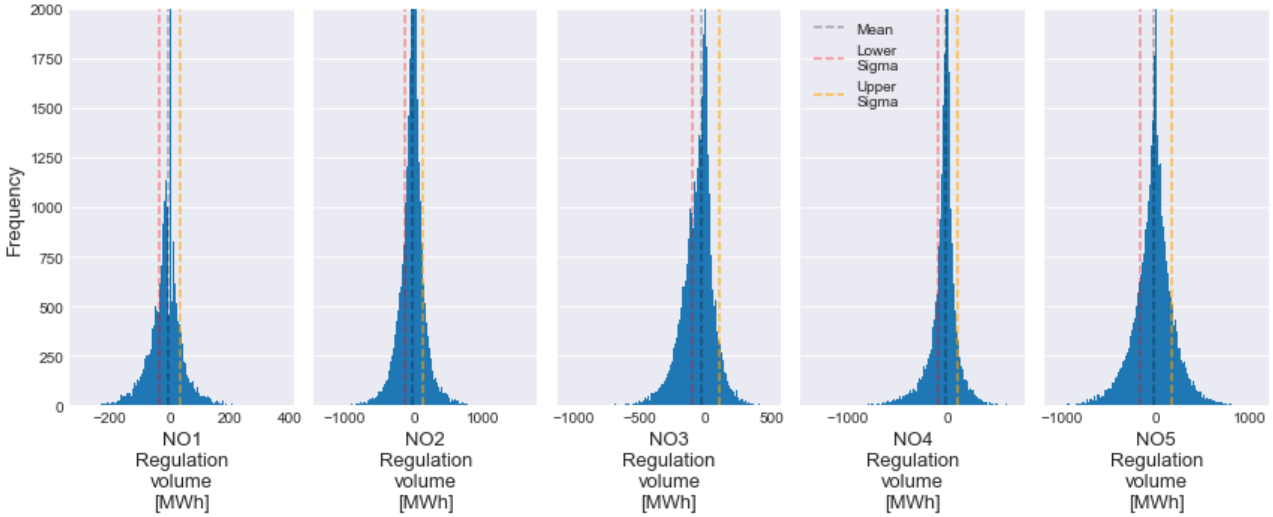


Figure 4.18: Regulation volume histogram zoomed for all bidding zones

	NO1	NO2	NO3	NO4	NO5
Median	0.0	0.0	0.0	0.0	0.0
Mean $[\mu]$	-5.24	-16.22	-32.31	-16.93	-16.05
Variance $[\sigma^2]$	1239.33	19252.42	10156.84	9684.75	29183.59
Standard deviation $[\sigma]$	35.20	138.75	100.78	98.41	170.83
Zero fraction [%]	67.55	44.43	43.31	59.70	40.48

Table 4.6: Regulation volume global statistics for NO1-NO5 in unit MWh for 2016-2022

4.11 Local statistics for regulation volume

When first looking at local statistics for regulation volume in figure 4.19, it reveals a relatively unchanged μ overtime for all zones. However, σ, especially 4σ, reveal some trends. This proves that σ change over time, as seen in the blue line in all subplots. The numbers plotted close to the green line representing μ are the number of wind farms in operation. The commissioning of the two export cables is plotted as vertical lines. It may not be entirely correct to draw finite conclusions solely from this figure, but it is tempting to view the change of direction to the 4σ slope in the context of the two export cables and the increasing wind power contribution to the grid.

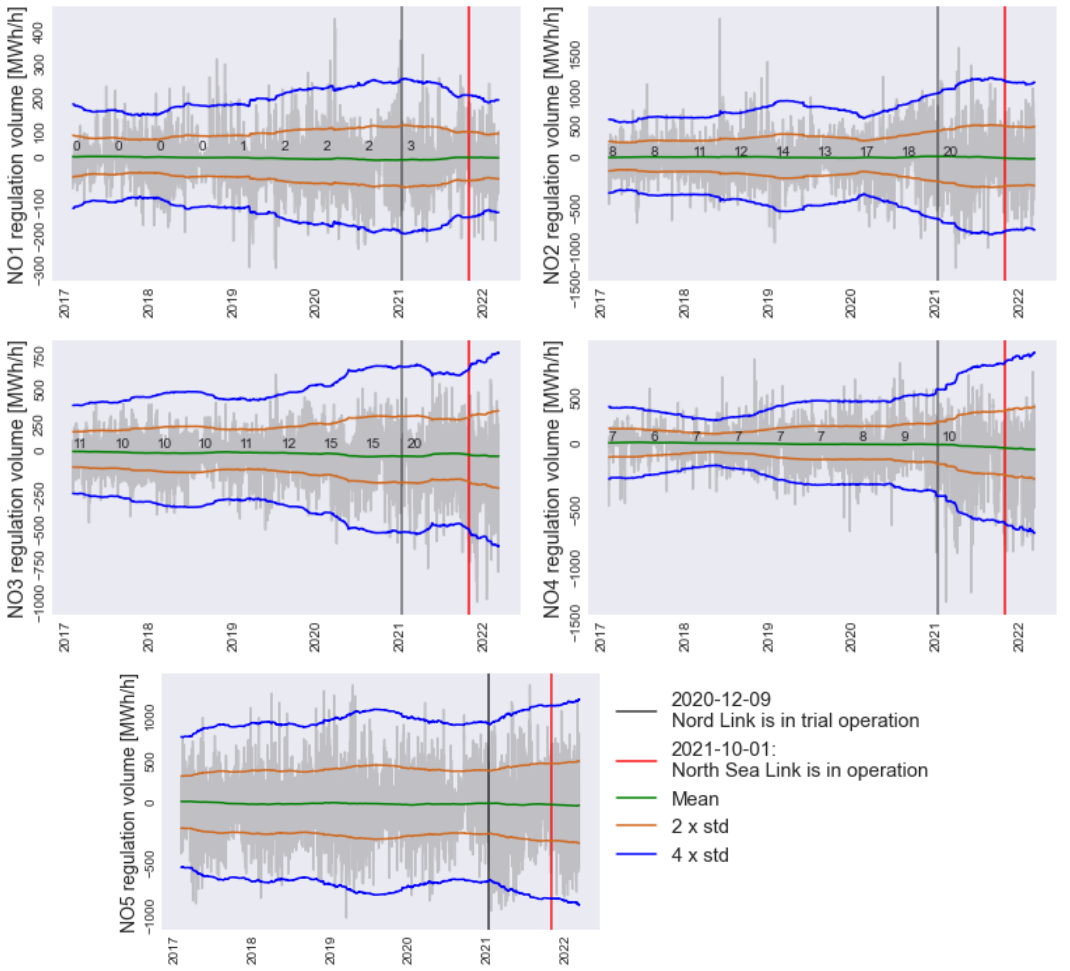


Figure 4.19: One year moving average and standard deviation for regulation volumes calculated using one hour incremental rolling from 2016 to February 2022

4.11.1 Statistical stationarity

A similar stationary test to what was conducted for regulation price can be shown for regulation volume in table 4.7 and 4.8. This test strongly prove that regulation volume is stationary for both time intervals 2016-2022 and 2020-2022 with a significance level of 0.01.

Regulation volume	Data range	Stationary	P-value	Significance level	Reject H0
NO1	2016-2022	Yes	0	0.01	Yes
NO2	2016-2022	Yes	0	0.01	Yes
NO3	2016-2022	Yes	0	0.01	Yes
NO4	2016-2022	Yes	0	0.01	Yes
NO5	2016-2022	Yes	0	0.01	Yes

Table 4.7: Statistically stationary test for regulation volume for NO1-NO5 on data between 2016-2022.

Regulation volume	Data range	Stationary	P-value	Significance level	Reject H0
NO1	2020-2022	Yes	0	0.01	Yes
NO2	2020-2022	Yes	0	0.01	Yes
NO3	2020-2022	Yes	0	0.01	Yes
NO4	2020-2022	Yes	0	0.01	Yes
NO5	2020-2022	Yes	0	0.01	Yes

Table 4.8: Statistically stationary test for regulation volume for NO1-NO5 on data between 2020-2022.

For comparison, same test was conducted on day-ahead volumes shown in table 4.9 and 4.10. Day-ahead volumes in all zones is non-stationary in the periode 2020-2022. This show that the non-stationarity characteristics and trending for day-ahead volumes in the day-ahead market do not influence the stationarity for regulation volumes in the regulation market.

Day-ahead volume	Data range	Stationary	P-value	Significance level	Reject H0
NO1	2016-2022	Yes	0.0021081	0.01	Yes
NO2	2016-2022	Yes	6.00E-07	0.01	Yes
NO3	2016-2022	Yes	1.30E-06	0.01	Yes
NO4	2016-2022	Yes	2.16E-05	0.01	Yes
NO5	2016-2022	Yes	0.0033048	0.01	Yes

Table 4.9: Statistically stationary test for day-ahead volume for NO1-NO5 on data between 2016-2022.

Day-ahead volume	Data range	Stationary	P-value	Significance level	Reject H0
NO1	2020-2022	NO	0.0908841	0.01	NO
NO2	2020-2022	NO	0.0487509	0.01	NO
NO3	2020-2022	NO	0.0152289	0.01	NO
NO4	2020-2022	NO	0.1120605	0.01	NO
NO5	2020-2022	NO	0.1241714	0.01	NO

Table 4.10: Statistically stationary test for day-ahead volume for NO1-NO5 on data between 2020-2022.

4.11.2 Autocorrelation

How correlated the present regulation volumes are relative to lagged versions is shown in figure 4.20. The figure shows that the present volume at hour y_t has a correlation coefficient slightly above 0.8 relative to the previous hour y_{t-1} for all zones. After 3 hours, the autocorrelation seems to be less significant. This indicates that autocorrelation in the previous 1-4 hours can be a significant feature for describing the present regulation volume. Thus, it motivates using machine learning models that can exploit previous hours using recurrence on time-dependent sequential data. Furthermore, the autocorrelation comprises both indirect and direct correlations between the current time and prior time steps[42]. This means that the correlation at lag 2 can be significantly influenced and dependent by the correlation for lag 1. The partial autocorrelation can be used to describe a direct association between the present time step and a given lag to cancel out the influence from other lags. *"The partial autocorrelation at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags."* [43]

One can see from the partial autocorrelation in figure 4.21 the direct association between the present time step and different lags. The figure shows that only the correlation at lag $k = 1$ significantly describes the present time step and that lags of $k > 1$ not contributing much in describing the present time step.

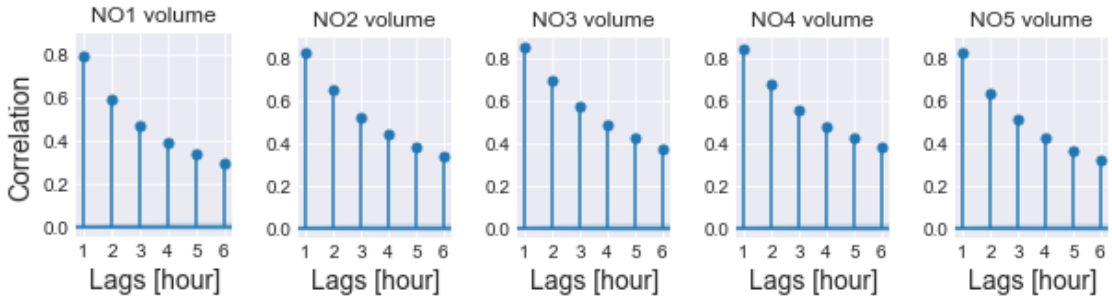


Figure 4.20: Autocorrelation of the regulation volumes over 2016-2021 data

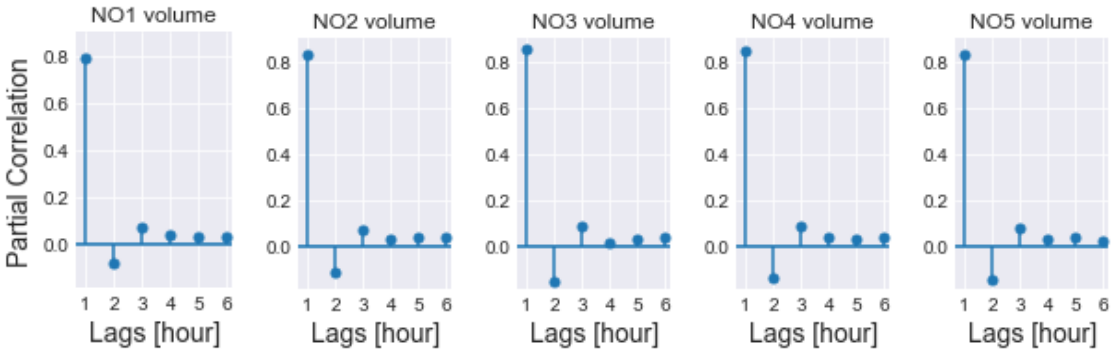


Figure 4.21: Partial Autocorrelation of the regulation volumes over 2016-2021 data

4.11.3 Periodicity behind tertiary volume activation

Since the periodic and seasonal patterns in the regulation volume are vague comparing regulation price, further investigation of periodicity is needed for regulation volume. A commonly used method for finding and extracting features that fluctuate periodically is the Discrete Fourier Transform (DFT). This method transforms the data from the time domain to the frequency domain. The data is then no longer considered a function of time,

and different individual frequencies may reveal periodic patterns based on the strength of the amplitude at each frequency.

Figure 4.22 show the DFT of regulation volume and day-ahead volume for NO2 and NO3 between 2016-2022. Day-ahead volumes are included in the plots for comparison. The Tensorflow.signal.rfft method is used to calculate the DFT[12]. For both plots the amplitude and y-axis are normalized based on largest observed frequency spike, which is revealed as the year frequency for all plots. Thus, a yearly periodic fluctuation is present for both day-ahead and regulation for NO2 and NO3. The day-ahead volume for NO2 and NO3 shows two clear spikes at day and for every twelfth hour. This shows that there is some periodicity on days, and the frequency component at every twelfth hour refers to two high peaks observed at 06:00 and 18:00 seen in previously discussed figure 4.5. Additionally, regulation volume for NO3 seems to have several periodic frequencies close to the yearly frequency, and every fourth month *(1/month*4)* seems to trigger a spike. Weeks may also be considered as a spike for NO3. It can be argued that day and every twelfth hour has a small spike for regulation volume. However, the regulation volume has much noise, and the majority of the frequencies cant be clearly distinguished from the noise observed. Zone NO1, NO4 and NO5 share similar characteristics as the ones plotted in the figure.

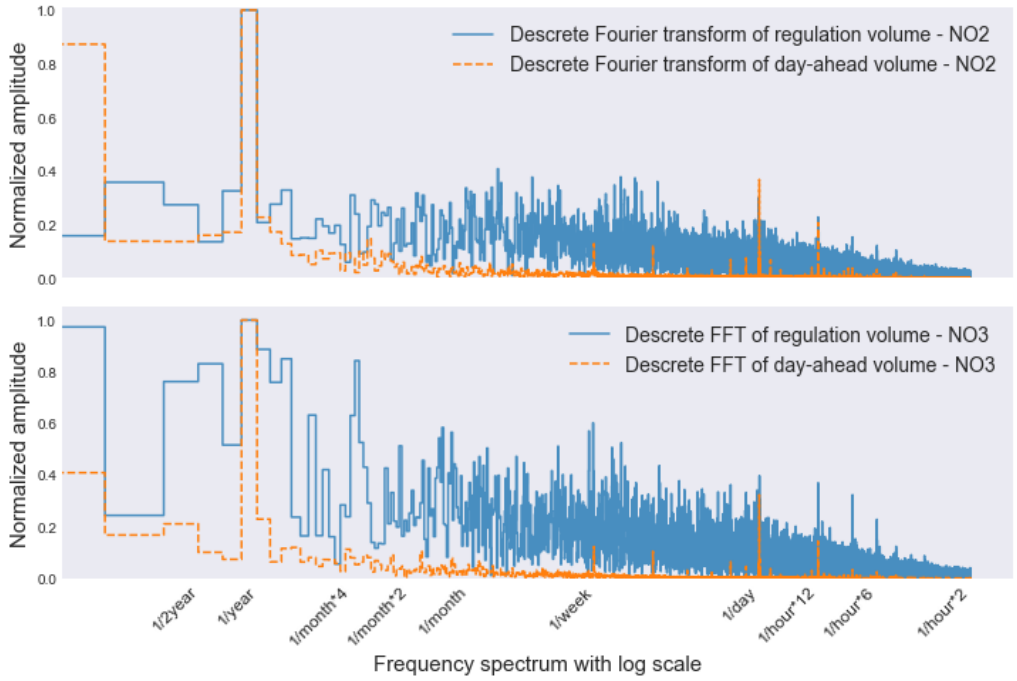


Figure 4.22: Discrete Fourier transform NO2 and NO3 for 2016-2022.

A five summary statistics of regulation volumes grouped by year is visualized in the boxplots in figure 4.23. Outliers are excluded in the plots for easier interpretation. The plots summarize the distribution of yearly grouped regulation volumes, and it does not seem like the year is static for the time-span investigated when excluding the median drawn with a black horizontal line. The mean value represents the red dot, the minimum and maximum (lower and upper whiskers) represents the Q_1 (0th percentile)

[12] https://www.tensorflow.org/tutorials/structured_data/time_series

and Q_4 (100th percentile), and the Q_1 and $Q2$ represents the lower and upper sides of the box (25th - 75th percentile). The most recent years seem to have slightly different characteristics than 2016-2018. Thus, backing up the Fourier transform that year has, to some degree, a periodic component when explicitly looking at NO2 and NO3.

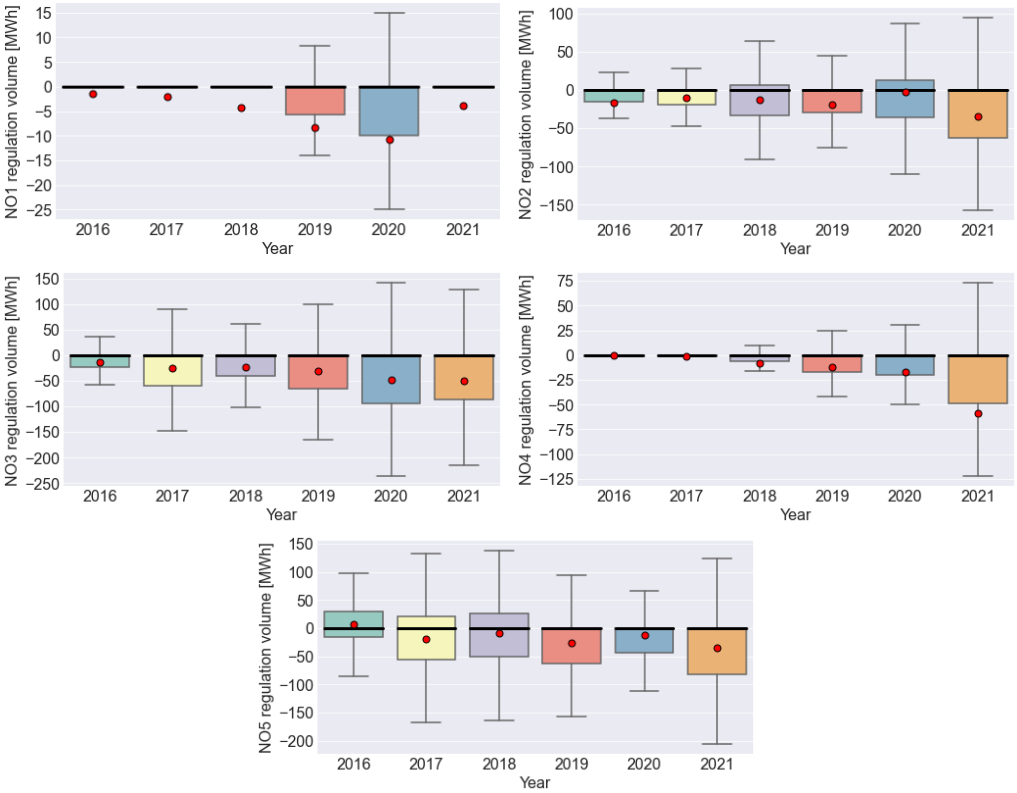


Figure 4.23: Yearly regulation volume for NO1, NO2, NO3,NO4, NO5 from 2016-2021

Moving on to investigating monthly fluctuations for regulation volume, the boxplot in figure 4.24 reveals that data between 2016-2021 grouped by month has a periodic component for NO5. NO1, NO2, NO3 and NO4 are not plotted, but they show similar monthly fluctuations.

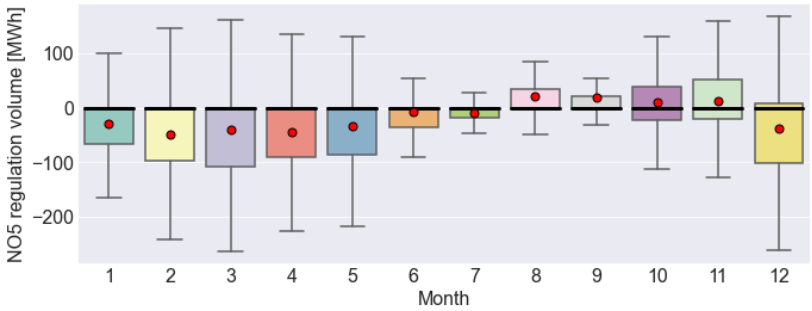


Figure 4.24: Monthly regulation volume for NO5 from 2016-2021

Daily and hourly fluctuations are summarized in the boxplots in figure 4.25 for NO5. The weekends (day 5 and 6 - Saturday and Sunday) seems to differ slightly from the business days 0-4. Some periodicity can also be observed when looking at the data grouped on hours between 2016-2021. The other zones share similar daily and hourly characteristics.

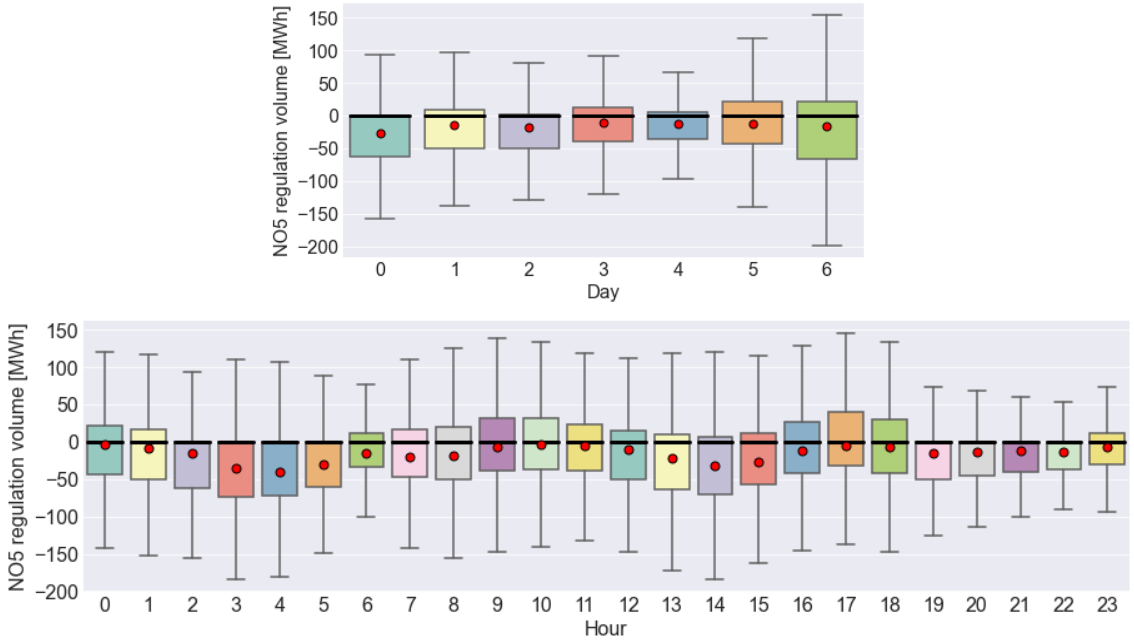


Figure 4.25: Daily and hourly regulation volume for NO5 from 2016-2021

4.12 Sibling balancing markets and main driving force

A strong correlation between the regulation price for different zones was previously shown. This phenomenon also applies to tertiary regulation volume activation as seen in figure 4.26. This is mainly due to the fact that regulating down in one zone occurs concomitantly with up regulation in another and contrary. Thus, a strong indirectly descriptor for activation of tertiary regulation volume could be considered regulation happening in another zone.

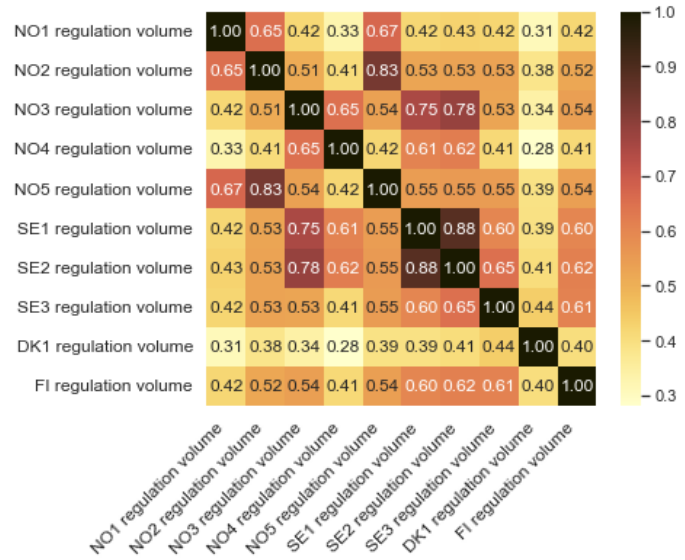


Figure 4.26: Spearman correlations between bidding zones for regulation (imbalance) volume

4.12.1 Price and production deltas

Skytte [6], Jaehnert et. al [29] and klæboe et al. [27] emphasized on price premiums (delta price) between day-ahead and regulation price as the most evident predictor for regulation volume. In the relatively scarce selection of previous research on tertiary reserves activation, the regulation price delta, Δpr surface as a obvious descriptor to investigate. However, no recent official research has been found to investigate if Δpr still holds significant descriptive properties for regulation volume on recent data.

Figure 4.27 indicates that Δpr still relates to regulation volume. The top plot in the figure shows a randomly selected time-interval in 2021 for NO4, where both the day-ahead price and the regulation price for 200 hours are plotted. The bottom plot shows the difference between the regulation price and day-ahead price (Δpr) together with the regulation volume. Price and volume have different y-axis; the left y-axis supports Δpr while the right y-axis supports the regulation volume for the selected time-span. One can see from the bottom plot that activation of regulation volumes (the orange line) triggers the Δpr (blue line). However, the magnitude of change in Δpr does not seem constantly linear to how much regulation volume change. NO1, NO2, NO3, and NO5 share qualitatively similar patterns as NO4.

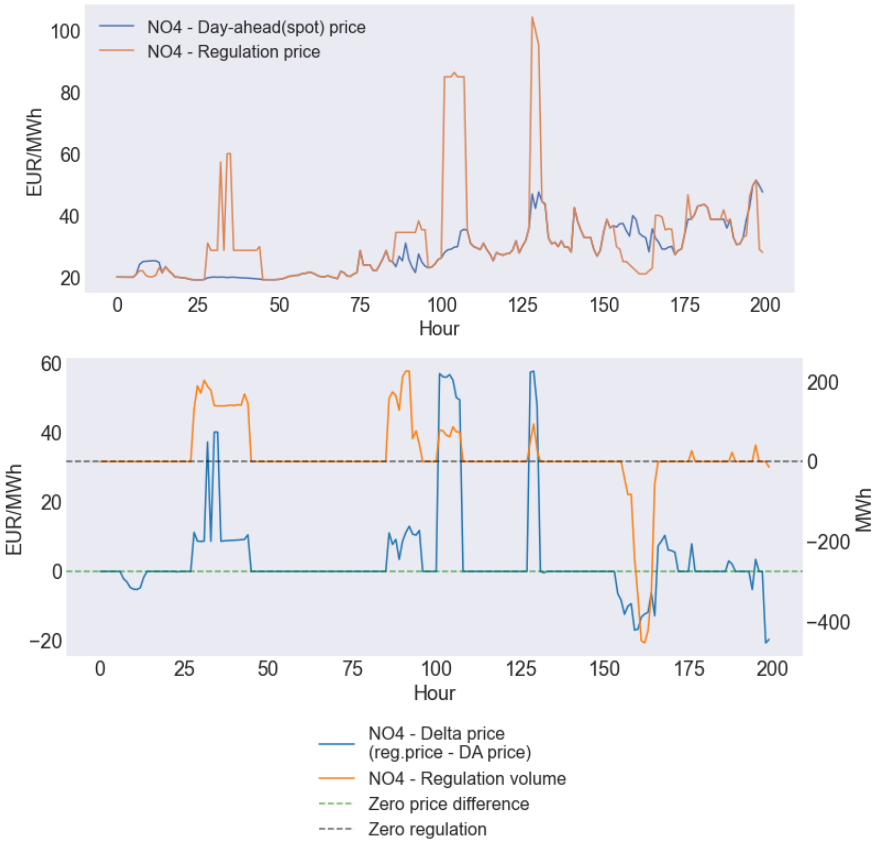


Figure 4.27: Day-Ahead price, regulation price and delta price over 200 hours between 2021-01-03 at 01:00 (UTC) and 2021-01-11 at 08:00 (UTC)

Besides the constructed feature Δpr, another strong predictor describes regulation volume denoted $\Delta prod$, where prod refers to production. This is the delta between actual hourly produced power (supply) and the production prognosis (expected or forecasted

supply) formulated as follows:

$$\Delta prod = prod_{actual} - prod_{prognosis}$$

The production delta was not investigated as potential predictor for regulation volume by Skytte [6], Jaehnert et al. [29] and Klæboe et al.[27]. When asking domain experts, it seems commonly known that the triggering of tertiary reserves, to a large extent, can be viewed as the deviation between planned production and actual production or the deviation between expected consumption and actual consumption. However, the formulated $\Delta prod$ has not been found in newer research, where its descriptive potential for regulation volume is investigated on recent data.

Figure 4.28 shows the relationship between actual power production, production prognosis, $\Delta prod$ and regulation volume for the exact same time-span investigated for Δpr in figure 4.27. One can see in the bottom plot in figure 4.28 that $\Delta prod$ follows the regulation volume relatively precise. The difference between the hourly actual production and the production prognosis can be seen in the top plot. The portion of the difference between production and production prognosis can be seen in the upper plot, and the amount of regulation volume activated seems on par with the difference in MWh. This indicates a strong linear relationship between $\Delta prod$ and regulation volume.

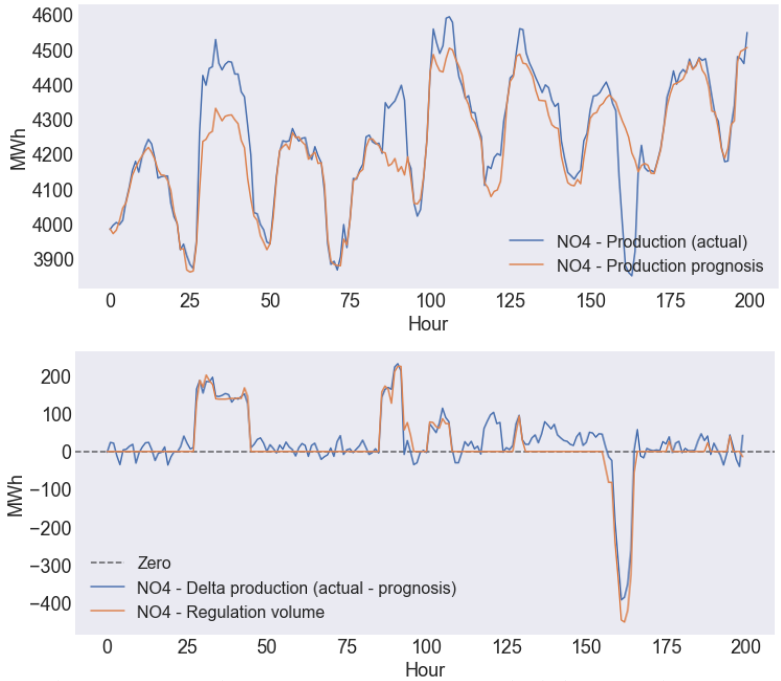


Figure 4.28: Production, production prognosis, and delta production over 200 hours between 2021-01-03 at 01:00 (UTC) and 2021-01-11 at 08:00 (UTC)

Figure 4.29 shows both the Pearson and Spearman correlation relationship between Δpr, $\Delta prod$, and regulation volume (vol_{reg}). As mentioned in related work, Jaehnert et al. found in 2009 in the 2003-2007 data set that $\varrho(vol_{reg}, \Delta pr) = 0.7811$ for NO1. Furthermore, Klæboe et al. [27] found that the *pearson correlation coefficient had declined from 0.78 in the 2003–2007 NO1 data set to 0.47 in the 2010–2012 NO2 dataset* [27]. Klæboe et al. [27] compared using NO2 since the current NO2 bidding zone was formerly a part of NO1.

When looking at Pearson correlation for NO2 in 2021 data between Δpr and regulation volume, it computes to 0.42 in 4.29, which is lower than what Klæboe et al. [27] found for 2010-2012 and significantly lower than what Jaehnert et al. discovered in the 2003-2007 data set. The Pearson correlation coefficient between Δpr and regulation volume is shown in figure 4.29 to be 0.36 for NO1. The bottom line is that the Pearson correlation between Δpr and regulation volume seems to decrease over time. However, the compared numbers may not be completely solid due to the market change for NO1 and NO2.

Moreover, looking at the ranked correlation in 4.29, the **Spearman** correlation, reveal a correlation of 0.91 for NO2 in the 2021 data (Δpr vs regulation volume). This is significantly higher than what both Jaehnert et al. and Klæboe et al. [27] found using Pearson correlation and comparing the Pearson correlation of 0.42 in the 2021 data. Due to the Spearman correlations ability to evaluate a monotonic relationship, the results indicate that Δpr has adopted some non-linear attributes comparing results from previous research and the differences between Pearson and Spearman for 2021. The Spearman correlation coefficient seems to be consequently higher than Pearson correlation when comparing the relationship between Δpr and the regulation volume for all zones (NO1-NO5). The Spearman correlation incorporate the same range as Pearson, $\varrho(vol_{reg}, \Delta pr) \in \{-1.0, ..., 0, ..., 1.0\}$

When looking at the correlation of the relationship between $\Delta prod$ and regulation volume, one can see from figure 4.29 that Pearson correlation yields a higher correlation coefficient than Spearman for all zones. For example, NO2 has a correlation of 0.91 between $\Delta prod$ and regulation volume, while the Spearman correlation shows a correlation of 0.77. This indicates that the Pearson correlation emphasizing a linear correlated relationship is more likely to represent better the relationship between $\Delta prod$ and regulation volume. The way the relationships between regulation volume and price and production are developing over time motivates the use of deep learning for time-dependent modeling of regulation volume.

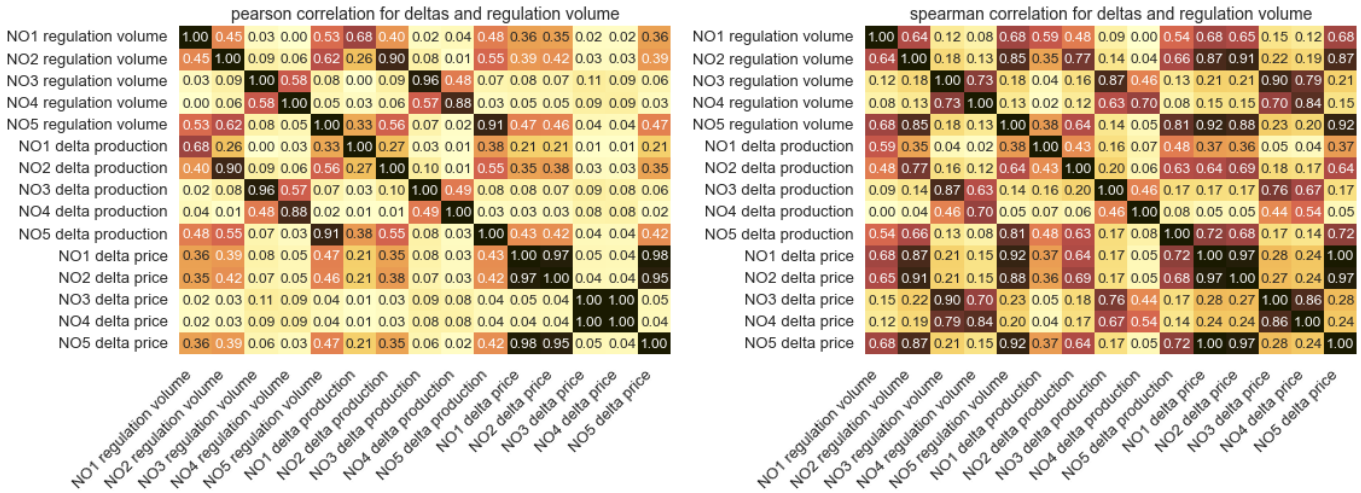


Figure 4.29: Pearson and Spearman correlations between regulation volumes and delta price and delta production for 2021 data.

Comparing Figure 4.30 against Figure 4.31 visualize the change in relationship between

delta price and regulation volume comparing 2016 with 2021. The observed increased complexity motivates to use advanced deep learning methods over the classical statistical methods used in previous research. Methods such as deep recurrent neural networks are able to model the changes happening over time when a lot of data is fed to the model during training. In contrast the relationship between production delta and regulation volume is still linear today as seen in Figure 4.32 and 4.33.

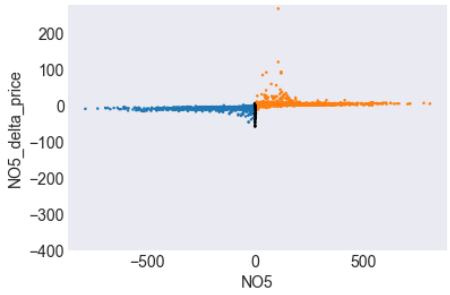


Figure 4.30: NO5 delta price against regulation volume - 2016

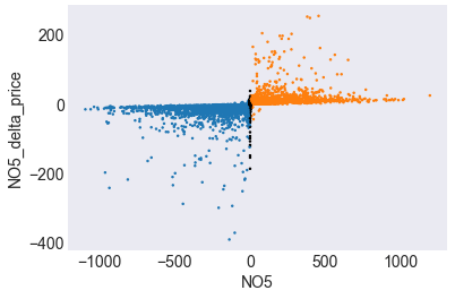


Figure 4.31: NO5 delta price against regulation volume - 2021

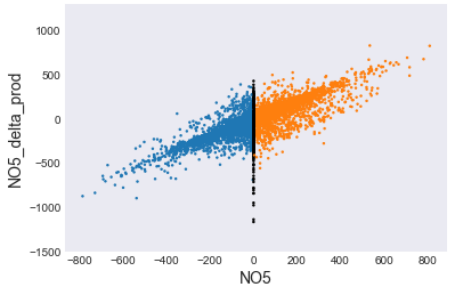


Figure 4.32: NO5 delta production against regulation volume - 2016

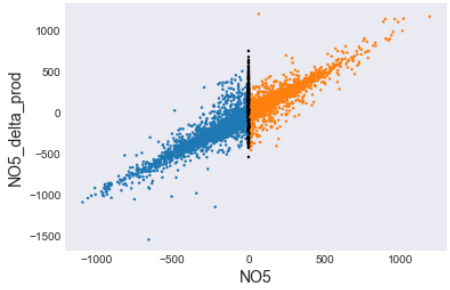


Figure 4.33: NO5 delta production against regulation volume - 2021

4.13 Tertiary reserves activation - case studies

Two case studies were conducted to observe specific events and their effect on the activation of the tertiary reserves (regulation volumes). One case observes the GYDA storm in the context of weather influence on the grid. The other study observes an unforeseen failure of the North Sea Link HVDC (High-Voltage Direct Current) cable connecting England with the Norwegian markets as a disruptive market coupling event.

4.13.1 Gyda storm

The GYDA storm was swiping through Norway in the middle of January 2022, with a wind said to occur every ten years potentially[13]. With todays technology, it is not difficult to tell that a storm is coming and where it will strike Norway. However, it is more difficult to know precisely how the weather impacts locally from hour to hour. Therefore, to some degree, the power sector can make general plans for the storm to avoid power production, trade, or grid balancing issues, but local effects and events happening secondary to the weather is mostly not feasible to foresee.

Figure 4.34 show the activation of regulation volumes for all zones from January 15-21. The weather prognosis from Netweather for 16-17 January is plotted inside the graph. The image shows the wind jet-streams several kilometers up in the atmosphere[14]. Several power blackouts in the grid were observed in different regions in Norway during the storm. At 2022-01-17 03:30 CET, a major power blackout occurred in NO3. The blackout event forced the TSO to continuously balance the grid for NO3 and NO4 until evening 2022-01-19.

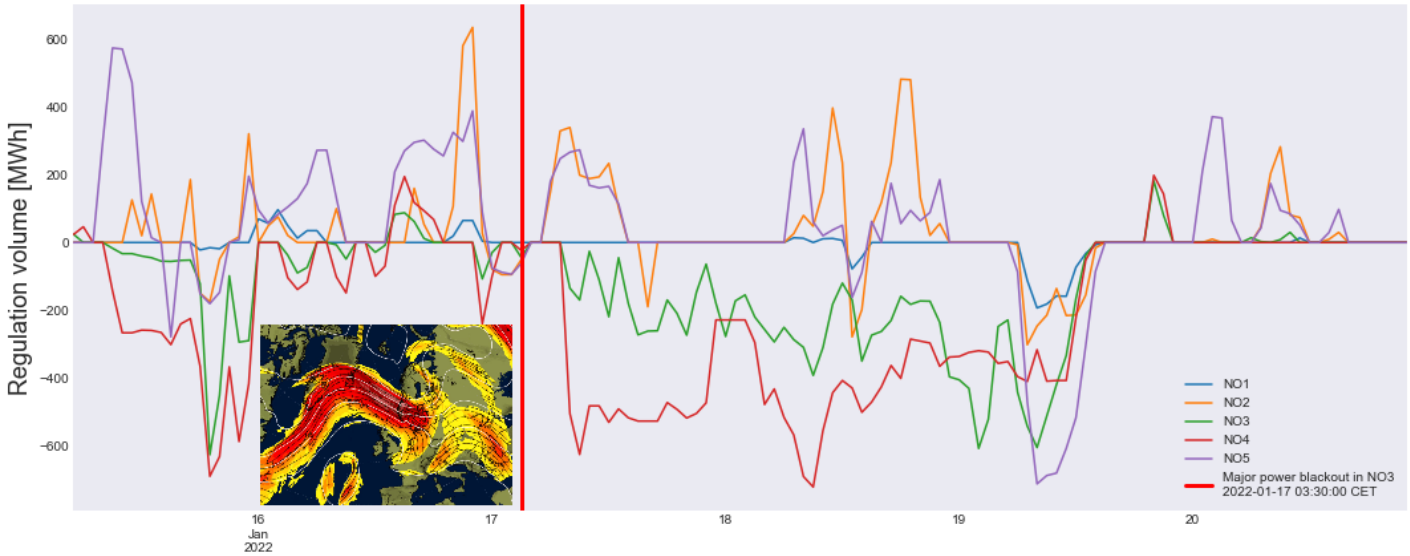


Figure 4.34: GYDA storm

Figure 4.35 show the local statistics in the same duration as the power blackout. The mean and standard deviation clearly change during the course of the event and normalizing post the event.

[13] https://www.vg.no/nyheter/innenriks/i/0GOKyM/varsler-full-storm-kan-skje-hvert-tiende-aar
[14] https://www.netweather.tv/charts-and-data/jetstream

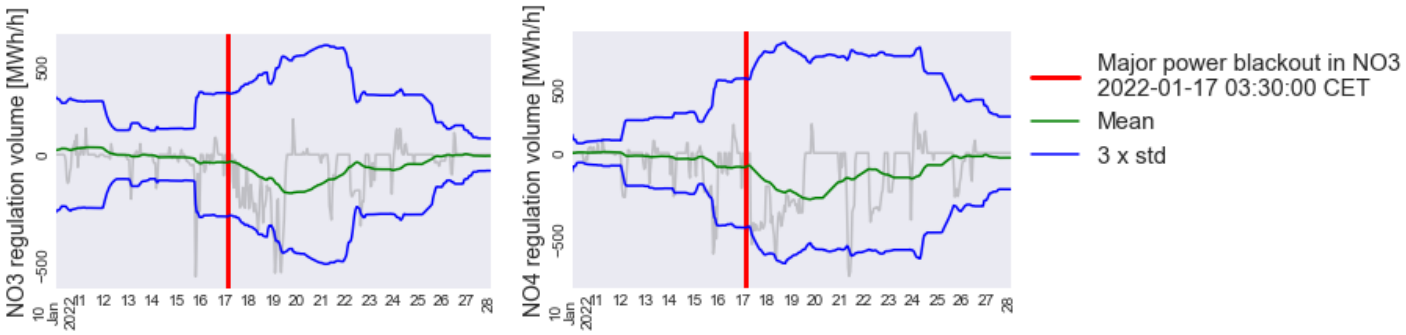


Figure 4.35: Local statistics for mean and 3 standard deviations during the Gyda storm.

Figure 4.36 intends to compare January month in 2022 (the Gyda storm month) with January over several years. No clear patterns can be seen in whether regulation volumes in January 2019-to 2021 are significantly different from 2022 when considering events triggered by the Gyda storm. Another extreme weather happened in the middle of January 2021 named Frank. Characteristics of Frank were extremely strong gusts of wind. The Gyda storm also had strong wind gusts but was primarily known for its extreme amounts of precipitation. It is not possible to get an intuition for the degree to which weather influence the grid directly on indirectly from a historical view shown in Figure 4.36. It may be that the weather itself is not causing the need for regulation, but rather the deviation between the weather forecast, used for planning and strategies, and the actual weather occurring.

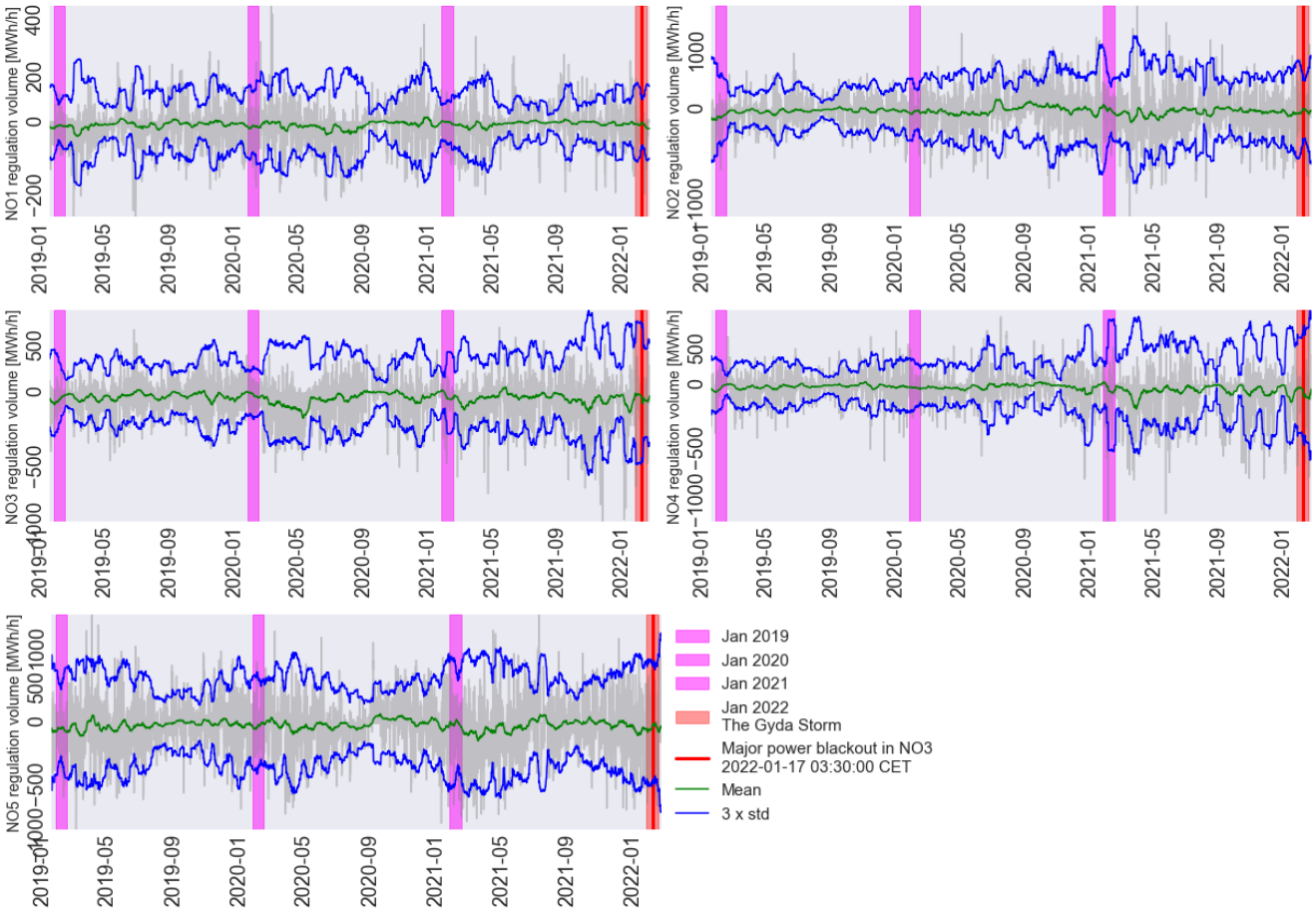


Figure 4.36: Comparison of local statstics during the Gyda Strom to previous years

4.13.2 North Sea Link failure

The previously discussed North Sea Link (NSL) cable, co-owned by Britain's National Grid and Statnett, experienced an unforeseen failure in 2021. A fault detected in the converter at the UK end led to a power transmission malfunction[15]. The top plot in figure 4.37 reveal a major response to grid balancing for several zones during the event. NO4 seems to be the least affected, which is logical since NO4 is geographically furthest away from NO2, which hosts the Norwegian connection side of the cable. NSL has a max capacity of 1400MW, representing a potential massive load connected to the grid. When a load at this scale suddenly disappears from the grid, down balancing (regulation) on the grid is inevitable in the regulation market. After NSL came back into operation (the green vertical line), the cable again put a massive load to the grid resulting in grid balancing through up-regulation and changes to the power production. An Urgent Market Message (UMM) was published by Nord pool at 19:30, one hour after the event started. The cable is back in operation at 17:00 November 10, and a UMM was published at 15:00, indicating that the cable will be back in operation15. A UMM is a message that is used to present urgent market information, such as planned or unexpected events concerning; transmission, generation, and consumption. It is hypothesized that UMM messages can be a potential predictor for tertiary reserves activation. However, humans manually type some messages, and utilizing UMM requires a deep dive into the Natural Language Processing field. Therefore, UMM is out of scope for this thesis.

The middle plot in figure 4.37 shows power production with an hourly resolution. A drop from approximately 5500MW down to about 3500MW of aggregated power production is observed from the start of the event until early next morning, November 10th, in NO2. This drop is a lot more than the max capacity of NSL, which reflects how sudden events can disrupt the market. After 12:00 November 10th, a production is ramp-up can be observed, even though the cable is not in operation. This is because the situation and its circumstances are included in the closing of the Day-Ahead market at 12:00.

In the bottom plot in figure 4.37, both the regulation price and day-ahead price can be observed during the event. It seems that only the regulation price in zone NO1, NO2, and NO5 is significantly affected by the event.

[15] https://umm.nordpoolgroup.com/#/messages/c5e25a9c-9863-44d3-ac9d-02170ec17bbe/6

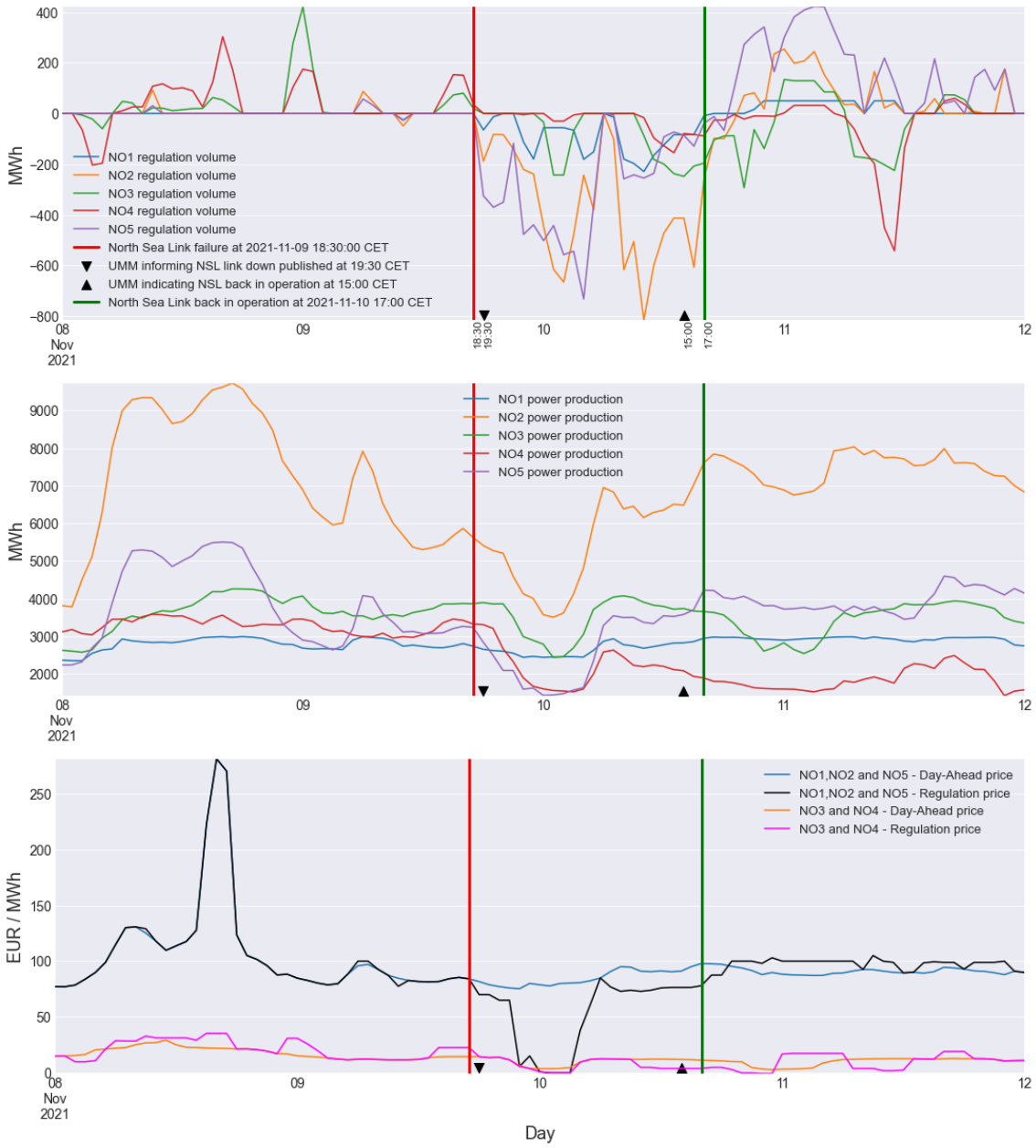


Figure 4.37: NSL failure, UMM and market reactions

4.14 Inaccuracy and uncertainty in weather forecasts

This section serves as a pre-study on weather forecast inaccuracy and uncertainty. The pre-study outline in this section was initially conducted to indicate if weather forecast uncertainties may affect regulation volumes. Furthermore, this study intends to contribute to determining if weather forecast uncertainties influence the activation of tertiary reserves. In addition, downstream of this pre-study, an entire experiment, experiment 2 in Section 5.4 was conducted as a continuation of this pre-study. This pre-study and experiment 2 form the foundation to test the research question hypothesis that weather forecast uncertainties influence the need for grid balancing through the activation of tertiary volumes.

The term inaccuracy is used to describe the difference between two images $\Delta image$, e.g., between actual weather and weather forecast. The term uncertainty is later used to represent the image statistics computed from the $\Delta images$. For the machine learning part and experiment 2 in Section 5.4, only the term $\Delta images$ and uncertainty are used. Beside the $\Delta image$ dataset created, a MAD dataset is constructed from the $\Delta images$ and consists of data from 6 different areas within the $\Delta images$ for every type of weather image (eight types in total). The MAD dataset is then a series of 8760 records (hours in 2021) having $6 \times 8 = 48$ features per observed hour of 2021. Analysis of $\Delta images$ relative to regulation volume is performed in Subsection 4.14.3, and analysis of uncertainties relative to regulation volume is performed in Subsection 4.14.4. In addition to performing data analysis on weather uncertainties, experiment 2 includes machine learning methods using calculated image statistics (estimated uncertainties), and modelling using $\Delta images$ directly.

4.14.1 Decisions on data gathering

Estimating inaccuracy in weather forecasts is done in two ways. One uses the difference between actual weather against the weather forecast of the same hour. The other is by skipping the actual weather by estimating the difference between old and updated weather forecasts of the same hour. The motivation of only using weather forecasts is that it enables to estimate the forecast inaccuracy and uncertainty into the future, not just for the current or historical hours. Working with only weather forecasts is more complex in terms of data utilization and is therefore left out of the exploratory data analysis and included as part of experiment 2 in Section 5.4.

A simplification of the problem of constructing the differentiated weather images was made since new weather forecasts 60 hours into the future are published at 00:00, 06:00, 12:00, and 18:00 every day. Therefore, only one of the four published forecasts mentioned was chosen, and given that forecasts contribute to planning the day-ahead market prior to its clearing, it is logical to focus on weather forecasts performed prior to the market-clearing at 12:00. This means that the weather forecast published at 00:00 or 06:00 is most likely the forecast used for planning the day-ahead market. For the analysis in this section, it was decided to focus on the 24 hours forecast starting at 00:00 (UTC) to make the analysis more feasible since it avoids tackling hours and days with overlapping forecasts when looking at a 24-hour sequence (one day). However, using the 00:00 weather forecast over the 06:00 may be suboptimal since the 06:00 published weather forecast is closer to the day-ahead market clearing. Therefore, the more advanced method of examining the weather forecast in Section 5.4 is based on the 06:00 forecast.

4.14.2 The data creation process

The constructed data set of $\Delta images$ consists of eight weather type images for each hour in 2021, a total of 69368 images (35GB of data). The difference between the weather forecast and actual weather computed is formulated as follows:

$$\Delta W^{i,j}_{weather\ type,h} = \forall_i \forall_j \sqrt{(A^{i,j}_{weather\ type,h} - F^{i,j}_{weather\ type,h})^2} \tag{4.1}$$

Where ΔW represents the difference or inaccuracy weather image also referred to as the delta image denoted $\Delta image$, A the actual weather image, F the weather forecast image, and i, j represents the pixel position. h represents a given hour (date and time) in the time-series sequence in 2021, and the *weather type* is defined as $\in \{$ *air pressure at sea level*, *air temperature*, *wind speed*, *wind direction*, *relative humidity*, *cloud fraction area*, *precipitation amount*, *surface downwelling shortwave*$\}$.

A vectorized method of calculating the delta image for a given weather type at a given hour is shown in listing 4.1.

```
    def delta_image(actual_image, forecast_image):
        return np.sqrt((actual_image - forecast_image) ** 2)
```

Listing 4.1: Python calculation of delta images

Wind direction is computed differently than the other weather type images, since the pixel values represents direction stemming from the U and V componts, where U is the horizontal wind component towards east parallel to the x-axis (longitude), and V is the horizontal wind component towards north parallel to the y-axis (latitude). Since the pixel values reflects the angle of the wind direction from the components, the 8bit image range is first converted into a range between $0 - 2\pi$. After that the sinus and cosinus of the difference between actual and forcasted image is computed. Lastly the absolute value of the *arctan*2 is computed to get the resulting pixel value in radians in positive values only. To make the computation correct for all 4 quadrants in the unite circle and to to avoid ambiguity in the returned value, the *arctan*2 function is used instead of the regular *arctan*[16]. Lastly the computed wind direction delta or differentiated wind direction image is converted from radians to an 8 bit unsigned image between 0-255. The code covering the operation can be seen in listing 4.2

```
def pixel_values_in_radians(img):
    # Pixel values scaled between 0 to 2pi
    return img * ((2*np.pi)/255)

def delta_image_wind_direction(actual_image, forecast_image):
    a_transformed = pixel_values_in_radians(actual_image)
    f_transformed = pixel_values_in_radians(forecast_image)

    # Sinus and Cosinus images computed
    sinus_diff_image = np.sin(a_transformed - f_transformed)
    cosinus_diff_image =  np.cos(a_transformed - f_transformed )

    # Absolute value of the invers tan2 of the sinus and cosinus image
    return np.abs(np.arctan2(sinus_diff_image, cosinus_diff_image))
```

Listing 4.2: Python calculation of delta image for wind direction

[16] http://tornado.sfsu.edu/geosciences/classes/m430/Wind/WindDirection.html

As previously mentioned, the estimates of uncertainties in this section were based on the difference between weather images for the entire 2021. From the $\Delta image$ process, the mean value and the Mean Absolute Deviation (MAD) were computed, representing extracted first-order image statistics. MAD is used for quantifying the deviation by measuring the dispersion of data within each region in the image. MAD uses the L1 distance (Manhattan distance) between the data (regions in the image) and the mean of that region. MAD is chosen over standard deviation because of its robustness to outliers since MAD calculates the absolute value while standard deviation computes the squared value.

$$MAD = \frac{1}{n}\sum_{k=1}^{n}\left|x_k - \frac{1}{n}\sum_{k=1}^{n}x_k\right| \tag{4.2}$$

Where n stand for number of pixels in the image, or sub regions of the image, x represents the pixel value at pixel k. The MAD calculations were performed for the entire global space of the image (NO) and locally for each bidding zone NO1-NO5. At the same time, NO1, NO2, NO3, NO4, and NO5 were computed using latitude, and longitude coordinates from the Norwegian Meteorological Institute projected down onto the image using the Lambert conformal conic projection. Projecting the coordinate onto the image constructs a non-convex polygon, forming a pixel position mask representing each bidding area. The latitude and longitude coordinates were downloaded from Meteorological Institute by the KoBas project as geojson files and can be found in the Meteorological Institute data download services. The Lambert conformal conic proj4 string used as a parameter for the projection can be found in the Meteorological Institute Github repository[17]. Figure 4.38 shows the polygons representing pixel positions for each bidding zone plotted on top of a $\Delta images$. The yellow dotted line around the image represents the global image referred to as zone NO. MAD is computed relative to all the areas (local and global) shown in the figure.

[17] https://github.com/metno/NWPdocs/wiki/Examples

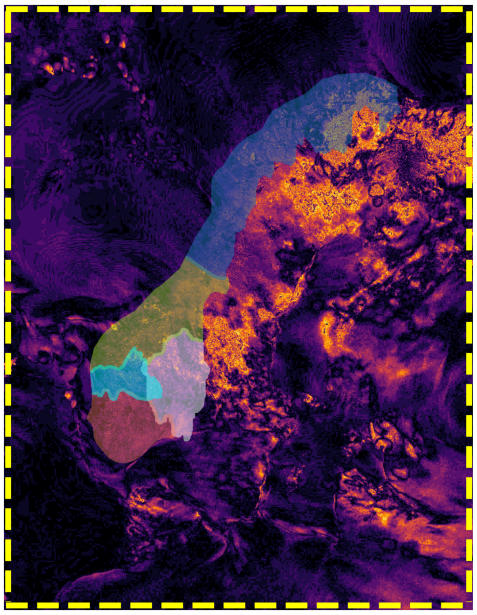


Figure 4.38: Delta image for air temperature at 2021-12-04 **01:00** overlayed with polygons covering each bidding zones projected onto the image using latitude and longitude coordinates from the Norwegian Meteorological Institute and the Lambert conformal conic projection [3]

If zooming in on Figure 4.38, one can see some bright orange-yellow areas in the background through the transparent polygons. The bright pixels values represents the mentioned inaccuracy. The MAD calculation measures the mean aggregated dispersion for all pixels covered by the mask relative to the mean of the entire area in the background covered by the polygon. Thus, the calculated deviation represents an aggregated measured quantity relative to its bidding zone which is regarded as being the weather forecast uncertainty in that area. This method is used to represent the prediction uncertainty of the weather forecast and not the model uncertainty of the MEPS model. Estimation of weather forecast uncertainty at a given hour using MAD is generally formulated as follows:

$$\textit{weather forecast uncertainty} = MAD\Big((\Delta W_{\textit{weather type},h})_{\textit{zone}}\Big) \tag{4.3}$$

Where the MAD statistics is computed relative to the area reflecting the mentioned global or individual bidding zones specified as: $zone \in \{NO, NO1, NO2, NO3, NO4, NO5\}$. The ΔW, *weather type* and h was specified for equation (4.1).

4.14.3 Analysis of weather forecast inaccuracy

Figure 4.39 shows an example of a ΔW on the rightmost image computed from the actual and forecasted images. Note that upper three plots in the figure represents air temperature in 2021-12-04 **01:00** (UTC), while the three lower plots represents air temperature at 2021-12-04 **01:00** (UTC). All plots are visualized using a the inferno colormap to highlight the differences between the images. The zero value in the color bar indicate no difference between the actual weather and the weather forecast. Brighter colors indicate significant differences. It can be observed that the ΔW image from 2021-12-04 **16:00** has more bright areas than the ΔW image from 2021-12-04 **01:00**. This shows

that the weather forecast for air temperature at 16:00 deviate a lot more from actual weather than what is observed for 01:00, which designates hour 16:00 to comprise a greater inaccuracy than observed for hour 01:00. Similar situation can be observed for the wind speed ΔW image in figure 4.40. It can be visually observed that there is zero difference between wind speed weather forecast and actual wind speed at 2021-12-04 **01:00**, while 2021-12-04 **01:00** revels a significant difference in the ΔW image. Interestingly, table 4.11 reveals that activated tertiary reserve volumes at 2021-12-04 **01:00** are zero for all zones. In contrast, 2021-12-04 **16:00** had a lot of regulated power at the same time as weather forecasts of several weather types deviated significantly from actual weather. Other weather types were also investigated at hour 01:00 and 16:00 showing weather forecasts at hour 01:00 not deviating much from actual weather, while hour 16:00 deviated significantly more. However, this observation calls for a statistical analysis rather than an image-by-image investigation, which is done in Subsection 4.14.4.

Regulation at hour	**NO1**	**NO2**	**NO3**	**NO4**	**NO5**
2021-12-04 01:00	0	0	0	0	0
2021-12-04 16:00	0	50	-288.2	-183.1	147

Table 4.11: Regulation volumes in MWh at 2021-12-04 **01:00** and 2021-12-04 **16:00**.

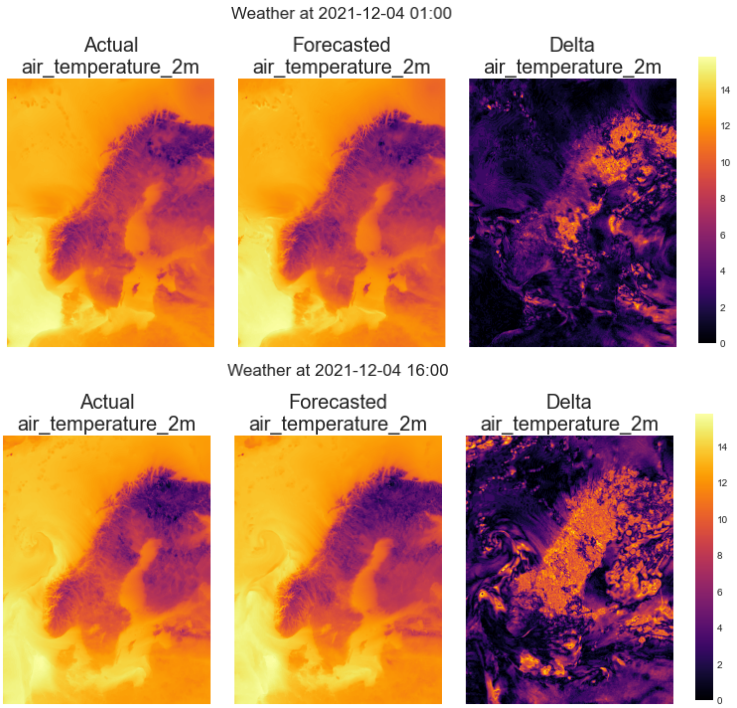


Figure 4.39: Actual, forecast, and delta image of air temperature at 2 meters above surface in 2021-12-04 **01:00** and 2021-12-04 **16:00**.

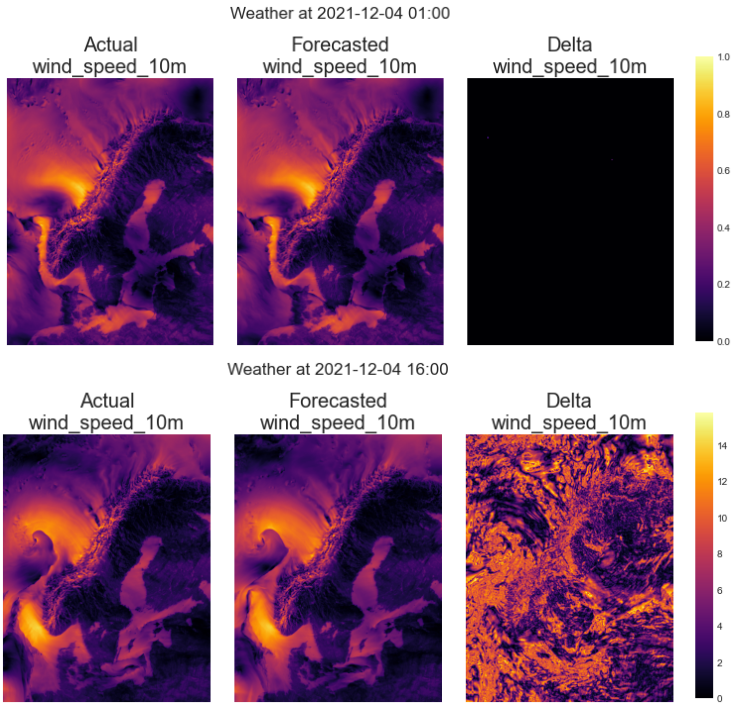


Figure 4.40: Actual, forecast, and delta image of wind speed at 10 meters above surface in 2021-12-04 **01:00** and 2021-12-04 **16:00**.

4.14.4 Analysis of weather forecast uncertainty

Figure 4.41 show boxplot statistics of the quantified uncertainties for air temperature and wind speed grouped on hours. The uncertainties of wind speed forecasts seem to be reasonably low the first three hours after the forecast conducted at hour 00:00 and increase considerably after that. Air temperature seems to increase more gradually. It seems that the weather forecast uncertainties reach a plateau at some point. The other weather type are not shown, but they indicate a similar saturation of uncertainties as observed for wind speed in figure 4.41

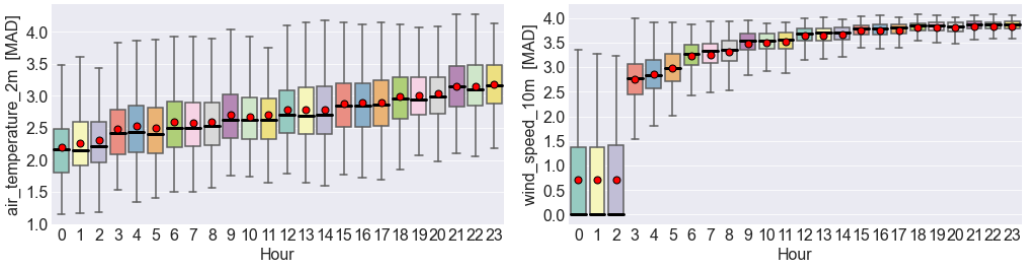


Figure 4.41: Weather uncertainties for air temperature, and wind speed calculated from delta images using aggregated MAD from each zone mask and globally.

Figure 4.42 illustrates the relationship between weather forecast uncertainties and regulation volumes. The y-axis represents the absolute Spearman correlation between wind speed and direction. The x-axis corresponds to the weather type and the region from which the uncertainty is calculated. The correlation coefficient value is not very strong maxing out at a correlation of 0.25 at most. The plots in Figure 4.42 are the plots of the two hours having the highest peak correlations found in the investigated data.

Bidding zone NO3 had the highest correlations at hour 21 for wind speed and direction, while other zones dominated other hours for different weather types. The power source mix analysis in the Section 4.4 revealed that NO3 had the highest portion of power from the onshore wind of all zones, as was shown in Figure 4.3 which may describe why NO3 has the highest correlation seen in Figure 4.42. All correlations observed were generally low. However, given that the Spearman correlations measure a non-linear relationship and that it is not zero, a machine learning model, particularly a deep learning model, could potentially utilize deep associations between regulation volume and weather forecast uncertainties. Another takeaway to consider from the wind speed plot in Figure 4.42 and other plots not included in the report was that uncertainties calculated globally (NO) from the ΔW often had higher correlations than individual zones NO1-NO5. This motivates a machine learning model to prioritize the entire spatial domain of the weather image rather than single individual zones. It might be that weather and weather forecast uncertainties occurring outside a particular bidding zone still influence the regulation within the bidding zone. Additionally, available transmission capacity between zones also plays a role in the regulation market since it enables power to flow. Therefore, uncertainties must be connected and shared between zones to some degree. Weather forecast uncertainties influencing the need for regulation in one particular zone may indirectly trigger grid balancing in another zone through available transmission cables.

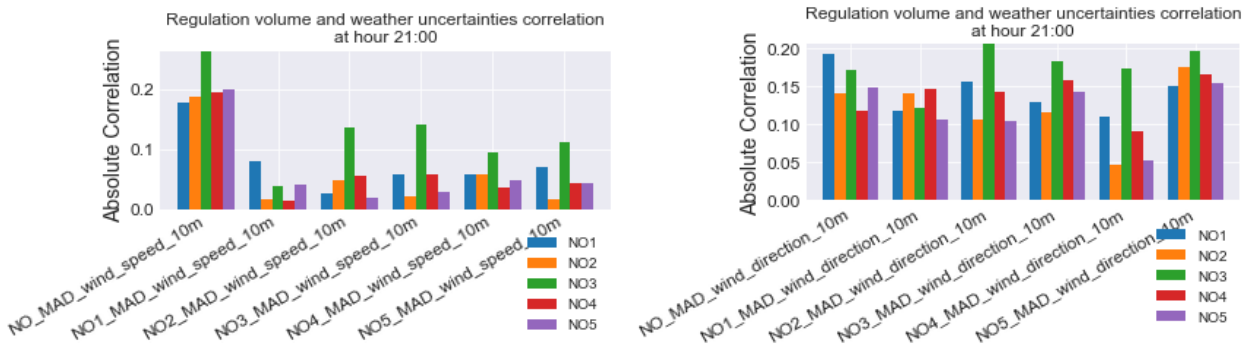


Figure 4.42: Correlation between regulation volume for every bidding zones and weather forecast uncertainties calculated using MAD from delta images of wind speed and wind direction.

Figure 4.43 is a high-level overview over all correlations found in the correlation analysis. The purpose of the heatmap is to get an impression of which hour (0-23) at which weather type that has the highest correlations. The heatmap shows hourly correlations between uncertainties and regulation volume with the mean correlation of each individual zone and the mean across all zones after that. Thus, the value in the heatmap is a mean of a mean quantity and is lower than the observed peak correlations. The aggregated correlation information in the heatmap is plotted relative to hours on the y-axis and the weather type on the x-axis. One can see from the plot that the correlations vary between different weather types and for different hours. This plot indicates that weather forecast uncertainties may contribute to activating tertiary reserves differently from hours and weather types. However, the influence of weather forecast uncertainties seems fractional if just a single weather type at a given hour is considered. Therefore, fragments of associations between different weather types and regulation volumes may add up to together be a significant descriptor of tertiary reserves activation when utilizing deep

learning methods. It was found that by computing the correlation for some specific months (not the entire year 2021), the situation in figure 4.43 completely changes, and air pressure, precipitation, and wind direction may not yield the highest correlation as seen in the figure.

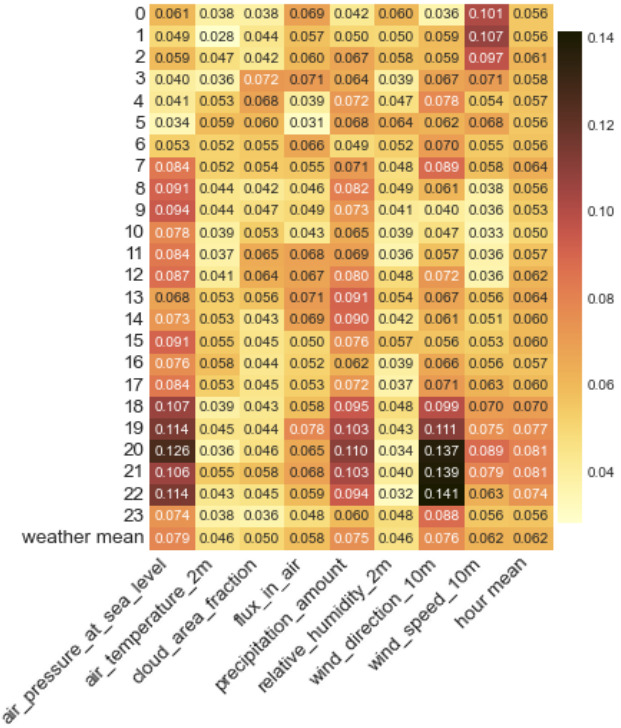


Figure 4.43: Mean of correlation between regulation volume and weather uncertainties over all bidding zones and weather types, sorted by the hour of the day.

4.15 Data analysis takeaways

This section partially concludes on some of the research questions and sets the direction for the remaining part of the thesis. The full conclusion will be given in Section 6.1

4.15.1 Noise and stationarity

The related work and the data investigation show that the regulation volume is noisy and very complex. The Universal approximation theorem state that, "*any function can be approximated by an appropriate neural network method*" [44, 45]. However, the Universal approximation theorem is theoretical, and the data utilized sets the boundary for what is feasible to achieve in practice. The exploratory data analysis reveals that regulation volume is not white noise. While the mean value is very close to being constantly zero through time, the standard deviation is not constant, and the autocorrelation does not equal zero when considering all lags. The regulation volume is stationary, which is a requirement for white noise but some seasonal components were found for the regulation volume in all zones; NO1, NO2, NO3, NO4, and NO5.

4.15.2 Outliers

In the light of regulation volume not being white noise, it can theoretically be modeled given sufficiently informative input features, and machine learning experiments are therefore relevant to pursue. Extreme outliers in the data may influence the model training such that it becomes biased toward the abnormal data. In general, a good model must be able to generalize to new, previously unseen input data without having any strong biases towards abnormal conditions. However, defining anomalies is not always straightforward. In the case of power regulation volumes, it is driven by situations deviating from expected plans. Thus, the power regulation is already in itself an abnormal condition that this thesis intends to describe using machine learning. Therefore, removing or changing data points for regulation volume marked as outliers is not considered an option in this case.

4.15.3 Correlations, features and model suggestions

Correlations for Δpr and $\Delta prod$ show a strong linear relationship between $\Delta prod$ and regulation volume on recent data, while Δpr looks to have an increasing amount of non-linear components in its relationship with regulation volume over time. This indicate a gain in price complexity over time. A deep recurrent neural network e.g., LSTM is therefore suggested, due to its ability to capture time-dependant patterns seen in some of the market data. Moreover, it is not advised to use Δpr or $\Delta prod$ directly as has been tried for Δpr in some previous work [6, 29, 27]. This is because the deltas can, in some ways, be seen as "actual" regulation volume for the current hour and do not necessarily describe the root cause of why the regulation volume appears both considering the current hour and the next hour. Therefore, it is more sensible to use day-ahead price, regulation price, actual production, and production prognosis directly such that a deep learning model can learn how the different time series relate. The regulation volume is clearly correlated with the previous hours but with a decreasing trend for increasing time difference. This motivates the use of lags as features for modeling regulation volume instead of using the computed Δpr and $\Delta prod$. Therefore, an experiment will be set up to include the features from target lags together with the actual day-ahead price, regulation price, actual

production, and production prognosis. The day-ahead market features are also known at least 12 hours forward in time, which enables using some future market data when forecasting regulation volumes. The impact of seasonal features is also interesting to investigate. However, given that the seasonal components found for regulation volumes are relatively weak, it is believed that a deep learning method would still be able to make use of the information. In addition, information regarding transmission flow and transmission capacities is relevant. The market coupling and connections between markets with the Nord Link and North Sea Link impacts the production, demand and regulation. However, only the transmissions between markets internal to Norway (NO1-5) are considered within the scope of the thesis. Additionally, the regulation volumes in the different bidding zones were discovered to be correlated, thus motivating to forecast the regulation volume for one specific zone while using previous hours of regulation volume for other zones as input features. The observed strong autocorrelation for the previous hour of regulation motivates an experiment using a univariate model that only use on time-series variable, the regulation volume.

Regarding model selection, pursuing modeling using different Autoregressive Models (AR) and Markov Models are considered out of scope and not relevant since this thesis focuses on deep learning using a large amount of data. Adding to this decision is that the power markets have changed a lot since the 2010-2012 data used by Klæboe et al. [27] and the 2014 data used by Dimoulkas et al. [31] on the Swedish market, which rules out the motivation for comparing the deep learning results with models used in related work for the univariate modeling.

The investigated weather and weather uncertainties do not indicate a clear or statistically significant relationship with tertiary volume activation. However, the correlations found were not zero. They were computed using the non-linear spearman correlation, which motivates the use of deep learning image models to find deep and complex patterns and relationships.

Chapter 5

Machine Learning experiments

5.1 Experiments overview

5 Machine Learning experiments were conducted which all are motivated by the raised research questions. Experiment1 is the machine learning part coupled to the research questions asking if there are any drivers of regulation volume, and to which degree regulation volumes are predictable or not. Experiment 2-5 is coupled to the research question about weather influence on the regulating power markets. The weather experiments also relate to the question on whether there are any significant drivers behind regulation volumes. The experiments and modeling mainly focus on regulation volume for the bidding zone NO5 to simplify the experiments and the model tuning for one particular zone. NO5 is chosen both because the partners in the KoBas project are interested in bidding zone NO5 and since NO5 is found to be the most complex and difficult bidding zone to model and forecast. With some tuning, the model performance for the bidding zones NO1-NO4 is found to yield lower MSE values more easily than what is discovered for NO5. Additionally, as was discovered in the exploratory data analysis in figure 4.15, NO5 has more activation of tertiary reserves and noise (the orange residuals in the figure) than the other zones making it the hardest bidding zone to forecast. Figure 5.1, 5.2, 5.3, 5.4 show an overview over the major content explored during experiment 1-4. Experiment 5 is shown both as an overview and in more detail in figure 5.40. Experiment 1 is conducted using only marked and seasonal data, while experiment 2-5 include 2-dimensional weather data. The experiments are conducted using knowledge gained through related work and exploratory and extensive data analysis. Previous relevant work serves as inspiration on how to approach the problem using a different methodology using state-of-the-art methods. Experiment 1 is more comparable with previous work, while experiment 2-5 is known as completely unexplored ways of forecasting the regulation volume. Additionally, no research tackling NO5 in a similar manner was found, and no research was found to be directly comparable. The experiments use the most recent available data, which no previous research has been found to use. The experiments also serve to address further the research question, which was also addressed during the exploratory data analysis. Experiment1-4 is evaluated using the same validation and test data for comparability reasons. Experiments 2 and 3 only have training data for 2021 due to storage and download complexity, while experiments 1 and 4 include training data from 2016 to 2021. The calculated market delta values (production delta and price delta) were not

used directly as input features to the models since target lags were used. However, the actual data behind the deltas (day-ahead price, regulation price, actual production, and production prognosis) were used in their original form. This choice intends not to include delta features that are almost equal to the regulation volume lags but rather to make it up to the models to learn the delta combinations and how it relates to the regulation volume targets.

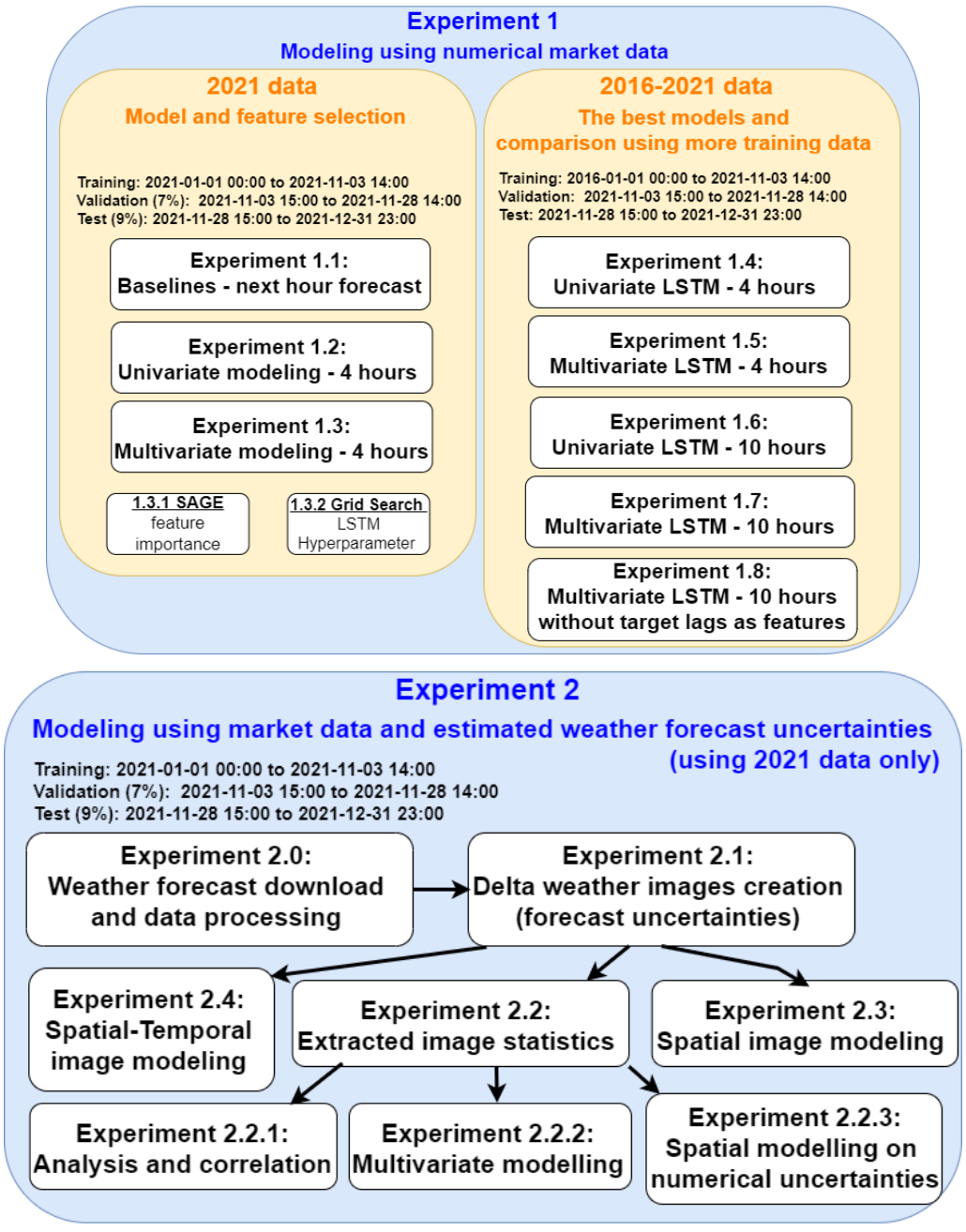


Figure 5.2: Overview over Machine Learning Experiment 2

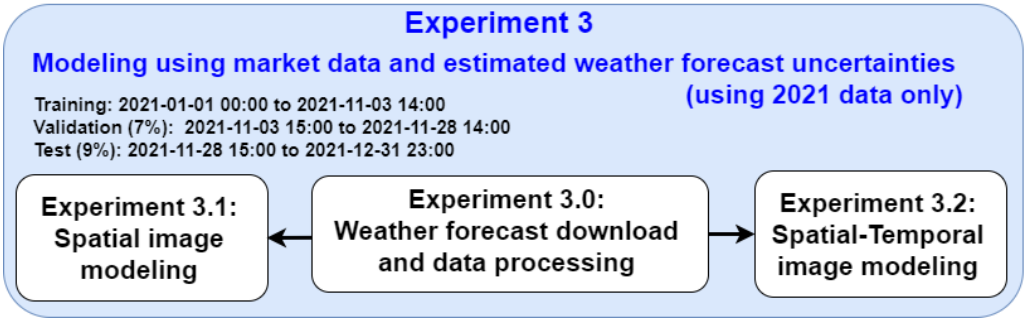


Figure 5.3: Overview over Machine Learning Experiment 3

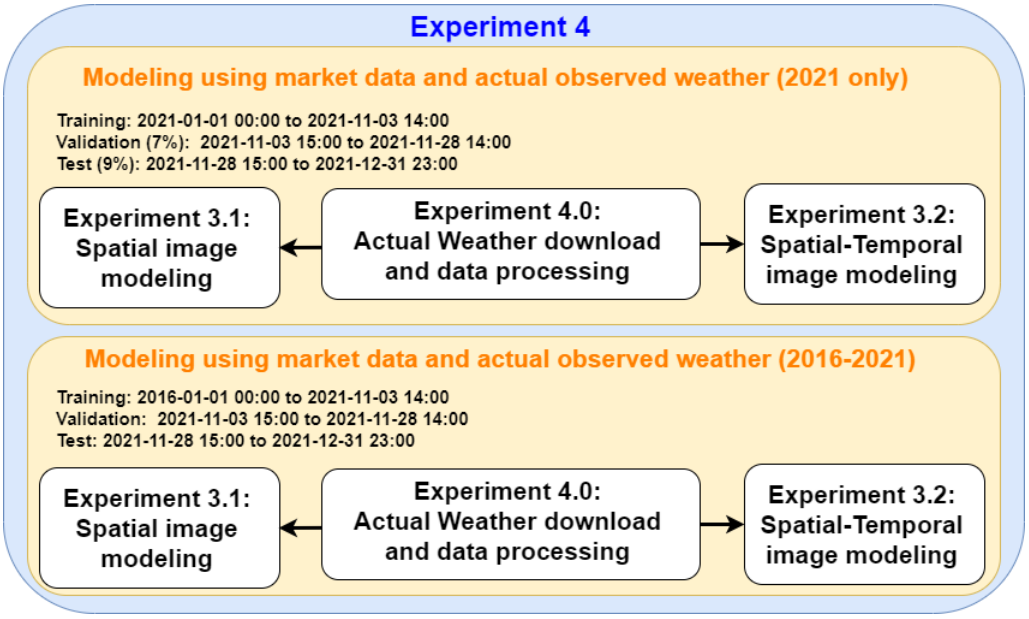


Figure 5.4: Overview over Machine Learning Experiment 4

5.2 Data pre-processing

5.2.1 Data cleaning

Python scripts were made in order to read in the downloaded raw data and to create a main dataset. One process for reading and cleaning data follows each experiment. For experiment 1, the numerical market data were read from the downloaded raw data and processed. All market data discussed in this thesis is processed and added to the main dataset. Empty or NaN (Not a Number) values were mainly replaced with zero values, and some few removed. Very few record of regulation volumes had NaN, and those found were removed instead of guessing the correct value. Seasonal features were created directly from the hourly time stamped data. Some consumption data were missing, and the the last hour were used to fill the missing data. Consumption delta, production delta, price delta were calculated and added to the dataset for analysis purpose. Some market data that were missing at NordPool were replaced from data in the entsoe database. Missing weather forecast images and actual weather images were marginally. Therefore, the missing weather forecasts data were note not replaced.

5.2.2 Data scaling and handling of outliers

No outliers were removed from the regulation volume target data based on the exploratory analysis part. Different types of scaling were investigated both for numerical market data and for target data. The MinMax scaler from sklearn[1] was finally chosen for target scaling. The MinMax scaler scales the data within 0-1 in floating value. The MinMax scaler is sensitive to outliers, but given that outliers constitute a significant portion of regulation volumes, scaling the regulation volume is not considered an issue using a MinMax scaler. Using a scaler sensitive to outliers may even be more appropriate in this time-series context, considering that the major part of the regulation volumes being modeled is defined as statistical outliers discovered during the exploratory data analysis. For simplicity, the input data were also scaled using a MinMax scaler. The standard scaler[2] where the mean is subtracted from the data folllowed by a division of the standard deviation was tried out. Standard scaling of both the input data and the targets did not reveal any noticeable improvements over the MinMax scaling. Another Scaler that was investigated was a scaler robust to outliers (Robust scaler)[3]. Scaling the data using the Robust scaler did not show any sign of improved model performance or faster convergence during training.

5.2.3 Sliding window and dataset creation

A windowing method named sliding window is utilized on the data to transform one single time-series data stream into a supervised learning problem. This concept exploits previous (historical) time steps while using subsequent time steps as targets. The benefit of this process is that time-series data can be transformed such that both simpler standard linear models and nonlinear models can model the data[21].

Framing a single variable or stream of data as a supervised learning problem maps the data into a function such that an input sequence of data can be used to describe an output sequence. Let w_i and w_o denote the window input and the window output at a given size. A function that maps the inputs into outputs can then be viewed as follows:

$$f : \mathbb{R}^{w_i} \rightarrow \mathbb{R}^{w_o}$$

This function represents the regression problem framed as a supervised learning problem where the w_i sequence represents the input data to learn from, and w_o represents the output data to forecast. A machine learning method can then be used to learn the function f that models the relationship between the data in w_i and w_o. A sliding window algorithm was developed to be able to model the described function and to have full control over the windowing process. It was necessary to be able to adjust the input window size, the output window size, and the stride. Another motivation for having full control over the algorithm is to utilize it multiple times for different data streams to combine several variables and form a multivariate setup. The algorithm work on one single data stream (one single data variable) at a time, forming a univariate windowed dataset of one single variable input into the method. The stride relates to how many data points to slide between each time slice.

The sliding window algorithm slides through the data and constructs time-slices as illustrated in table 5.1. In the example, the stride s is set equal to 1, and the data d is

[1] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html
[2] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
[3] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html

the datapoint at time t, e.g., regulation volume at a given hour. The gray-colored cells represent the input window w_i of size 4, and the blue-colored cell represents the output window w_o of size 4. Each input and output slice is stored in temporary vectors.

time	0	1	2	3	4	5	6	7	8	...	...	...	...	...	...
data	d_0	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	...	...	...	...	...	...
		$\vec{X}_0$				$\vec{T}_0$									

time	0	1	2	3	4	5	6	7	8	9	...	...	...	...	...
data	d_0	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	...	...	...	...	...
			$\vec{X}_1$				$\vec{T}_1$								

time	0	1	2	3	4	...	...	t-4	t-3	t-2	t	t+1	t+2	t+3	t+4
data	d_0	d_1	d_2	d_3	d_4	...	...	d_{t-4}	d_{t-3}	d_{t-2}	d_{t-1}	d_{t+1}	d_{t+2}	d_{t+3}	d_{t+4}
									$\vec{X}_k$				$\vec{T}_k$		

Table 5.1: Time-series data windowing using the sliding window algorithm

The windowed processed datastream is then constructed into matrices:

$$\mathbf{X} = \begin{pmatrix} \vec{X}_0 \\ \vec{X}_1 \\ \vdots \\ \vec{X}_k \end{pmatrix} \qquad \mathbf{T} = \begin{pmatrix} \vec{T}_0 \\ \vec{T}_1 \\ \vdots \\ \vec{T}_k \end{pmatrix}$$

Where the input data (features) in the design matrix is $\mathbf{X}$ and the target matrix $\mathbf{T}$ constitutes the vectors or samples from all time-slices $\{1, 2, 3...., k\}$ defining the dataset $\mathbf{D} = \{\mathbf{X}, \mathbf{T}\}$. The data is then later split into training, validation, and test, where some samples (time-slices) are reserved for training, some for validation, and some for test data.

Modeling using different input and output window sizes was tried. However, it was standardized using an output window of 4 since the KoBas project desired a 4-hour forecast horizon. When forecasting 4 hours into the future, it is logical to have an input window size of 4 or more. The size of the input window relates to historical data or lags. As investigated in the exploratory data analysis, lag1 is highly correlated with the next hour of regulation volume, while lag2 and below are marginally correlated with the next hour. Having a larger input window size than 4 may not be suitable since it may invoke even more noise in the modeling process. However, a 4-hour historical input window was used as a benchmark when modeling using other sizes for the input window (lags) were investigated. The sliding window algorithm is designed to take in a pandas dataframe and a key representing the column, the data, to apply the windowing process to. The returned dataframe is the processed data where the columns represent the different input window features and columns represent the output window to be used as target values. A prefix with the name Lag1, Lag2, Lag3, Lag4, up to LagN is used in the column to reflect the time $t-1, t-2, t-3, t-4, ..., t-N$. Similary a prefix t1,t2,t3,t4 upto tM is introdiced to the columns representing targets into the future $t+1, t+2, t+3, t+4, ..., t+M$. Where N and M represent the size of the input and output window. Note that using the sliding window method with an input window of size 4 makes it impossible to forecast the

first 4 hours in the data stream since a input window of 4 records must be used in order to forecast the fifth, sixths, seventh, and eight output in the output sequence. An input window size of 4 means that the new windowed dataset is shifted equally to the input window size compared the unprocessed data.

The process above is an example of a univariate setup using, e.g., only the regulation volume of NO5. This method is directly used for the univariate modeling in experiment1 - step1.2. Furthermore, several streams of time-series data must be included for a multivariate setup. This is done by using the sliding window method first on the regulation volume to generate a dataset of lags and target values of the regulation volume. Additionally, the sliding window is used on other data streams that are concatenated with the dataset constructed from regulation volume. The final dataset then can be described as:

$$\mathbf{D}_{\textbf{final}} = \{(\mathbf{X}_{\textbf{other market features}}, \mathbf{X}_{\textbf{seasonal features}}, \mathbf{X}_{\textbf{reg. volume lags}}), \mathbf{T}_{\textbf{reg. volume targets}}\}$$

Note that only the target matrix $\mathbf{T}$ for regulation volume is included as a target since the output sequence of the other data streams is not considered a target for modeling. To align the data for all data streams, they must be processed with the same input window size, such that hours are aligned and matched for all the window processed data streams, given the shift in hours that the input window introduces to the windowed regulation volume targets. Moreover, using a window size of 4 to process, e.g., a seasonal feature result is 4 features with prefix Lag1, Lag2, Lag3, and Lag4. One could use lagged versions of the input data for all other features. However, only the previous hour (Lag1) is kept to reduce the overall number of additional features in the multivariate setup.

The data into the future is already known for day-ahead price and the power production prognosis data since they both originate from the day-ahead market clearing. Therefore the next hour columns, the target1 (*t1*) from the $\mathbf{T}$ matrix from the data windowing process is added to the $\mathbf{X}$ matrix only for day-ahead price and power production prognosis. This feeds the model with some known future data for the features rooted in the day-ahead market. Giving the model only some future information from the day-ahead market and no future data of the regulation market results in giving the model only parts of valuable future information relevant for the regulation volume forecasts. However, the motivation for including future hours from the day-ahead market is to help the model learn how the day-ahead data trends into the future. Ideally, this also helps the model learn how the regulation market may trend when having historical day-ahead data, historical regulation market data, and some added information on future day-ahead data. The siding window process was also used for picking the correct images during spatial-temporal modeling. Furthermore, a significant work in hours was put into both downloading and organization of images for modeling, see Section 7.2.2 in Appendix.

5.2.3.1 Data for RNN and LSTM models

Recurrent Neural Networks, such as the LSTM model, require data to have a specific dimensional form. The LSTM model expects the training, validation, and test data to be structured as a 3D array on the form (batch size, sequence length, features), where the batch size is the number of samples, the sequence length is the number of historical time steps, and features is the number of features used. Therefore, the data created as described in the sliding window section is reshaped to be compatible with the LSTM model. This is done by grouping several time steps into a batch. The number of time steps (the sequence length) used for each batch represents how many time steps back

into the past the LSTM model looks when it learns how to forecast the future. The most optimal sequence length both for the univariate and the multivariate LSTM was found during a grid search covered in the multivariate modeling section.

5.2.4 Image normalizations and transformation

The image normalization is performed on the fly from the dataloader iteration. The pixel values are first normalized to be within the range 0-1 in floating-point value. After that, the images are scaled by subtracting a global mean value followed by a division of a globally calculated standard deviation according to the method:

$$Standardized\ image = \frac{Image - \hat{\mu}}{\hat{\sigma}}$$

Where $\hat{\mu}$ and $\hat{\sigma}$ is the sample mean and standard deviation estimated from a sample size of 40% of all images within each individual weather type (8 estimated $\hat{\mu}$ and $\hat{\sigma}$ values in total). This results in the images being standardized equally relative to the given weather type. The calculation of $\hat{\mu}$ and $\hat{\sigma}$ is done prior to model training, and the estimated statistics are fed to the dataloader prior to the modeling process.
Gaussian blur was also added as part of the image transformation process. However, whether or not to include Gaussian blur was included as a parameter to measure the effect with and without. The motivation behind introducing Gaussian blur is to apply image noise reduction as a low pass filter to filter out high-frequency noise. Another motivation was to see if the effect on Gaussions blur and its impact on spatial relations when the images are sequenced as time-series data. The kernel size of the Gaussian blur filter was set to default as a 3x3 kernel while also leaving an additional parameter enabling filter size adjustments during experiments. Image augmentations was considered not relevant. Augmentation can be used both to increase the size of the training data and by augmenting the images with the purpose of generalize better during training. It was decided not to augment the weather images during training to stay true to the nature of weather and its topographical location relative to each bidding zone. Another motivation not to augment the images is that the weather images are created artificially from the MEPS models (not by a camera) and incorporate underlying physics related to weather.

5.2.5 image resize and loss of statistical image properties

In the computer vision field, computationally available resources or the image size of what pre-trained image models are trained on often sets the bar for which image size to use for modeling. During the exploratory data analysis, the relationship between regulation volume and weather uncertainties, and weather in general, were found to be complex. When all 8 weather types are used for modeling, the original images must be significantly resized to have a computationally feasible problem. Standard image analysis techniques were utilized to compute image statistics for different image sizes. The motivation for this is to know how much the statistical properties within each weather type are affected when downscaling the images and which size that is suitable during modeling. The images size of 96×96 were found to be a optimal image size considering both the loss of image properties and computationally demands. Bicubic interpolation was used for the image resizing. See Section 7.2.1 in Appendix for more information.

5.3 Experiment1

Experiment1 contribute to answering the research questions that are not directed towards the influence of weather. Figure 5.5 outlines experiment1 and the different sub steps from baseline models and predictions in step 1.1 up to time-series modelling. Substep 1.2 trains univariate data models with the last 4 historical target lags as the only input features. Both the univariate and multivariate modelling use the sliding window method to construct the time-series data in a windowed manner. The best univariate models (BiLSTM and XGBOOST) are carried over to multivariate modelling in step 1.3. The linear regression model is replaced by ridge regression in step 1.3 as the increase number of features calls for regularization. The features used in the multivariate models are listed under substep 1.3 in Figure 5.5. Investigating feature importance is an essential part of the machine learning process to give the model optimal conditions for learning patterns and relationships. The global importance was investigated for a range of features using SAGE (Shapley Additive Global importancE)[4]. SAGE was used to determine global feature importance since the coefficients from e.g., Ridge regression and the local feature importance function of XGBOOST measure model-specific feature importance. SAGE measures a global data-oriented feature contribution to the model [46]. The grid search conducted under step 1.3 serves to find optimal hyperparameters primarily aimed at the more complicated multivariate BiLSTM model. However, the hyperparameters also improved the performance of the univariate model, as well as other BiLSTM models. In experiments 1.4 and 1.5 the best univariate BiLSTM and multivariate BiLSTM models trained on 2021 data were re-trained on data from 2016-2021 to see if model performance increases or decreases with additional data, considering the inherent stochastic noise in the regulation volume data. The seasonal feature "year" is added for the multivariate BiLSTM in step 1.5 since the data spans several years. Lastly, the univariate and multivariate models are pushed to their limit in experiments 1.6 to 1.8 using a forecasting horizon of 10 hours into the future. In experiment 1.8, the target lags are excluded to see if the features found during the exploratory data analysis carries enough information without the dominating information in target lags. Other power market related features than those listed were tried but were not included since they did not improve the models.

[4] https://iancovert.com/blog/understanding-shap-sage/

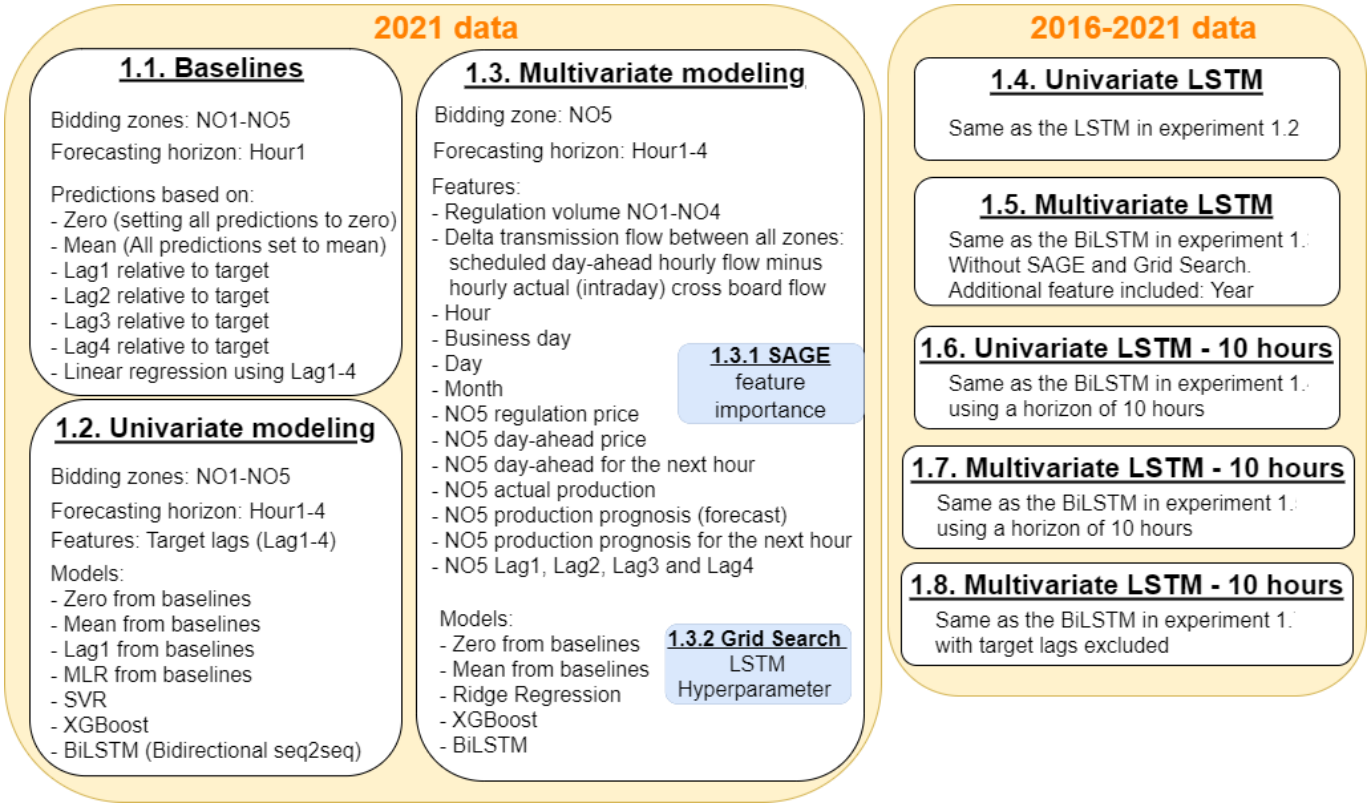


Figure 5.5: Machine Learning Experiment1

5.3.1 Step 1.1 - Baselines

Figure 5.6 shows the normalized MSE performance of the considered baseline models. The baseline models only predict one hour ahead. The Figure 5.6 shows the MSEs of regulation volumes for NO1-NO5 evaluated on the last 800 hours of observations in 2021 for the Zero, Mean, Lag1-4, and multiple linear regression models with lag1-4 as input features.

5.3.1.1 Results - Step 1.1 Baselines

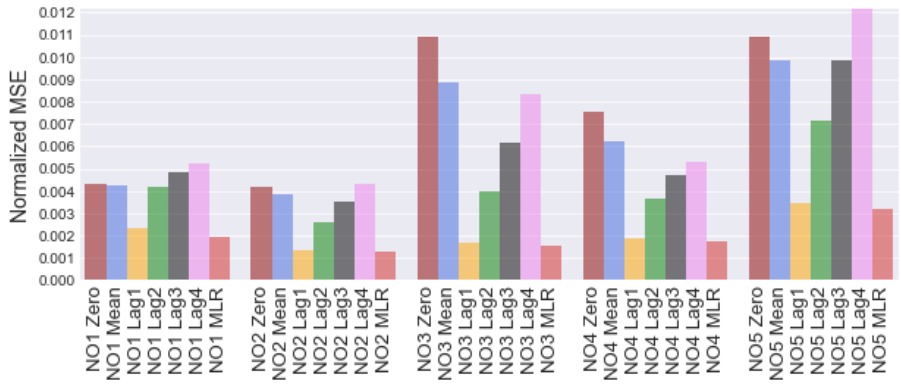


Figure 5.6: Performance of baseline models for regulation volume for all bidding zones predicting one hour into the future: Fixed predictions and multiple linear regression (lag1-4 as features)

5.3.1.2 Discussion - Step 1.1 Baselines

The baselines were all evaluated on the last 800 hours of 2021 to have a reference for other models when evaluated over the same time span. The normalized MSE in Figure 5.6 show that the multiple linear regression (MLR) model is comparable to lag1 of regulation volume for all zones NO1-NO5. This is expected due to the significant partial autocorrelation for lag 1 (Figure 4.21). One can see from Figure 5.6 that predicting zero or the mean of the target for every predicted hour is a decent baseline. This stems from the nature of the regulation volume (Section 4.10), where a significant portion of the data is zero for all zones (Table 4.6).

5.3.2 Step 1.2 - Univariate modeling

This univariate step is the first step in modeling the regulation volume four hours into the future. In the univariate modelling setting only one regulation volume variable is utilized. The modeling process was repeated for all each of the five zones NO1-NO5. Modeling all five zones is time-consuming, and it was decided for the other sub-steps in experiment 2 and for all other experiments to focus solely on modeling regulation volume for NO5. The initial intention for step 1.2 was to find a good univariate forecasting model that could be carried over to the multivariate modeling, which uses multiple market features as input. Four baselines were carried over to the univariate step from the initial baseline step 1.1 in Subsection 5.3.1; zero, mean, and lag1 prediction, and the multiple linear regression (MLR). In addition, a Support Vector Regression (SVR), eXtreme Gradient Boosting (XGBOOST), and Bidirectional Long short-term memory (BiLSTM) models were added. Four models were trained for each of the MLR, SVR, and XGBOOST models, one model for each future hour for each model type. The BiLSTM model was made as a sequence-to-sequence model where all four hours are predicted using a single model. All models were fed with lag1, lag2, lag3, and lag4 as input features. The BiLSTM is designed for processing sequences of data and was created using a sequence length of 6, having four lags (features) of historical regulation volumes for each sequence. The sequence length is considered a hyperparameter, and several lengths were tried, with 6 being the most optimal length found during the grid search in step 1.3 (Subsection 5.3.3). Additional hyperparameters were found during an extensive grid search in step 1.3 - multivariate modeling in Subsection 5.3.3. The best hyperparameters and the best BiLSTM architecture found in step 1.3 in Subsection 5.3.3 and shown in Figure 5.3 turned out to also perform well in the univariate setting. Therefore, the univariate modeling step was redone using the same configuration from the multivariate modeling, except that the univariate modeling only uses one regulation volume variable without extra market data. Several BiLSTM architectures were investigated, and the best model used for both the univariate modeling and the multivariate modeling in step 1.3 is shown in Figure 5.7. The hyperparameters for the SVR and the XGBOOST are not discussed due to the superior performance of the BiLSTM. However, some tuning was done for the SVR and XGBOOST, which did not improve their performance much.

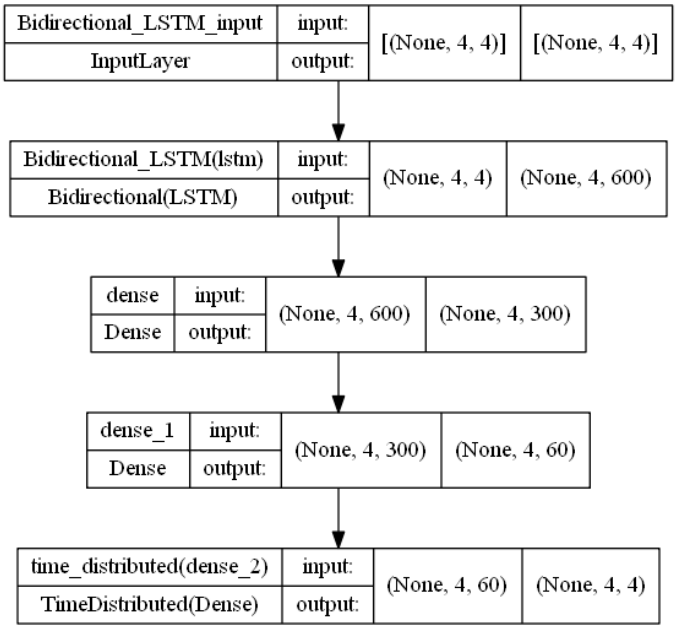


Figure 5.7: The univariate BiLSTM model with 930604 trainable parameters.

5.3.2.1 Results - Step 1.2 Univariate modeling

Figure 5.8 shows the performance of the univariate modeling on the 2021 data modeled using regulation volume for NO5. The figure shows the normalized MSE for all models, including the baselines, four hours into the future. Figure 5.9 shows a qualitative plot of the second forecasted hour from the MLR for 200 predictions in the test data. Figure 5.10 shows a qualitative plot for the forecasted second hour from the BiLSTM model evaluated on the same data as the MLR model.

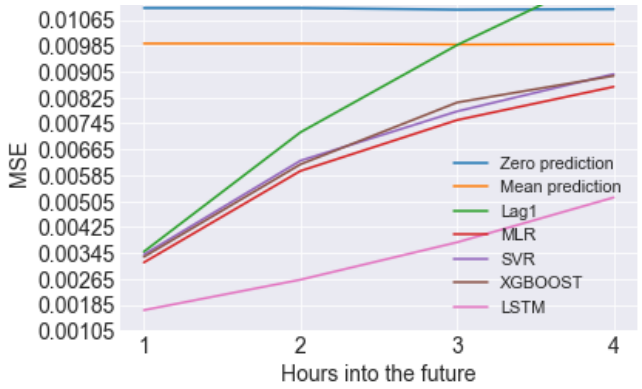


Figure 5.8: Univariate performance over 4 future hours for Multiple Linear Regression (MLR), Support Vector Regression (SVR), XGBOOST, and BiLSTM trained on 2021 data on bidding zone NO5

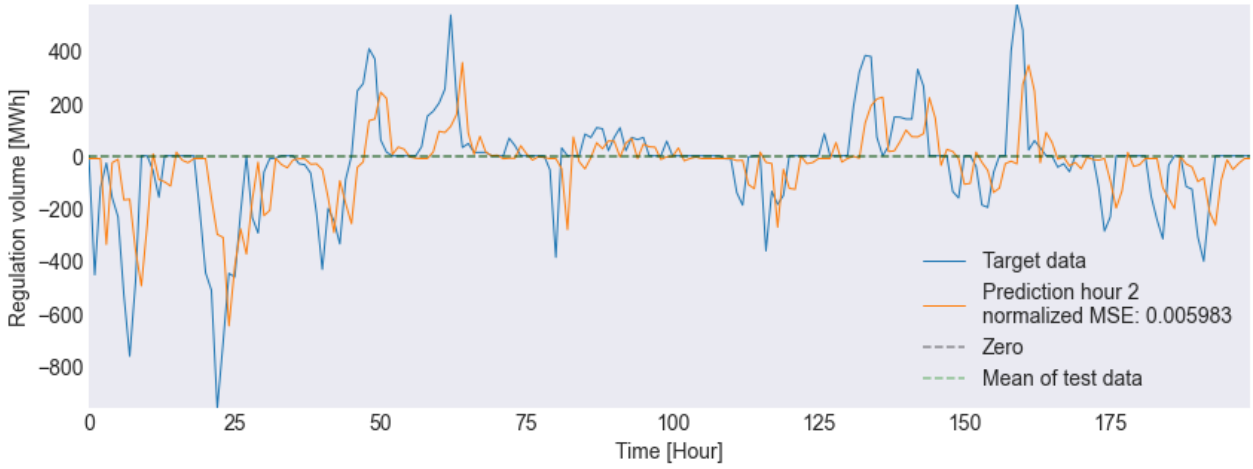


Figure 5.9: Univariate results at hour 2 for Multiple Linear Regression trained on regulation volume for NO5 using the 2021 data.

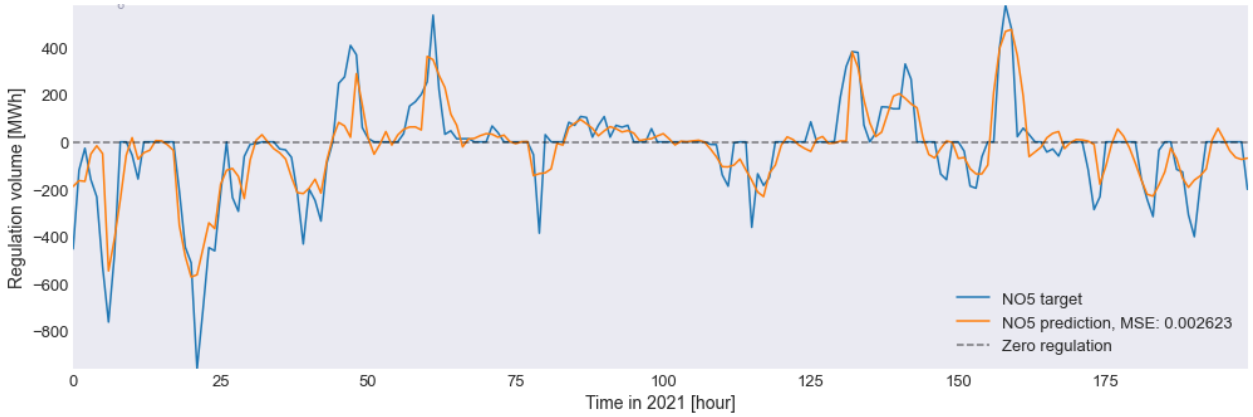


Figure 5.10: Univariate results at hour 2 for the Bidirectional Long short-term memory BiLSTM model trained on regulation volume for NO5 using the 2021 data.

5.3.2.2 Discussions - Step 1.2 Univariate modeling

The performance in Figure 5.8 of the MLR, SVR, and XGBOOST models are almost similar to predicting the mean value at hour4. The LSTM (which is the Bidirectional LSTM), has a significantly lower normalized MSE value for all four hours. The MSE value for hour4 from the LSTM model is significantly lower than the hour2 forecast from the other models. Using Lag1 for hour3 results in an MSE score equal to setting the mean value.

One significant characteristic for the MLR model in Figure 5.9 is the temporal shift in the predictions relative to the target data. This behavior stems from the input data being the previous lags of the target. Thus, the model predictions are biased toward the lagged hours of regulation volume and repeat something similar to what was known from recent history. However, the MLR model consistently predicts a value closer to zero than what was known in the previous lags. It seems that the model has learned that the spikes from regulated power always go back to zero after some hours. This reflects the nature of the regulation volume, where the effect of activating regulating power volumes results in the grid being balanced, and eventually, zero regulating power is attained. Forecast-

ing further into the future resulted in more predictions around zero regulation volume. Comparing the MSE from the MLR model in Figure 5.9 with the BiLSTM model in Figure 5.10, the normalized MSE from the MLR model is almost 2.3 times higher than the BiLSTM model. One can see from Figure 5.10 that the BiLSTM model corrects much of the temporal shift seen from the MLR model in Figure 5.9. The LSTM model, in general, is designed to process sequences of temporal data, which makes it more potent at learning the signature of regulation volume fluctuations over time. In addition, during model training, the bidirectional LSTM model was trained on the time series in both forwards and reverse order, learning the signature of regulating power events even better.

5.3.3 Step 1.3 - Multivariate modeling

In addition to using historical regulation volumes to model future regulation volumes, the multivariate modeling takes in additional market data with the intention of improving the results from the univariate modeling. The setup for the multivariate modeling is listed in **1.3. Multivariate modeling** in Figure 5.5. All features have the unit Mega-Watt hour [MWh] except for the day-ahead price, imbalance price, and seasonal features. Analysis of cross-border transmission flow and capacities was not presented in the exploratory data analysis in Chapter 4. However, an investigation of correlations between regulation volumes and flow revealed some relationships. It is logical to include cross-border flow between bidding zones since activation of regulation reserves may involve transmitting or receiving power via available transmission cables in the grid. The calculated delta flow takes the day-ahead planned flow and subtract the hourly intraday flow.

The Zero and Mean baselines are carried over for benchmark purposes. The same XGBOOST model and BiLSTM model from the univariate modeling is used for the multivariate modeling. Additionally, the Ridge regression model is included over the linear regression model due to the high number of predictors or features and to give the model the ability to regularize coefficients to achieve better predictions. Different architectures of the BiLSTM model were investigated, and the final architecture used is the same model presented in the univariate modeling in Figure 5.7. However, the InputLayer dimension is increased to take 42 features instead of the 4 features used for the univariate BiLSTM in Figure 5.7. The hyperparameter space used in the grid search can be seen in Table 5.2, and the most optimal parameters from the grid search can be seen in Table 5.3.

Learning rate	RNN Units	Batch size	Sequence Length	Target lags	Regularization
0.0001	64	100	3	4	1.e-07
0.0005	88	200	4	5	1.e-05
0.001	128	300	6		1.e-03
0.005	224	400			1.e-01
0.01	300				1.e+01

Table 5.2: Grid Search parameters for the BiLSTM model

Learning rate	RNN Units	Batch size	Sequence Length	Target lags	Regularization
0.0005	300	100	6	4	1.e-07

Table 5.3: The best found combination of parameters for the BiLSTM model

5.3.3.1 Results - Step 1.3 Multivariate modeling

The global feature importance described by SAGE[5] can be seen for the forecasted hour1 and hour4 in Figure 5.11 and Figure 5.12.

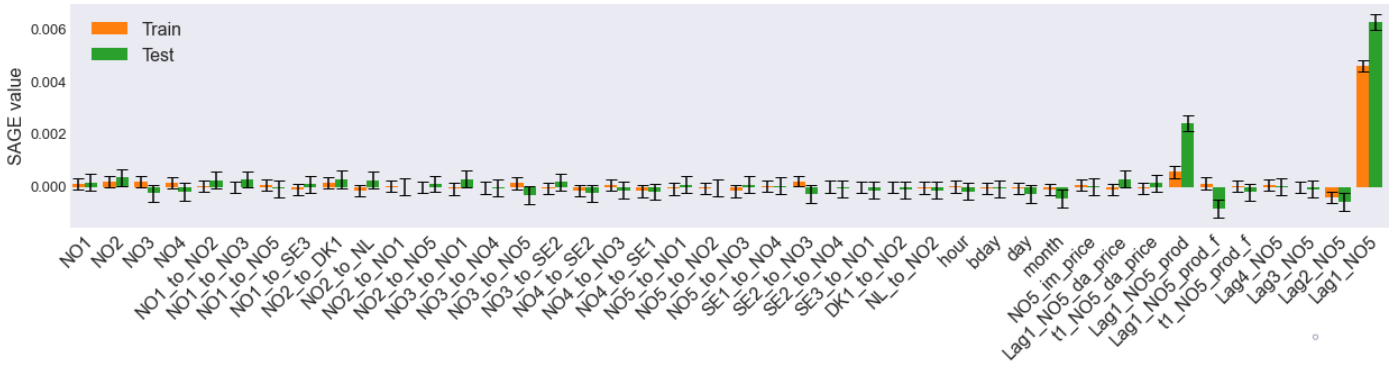


Figure 5.11: Feature importance at hour1 using SAGE on Ridge Regression on data from 2021 on regulation volume for bidding zone NO5

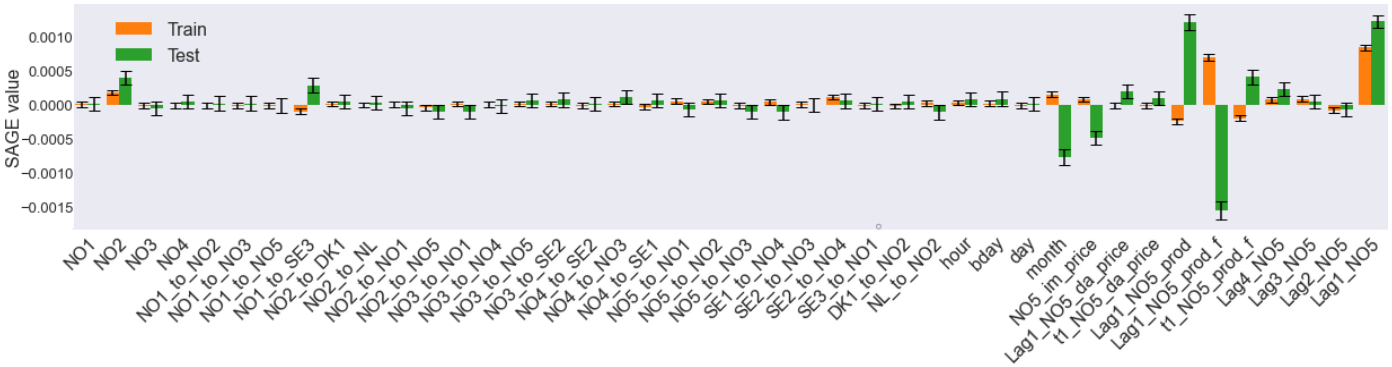


Figure 5.12: Feature importance at hour4 using SAGE on Ridge Regression on data from 2021 on regulation volume for bidding zone NO5

Figure 5.13 shows the performance of the performance on the 2021 data modeled using regulation volume for NO5. The figure shows the normalized MSE for all models, including the baselines for zero and mean predictions, four hours into the future. Figure 5.14 shows a qualitative plot of the second forecasted hour from the multivariate BiLSTM evaluated on the same test data as the univariate models.

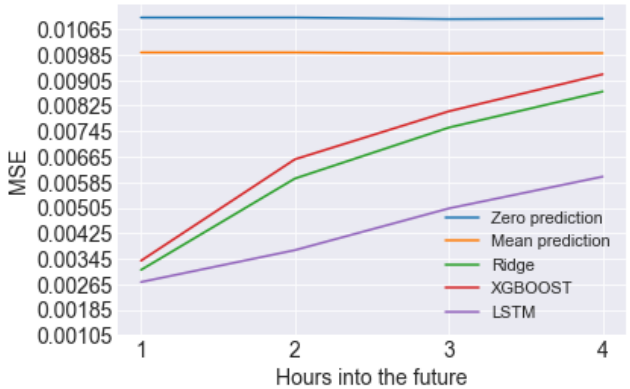


Figure 5.13: Multivariate performance over 4 future hours for Ridge Regression (Ridge), XGBOOST, and BiLSTM trained on 2021 data on bidding zone NO5

[5] https://iancovert.com/blog/understanding-shap-sage/

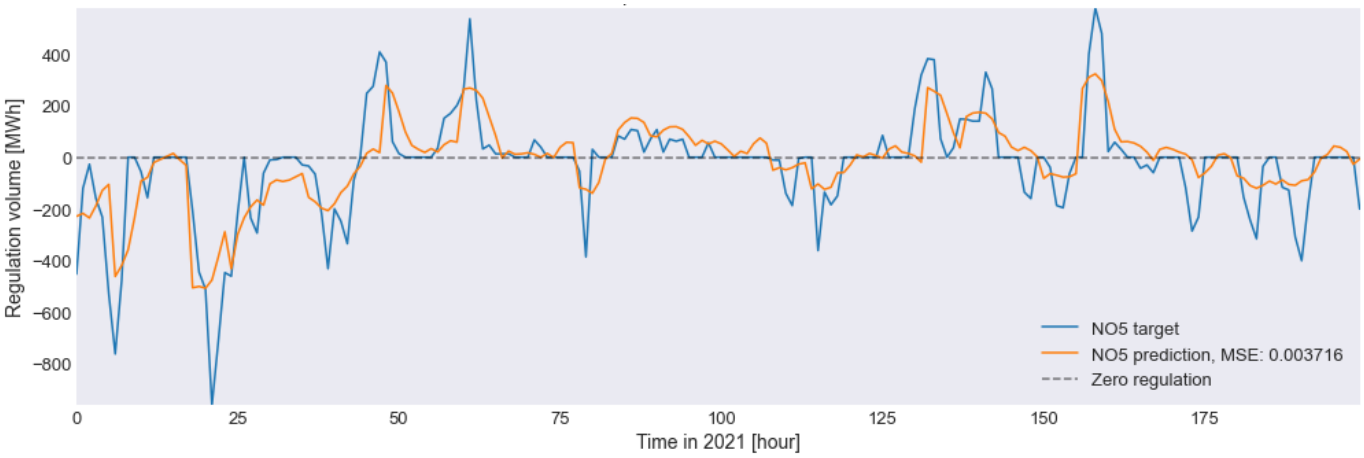


Figure 5.14: Multivariate performance at hour 2 for the Long short-term memory BiLSTM model trained only on 2021 data on bidding zone NO5

5.3.3.2 Discussions - Step 1.3 Multivariate modeling

The global feature importance for the first forecasted hour in Figure 5.11 reveals the dominating influence from lag1 of NO5 (*Lag1_NO5*) regulation both for the training and test data. The feature importance on the fourth forecasted hour in Figure 5.12 reveals that lag1 of NO5 regulation volume is significantly less influential. The lagged features contribute less when forecasting further into the future. In the absence of the lagged features dominating the feature space, other features get more room for influencing the modeling, as seen in Figure 5.12 when compared with hour1 in Figure 5.11.

Figure 5.13 shows that the BiLSTM overall performance is significantly better than the Ridge regression and XGBOOST model. However, none of the multivariate models beats the results from the univariate BiLSTM in Figure 5.8. The regularization parameter for the BiLSTM layers turned out to be very small, as seen in Table 5.3. Therefore, I discarded the regularization of the BiLSTM layers. The performance was indistinguishable with and without the weight regularization of the BiLSTM model. The predictions for hour2 in Figure 5.14 reveal that the predicted results from the multivariate BiLSTM are not better than the predicted results from the univariate BiLSTM in Figure 5.10. Thus, so far the univariate BiLSTM model from Figure 5.8 trained and tested on 2021 is the best performing model baseline when increasing the data range (2016-2021) in step 1.4 and step 1.5 in Subsection 5.3.4.

5.3.4 Step 1.4 and 1.5 Uni and multivariate on 2016-2021 data

Step1.4 and step1.5 consist of re-modelling two previous univariate and multivariate models (steps1.2 and step1.3) using data (2016-2021) while keeping the same test data from step1.2 and step1.3 for evaluation. Comparing the results from modeling using data for one year versus several years was not found in previous work (Section 3.1), which serves as a motivation for this subsection.

5.3.4.1 Results - Step 1.4 and 1.5 Uni and multivariate on 2016-2021 data

Figure 5.15 shows the performance of the univariate models from step1.4, and Figure 5.16 shows the performance of the multivariate models from step1.5. Table 5.4 summarize the performance of all BiLSTM models from step1.2-step1.5. Model performance during

training is shown in Figure 5.17 and Figure 5.18 for the best univariate and multivariate BiLSTM models. The Figures 5.19, 5.21, 5.20, show some representative forecasts for hour1 and hour2 for the univariate and multivariate BiLSTM from step 1.4 and step1.5. The performances of the best univariate and multivariate BiLSTM models are shown in Figure 5.22 with confidence intervals for each forecasted hour.

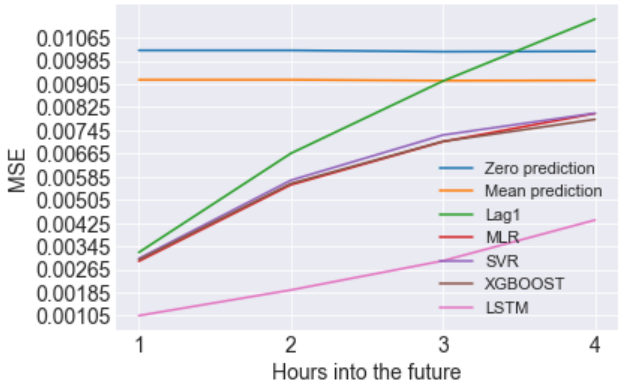


Figure 5.15: Step1.4 results (step1.2 with more data): Univariate performance from MLR, XGBOOST and BiLSTM trained on 2016-2021 data on bidding zone NO5.

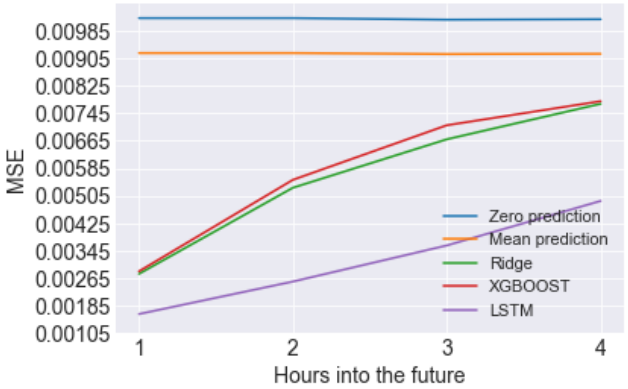


Figure 5.16: Step1.5 results (step1.3 with more data): Multivariate performance from Ridge, XGBOOST and BiLSTM trained on 2016-2021 data on bidding zone NO5.

BiLSTM performance	MSE hour1	MSE hour2	MSE hour3	MSE hour4
Univariate on 2021	0.001683	0.002623	0.003778	0.005163
Univariate on 2016-2021	0.00105	0.001935	0.002944	0.004353
Multivariate on 2021	0.002715	0.003716	0.005025	0.006023
Multivariate on 2016-2021	0.001677	0.002525	0.003617	0.004942

Table 5.4: MSE comparisions between using 2021 or 2016-2021 data for both univariate and multivariate BiLSTM.

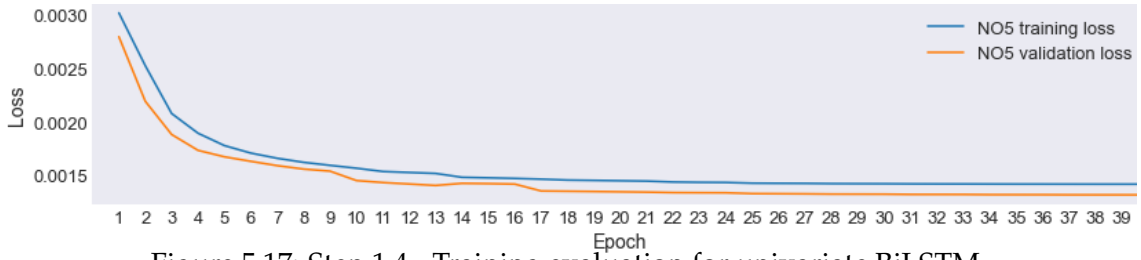


Figure 5.17: Step 1.4 - Training evaluation for univariate BiLSTM.

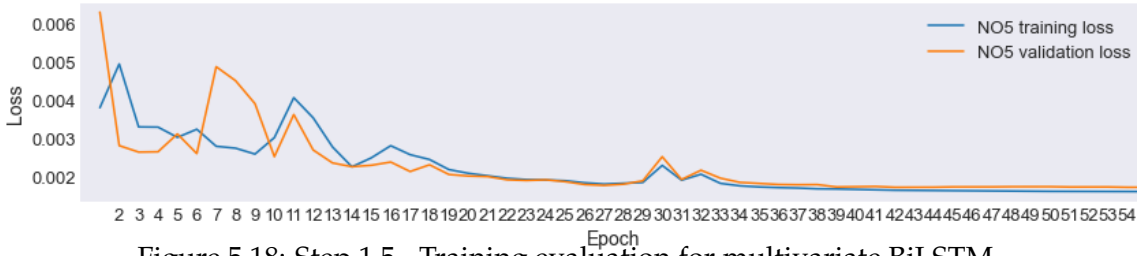


Figure 5.18: Step 1.5 - Training evaluation for multivariate BiLSTM.

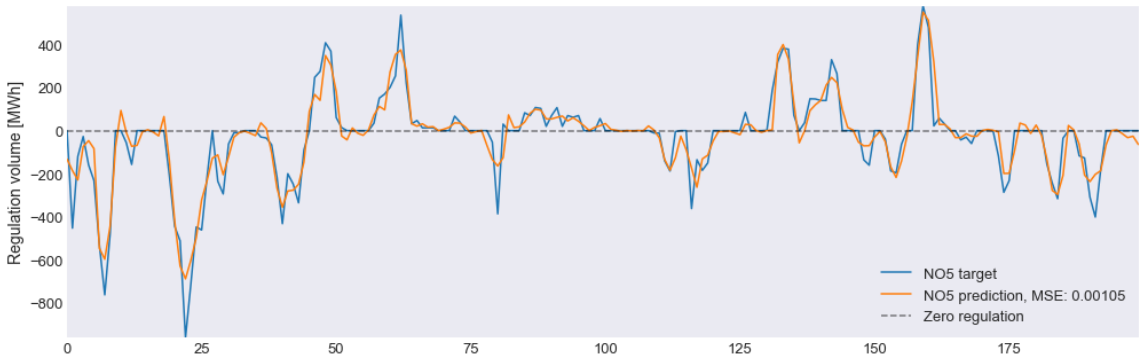


Figure 5.19: Results from step 1.4 - univariate BiLSTM: forcast of hour 1.

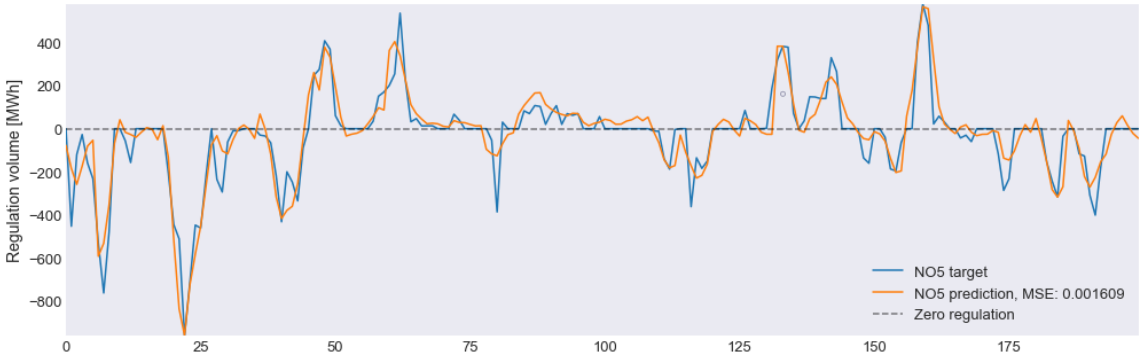


Figure 5.20: Results from step 1.5 - mulitvariate BiLSTM: forcast of hour 1.

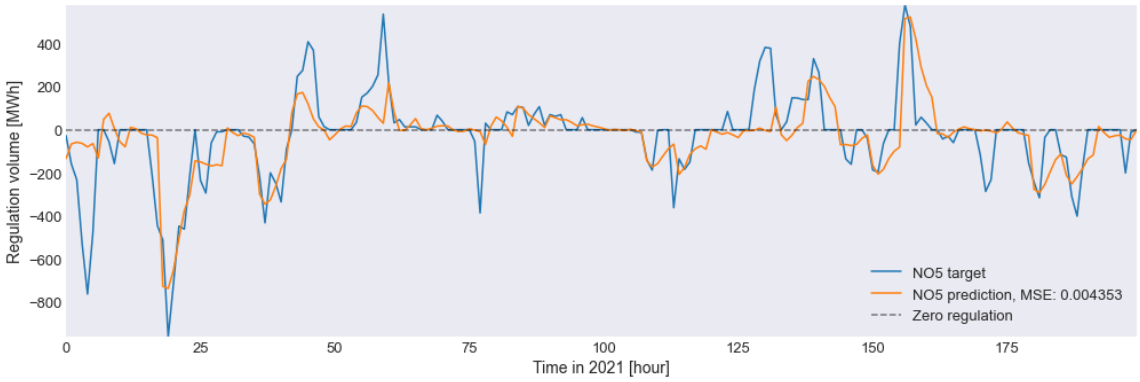


Figure 5.21: Results from step 1.4 - univariate BiLSTM: forcast of hour 4.

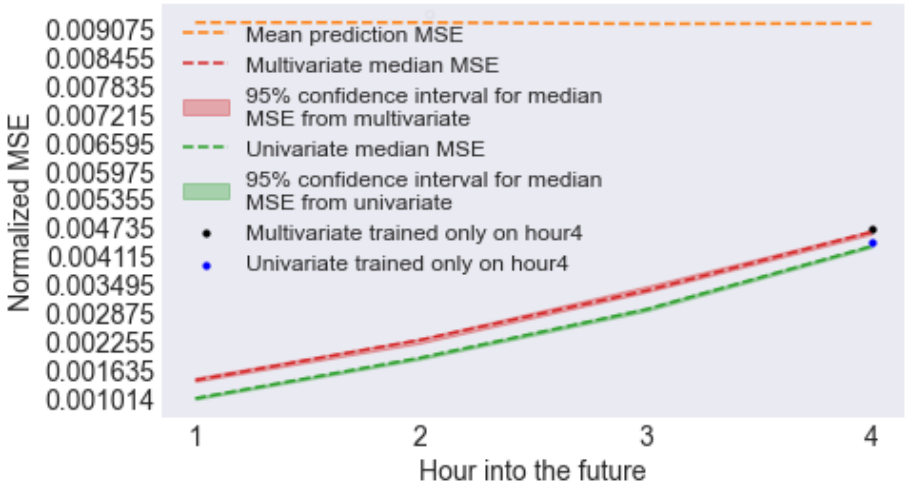


Figure 5.22: The best performing BiLSTM models from experiment 1.4 and 1.5 trained and evaluated on data from 2016-2021. Median MSE with 95% confidence interval centered around the median.

5.3.4.2 Discussion - Step 1.4 and 1.5

When using more data, the Figures 5.15 and 5.16 reveals that the LSTM model (BiLSTM) significantly outperforms the other models both in the univariate and multivariate settings. This rules out all other models but the BiLSTM in pursuit of the best performing model. Table 5.4 summarizes the performance of all BiLSTM models, and it shows that both the univariate and multivariate models trained on data from 2016-2021 outperform the models that are only exposed to the 2021 data. Considering the MSE score for all four hours, the best performing BiLSTM model is the univariate model trained on data from 2016-2021. Figure 5.17 and 5.18 show the training evaluation for the univariate and multivariate models exposed to the 2016-2021 data. All BiLSTM models are trained using validation data for learning rate reduction.

The learning rate is reduced on plateau[6] with the patience set to 3 epochs. The learning rate is reduced if the MSE on the validation data is not improving after 3 epochs to prevent overfitting and to guide the model into possible better generalization and per-

[6] https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/ReduceLROnPlateau

formance. One can see from the training of the univariate model in Figure 5.17 that the training goes rather smooth until the loss converges. One can see in Figure 5.17 the learning rate reduction at epoch 16 after not improving on the validation loss since epoch 13, and this guides the model into a lower validation loss from epoch 17. In Figure 5.18 one can see that the multivariate model struggles slightly in the optimizing process, which is most likely due to the higher number of features and it needs to take several relationships into account when updating weights through the backpropagation process. Upon converging, the training and validation loss for both the univariate and multivariate training seems to be at the lowest value and stable, indicating that the models do not overfit to the training data.

Investigating Figure 5.19 and 5.20 reveals some predicted results for hour 1 for the univariate and multivariate BiLSTM. It is previously mentioned that all models are evaluated on the same test data, the last 800 observations in 2021. However, a detailed presentation of actual predictions is done using the first 200 predictions of the test data, while the MSE value is from the whole test data. The univariate predictions in Figure 5.19 generally follow the target but struggle to reach some of the peaks, e.g., the largest peak close to hour 25 on the x-axis. The multivariate predictions had a higher MSE value than the univariate as summarized in table 5.4. The multivariate BiLSTM occasionally managed to forecast some peak values better than the univariate BiLSTM model e.g. around hour 25 in Figure 5.20. In comparison, the univariate model was not able to forecast the same peak value in Figure 5.19. Figure 5.21 show the fourth forecasted hour of the best univariate BiLSTM model. It is seen that the predictions are generally closer to zero compared with the hour1 forecast from the univariate and the multivariate. However, the univariate model still captures some peak values when forecasting four hours into the future.

Figure 5.22 shows the best performing univariate and multivariate models with a computed 95% confidence interval for each forecasted hour. The median MSE of 30 trained and evaluated models are shown with a band corresponding to the confidence interval. The confidence interval is computed by bootstrapping the 30 evaluated MSE values 100000 times and then calculating the upper and lower percentile corresponding to a 95% confidence interval of the bootstrapped MSE values. The calculated confidence intervals shown in the figure reveal that both the univariate and multivariate models perform consistently. In addition, Figure 5.22 also shows the two univariate and multivariate BiLSTM trained on only using the fourth future hour as the target. It reveals that BiLSTM performs equally on the fourth forecasted hour when only trained on hour4 compared to the entire output sequence of hour1 up to hour4.

5.3.5 Step 1.6 and 1.7 - Uni and multivariate with 10 hour forecast

The best univariate and multivariate BiLSTM models are considered the two superior models in this thesis for forecasting regulation volumes. These two models were stretched to forecast 10 hours into the future to investigate their potential. The univariate and multivariate modelling steps (step 1.4 and step1.5 in Subsection 5.3.4) on data from 2016-2021 was redone using a 10-hour forecast.

5.3.6 Step 1.8 - Multivariate BiLSTM with target lags excluded

The BiLSTM performance with target lags excluded was measured in step1.8. This step serves to measure how the BiLSTM can utilize the other features without the dominating

lagged target features. This test was conducted using the same BiLSTM architecture as in the uni and multivariate setup with the same hyperparameter combination. Additionally, another BiLSTM architecture, including convolution, was testet to see the effect of using a 1-dimensional convolution across time-series to further extract features from the processed time-series sequence coming from the bidirectional LSTM layer. The BiLSTM architecture including convolution is visualized in figure 5.23. For simplicity, the best-found hyperparameter combination from the grid search was used for the BiLSTM architecture that includes convolution. Other parameters were manually investigated, but the combination from the grid search turned out to supersede those that were manually tested 5.3. The results with or without the added convolution can be seen in figure 5.24. Due to increased complexity, the regularization parameter was included in all the convolutional BiLSTM model layers to ensure some weight punishment. The model with convolutional layers has 2604904 trainable parameters. However, the regularization parameter from the grid search (model without convolution) was used and resulted in a relatively good option out of the manually randomly tried out parameters.

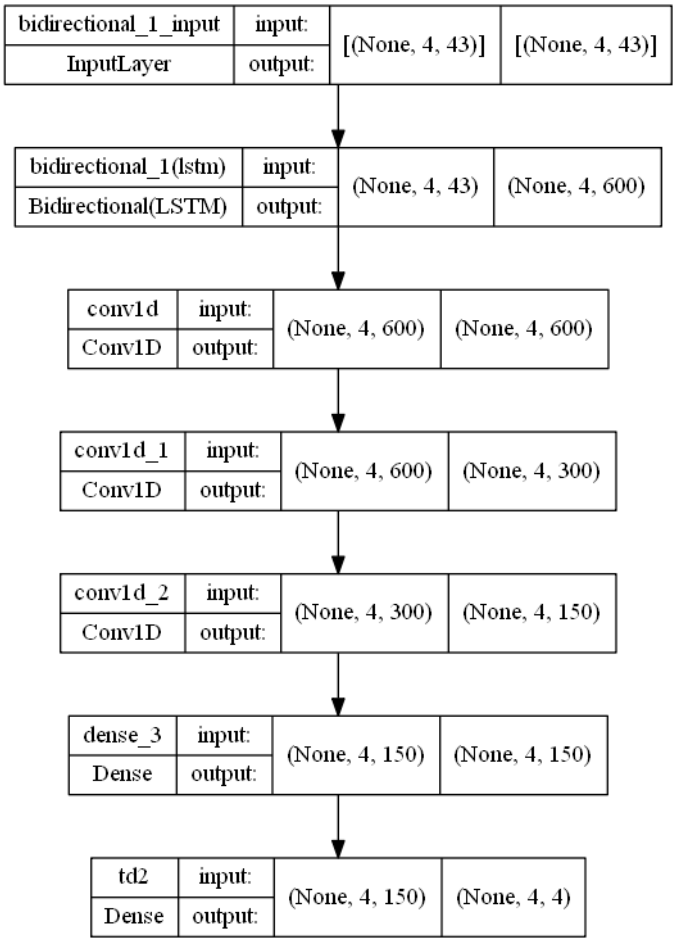


Figure 5.23: The multivariate BiLSTM with convolution having 2604904 trainable parameters.

5.3.7 Results - Step 1.6, 1.7 and 1.8

The resulting performance from step1.6 - step1.8 is shown in Figure 5.24 with a 95% confidence interval.

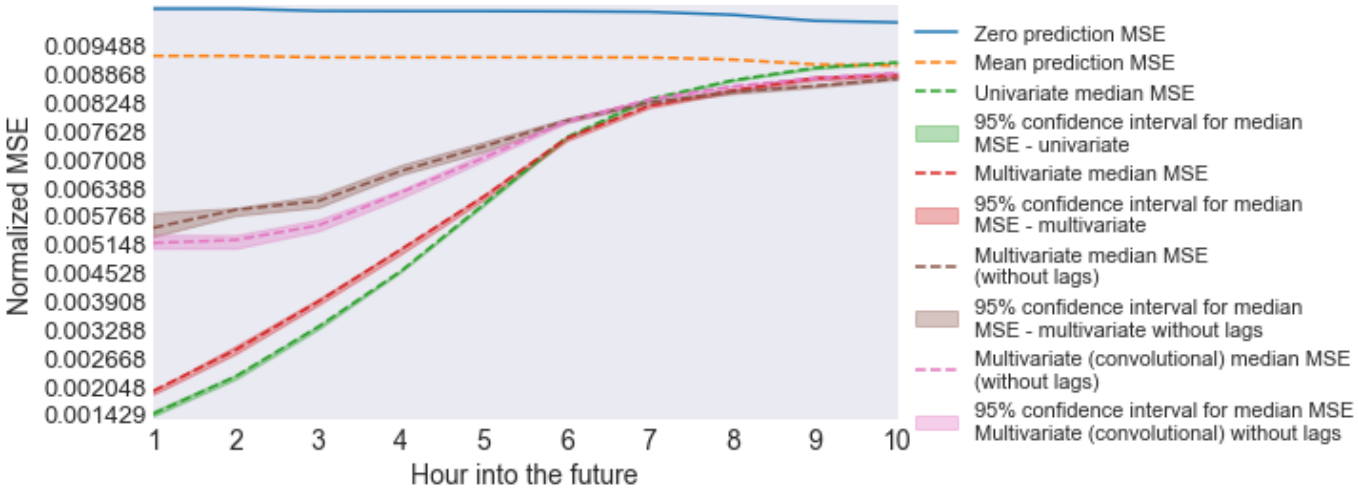


Figure 5.24: Results from step 1.6-1.8. Univariate BiLSTM (step1.6), multivariate BiLSTM(step1.7), multivariate BiLSTM without lags(step1.8), and multivariate BiLSTM with convolution(step1.8) pushed to its limits with 10 hour forecast using mean prediction as common reference. Bidding zone: NO5 - Data: 2016 to 2021.

5.3.8 Discussions - Step 1.6, 1.7 and 1.8

All the models in Figure 5.24 outperform the mean prediction baseline up to the ninth forecasted hour. In contrast, the Ridge regression and XGBOOST in Figure 5.13 were close to the mean predicted MSE at hour 4. The univariate model in Figure 5.24 is best when forecasting up to about hour 5 or 6. After that, the other multivariate models perform slightly better than the univariate model. This shows that the extra features used in the multivariate modeling affect forecasting longer into the future. This is aligned with the investigated SAGE plots in Figure 5.11 and 5.12, which showed the extra features gaining room to influence the forecast when the lagged features were less dominating for the fourth forecasted hour. Moreover, the BiLSTM models trained without including lags performed poorly as expected and can be seen in Figure 5.24. However, training completely without lags supersedes the univariate model in performance when forecasting the seventh hour into the future.

5.4 Experiment2

Experiment 2 and its sub steps are carried out as shown in Figure 5.25. The experiment is a continuation of the exploratory data analysis part that looks at uncertainties in weather forecasts in Section 4.14. Experiment 2 intends to investigate weather forecast uncertainty further to better determine its influence on regulation volume forecasts and test the hypothesis that weather forecast uncertainties influence the need for grid balancing. This experiment takes a different approach in estimating the forecast uncertainties than what was conducted in the exploratory data analysis. Instead of subtracting the actual weather data, this experiment subtracts from the 24 hour forecast the updated forecasts within each day. The intraday forecasts that are made available at 00:00, 06:00, 12:00 and 18:00 are than substituted with the actual weather data used in the exploratory data analysis. The motivation for this approach is that by working solely with weather forecasts, one have access to future data which in short-term (a 6 hours time span) is considered relatively similar to actual weather. Another significant difference between the data ana-

lysis and this experiment is that the 24-hour lead weather forecast is taken from the 06:00 published weather forecast and not the 00:00 as was done in the exploratory data analysis. Experiment 2 consists of several substeps as shown in Figure 5.25 which will be described in the subsequent subsections. All steps involving machine learning uses regulation volume for NO5 as target with December 2021 being reserved for test similar to what was used in experiment 1 5.3.

Based on the weather analysis and numerous testing in experiment2, it was standardized using all 8 weather types simultaneously merged into images with 8 channels. Both the analysis and tests performed indicated that all weather types contribute differently to different situations. The results showed that merging all weather types into 8 channels made slightly more sense, and it is believed that both global conditions and local circumstances contribute differently from different weather types for different hours.

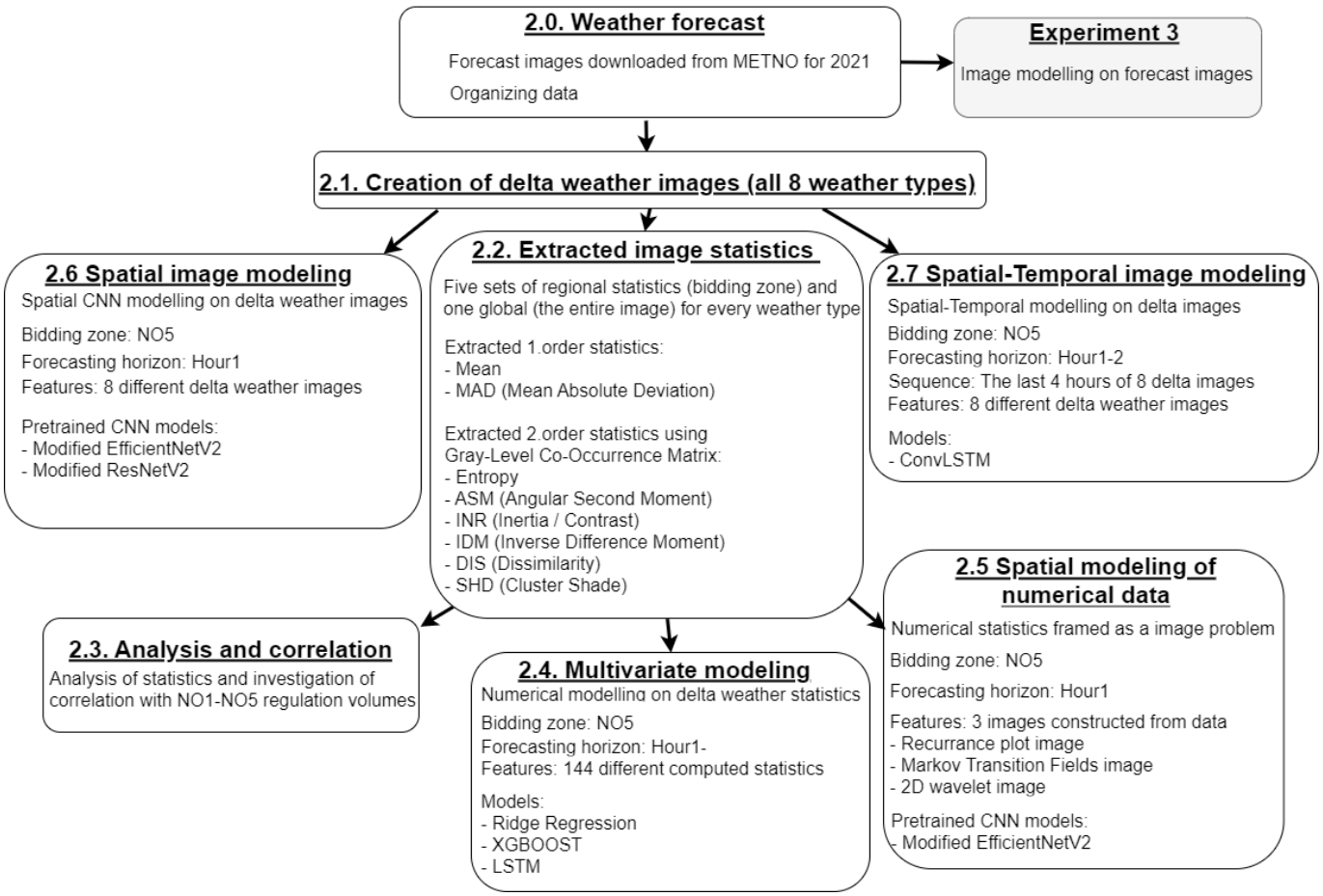


Figure 5.25: Machine Learning Experiment2

5.4.1 Step 2.0 and 2.1

Download and $\Delta images$ creation for experiment 2 is the second method already mentioned in the Decision on data gathering Subsection 4.14.1. Details on the more advanced creation of $\Delta images$ is moved to Section 7.3 in Appendix to leave space and focus on for machine learning and analysis.

5.4.2 Step 2.2, 2.3, 2.4

In step2.2, image statistics were derived from the created $\Delta images$ from step 2.1. The image statistics listed in step2.2 in figure 5.25 were used and computed for each bidding zone, as well as for the entire global image, similar to what was done in the exploratory data analysis part. Additionally to what was done in the exploratory data analysis (mean and MAD statistics), some second-order statistics were derived using the Gray-Level Co-Occurrence Matrix. The second order statistics were derived from the GLCM representation of the entire global $\Delta image$ for each weather type. In addition to the MAD which were shown in equation (4.3), the second order statistics are computed as follows:

$$weather\ forecast\ uncertainty_{INR} = INR\left(GLCM\left((\Delta W_{weather\ type,h})_{NO}\right)\right) \tag{5.1}$$

Where INR is an example on one of the used second order statistics computed from the GLCM representation of the entire global $\Delta image$ for each weather type.

Correlation analysis was performed between the derived statistics and the regulation volume in step2.3, and in step2.4 the multivariate BiLSTM model from experiment 1 was used directly on the derived statistics to predict the regulation volume.

5.4.2.1 Results - Step 2.3

The Figures 5.26 show the results with the two highest observed correlations between regulation volume and estimated weather uncertainties found when computed using January-March data. The highest correlation observed in the right plot reflects the correlation between regulation volume for NO5(the purple bar) and Entropy on the global $\Delta image$ (NO) of air temperature. The highest correlation observed in the left plot reflects the correlation between regulation volume NO1(the blue bar) and MAD on the global $\Delta image$ (NO) of surface downwelling.

Figure 5.27 show the correlations for the same weather types with correlations computed using all twelve months in 2021.

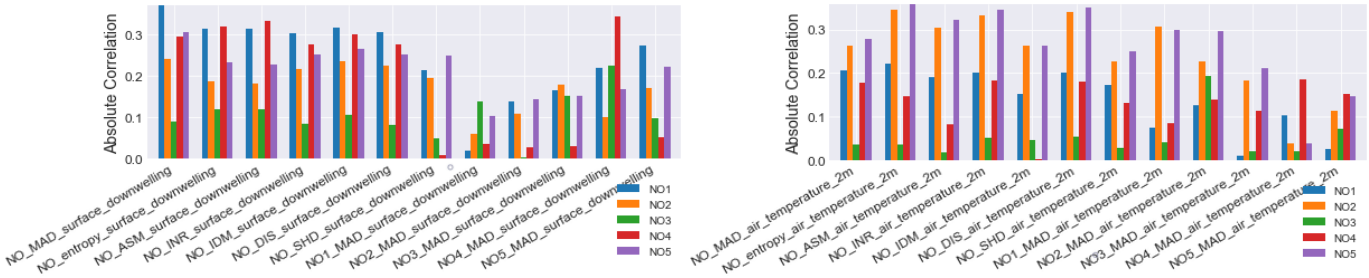


Figure 5.26: Correlation between regulation volume for every bidding zones and weather forecast uncertainties from radiation from the sun and temperature. Both plots shows the highest correlation found using data for January to March from 2021.

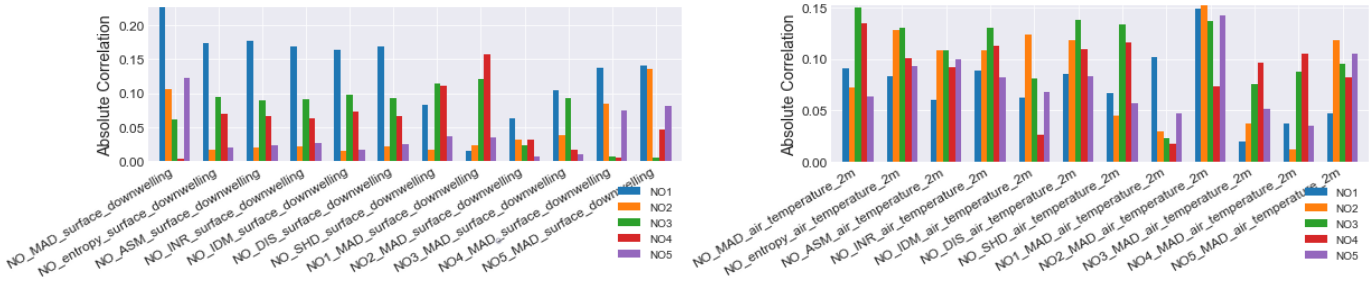


Figure 5.27: Correlation between regulation volume for every bidding zones and weather forecast uncertainties from radiation from the sun and temperature. Both plots shows the correlations computed using all twelve months in 2021.

5.4.2.2 Discussions - Step 2.3

The correlations are generally observed to be higher when using the more advanced method of computing uncertainties than what was done in the exploratory data analysis in Subsection 4.14.4. Comparing the results between correlation for January-March and using all twelve months, one can see that the correlation drops significantly using a sample size of the entire year. Additionally, it was discovered that using a sample size of only one month further increases the correlations, and some months are significantly more correlated than others. The highest observed correlation was between 0.4 - 0.5 from surface downwelling when correlations were computed using only January month. Upon investigating correlations for single months only, it was revealed that different weather types had the highest correlations for different months. There is no clear pattern as to which weather type is most correlated with regulation volume for spesific months. However, further analysis on one month only is not considered relevant since the entire year is considered for modeling, not single months. It is believed that the loss in correlation using more data is related to more noise being introduced and that regulation varies slightly between different months for different weather types.

5.4.2.3 Results - Step 2.4

Figure 5.28 shows results when using the Ridge, XGBOOST, and BiLSTM models from multivariate modeling Subsection 5.3.3 with all the extracted image uncertainties as input features, 144 features in total.

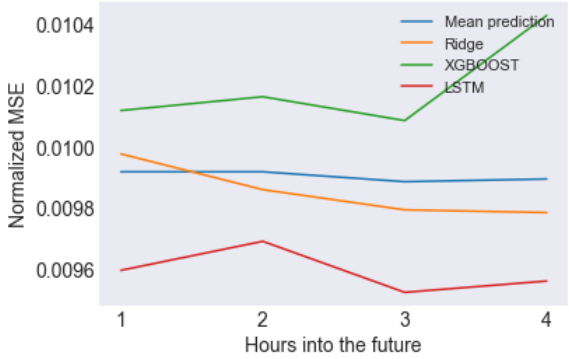


Figure 5.28: Multivariate model performance on delta weather statistics on NO5 regulation volume

5.4.2.4 Discussions - Step 2.4

Numerous tries were made to enhance the performance when modeling the regulation volume for NO5 were none gave good results. Figure 5.28 shows that all models perform close to the blue mean prediction line in terms of normalized MSE value. Presenting statistics on the MSE results were not considered since the results were generally poor. The figure shows that the BiLSTM model had a marginally lower MSE value than the other models. However, the model performances are considered indistinguishable from predicting a fixed mean value for all hours.

5.4.3 Step 2.5

In step2.5, the derived image statistics are formulated as an image problem by creating Recurrence plot images, Markov Transition fields images, and 2D wavelet images directly from the numerical and fed into a CNN model. Reformulating time series data into an image problem using recurrence plot and 2D wavelet representations has been proved fruitful in other domains, e.g., Bordvik et al. [47], which serves as motivation for trying it on weather forecast uncertainties. Figure 5.29 shows an example of the constructed images. The images are made directly, on the fly, from the one-dimensional numerical data by the dataloader during training, validation, and test phases. Each constructed image has the dimensions of 100×100, and the three images are merged into one image having three channels. The 2D wavelet transformation image incorporates a 2D wavelet transformation of the regulation volume for NO5 over a window of the last 100 regulation volume observations and therefore captures temporal information from historical regulation volumes of NO5. The Recurrence Plot images are made using 100 weather uncertainty features along the x-axis and the last 100 historical weather uncertainties along the y-axis. The patterns reflect conditions or states that reoccur in the data. The idea behind the Recurrence Plot is similar to the Markov Transition Field image. However, the Markov Transition Field image is based on the first-order Markov chain and incorporates floating-point pixel values, while the Recurrence Plot image is an image with binary values. The images were fed into a pre-trained CNN model, and a simple custom-built model was also briefly investigated.

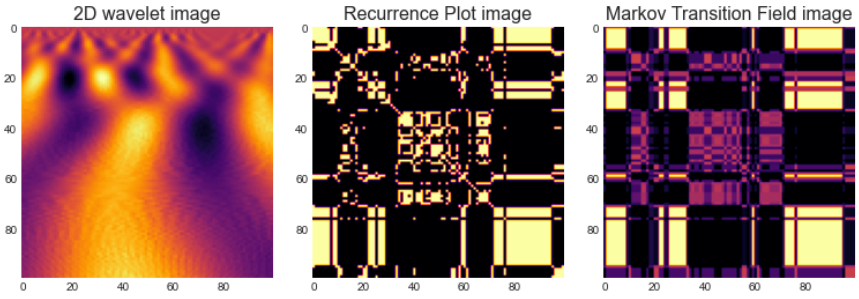


Figure 5.29: 2D wavelet representation of regulation volume and Recurrence Plot and Markov Transition Field representation of weather uncertainties.

The pre-trained model used is a version of the EfficientNet model pre-trained on the imagenet dataset. Additional layers were added at the output of the pre-trained model. Figure 5.30 show the pre-trained EfficientNet model with the added layers. The EfficientNet model is made without its top, and the final output layer is substituted by

the layers shown below the EfficientNet model. Details on the EfficientNet model is excluded and can be found in the EfficientNet paper by Tan et al [48].

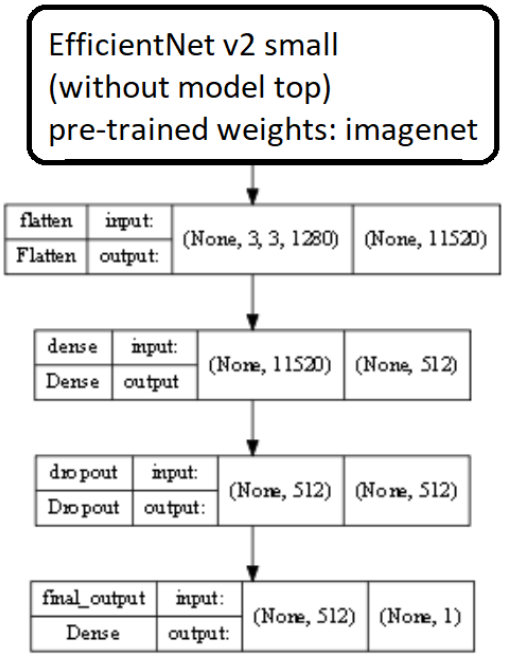


Figure 5.30: Pre-trained EfficientNet version2 small with additional layers added

5.4.3.1 Results - Step 2.5

Figure 5.31 shows predicted results from the spatial model that uses images constructed from one-dimensional data of regulation volume, and weather uncertainties.

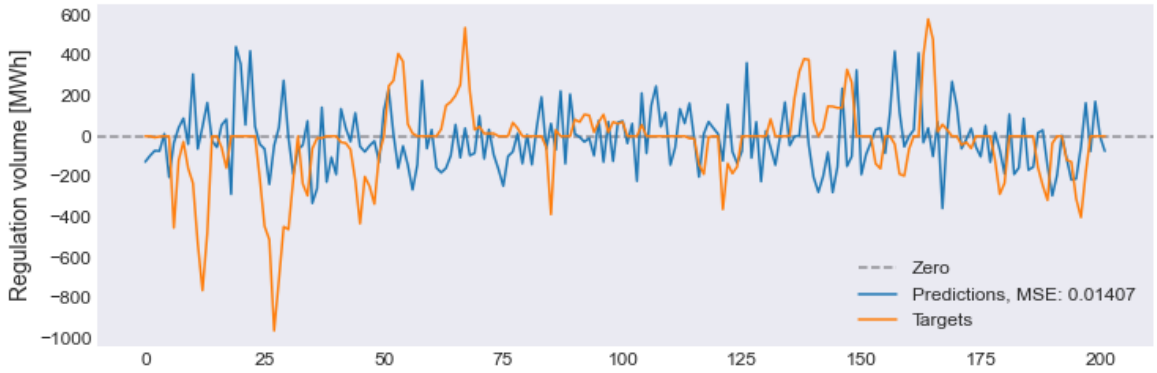


Figure 5.31: Results after training convergence. One hour predictions of NO5 regulation volume using estimated weather uncertainties formulated as an image analysis problem using constructed recurrence plot, Markov transition fields, and 2d wavelet images. The x-axis represents the first 200 hours in the test set used

5.4.3.2 Discussions - Step 2.5

The model only predicts one hour into the future and seems to perform worse than the first hour of forecast from the models that predicted only using one-dimensional data in the previous sub-step 2.4. Several tries using different hyper-parameters, e.g., learning rate, batch size, and dropout percentage, were tested. The prediction observed for this method mainly resembles noise around zero regulation. From the plot in figure 5.31 and the tests performed, it was not possible to determine if the model had learned anything. Thus this approach was not considered for further investigation. Since this experiment sub-step did not yield good results, the details on constructing the images were kept to a minimum to reduce thesis content. More information on recurrence plots and 2D wavelet representations images can be found in the paper by Bordvik et al. [47]. Additional theory and examples on converting sequenced one-dimensional data into an image of two dimensions can be found on the codestudyblog.com[7]. Besides the results using the EfficientNet version 2 small, some smaller and lightweight versions of the EfficienetNet architecture were tried out. However, they did not perform any differently from the model presented.

5.4.4 Step 2.6 - spatial modeling

Besides the deep investigation of extracted image statistics, spatial modeling using the $\Delta images$ directly using transfer learning is done in this step. The transfer learning part is enabled by fine-tuning (re-training) an already pre-trained model as was done in step 2.5 in Subsection 5.4.3. The pre-trained EfficientNet model was introduced in Figure 5.30. Pre-trained ResNet models (ResNet18 and ResNet50) was also tried were the ResNet [49] substitutes the EfficientNet box in Figure 5.30.

5.4.4.1 Results - Step 2.6

Figure 5.32 shows the 300 next hour predictions when using $\Delta images$ as input to the pre-trained EfficientNet model using a learning rate of 0.001 and dropout procentage of 40% in the dropout layer in Figure 5.30.

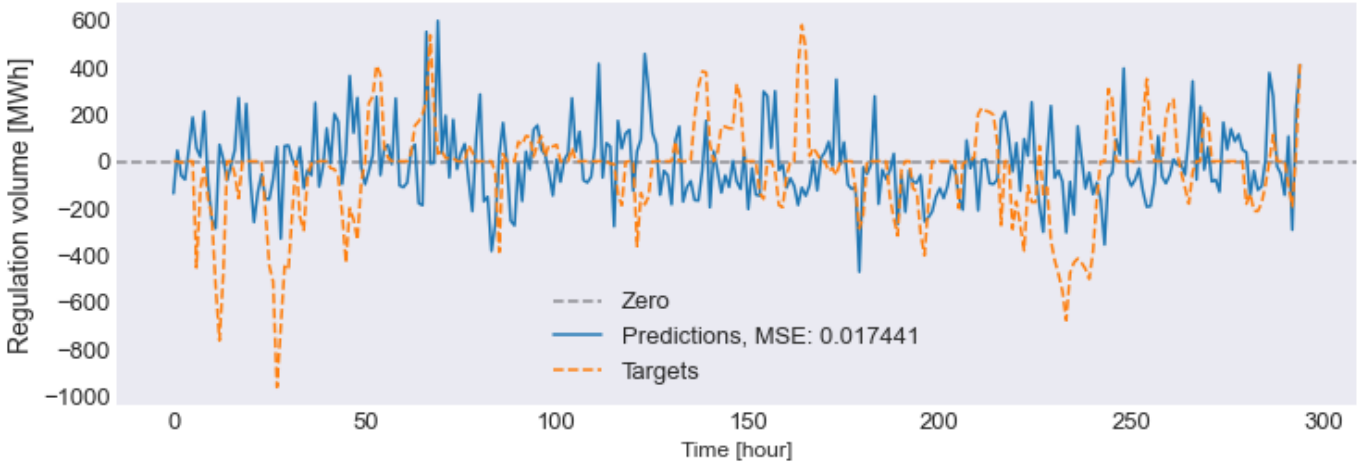


Figure 5.32: Predictions of NO5 regulation volume using delta weather forecast images

[7] https://www.codestudyblog.com/cs2201py/40119155307.html

5.4.4.2 Discussions - Step 2.6

The performance of the pre-trained model using $\Delta images$ directly has a higher normalized MSE value than what was discovered in the previous steps, step2.4 and step2.5. The MSE value shown in 5.32 is 0.017441, which is higher than all other models in Experiment1 and Experiment2. Regardless of the high MSE value, some predicted peaks, e.g., at around 300, indicate that the model has learned something. However, the level of noise in the predictions makes it challenging to analyze to which degree the model has learned anything. Based on the observed situations in Figure 5.32 which looks like the model has learned something, the model seems less greedy in optimizing only for the lowest possible MSE value and that it has sacrificed a lower MSE over the ability to follow some patterns. This is likely due to a better generalization from using a dropout probability of 40% before the output layer. Without the dropouts, the predictions were found to be more flat and close to zero. Predicting values close to zero or the mean was discussed in the data analysis as a relatively decent prediction, given that the data is dominated by zero values and the mean being close to zero.

I investigated the difference of fine-tuning only the added layers (all pre-trained layer locked) versus fin-tuning all pre-trained and added layers. This played out differently in therm of model performance. When the pre-trained part is locked, the model seems to converge faster but towards a more flat prediction close to zero or the mean value. Pre-trained models that have all the pre-trained layers locked during fine-tuning were generally found to struggle with predicting regulation volume fluctuations and spikes. It was found to be slightly more successful in fine-tuning all layers.

The image types in the imagenet dataset are generally very different from topographical weather data. Additionally, the pre-trained models are pretrained on a classification problem with categorical labeled target values. In many ways, classification problems and classified data are thresholded values representing the probability of an input belonging to a certain label or class. In contrast, stochastic regression problems have continuous floating targets. It is therefore believed that when modeling using the weather data on regulation volumes the pre-trained part requires some additional adjustments to weights and filters to capture more details of different weather scenarios and their relationship to floating target values. For example, it may be that one hour and the next subsequent hour must describe totally different regulation volumes since regulation volumes can spike one hour and have almost zero regulation volumes the next hour. Therefore, subtle differences in the weather data must be captured, and it was decided to standardize fine-tuning all layers.

5.4.5 Step 2.7 - spatial-temporal modeling

Since the images are sequenced in time, modeling the images using a spatial-temporal approach is also done in step 2.7, where a Convolutional LSTM (ConvLSTM [50]) is used. The idea of modeling the sequence of images is to capture the spatial-temporal correlations between image frames and regulation volumes in time. In step 2.7, the delta images are treated as a video stream (sequences of time-series delta images). Inputting a single weather type and multiple weather types in different channels were tried out. Two architectures of the ConvLSTM was used, one simple made from scratch using Keras[8] in Figure 5.33 and another proven and advanced ConvLSTM presented in an

[8] https://keras.io/

medium.com blog post[9]. Figure 5.34 shows the advanced model, which was slightly modified to fit the weather data used. The advanced model does support multiple heads, e.g., one head per target for multiple targets or labels, and this part was not put to use. It was also necessary to change the order of dimensions since the initial architecture used channels first, while I have designed all dataloaders for weather data to have channels as the last dimension. Both models were tried using different hyperparameters. The hyperparameters presented in Table 5.5 represents the parameters used for the results presented in the results Subsection 5.4.5.1.

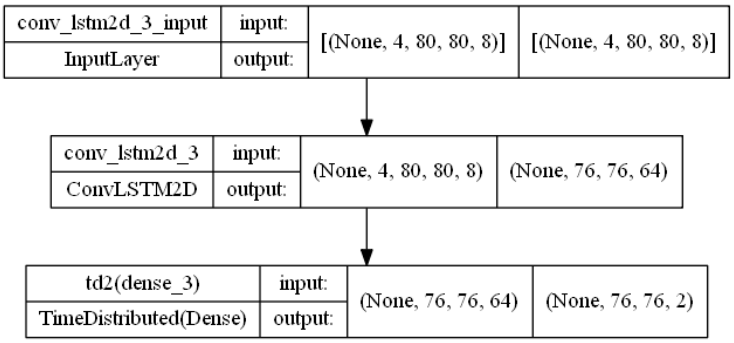


Figure 5.33: Simple ConvLSTM model used

[9] https://medium.com/neuronio/an-introduction-to-convlstm-55c9025563a7

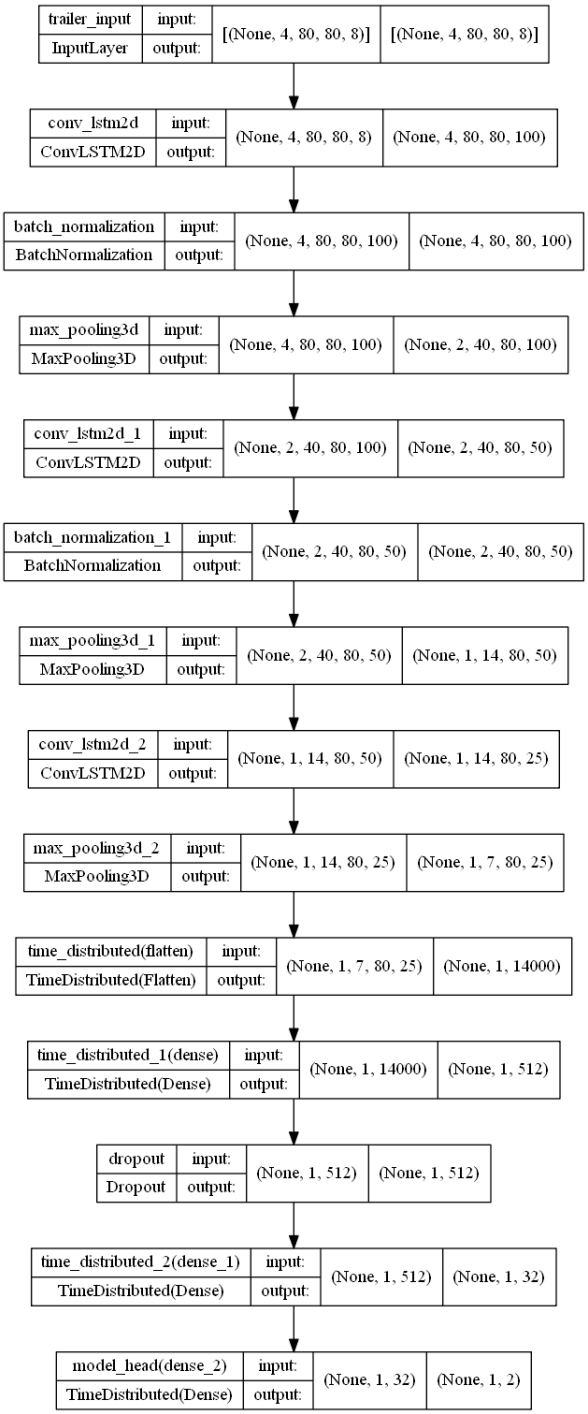


Figure 5.34: Advanced ConvLSTM model based on a modified version of a medium.com blog post[10]

Learning rate	RNN Units	Batch size	Sequence Length	image size
0.0001	64	100	4	80x80

Table 5.5: ConvLSTM parameters used for the simple model

5.4.5.1 Results - Step 2.7

Figure 5.35 shows predicted result for hour1 forecast from the spatial-temporal modeling using Δ*images* and the advanced ConvLSTM model architecture in Figure 5.34.

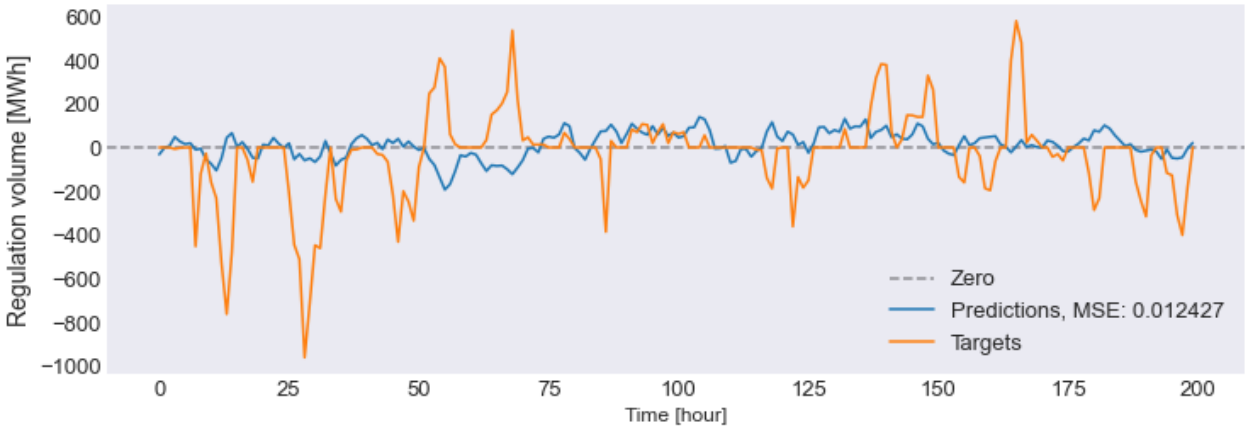


Figure 5.35: Predictions from convLSTM of NO5 regulation volume using delta weather forecast images. Trained using 60 epochs

5.4.5.2 Discussions - Step 2.7

After comparing two poorly performing models, it was concluded that the results from the advanced ConvLSTM model5.34 looked slightly more interesting. However, the model is not able to find spatial-temporal relationships between Δ*images* and regulation volume for NO5. It does not come as a surprise that the model is not able to relate spatial-temporal patterns with regulation volume, given the weak descriptive power generally found from Δ*images* and for the estimated weather forecast uncertainties.

5.5 Experiment3

Experiment 3 intends to model using the weather forecast images directly. Figure 5.36 outlines the experiment conducted. Only results and discussions are included since the modeling is done by re-using the setup from the spatial and spatial-temporal steps experiment 2, Section 5.4.4 and Section 5.4.5. Only results and discussions are included.

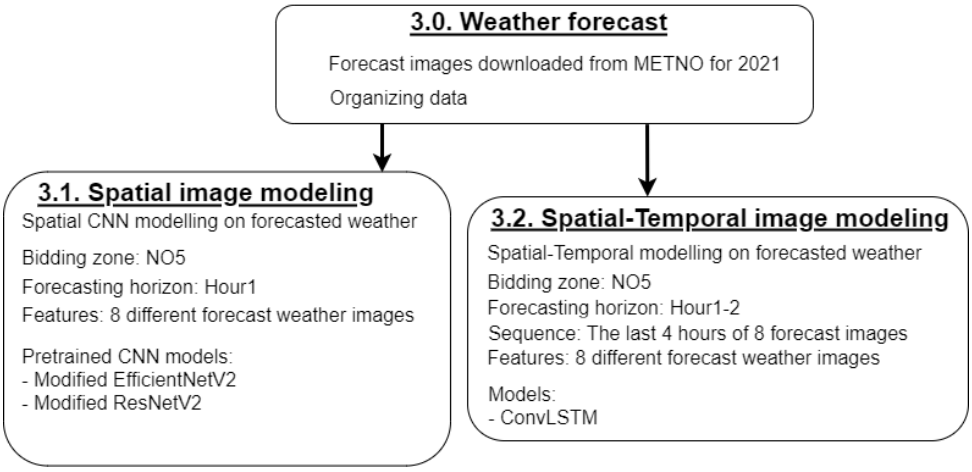


Figure 5.36: Machine Learning Experiment3

5.5.1 Results - Step 3.1

Figure 5.37 shows the 300 next hour predictions when using weather forecast images as input to the pre-trained EfficientNet model using a learning rate of 0.001 and dropout procentage of 40% in the dropout layer in Figure 5.30.

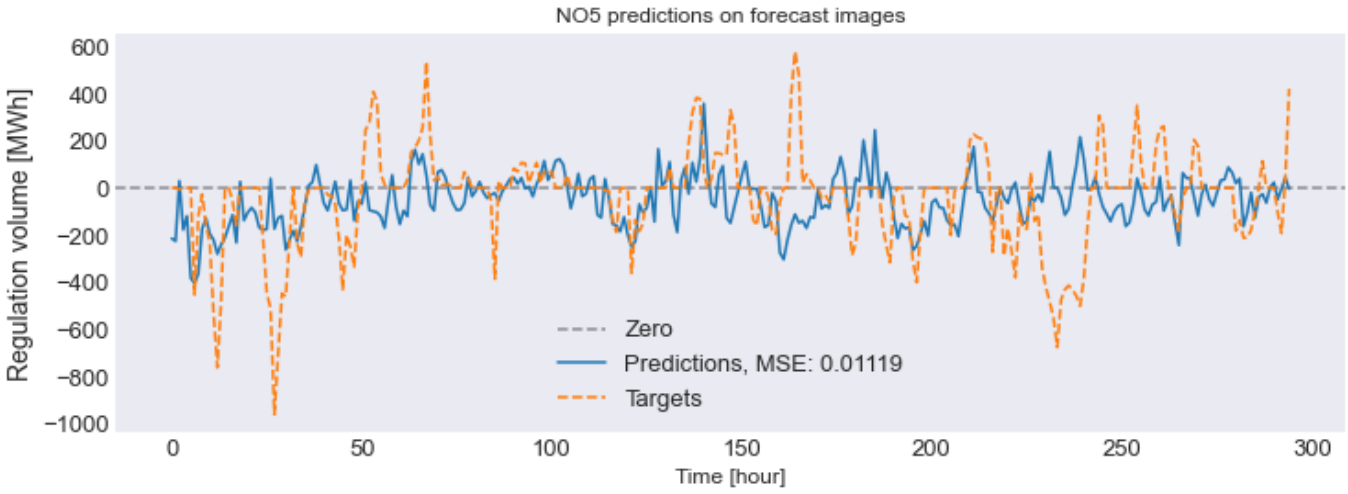


Figure 5.37: Predictions of NO5 regulation volume using weather forecast images

5.5.2 Discussions - Step 3.1

Modeling using weather forecast images does not reveal promising results in Figure 5.37. Other tries with different hyperparameters were neither promising. When comparing the

predicted results from the Δ*images* in Figure 5.32 to the weather forecast images in Figure 5.37, the prediction using weather forecast images has a lower MSE value. However, the prediction using weather forecast images seems to be less noisy and closer to a zero prediction than for the Δ*images*. It may be that Δ*images* are slightly more correlated with the regulation volume of NO5 than the weather forecast images. Moreover, this consideration may not hold since the predictions from both models are generally very noisy, and no statistical confidence interval is computed on the results. It is infeasible computationally wise to crate statistics on predictions using weather data since it can take 1-2 working days to train one model.

A modeling trade-off is discovered when using transfer learning for predicting regulation volume. Due to the low influence from the weather features and images in general, the modeling often struggled to prioritize between a greedy optimization towards the lowest possible MSE over the few situations where weather plays some part. Optimizing for the lowest possible MSE using transfer learning often results in a more flat prediction that centers around the zero or mean value of the regulation volume target. Introducing techniques to generalize more during modeling comes at the cost of a higher MSE and a more noisy prediction where some rare predictions seem to follow the target value.

5.5.3 Results - Step 3.2

Figure 5.38 shows predicted result for hour1 forecast from the spatial-temporal modeling using weather forecast images and the ConvLSTM model.

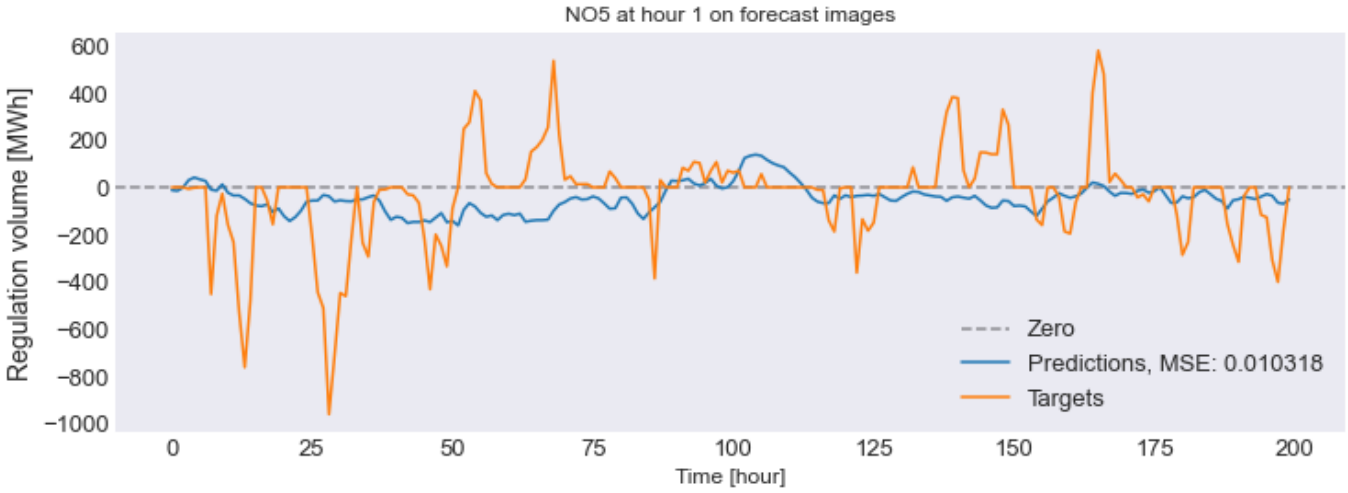


Figure 5.38: Spatial-temporal predictions of NO5 regulation volume using weather forecast images. Predictions show predictions one hour into the future.

5.5.4 Discussions - Step 3.2

Similar to the spatial-temporal modeling on Δ*images* in Section 5.35, modeling using weather forecast images does not show promising results when forecasting hour1 in Figure 5.38.

5.6 Experiment4

This experiment is outlined in Figure 5.39. Results and discussions from experiment 4 were excluded from the report since it did not contribute to additional insight for answering the research question. In addition, step 4.1 was found to overlap with the spatial modeling in experiment 3, and step 4.2 was found to overlap with the spatial modeling done in experiment 5.

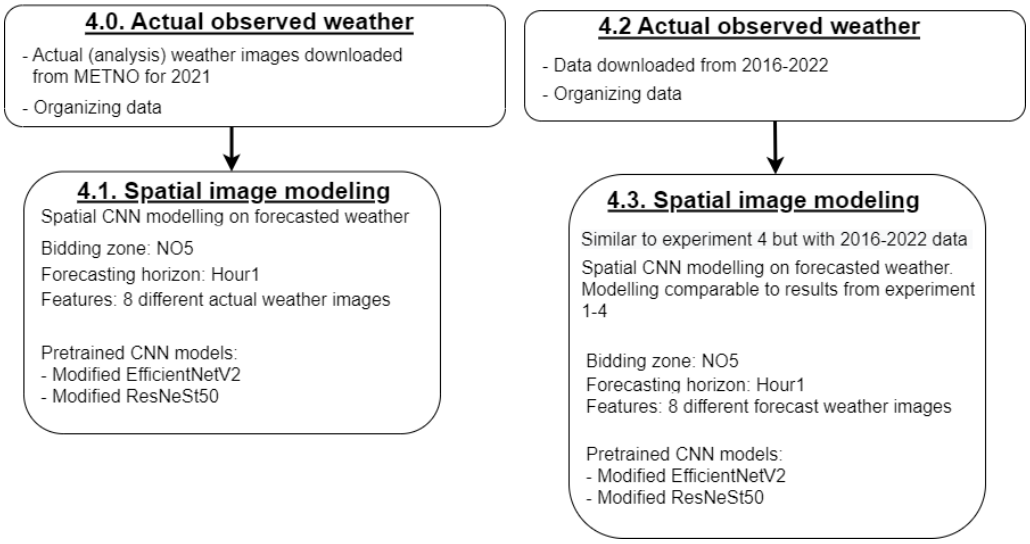


Figure 5.39: Machine Learning Experiment4

5.7 Experiment5

Figure 5.40 outlines experiment 5. Experiment 5 is framed as an experiment in the report but was not planned and structured as an experiment. Experiment 5 is rather a "sandbox" or a summary of many things that were tested and investigated using actual weather data and market data from 2016-to 2021. Therefore, it was decided to keep the work done in experiment 5 separated from the other experiments and to prioritize space in the report for experiments 1-2. All modeling in experiment 5 was done using Pytorch, whereas TensorFlow was used for the other experiments.

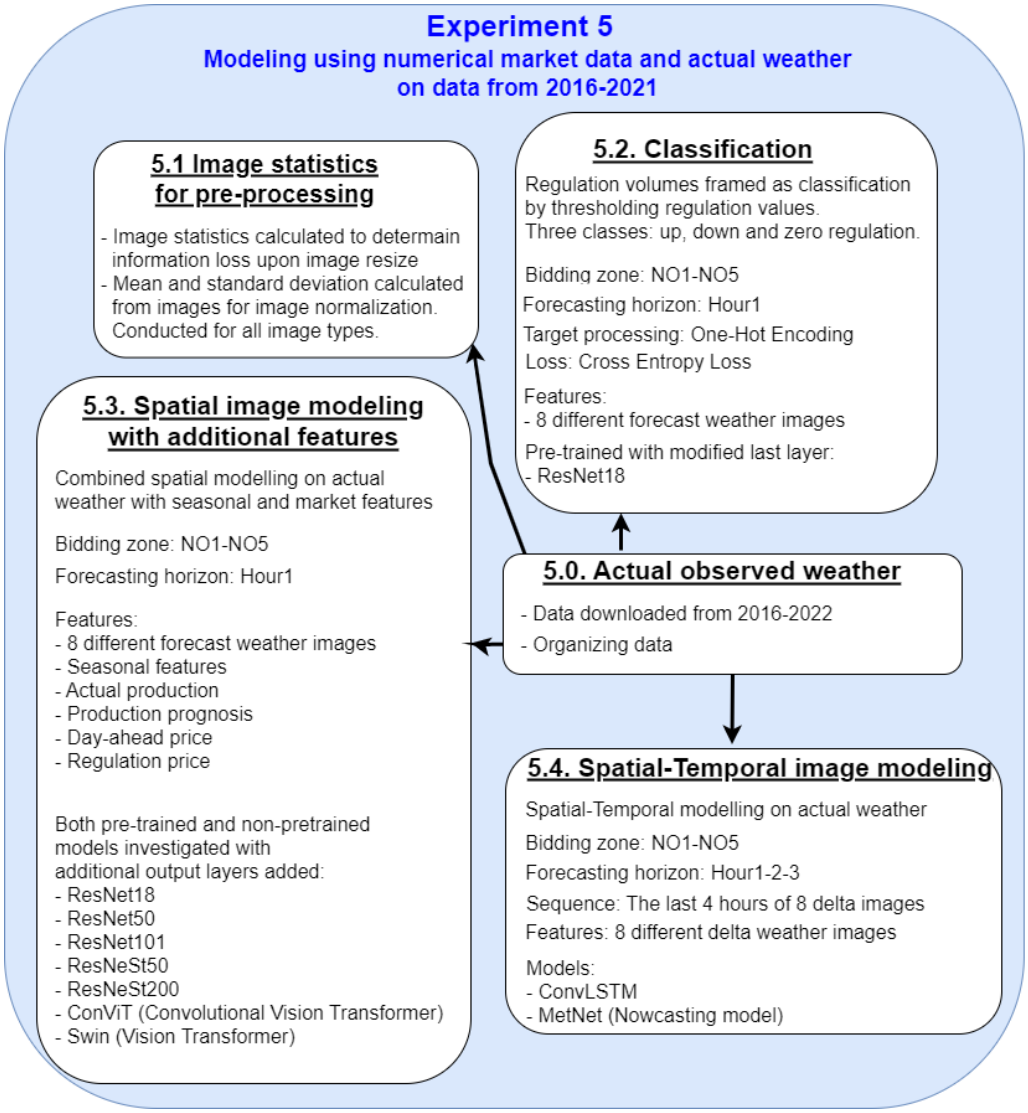


Figure 5.40: Machine Learning Experiment5

5.7.1 Classification

The classification tests done on actual weather data were performed using three classes. Up, down and zero regulation with thresholded values of regulation volume. The results of the classification was unsatisfying even using weighted loss for the different classes. The results from the classification test resulted in not pursuing a path of classifying the

regulation volumes.

5.7.2 Spatial-temporal modeling

The spatial-temporal tests performed during experiment 5 were done using a ConvLSTM model and the MetNet[51] model, which is primarily built as a Neural Weather Model for precipitation forecasting. The model was cloned from a GitHub repository[11] and then configured to forecast regression values instead of new weather forecast images. The training of the model looked relatively normal. However, the model was strongly biased towards predicting zero values for any tried parameters. Furthermore, the ConvLSTM[12] used did neither perform well on the actual weather data.

5.7.3 Spatial modeling

Several state-of-the-art models were tried out on the actual weather data. Two vision transformer models were investigated; ConViT and the Swin transformers. The ConVit model was released June 2021 by Dascoli et al. [52], and the Swin in August 2021 by Liu et al. [53]. Modeling using the vision transformers was a tedious process and was found to be relatively difficult to use for modeling the weather data. When the models were trained on randomly sampled data, the performance of the vision transformer models was sub-par the split-attention network ResNeSt200 by Zhang et al. [54] published April 2020. It may be that the ResNet type model with attention mechanismn[54] and the inherent inductive bias[13] of being a CNN model outshines the vision transformers using the weather data. The vision transformer model must learn inductive biases[52], such as translation invariance in images and pixel localities, while CNN models already have these mechanisms intact from the convolutional operator. The vision transformers require much more data than CNN models to overcome the lack of inherent inductive bias [52]. The vision transformer and CNN models were utilized with image 8-channels and as pre-trained transfer learning models.

5.7.3.1 Results

All results shown come from the fin-tuning of the pre-trained ResNeSt200 model using all weather types (8 channels) with image size 96×96. The model is fine-tuned using the Adam optimizer[2] with the following hyperparameters: learning rate = 0.001, weight decay regularization = $1e^{-6}$, and batch size = 200. Figure 5.41 shows the prediction of NO5 regulation volume where training, validation, and test data is randomly sampled from the 2016-2021 dataset of actual weather images. The results shown are predictions one hour into the future. Figure 5.42 show the same predicted hours as in Figure 5.41 sorted from the lowest target value to the highest. Figure 5.43 shows how the weights in different layers (bias excluded in the figure) have changed from before the fine-tuning.

[11] https://github.com/tcapelle/metnet_pytorch
[12] https://github.com/KimUyen/ConvLSTM-Pytorch
[13] https://towardsdatascience.com/vision-transformers-or-convolutional-neural-networks-both-de1a2c3c62e4

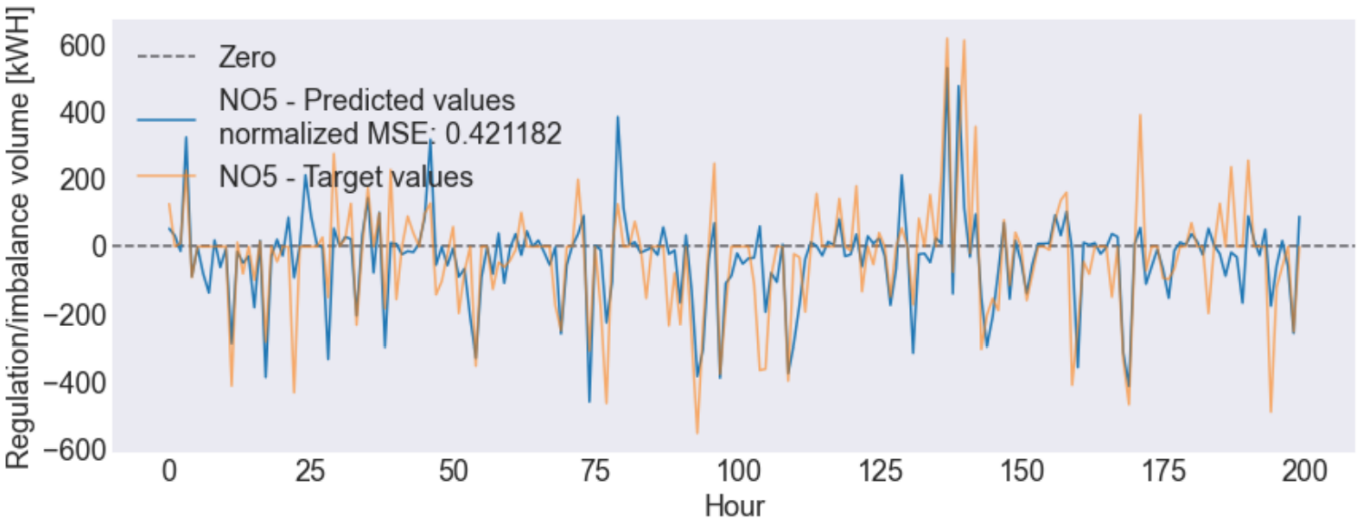


Figure 5.41: Spatial predictions on random sampled data of actual weather images. Target is regulation volume for NO5

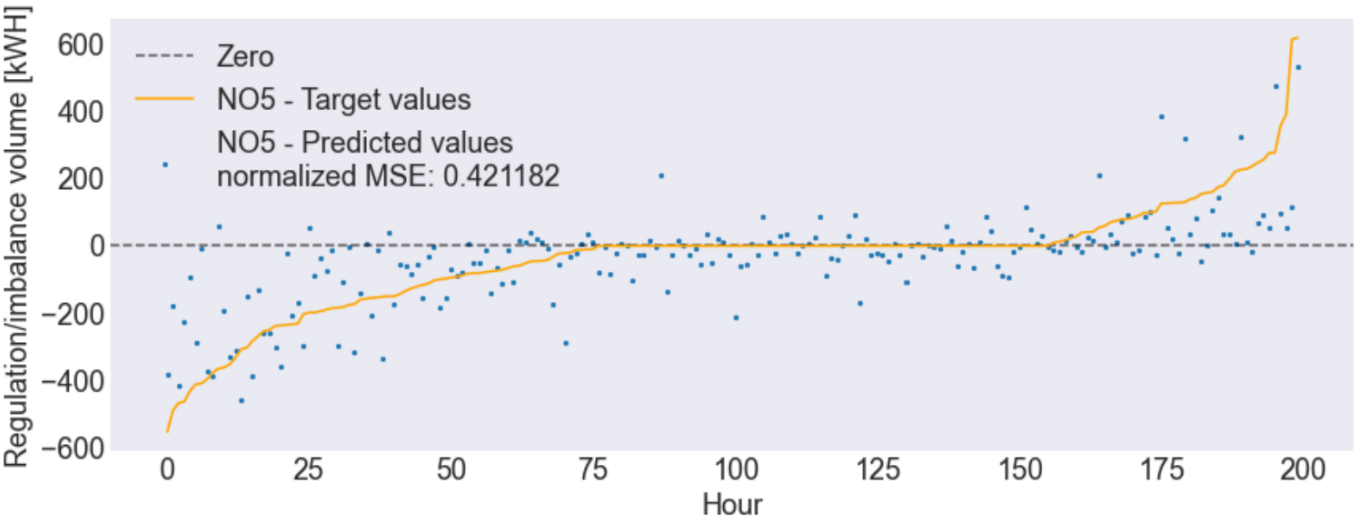


Figure 5.42: Spatial predictions on random sampled data of actual weather images. The plot is sorted from lowest to highest NO5 regulation volume target

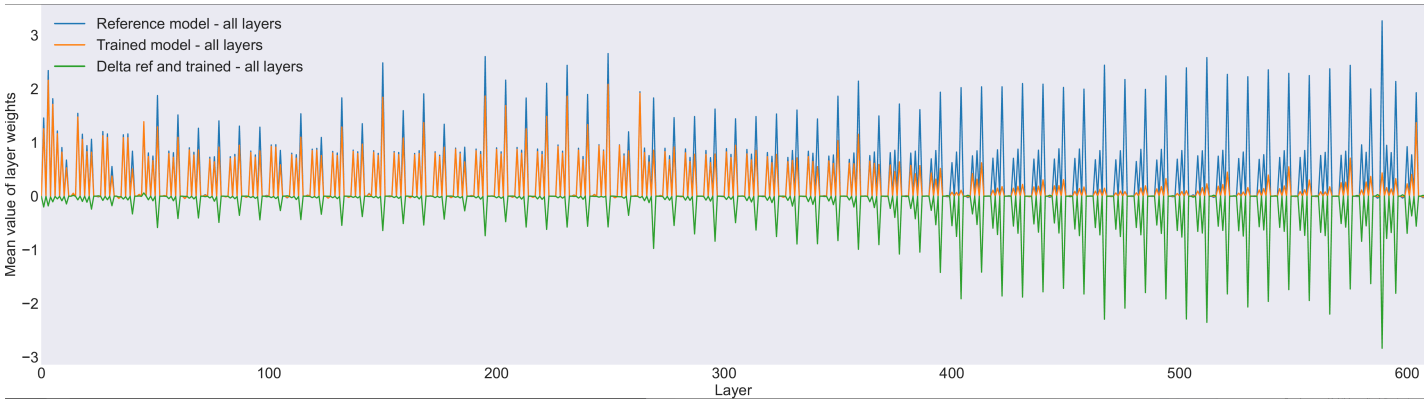


Figure 5.43: Mean change in trained model weights relative to before training starts

5.7.3.2 Discussions

When training, validation, and test images were randomly sampled, the predicted results were relatively good for all the spatial models that were tried out. This was demonstrated in Figure 5.41 using the pre-trained ResNeSt200 model. With the images split into fixed sequences, the performance became very noisy and unsatisfying, and the predictions were similar to what was shown for the weather forecast modeling in Figure 5.37. The weather images used in this thesis are time-series images. This means that the results presented in Figure 5.41 and 5.42 must be taken with a grain of salt since the correct approach for time-dependent modeling images is by first sampling the images in fixed continuous sequences. Figure 5.42 is included to show a different view of the predictions in Figure 5.41 with emphasis on the tails (down-regulation and up-regulation).

The weather images are very dissimilar in structure to the ImageNet images. However, deep learning models trained on images learn features similar to well-known image-processing filters. This is interpreted as the first layers describing general information and the last layers being task specific [55]. Hence it is expected that a pre-trained model will converge faster than a model trained from scratch. It is interesting to study the transfer learning process. In Figure 5.43 I compare the average of the weights (within all layers) for the original pre-trained model with the average of the fine-tuned (transfer learned) model. Post fine-tuning, one can see that the initial layers change less than the end-layers as supported by theory.

Chapter 6

Conclusion and future work

6.1 Conclusion

The data analysis in Section 4 demonstrated the complexity of the power market and identified the challenges with predicting power regulation several hours ahead in time. Additionally, the lack of previous comparable work influenced the decision on thoroughly analysing the data and to search for possible drivers behind regulation volumes. The exploratory data analysis revealed much of the complexity in regulation volumes for the Norwegian regulation market, and gave an indication on the level of noise inherent in the data. This thesis focused only on modelling regulation for the bidding zone NO5, but the applied methods are directly transferable to the other bidding zones.

6.1.1 How does the regulation market relate to other markets and are there any significant drivers of activation of tertiary reserves volume?

It is evident that the regulating market serves its own purpose of balancing, and the balancing volumes are significantly different from the traded volumes in the day-ahead market and the hourly actual produced power. However, the difference between hourly actual power production and production prognosis is strongly linearly correlated with the regulation volume, making power production and production prognosis essential features in a machine learning setting, Subsection 4.12.1. Furthermore, it has been suggested that the relationship between regulation volume and the difference between the price in the day-ahead market and price in the regulation market has been decreasingly correlated Klæboe et al. [27]. However, that study showed this only using the linear Pearson correlation. I used Spearman correlation and it revealed a strong relationship between the regulation volume and delta price in Subsection 4.12.1, which proves that the relationship has developed non-linear attributes over the years. The exploratory data analysis showed that regulation volumes relate to day-ahead volumes differently than how regulation price relates to the day-ahead price. Therefore, the relationship between the regulation market and the day-ahead market is compounded. The most influential force behind regulation volume at a given hour was found to be the regulation volume at the hour before. However, it can be debatable if lags of regulation volume are a driver for regulation volume or if it is more a subject for exploitation to enhance regulation volume forecasts.

6.1.2 Is regulation volume predictable or just stochastic white noise?

The data analysis takeaways Section 4.15 concluded that the regulation volume is not purely random or stochastic white noise and that modeling can benefit from using a suitable machine learning model with emphasis on deep learning. Experiment 1 in Section 5.3 further proves that regulation volume is predictable while also indicating how predictable it is. Predicting just zero values or the mean value of regulation volume were found to be good baselines since a major part of the data is zero and centered around a mean value close to zero, chapter 4. Therefore, all models were measured according to just using zero and the mean value as a fixed prediction. In terms of the predictability of regulation volumes, experiment 1 in Section 5.3 showed it being predictable both using data for one year (2021) and when using data over several years (2016-2021), which no other research has previously investigated.

6.1.2.1 Univariate modeling

In the univariate modeling, the models are only fed with historical regulation volumes limiting the models only to exploit information in the historical observed regulation volume. The motivation behind this was to exclude other input data, and possible extra added noise, thereby focusing the model on the signature of regulation volume fluctuations. Investigating a univariate setup was inspired by the exploratory data analysis that revealed a high autocorrelation of the previous regulation hour (lag1) in Figure 4.21.
Using only 2021 data revealed a significant leap in improved performance by the bidirectional LSTM model relative to the linear regression, the Support Vector Regression, and the XGBOOST models. The last four lags were used as input features in all models to forecast 4 hours into the future. No more than four lags were considered based on the partial autocorrelation from the data analysis showing only a strong correlation for lag1. The other lags are weakly correlated. However, it was logical to use at least four lags when forecasting 4 hours. All models were evaluated on the same test data, December 2021. The BiLSTM model significantly outperformed the other models with a lower MSE value at the fourth forecasted hour than the other models when forecasting the second hour 5.8. The MSE score alone does not necessarily reveal all about the model performance, and qualitative plots revealed that all models except the BiLSTM had a right-shifted offset in the predictions (figure 5.9 and 5.10). The BiLSTM generally performed better due to its strong ability to learn sequences that are both linear and non-linear. Additionally, the bidirectional BiLSTM architecture was able to correct much of the temporally shifted offset of other models by learning the regulation volume signature both in a forward and reverse order during training. Figure 5.15 shows the results from the univariate modeling trained on data from 2016-2021. The MSE is lower than for 2021 only data 5.8 indicating that more data further improved the results of the BiLSTM. The other models did not show significant improvement gains using more data.

6.1.2.2 Multivariate modeling

The multivariate models followed similar architectures to the univariate but with additional input features. The multivariate BiLSTM model did not outperform the univariate model for any of the four forecasting hours when trained on the 2021 data or the 2016-2021 data. Both cases were evaluated on the last 800 observations in 2021. The multivariate BiLSTM had a slightly higher MSE score. However, the multivariate model seemed

to be better at high peak predictions for some situations investigated. An example is seen around hour 25 in figure 5.19 versus 5.20.

The best models were the univariate and multivariate models trained on the 2016-2021 data. The models performed consistently over many random initializations, as seen in Figure 5.22. For the best models, I explored the capabilities for forecasting up to 10 hours. For the forecasted hours 6-10, the multivariate model outperformed the univariate model (Figure 5.24). The univariate BiLSTM best exploits historical lags when forecasting near future hours hour1-5 while forecasting longer into the future results in the historical lags becoming more distant and less useful. Therefore, from hours 6-10 the multivariate model is preferred since it draws on the additional features to influence the predictions positively. The shift in feature importance was investigated using SAGE plots in Subsection 5.3.3.1, where Figure 5.12 shows the additional feature still contributing at hour4. At the same time, lags have decreased compared with the hour1 Figure 5.11.

6.1.3 Does weather influence the regulating power markets in Norway, and do certain weather conditions drive the need for down or up-regulation?

The case study on the Gyda storm in January 2022 revealed unusual activities in regulating power activation for the duration of the storm, which was also seen during the brief investigation of the Frank storm occurring in January 2021 (Subsection 4.13.1). The two events influenced the regulating power markets in a secondary manner, meaning that extreme weather may lead to events, e.g., power blackouts, which again leads to grid balancing, which adds another stochastic event to the regulating power data.

Besides what has been tested in this thesis and the KoBas project, no published work was found to use similar 2-dimensional weather images to forecast regulation volumes. In experiment5 5.7, modeling using weather images of actual weather data was conducted in two ways; one using random shuffled image selection and one using fixed sequences for training, validation, and testing. Using randomly shuffled data resulted in a relatively good performance 5.41, indicating some associations in the data. However, it was not possible to generalize these associations when the test data was not correlated with the training data when the data was split into fixed continuous sequences. This indicates that the relationship between weather and regulation volumes was more complex than we had data available to describe in the KoBas project. Thus it is not possible to draw a hard conclusion on whether the regulation can be related to specific weather conditions. It is believed that more data would have improved the modeling when using temporal cutoffs to splitting the data into training, validation and test samples

Combining weather and market data was explored during weather data modeling in experiment5 (Section 5.7) and was briefly tested in the other weather experiments. Model performance of the weather models in experiment5 with or without the added market data features is indistinguishable when the model was evaluated on continuous intervals of unseen test data. The performance when data is not shuffled did not reveal a noteworthy direct influence on regulation volume from weather or when the weather was combined with market data. Examples of market features that were combined with weather are; power production, production prognosis, consumption, consumption prognosis, day-ahead price, and day-ahead volumes.

6.1.4 Does weather forecast uncertainties influence the need for grid balancing trough the activation of tertiary volumes?

No significant relationship was found between weather forecast uncertainties and tertiary volume activation in Subsection 5.4.2.1 and 4.14.4. Even though the analysis showed low correlations, the Spearman correlations were not zero. The correlation analysis in Experiment 2 in Subsection 5.4.2.1 showed slightly higher correlations than what was discovered in the data analysis chapter in Subsection 4.14.4. This indicates that the 06:00 published weather forecast is more likely used in the day-ahead planning than the 00:00 forecast since 06:00 is closer to day-ahead market clearing. Furthermore, the work done in experiment2 in Subsection 5.4.2 show that it is possible to estimate weather forecast uncertainties into the future only by using old and updated weather forecasts.

The machine learning experiments on $\Delta images$ in Subsections 5.4.4 and 5.4.5 revealed that weather forecast inaccuracies (the $\Delta images$ specifically) did not significantly influence the activation of regulation volume in general. However, the modeling results in 5.4.4 indicated that the $\Delta images$ play a role in some rare occasions. Moreover, extracting these situations through modeling is found to come at the cost of more noisy predictions in general, resulting in overall worse performance with a higher overall MSE score, Figure 5.32. It is believed that the noise level overshadows the few situations where weather forecast inaccuracies contribute to grid balancing and that more data is needed to better model the relationship.

6.1.5 Can information from a domain with images of animals, humans, etc, be transferred to better forecast the power regulation volumes from weather data?

This research question heavily depends on the initial potential in forecasting regulation volume using weather data. Weather in general or weather forecast uncertainties was not found to be significantly influential. However, I plotted how the layers change before and after fine-tuning a pre-trained model in Figure 5.43. This plot revealed some unchanged layers and weights, meaning some information being kept, especially those describing general information at the beginning of the architecture. I also experimented with loading in default weights instead of the pre-trained, resulting in worse performance. The models that were not available for transfer learning generally performed poorly for different hyperparameters. These models are ConvLSTM, MetNet, and the custom-built CNN. In conclusion, some general information can be transferred from other domains for marginally improved modeling of power regulation volumes using weather images.

6.2 Future work

I was not able to get satisfying results by modeling the regulation volume using weather data and weather forecast uncertainties. Therefore, further investigations using weather data is proposed for future work. In the long term power marked analysis forecast for 2021-2040[1] conducted by the Norwegian Energy Regulatory Authority NVE[2] the power markets are considered to become more weather dependent. Thus the potential in weather data to forecast power regulation volumes may increase. However, it may be that weather forecast accuracy improve in the future, making the potential for forecasting regulation volumes using weather data uncertain, and a topic for future work.

Due to the demanding work of modeling the relationship between regulation volumes and marked and weather data, an autonomous system that uses the models to give forecasts for grid balancing autonomously was placed out of scope for this thesis. Therefore, this part is reserved for future work and is an overall goal of KoBas.

[1] https://publikasjoner.nve.no/rapport/2021/rapport2021_29.pdf
[2] https://www.nve.no/energi/analyser-og-statistikk/langsiktig-kraftmarkedsanalyse/

Chapter 7

Appendix

7.1 Delta consumption

The delta consumption (prognosis - actual) correlation with regulation volume is show in figure 7.1

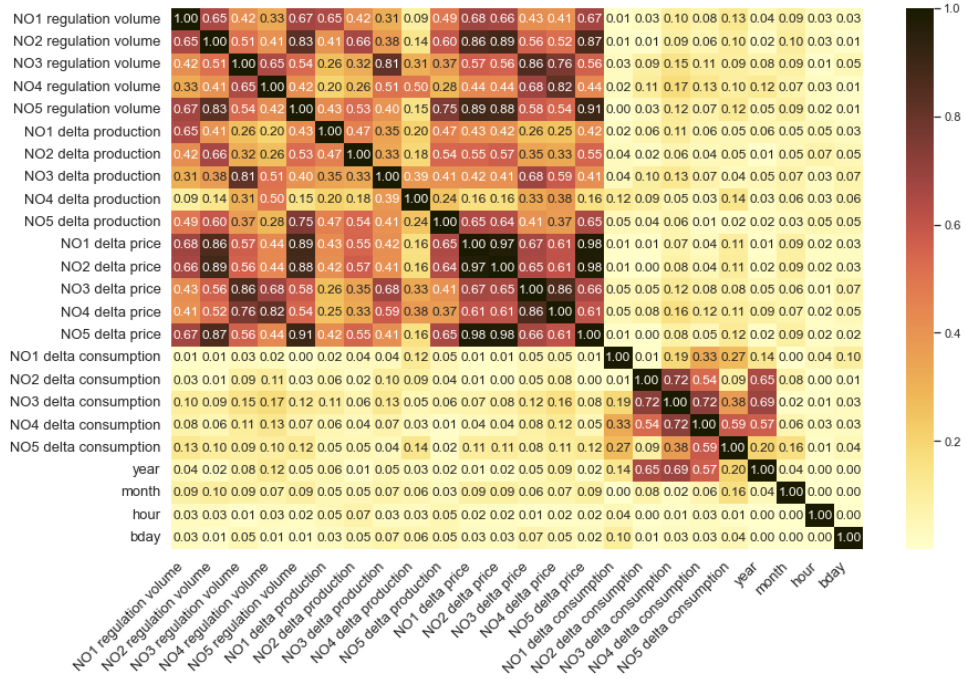


Figure 7.1: Spearman correlation heatmap of regulation volume predictors

7.2 Weather data

7.2.1 Image resize and statistics

For the modeling using PyTorch, the package torchvision was used. Listing 7.1 represents the image transformations utilized. In the listing, one can also see a list of several interpolation methods that were investigated for the resize transformation. The bicubic

interpolation is generally considered to produce high-quality results at the cost of a high computationally complexity[1], which impacts the training time. However, it was concluded to use bicubic interpolation at the cost of some training time to preserve image quality. Preserving the image quality and statistics was considered important since the images at their original shape of 2321×1796 had to be significantly resized down to make it feasible to train all 8 images types using 8 channels in a deep learning model.

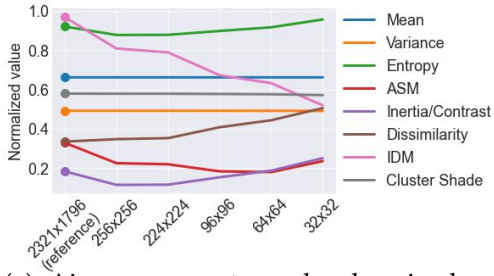
```
from torchvision import transforms

interp = [transforms.InterpolationMode.NEAREST,
          transforms.InterpolationMode.BILINEAR,
          transforms.InterpolationMode.BICUBIC,
          transforms.InterpolationMode.BOX,
          transforms.InterpolationMode.HAMMING,
          transforms.InterpolationMode.LANCZOS][2]

def get_transforms(img_size=[96,96], mean=0.5, std=0.5,
                   interpolation=interp,
                   kernel_size=P["kernel_size"],
                   apply_low_pass=P["apply_low_pass"]):

    if apply_low_pass:
        print("Gaussian lowpass is enabled")
        transform_1D = transforms.Compose([
                    transforms.Resize(size=img_size, interpolation=interp),
                    transforms.GaussianBlur(kernel_size=kernel_size,
                                            sigma=(0.1, 2.0)),
                    transforms.ToTensor(),
                    transforms.Normalize(mean=[mean,], std=[std,])])
    else:
        transform_1D = transforms.Compose([
                    transforms.Resize(size=img_size, interpolation=interp),
                    transforms.ToTensor(),
                    transforms.Normalize(mean=[mean,], std=[std,])])
    return transform_1D
```

Listing 7.1: Image normalization and transformation

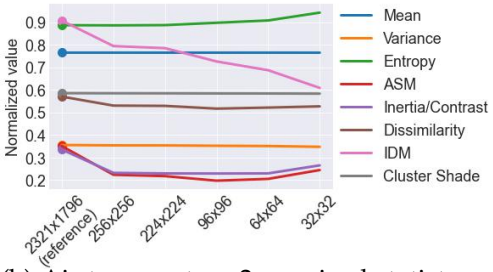
Figure 7.2 show the effect on the computed image statistics for all weather types for different image dimensions. The computed image statistics are mean, variance, entropy, Angular Second Moment (ASM), Inertia/Contrast, Dissimilarity, Inverse Difference Moment (IDM), and Cluster Shade (SHD). Most of the pre-trained image models utilized are pre-trained on images of dimensions 256x256 or 224x224. The image statistics are normalized. Some changes in the statistics after downscaling the images from their reference (original size) can be seen in the figure. A minor change in Dissimilarity and Inertia can be seen for Cloud area fraction in 7.2c from 256x256 down to 96x96. However, it is not known if Dissimilarity and Inertia are important for Cloud area fraction during model training. The overall look does not indicate major changes in the image statistics from 256x256 downwards. This indicates that the statistical properties of an image at, e.g., 96x96 are not significantly changed over 256x256, making the image size 96x96 a potential candidate after the commonly used 256x256 and 224x224. Going below 96x96 leads to the images being less complex to model, which may invoke more complex image models to overfitting at early epochs, thus making 96x96 a reasonable compromise to try out. The statistics are calculated from a sample size of 8000 ($\sim$ 20%) randomly selected timestamps between 2016-2021 from the actual weather images. Another effect

[1] https://en.wikipedia.org/wiki/Bicubic_interpolation

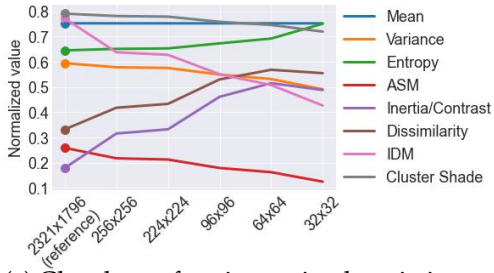
of resizing images besides what is computationally suitable is that images resized down into a lower dimension results in greater distances between pixels for different images, which may lead to a better generalization during modeling.



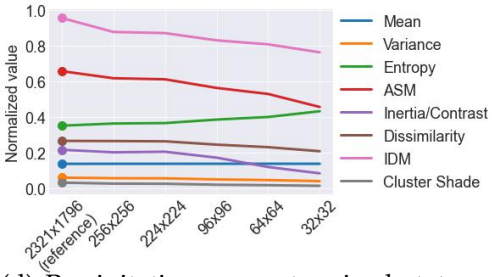
(a) Air pressure at sea level resized statistics



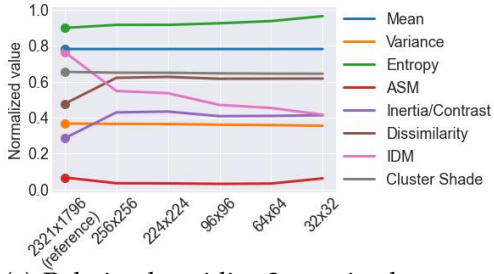
(b) Air temperature 2m resized statistics



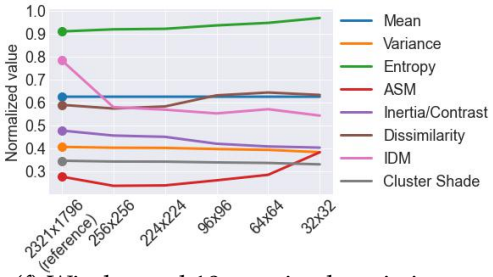
(c) Cloud area fraction resized statistics



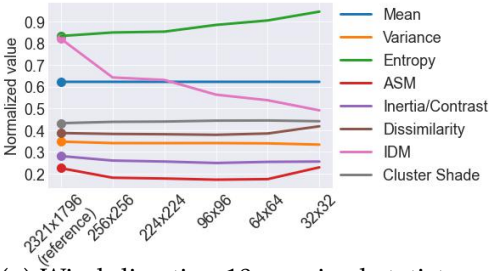
(d) Precipitation amount resized statistics



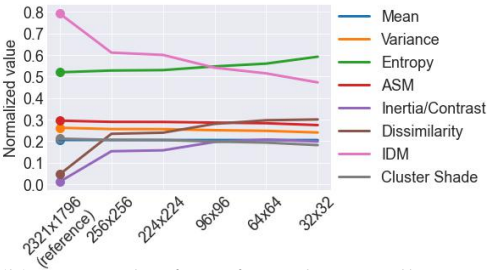
(e) Relative humidity 2m resized statistics



(f) Wind speed 10m resized statistics



(g) Wind direction 10m resized statistics



(h) Integral of surface downwelling shortwave flux in air wrt time resized statistics

Figure 7.2: Global weather image statistics for different downscaled sizes

7.2.2 Organizing weather images for modeling

The weather images for actual weather, weather forecasts, and Δ*weather* are all located in folders named according to their date and time YYYYMMDDHH; where YYYY represents the year, MM the month, DD the day, and HH for an hour. The naming

of the folders is done during the download of the data. Each folder includes weather images for all the 8 weather types for that particular hour. The image data is organized such that the folder name must match the hour of regulation volume to align the images with the regulation volumes at an hour-by-hour resolution. Aligning of all market data utilized and the path at which the images for all 8 weather types are located are made using a dedicated python script. 3 different scripts create 3 different datasets; one for the actual weather images, weather forecast images, and the $\Delta images$. The market data used is similar for all the 3 datasets. This process results in datasets that include the directory and its path for all relevant images for a given hour. The dataset created is then fed into custom-built dataloaders, which iterate the dataset during training, validation, and testing. This process is necessary to pick the right image paths, read images, apply image transformations, and collect regulation volume and other desired features on the fly during training, validation, and testing. The dataloader for the spatial models considers the hourly snapshots of weather and regulation volumes. The dataloader for spatial-temporal models utilizes the sliding window method to correctly create sequences (input windows) of image paths aligned with the correct regulation volumes to predict.

7.2.3 Image statistics - More details

The GLCM matrix is squared, and each dimension is equal to the number of gray level values. For 8 bit images, the GLCM matrix is then a 256×256 matrix. Constructing the GLCM matrix is done by looping through a source image, and for every iteration, pixel pairs are matched starting at a pixel i. The paired destination pixel j is then at a distance d with a direction θ from pixel i. The number of times the pixel pair at gray level i and j has occurred is counted. By counting the number of times, two considered neighboring pixels change (or do not change) from an intensity value x in pixels i to an intensity value y in pixels j in the image, the spatial relationship between these grayscale intensity pairs is stored.

For second order image statistics, they are derived from the Gray-Level Co-Occurrence Matrix (GLCM). The GLCM is calculated using four different angles since the weather data in the image grid stems from measurements having latitude and longitude coordinates. Weather is not bound to a specific direction, and considering statistics and pixel relations covering the span of 360 degrees is therefore relevant. Thus, a resolution using four angles covering the span of 360 degrees is used. Computing GLCM using four angles results in four different GLCMs, and since it is not practical to evaluate four different GLCMs, the mean value between all four matrices is calculated as the final GLCM. Moreover, the averaging of all GLCMs results in a Isotropic GLCM where the image texture is considered equal in all directions where $\theta \in [0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}]$. Contrary, if the textures in the weather images had fixed orientation for all images, the angle θ should have been selected in allignment to the static texture orientation. Additionally, the GLCM is computed as a symmetric GLCM. A symmetric GLCM regard the two pixel pairs in both direction. For a given angle and distance, the counting part is done both for $P(i,j)$ and $P(j,i)$, which regard not only the pixels relations, but also the pairwise relations.

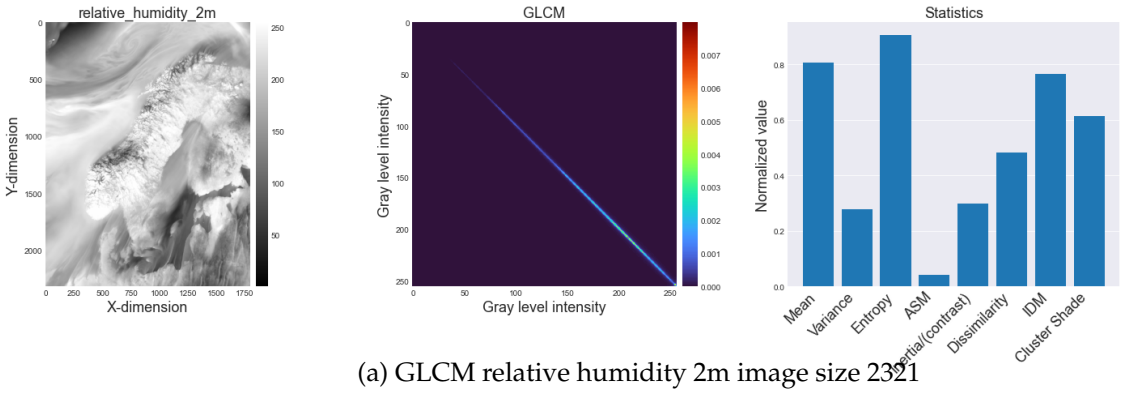
$$P\Big(d,0\Big) = P\Big(d,\pi\Big)$$

$$P(d,\frac{\pi}{4}) = P\Big(d,\frac{5\pi}{4}\Big)$$

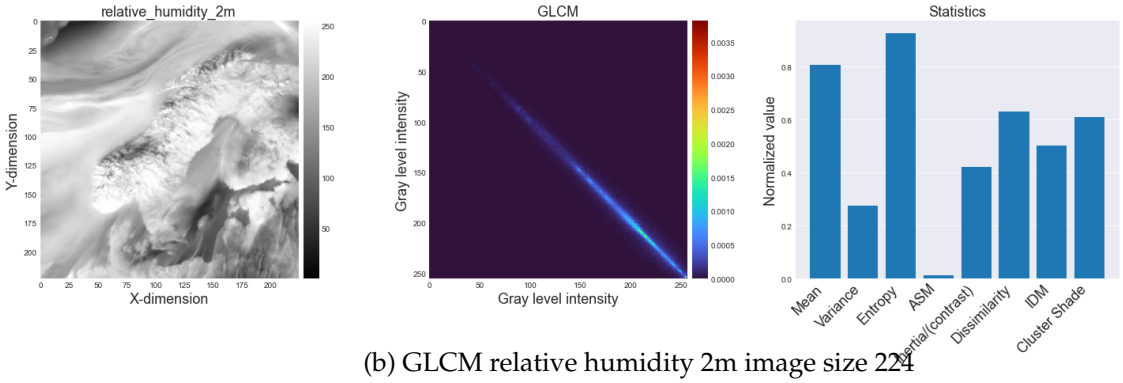
$$P\Big(d,\frac{\pi}{2}\Big) = P\Big(d,\frac{3\pi}{2}\Big)$$

$$P\left(d, \frac{3\pi}{4}\right) = P\left(d, \frac{7\pi}{4}\right)$$

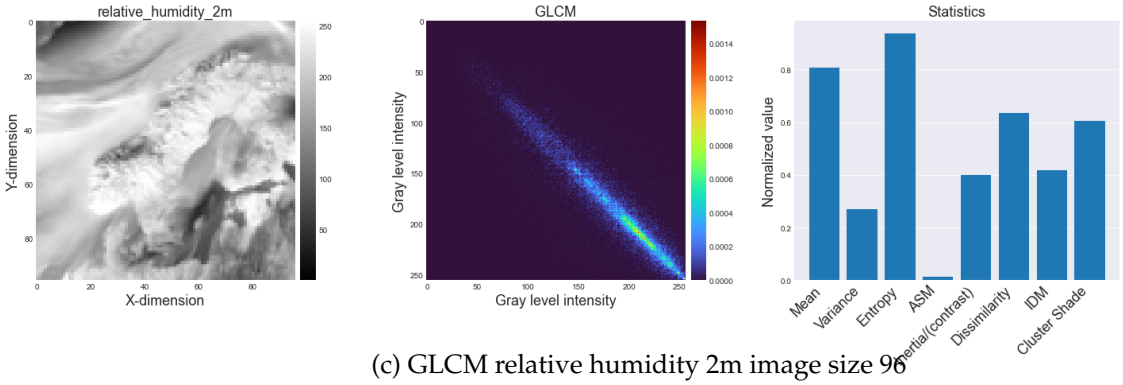
Only the closest neighbouring pixels are considered when computing the GLCM, resulting in a distance between the two pixel pairs to be $d = 1$. Furthermore, in addition to compute the GLCM as a symmetric matrix, the GLCM is also normalized.



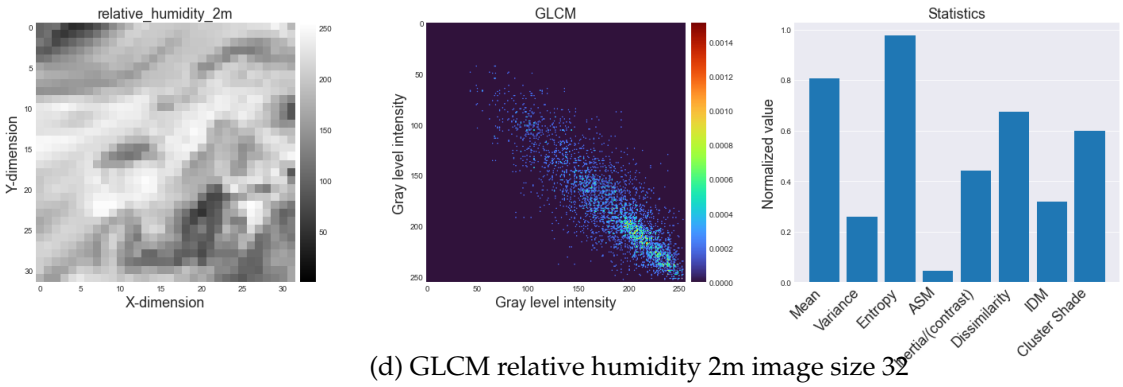
(a) GLCM relative humidity 2m image size 2321



(b) GLCM relative humidity 2m image size 224

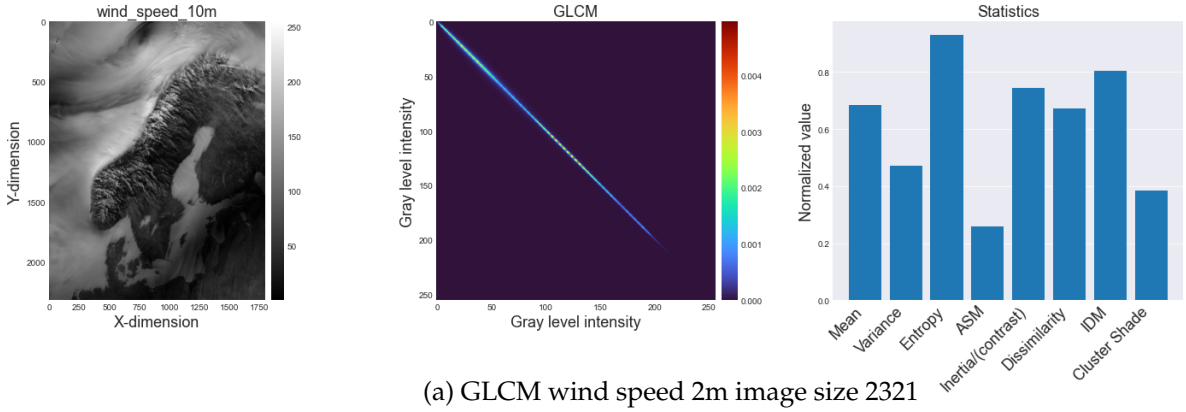


(c) GLCM relative humidity 2m image size 96

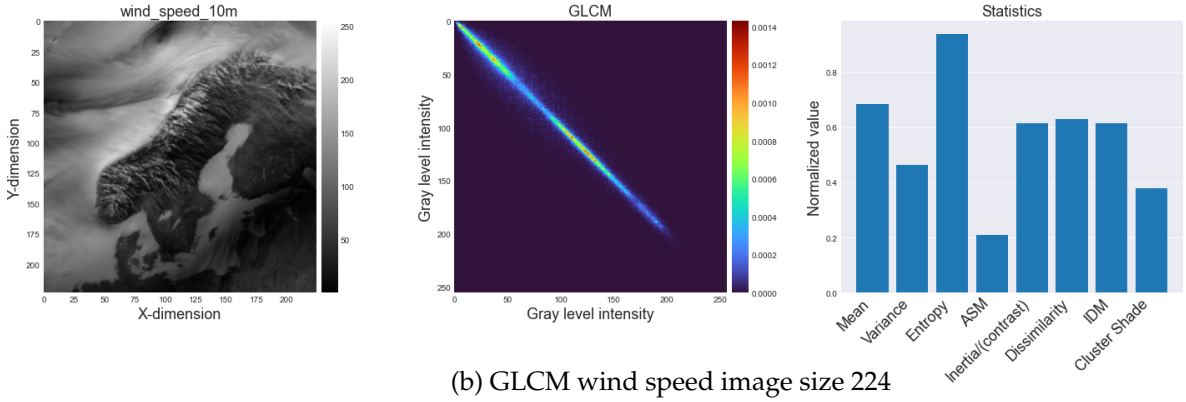


(d) GLCM relative humidity 2m image size 32

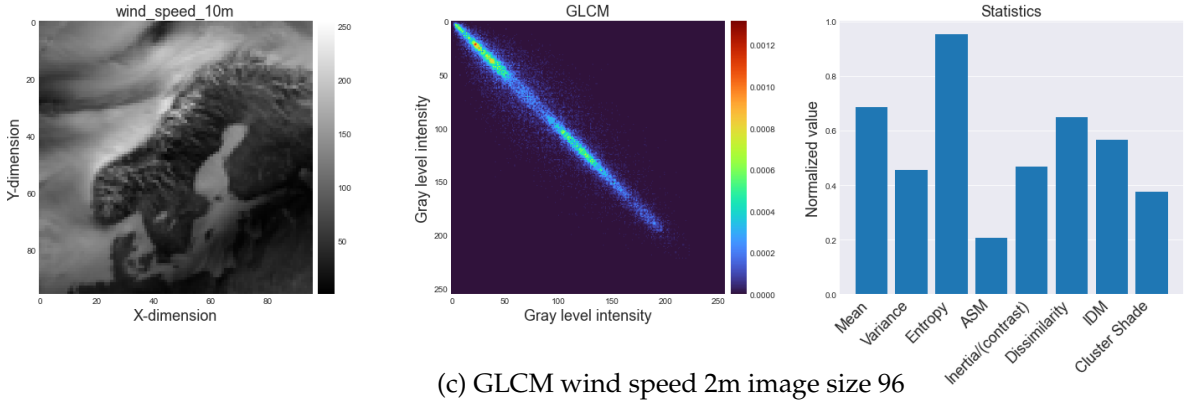
Figure 7.3: Relative humidity example image, its GLCM, and some globally aggregated statistics



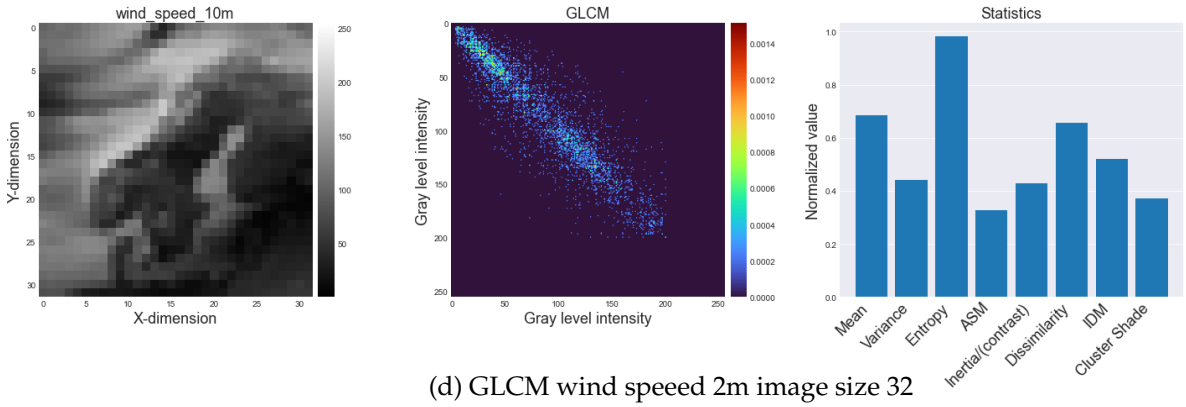
(a) GLCM wind speed 2m image size 2321



(b) GLCM wind speed image size 224

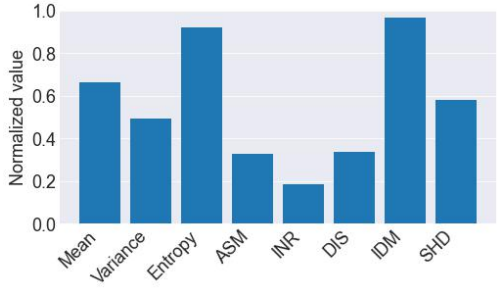


(c) GLCM wind speed 2m image size 96

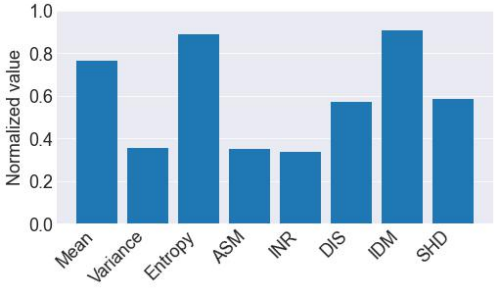


(d) GLCM wind speeed 2m image size 32

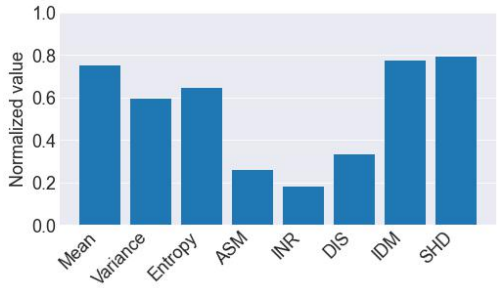
Figure 7.4: Wind speed example image, its GLCM, and some globally aggregated statistics



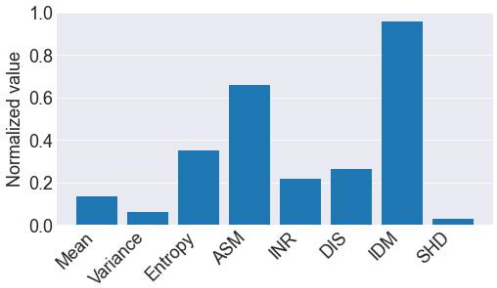
(a) Air pressure at sea level statistics



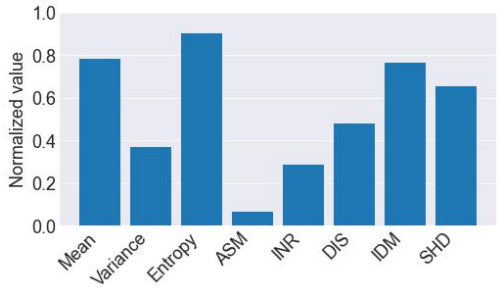
(b) Air temperature 2m statistics



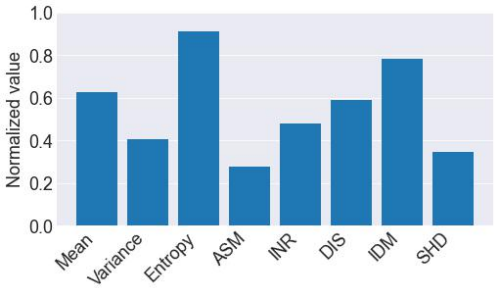
(c) Cloud area fraction statistics



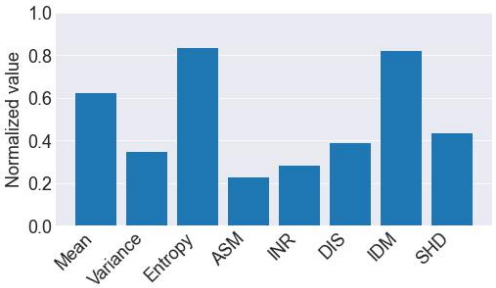
(d) Precipitation amount statistics



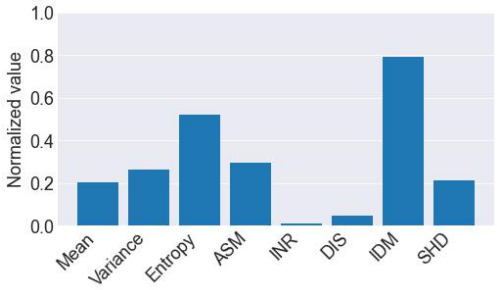
(e) Relative humidity 2m statistics



(f) Wind speed 10m statistics



(g) Wind direction 10m statistics



(h) Integral of surface downwelling shortwave flux in air wrt time statistics

Figure 7.5: Global mean weather image statistics for the original sized image

Weather type	**Mean** $[\mu]$	**Standard deviation** $[\sigma]$
Air pressure at sea level	0.5580	0.2137
Air temperature 2m	0.6592	0.1409
Cloud area fraction	0.7307	0.3197
Integral of surface downwelling shortwave flux in air wrt time	0.2042	0.1419
Precipitation amount	0.0073	0.0217
Relative humidity 2m	0.6986	0.1881
Wind direction 10m	0.5253	0.2138
Wind speed 10m	0.2955	0.1742

Table 7.1: Mean and sigma statistics used for normalization of images of 96×96 dimensions

7.3 Delta images for experiment2

Prior to creating the delta images the 00:00, 06:00, 12:00, and 18:00 forecasts were downloaded separatly as illustraded in the green, blue, red, and gray-colored forecast images boxes in figure 7.6. The day-ahead market aligned weather forecast was downloaded from the 06:00 the day before as illustrated from the black time-line in the figure. The yellow boxes and hours above the black time-line represents the regulation volume hour in UTC within the current day. The 6 hours of intra-day weather forecasts in the green, blue, red, and gray-colored forecast images boxes are considered "actual" weather shown in figure 7.6. These weather forecast images are substituted over the actual weather images that were used in the exploratory data analysis part. The process of calculating the delta images is described under the exploratory data analysis part and the calculation is done similarly using only forecast images. Referring to the code in listing 4.1 and 4.2, it takes in the actual weather image and the weather forecast image. In experiment2, the intra-day weather forecasts (00:00, 06:00, 12:00, and 18:00) replace the actual weather image input parameter of the function, while the weather forecast is considered the black timeline in figure 7.6. The black timeline covers the day-ahead planning schedule. As previously mentioned, the day-ahead market closes at 12:00. The most relevant weather forecasts for the day-ahead planning stem from the 06:00 forecast and cover the forecasts for the next day. The black time-line indicates this in figure 7.6. This process aligns the closing of the day-ahead market with the most relevant weather forecasts for the next day performed at 06:00. Additionally, the updated weather forecasts within the day measure how much the day-ahead-based weather forecast misses relative to the more accurate recent weather forecast. In short summary; short-term weather forecasts are used to determine how inaccurate the long-term weather forecasts are when the long term-forecasts are aligned with the day-ahead power market.

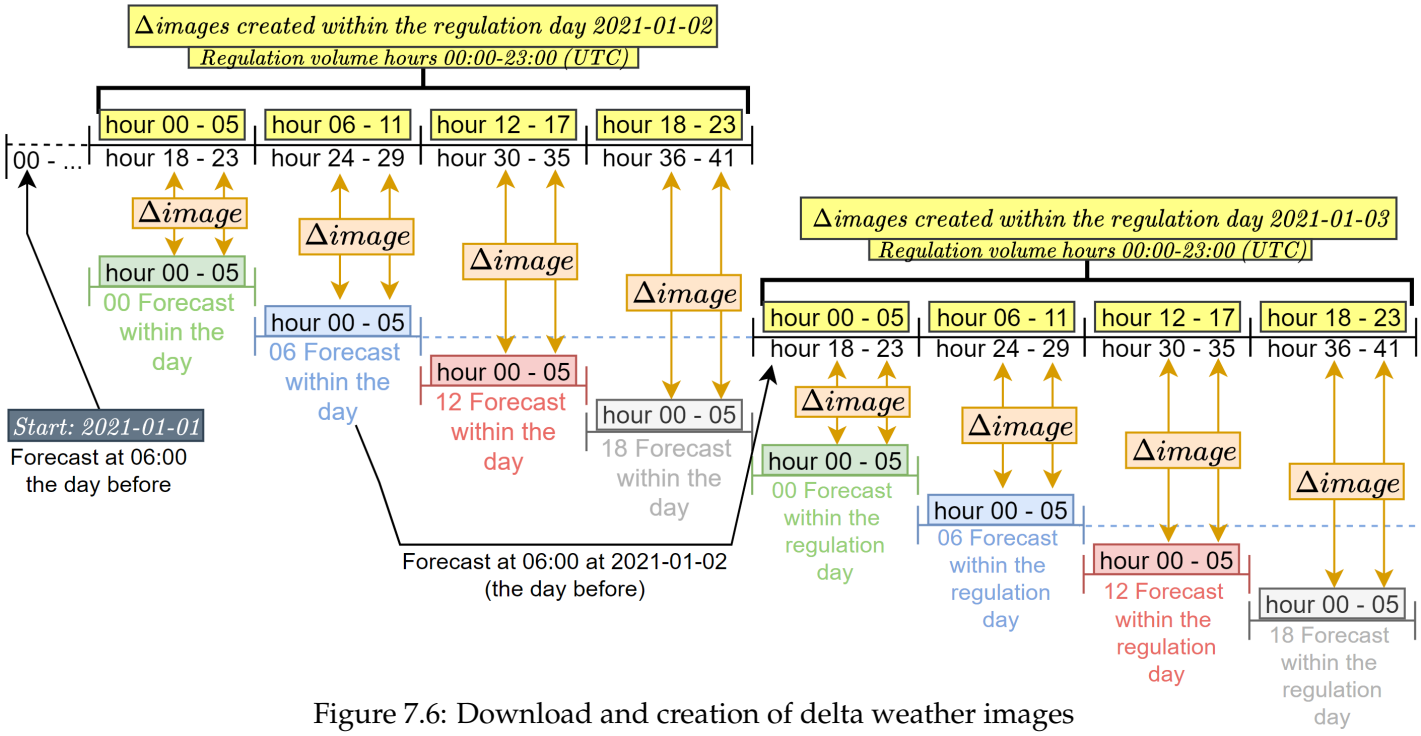


Figure 7.6: Download and creation of delta weather images

Figure 7.7 shows a resulting example of the process of calculating a wind direction delta image solely from weather forecasts. The plot is from the same hour as the regulation shown in the exploratory data analysis, where hour 01:00 had zero regulation in all zones at hour 01:00 and hour 16:00 had some regulation in table 4.11. The leftmost images in figure 7.7 is titled Actual but in this case it represents the intra-day weather forecast, where hour 01:00 is the hour 1 image from the 00:00 forecast (the green box in 7.6) and the hour 16:00 is the hour 4 image from the 12:00 forecast (the red box in 7.6). The images titled forecasted image in 7.7 is the day-ahead aligned weather forecast (the black timeline in figure 7.6). The rightmost images are the calculated Δ*wind direction*. One can see the highlighted differences between the forecasted images in the Δ*wind direction* at hour 16 to be significant comparing the hour 1 image. This proves that it is possible to estimate weather forecast uncertainties solely using old and updated weather forecasts.

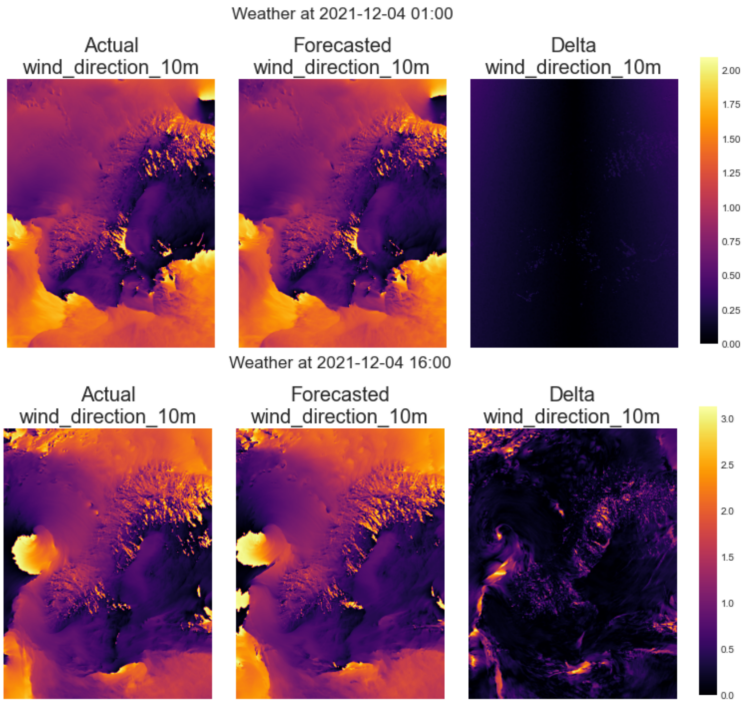


Figure 7.7: Intraday forecasted weather image ("actual weather"), day-ahead aligned forecast, and delta image of wind direction at 10 meters above surface in 2021-12-04 **01:00** and 2021-12-04 **16:00**

Bibliography

[1] Isuru Pamuditha. The connection between Deep learning & Real life learning, September 2019. URL: https://medium.com/analytics-vidhya/the-connection-between-deep-learning-real-life-learning-50de63d981ba.

[2] *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. URL: https://learning.oreilly.com/library/view/hands-on-machine-learning/9781492032632/.

[3] Post processed products · metno/NWPdocs Wiki. URL: https://github.com/metno/NWPdocs/wiki/Post-processed-products.

[4] https://energifaktanorge.no/en/norsk-energiforsyning/kraftmarkedet/ Energy facts norway. *The power market*. Publication Title: Energifakta Norge. URL: https://energifaktanorge.no/en/norsk-energiforsyning/kraftmarkedet/.

[5] nordpoolgroup nordpoolgroup. System price and Area price calculations. URL: https://www.nordpoolgroup.com/trading/Day-ahead-trading/Price-calculation/.

[6] Klaus Skytte. The regulating power market on the Nordic power exchange Nord Pool: an econometric analysis. *Energy Economics*, 21(4):295–308, August 1999. URL: https://www.sciencedirect.com/science/article/pii/S014098839900016X, doi:10.1016/S0140-9883(99)00016-X.

[7] https://prosjektbanken.forskningsradet.no/#/project/NFR/309315 Forskningsrådet. *Kollektiv for Balansetjenester*. Publication Title: Forskningsrådet, Prosjektbanken Type: https://prosjektbanken.forskningsradet.no/#/project/NFR/309315. URL: https://prosjektbanken.forskningsradet.no/#/project/NFR/309315.

[8] Tara Botnen Holm. The future importance of short term markets: An analyse of intraday prices in the Nordic intraday market; Elbas. *65*, May 2017. URL: https://nmbu.brage.unit.no/nmbu-xmlui/handle/11250/2468078.

[9] Opening and Closing hours XBID. Intra-day: Opening and Closing hours, September 2021. URL: https://www.nordpoolgroup.com/49ba3b/globalassets/download-center/xbid/nord-pool-sidc-gate-opening-times-gate-closing-times.pdf.

[10] Shai Shalev-Shwartz and Shai Ben-David. Understanding Machine Learning: From Theory to Algorithms, May 2014. ISBN: 9781107057135 9781107298019 Publisher: Cambridge University Press. URL: https://www-cambridge-org.ezproxy.uio.no/core/books/understanding-machine-learning/3059695661405D25673058E43C8BE2A6, doi:10.1017/CBO9781107298019.

[11] Jason Brownlee. Basic Concepts in Machine Learning, December 2015. URL: https://machinelearningmastery.com/basic-concepts-in-machine-learning/.

[12] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*. Springer, New York, NY, 2013.

[13] *Deep Learning For Dummies*. URL: https://learning.oreilly.com/library/view/deep-learning-for/9781119543046/.

[14] Aileen Nielsen. *Practical Time Series Analysis*. URL: https://learning.oreilly.com/library/view/practical-time-series/9781492041641/.

[15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009. ISSN: 1063-6919. doi:10.1109/CVPR.2009.5206848.

[16] Sinno Jialin Pan and Qiang Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, October 2010. Conference Name: IEEE Transactions on Knowledge and Data Engineering. doi:10.1109/TKDE.2009.191.

[17] Yuqing Gao and Khalid M. Mosalam. Deep Transfer Learning for Image-Based Structural Damage Recognition. *Computer-Aided Civil and Infrastructure Engineering*, 33(9):748–768, 2018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/mice.12363. URL: http://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12363, doi:10.1111/mice.12363.

[18] Robert H. Shumway and David S. Stoffer. *Time Series Analysis and Its Applications*. Springer International Publishing, 2017.

[19] Ben Auffarth. *Machine Learning for Time-Series with Python : Forecast, Predict, and Detect Anomalies with State-Of-the-art Machine Learning Methods*. Packt Publishing, 2021.

[20] Vishwas B V and Patel Ashish. *Hands-on Time Series Analysis with Python: From Basics to Bleeding Edge Techniques*. URL: https://learning.oreilly.com/library/view/hands-on-time-series/9781484259924/.

[21] *Machine Learning for Time Series Forecasting with Python*. URL: https://learning.oreilly.com/library/view/machine-learning-for/9781119682363/.

[22] M. Schuster and K.K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, November 1997. Conference Name: IEEE Transactions on Signal Processing. doi:10.1109/78.650093.

[23] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. The Performance of LSTM and BiLSTM in Forecasting Time Series. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 3285–3292, December 2019. doi:10.1109/BigData47090.2019.9005997.

[24] Digital Image Processing, 4th Edition. URL: https://www.pearson.com/content/one-dot-com/one-dot-com/us/en/higher-education/program.html.

[25] Xiaofeng Yang, Srini Tridandapani, Jonathan J. Beitler, David S. Yu, Emi J. Yoshida, Walter J. Curran, and Tian Liu. Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicity. *Medical Physics*, 39(9):5732–5739, September 2012. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3443195/, doi:10.1118/1.4747526.

[26] Zeenat Hameed, Seyedmostafa Hashemi, and Chresten Træholt. Applications of AI-Based Forecasts in Renewable Based Electricity Balancing Markets. In *2021 22nd IEEE International Conference on Industrial Technology (ICIT)*, volume 1, pages 579–584, March 2021. doi:10.1109/ICIT46573.2021.9453469.

[27] Gro Klæboe, Anders Lund Eriksrud, and Stein-Erik Fleten. Benchmarking time series based forecasting models for electricity balancing market prices. *Energy Systems*, 6(1):43–61, March 2015. doi:10.1007/s12667-013-0103-3.

[28] Lion Hirth and Inka Ziegenhagen. Balancing power and variable renewables: Three links. *Renewable and Sustainable Energy Reviews*, 50:1035–1051, October 2015. URL: https://www.sciencedirect.com/science/article/pii/S1364032115004530, doi:10.1016/j.rser.2015.04.180.

[29] Stefan Jaehnert, Hossein Farahmand, and Gerard L. Doorman. Modelling of prices using the volume in the Norwegian regulating power market. In *2009 IEEE Bucharest PowerTech*, pages 1–7, June 2009. doi:10.1109/PTC.2009.5281972.

[30] Hannele Holttinen. Estimating the impacts of wind power on power systems—summary of IEA Wind collaboration. *Environmental Research Letters*, 3(2):025001, April 2008. Publisher: IOP Publishing. doi:10.1088/1748-9326/3/2/025001.

[31] Ilias Dimoulkas, Mikael Amelin, and Mohammad Reza Hesamzadeh. Forecasting balancing market prices using Hidden Markov Models. In *2016 13th International Conference on the European Energy Market (EEM)*, pages 1–5, June 2016. ISSN: 2165-4093. doi:10.1109/EEM.2016.7521229.

[32] Tárik S. Salem, Karan Kathuria, Heri Ramampiaro, and Helge Langseth. Forecasting Intra-Hour Imbalances in Electric Power Systems. *arXiv:1902.00563 [cs, stat]*, February 2019. arXiv: 1902.00563. URL: http://arxiv.org/abs/1902.00563.

[33] M.P. Garcia and D.S. Kirschen. Forecasting system imbalance volumes in competitive electricity markets. *IEEE Transactions on Power Systems*, 21(1):240–248, February 2006. Conference Name: IEEE Transactions on Power Systems. doi:10.1109/TPWRS.2005.860924.

[34] Jari Miettinen and Hannele Holttinen. Characteristics of day-ahead wind power forecast errors in Nordic countries and benefits of aggregation. *Wind Energy*, 20(6):959–972, 2017. Publisher: Wiley. URL: https://cris.vtt.fi/en/publications/characteristics-of-day-ahead-wind-power-forecast-errors-in-nordic, doi:10.1002/we.2073.

[35] entsoe-py, March 2022. original-date: 2017-07-12T13:17:39Z. URL: https://github.com/EnergieID/entsoe-py.

[36] Magasinstatistikk - NVE. URL: https://www.nve.no/energi/analyser-og-statistikk/magasinstatistikk/.

[37] Nord pool market data. URL: https://www.nordpoolgroup.com/Market-data1#/nordic/table.

[38] Regulerkraftmarkedet. URL: https://www.statnett.no/for-aktorer-i-kraftbransjen/systemansvaret/kraftmarkedet/reservemarkeder/tertiarreserver/regulerkraftmarkedet/.

[39] Jan Horst Keppler, Sébastien Phan, and Yannick Le Pen. The Impacts of Variable Renewable Production and Market Coupling on the Convergence of French and German Electricity Prices. *The Energy Journal*, 37(3):343–359, 2016. Publisher: International Association for Energy Economics. URL: http://www.jstor.org/stable/44075661.

[40] Autocorrelation, December 2021. Page Version ID: 1059192357. URL: https://en.wikipedia.org/w/index.php?title=Autocorrelation&oldid=1059192357.

[41] Fatoumata Dama and Christine Sinoquet. Time Series Analysis and Modeling to Forecast: a Survey. *arXiv:2104.00164 [cs]*, September 2021. arXiv: 2104.00164. URL: http://arxiv.org/abs/2104.00164.

[42] Jason Brownlee. A Gentle Introduction to Autocorrelation and Partial Autocorrelation, February 2017. URL: https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/.

[43] Paul S.P. Cowpertwait and Andrew V. Metcalfe. Correlation. In Andrew V. Metcalfe and Paul S.P. Cowpertwait, editors, *Introductory Time Series with R*, pages 27–43. Springer, New York, NY, 2009. doi:10.1007/978-0-387-88698-5_2.

[44] G Cybenkot. Approximation by superpositions of a sigmoidal function. page 12.

[45] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, January 1989. URL: https://www.sciencedirect.com/science/article/pii/0893608089900208, doi:10.1016/0893-6080(89)90020-8.

[46] Ian Covert, Scott Lundberg, and Su-In Lee. Understanding Global Feature Contributions With Additive Importance Measures. *arXiv:2004.00668 [cs, stat]*, October 2020. arXiv: 2004.00668. URL: http://arxiv.org/abs/2004.00668.

[47] David Andreas Bordvik, Jie Hou, Farzan M. Noori, Md Zia Uddin, and Jim Torresen. Monitoring In-Home Emergency Situation and Preserve Privacy using Multi-modal Sensing and Deep Learning. In *2022 International Conference on Electronics, Information, and Communication (ICEIC)*, pages 1–6, February 2022. ISSN: 2767-7699. doi:10.1109/ICEIC54506.2022.9748829.

[48] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv:1905.11946 [cs, stat]*, September 2020. arXiv: 1905.11946. URL: http://arxiv.org/abs/1905.11946.

[49] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016. ISSN: 1063-6919. doi:10.1109/CVPR.2016.90.

[50] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *arXiv:1506.04214 [cs]*, September 2015. arXiv: 1506.04214. URL: http://arxiv.org/abs/1506.04214.

[51] Casper Kaae Sønderby, Lasse Espeholt, Jonathan Heek, Mostafa Dehghani, Avital Oliver, Tim Salimans, Shreya Agrawal, Jason Hickey, and Nal Kalchbrenner.

MetNet: A Neural Weather Model for Precipitation Forecasting. *arXiv:2003.12140 [physics, stat]*, March 2020. arXiv: 2003.12140. URL: http://arxiv.org/abs/2003.12140.

[52] Stéphane d'Ascoli, Hugo Touvron, Matthew Leavitt, Ari Morcos, Giulio Biroli, and Levent Sagun. ConViT: Improving Vision Transformers with Soft Convolutional Inductive Biases. Technical Report arXiv:2103.10697, arXiv, June 2021. arXiv:2103.10697 [cs, stat] type: article. URL: http://arxiv.org/abs/2103.10697, doi:10.48550/arXiv.2103.10697.

[53] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. Technical Report arXiv:2103.14030, arXiv, August 2021. arXiv:2103.14030 [cs] type: article. URL: http://arxiv.org/abs/2103.14030, doi:10.48550/arXiv.2103.14030.

[54] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Haibin Lin, Zhi Zhang, Yue Sun, Tong He, Jonas Mueller, R. Manmatha, Mu Li, and Alexander Smola. ResNeSt: Split-Attention Networks. Technical Report arXiv:2004.08955, arXiv, December 2020. arXiv:2004.08955 [cs] type: article. URL: http://arxiv.org/abs/2004.08955, doi:10.48550/arXiv.2004.08955.

[55] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? Technical Report arXiv:1411.1792, arXiv, November 2014. arXiv:1411.1792 [cs] type: article. URL: http://arxiv.org/abs/1411.1792, doi:10.48550/arXiv.1411.1792.

[56] Ulrich Focken, Matthias Lange, Kai Mönnich, Hans-Peter Waldl, Hans Georg Beyer, and Armin Luig. Short-term prediction of the aggregated power output of wind farms—a statistical analysis of the reduction of the prediction error by spatial smoothing effects. *Journal of Wind Engineering and Industrial Aerodynamics*, 90(3):231–246, March 2002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167610501002227, doi:10.1016/S0167-6105(01)00222-7.

[57] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081, October 2014. URL: https://www.sciencedirect.com/science/article/pii/S0169207014001083, doi:10.1016/j.ijforecast.2014.08.008.

[58] Sanjeev Kumar Aggarwal, Lalit Mohan Saini, and Ashwani Kumar. Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power & Energy Systems*, 31(1):13–22, January 2009. URL: https://www.sciencedirect.com/science/article/pii/S0142061508000884, doi:10.1016/j.ijepes.2008.09.003.

[59] Xing Deng, Haijian Shao, Chunlong Hu, Dengbiao Jiang, and Yingtao Jiang. Wind Power Forecasting Methods Based on Deep Learning: A Survey. *Computer Modeling in Engineering & Sciences*, 122(1):273–301, 2020. URL: http://www.techscience.com/CMES/v122n1/38246, doi:10.32604/cmes.2020.08768.

[60] Lu Peng, Shan Liu, Rui Liu, and Lin Wang. Effective long short-term memory with differential evolution algorithm for electricity price prediction. *Energy*, 162(C):1301–1314, 2018. URL: https://ideas.repec.org/a/eee/energy/v162y2018icp1301-1314.html.

[61] José R. Andrade, Jorge Filipe, Marisa Reis, and Ricardo J. Bessa. Probabilistic Price Forecasting for Day-Ahead and Intraday Markets: Beyond the Statistical Model. *Sustainability*, 9(11):1990, November 2017. URL: https://www.mdpi.com/2071-1050/9/11/1990, doi:10.3390/su9111990.

[62] Johannes Krokeide Kolberg. Artificial Intelligence and Nord Pool's Intraday Electricity Market Elbas: A Demonstration and Pragmatic Evaluation of Employing Deep Learning for Price Prediction. URL: https://www.academia.edu/37117717/Artificial_Intelligence_and_Nord_Pools_Intraday_Electricity_Market_Elbas_A_Demonstration_and_Pragmatic_Evaluation_of_Employing_Deep_Learning_for_Price_Prediction.

[63] Richard Scharff and Mikael Amelin. Trading behaviour on the continuous intraday market Elbas. *Energy Policy*, 88(C):544–557, 2016. URL: https://ideas.repec.org/a/eee/enepol/v88y2016icp544-557.html.

[64] Beáta Polgári, Dániel Divényi, Péter Sőrés, Ádám Sleisz, Bálint Hartmann, and István Vokony. Increasing the Flexibility of Continuous Intraday Markets in Europe. In Bálint Németh and Lambros Ekonomou, editors, *Flexitranstore*, Lecture Notes in Electrical Engineering, pages 23–34, Cham, 2020. Springer International Publishing. doi:10.1007/978-3-030-37818-9_3.

[65] Eran Raviv, Kees E. Bouwman, and Dick van Dijk. Forecasting day-ahead electricity prices: Utilizing hourly prices. *Energy Economics*, 50(C):227–239, 2015. URL: https://ideas.repec.org/a/eee/eneeco/v50y2015icp227-239.html.

[66] Kiran Karkera. Regression Models with multiple target variables, July 2020. URL: https://towardsdatascience.com/regression-models-with-multiple-target-variables-8baa75aacd.

[67] Swagger UI. URL: https://nvebiapi.nve.no/swagger/index.html.

[68] D.W. Bunn. Forecasting loads and prices in competitive power markets. *Proceedings of the IEEE*, 88(2):163–169, February 2000. Conference Name: Proceedings of the IEEE. doi:10.1109/5.823996.

[69] F.J. Nogales, J. Contreras, A.J. Conejo, and R. Espinola. Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, 17(2):342–348, May 2002. Conference Name: IEEE Transactions on Power Systems. doi:10.1109/TPWRS.2002.1007902.

[70] Markets divided into bidding areas. URL: https://www.nordpoolgroup.com/the-power-market/Bidding-areas/.

[71] Wei Li and Denis Mike Becker. Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling. *Energy*, 237:121543, December 2021. URL: https://www.sciencedirect.com/science/article/pii/S0360544221017916, doi:10.1016/j.energy.2021.121543.

[72] Jakub Nowotarski and Rafał Weron. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81:1548–1568, January 2018. URL: https://www.sciencedirect.com/science/article/pii/S1364032117308808, doi:10.1016/j.rser.2017.05.234.

[73] Machine Learning vs. Traditional Programming Paradigm | Data Science Blog. URL: https://datalya.com/blog/machine-learning/machine-learning-vs-traditional-programming-paradigm.

[74] Chao De-Yu. What is the difference between Traditional Programming and Machine Learning?, May 2021. URL: https://medium.com/mlearning-ai/what-is-the-difference-between-traditional-programming-and-machine-learning-f6128ed4f595.

[75] Alexandre Lucas, Konstantinos Pegios, Evangelos Kotsakis, and Dan Clarke. Price Forecasting for the Balancing Energy Market Using Machine-Learning Regression. *Energies*, 13(20):5420, January 2020. Number: 20 Publisher: Multidisciplinary Digital Publishing Institute. URL: https://www.mdpi.com/1996-1073/13/20/5420, doi:10.3390/en13205420.

[76] Derek W. Bunn, John N. Inekwe, and David MacGeehan. Analysis of the Fundamental Predictability of Prices in the British Balancing Market. *IEEE Transactions on Power Systems*, 36(2):1309–1316, March 2021. Conference Name: IEEE Transactions on Power Systems. doi:10.1109/TPWRS.2020.3015871.

[77] Shadi Goodarzi, H. Niles Perera, and Derek Bunn. The impact of renewable energy forecast errors on imbalance volumes and electricity spot prices. *Energy Policy*, 134:110827, November 2019. URL: https://www.sciencedirect.com/science/article/pii/S0301421519304057, doi:10.1016/j.enpol.2019.06.035.

[78] Jari Miettinen and Hannele Holttinen. Impacts of wind power forecast errors on the real-time balancing need: a Nordic case study. *IET Renewable Power Generation*, 13(2):227–233, 2019. _eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-rpg.2018.5234. URL: https://onlinelibrary.wiley.com/doi/abs/10.1049/iet-rpg.2018.5234, doi:10.1049/iet-rpg.2018.5234.

[79] Samir Avdakovic, Alma Ademovic, and Amir Nuhanovic. Correlation between air temperature and electricity demand by linear regression and wavelet coherence approach: UK, Slovakia and Bosnia and Herzegovina case study. *Archives of Electrical Engineering*, 62:521–532, December 2013. doi:10.2478/aee-2013-0042.

[80] Rolf Golombek, Sverre A. C. Kittelsen, and Ingjerd Haddeland. Climate change: impacts on electricity markets in Western Europe. *Climatic Change*, 113(2):357–370, July 2012. doi:10.1007/s10584-011-0348-6.

[81] Arslan Salim Dar and Lüder von Bremen. Short-Term Forecasting of Wake-Induced Fluctuations in Offshore Wind Farms. *Energies*, 12(14):2833, January 2019. Number: 14 Publisher: Multidisciplinary Digital Publishing Institute. URL: https://www.mdpi.com/1996-1073/12/14/2833, doi:10.3390/en12142833.

[82] Cong Feng and Jie Zhang. SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. *Solar Energy*, 204:71–78, July 2020. URL: https://www.sciencedirect.com/science/article/pii/S0038092X20303285, doi:10.1016/j.solener.2020.03.083.

[83] Simona Bigerna. Estimating temperature effects on the Italian electricity market. *Energy Policy*, 118:257–269, July 2018. URL: https://www.sciencedirect.com/science/article/pii/S0301421518302015, doi:10.1016/j.enpol.2018.03.068.

[84] Chiou-Jye Huang, Yamin Shen, Yung-Hsiang Chen, and Hsin-Chuan Chen. A novel hybrid deep neural network model for short-term electricity price forecasting. *International Journal of Energy Research*, 45(2):2511–2532, 2021. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/er.5945. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/er.5945, doi:10.1002/er.5945.

[85] Dive into Deep Learning — Dive into Deep Learning 0.16.2 documentation. URL: https://d2l.ai/.

[86] *Introduction to Time-Series with Python*. URL: https://learning.oreilly.com/library/view/machine-learning-for/9781801819626/Text/Chapter_1.xhtml.

[87] *Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR*. URL: https://learning.oreilly.com/library/view/advanced-forecasting-with/9781484271506/.

[88] Amanpreet Kaur, Lukas Nonnenmacher, Hugo T. C. Pedro, and Carlos F. M. Coimbra. Benefits of solar forecasting for energy imbalance markets. *Renewable Energy*, 86:819–830, February 2016. URL: https://www.sciencedirect.com/science/article/pii/S0960148115302901, doi:10.1016/j.renene.2015.09.011.

[89] Štěpán Kratochvíl. System Imbalance Forecast. 2016. Accepted: 2017-03-03T12:21:07Z. URL: https://dspace.cvut.cz/handle/10467/67662.

[90] Carolina Contreras. System imbalance forecasting and short-term bidding strategy to minimize imbalance costs of transacting in the spanish electricity market. 2016. Accepted: 2017-02-10T08:09:11Z. URL: https://repositorio.comillas.edu/xmlui/handle/11531/16621.

[91] Barry R. Masters, Rafael C. Gonzalez, and Richard Woods. Book Review: Digital Image Processing, Third Edition. *Journal of Biomedical Optics*, 14(2):029901, March 2009. Publisher: SPIE. URL: https://www.spiedigitallibrary.org/journals/journal-of-biomedical-optics/volume-14/issue-2/029901/Book-Review-Digital-Image-Processing-Third-Edition/10.1117/1.3115362.full, doi:10.1117/1.3115362.