Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

Інститут електроенергетики
(інститут)

Факультет інформаційних технологій
(факультет)

Кафедра Програмного забезпечення комп’ютерних систем
(повна назва)

ПОЯСНЮВАЛЬНА ЗАПИСКА
кваліфікаційної роботи ступеня

магістра
(назва освітньо-кваліфікаційного рівня)

студента *Railings Мединського Антона Геннадійовича*
(ПІБ)

академічної групи *121М-19-1*
(шифр)

спеціальності *121 Інженерія програмного забезпечення*
(код і назва спеціальності)

на тему: *Моделі, алгоритми та програмне забезпечення для прогнозування цін на електроенергію на основі статистичних методів та штучних нейронних мереж*

______________________ *Мединський А.Г.*

Керівники	**Прізвище, ініціали**	**Оцінка за шкалою**		**Підпис**
		рейтинговою	**інституційною**	
розділів кваліфікаційної роботи				
спеціальний	Доц. Сироткіна О.І.			
економічний	Доц. Касьяненко Л.В.			

Рецензент				

Нормоконтролер	Доц. Сироткіна О.І.			

Дніпро
2020

Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

ЗАТВЕРДЖЕНО:
Завідувач кафедри
Програмного забезпечення комп’ютерних систем
(повна назва)

_______________ І.М. Удовик
(підпис) (прізвище, ініціали)

« » _____________ 20 20 Року

ЗАВДАННЯ
на виконання кваліфікаційної роботи магістра

Спеціальності *121 Інженерія програмного забезпечення*
(код і назва спеціальності)

студенту *121М-19-1* Мединському Антону Геннадійовичу
(група) (прізвище та ініціали)

Тема кваліфікаційної роботи *Моделі, алгоритми та програмне забезпечення для прогнозування цін на електроенергію на основі статистичних методів та штучних нейронних мереж*

1 ПІДСТАВИ ДЛЯ ПРОВЕДЕННЯ РОБОТИ

Наказ ректора НТУ «Дніпровська політехніка» від 22.10.2020 р. № 888-с

2 МЕТА ТА ВИХІДНІ ДАНІ ДЛЯ ПРОВЕДЕННЯ РОБІТ

Об’єкт досліджень – процес оптимізації моделей для прогнозування часових рядів.

Предмет досліджень – методи та моделі прогнозування цін на електроенергію.

Мета роботи – зниження витрат підприємств на електроенергію за рахунок оптимізації графіків споживання та/або виробництва.

3 ОЧІКУВАНІ НАУКОВІ РЕЗУЛЬТАТИ

Наукова новизна результатів кваліфікаційної роботи полягає в удосконаленні методів прогнозування цін на електроенергію.

Практична цінність полягає в тому, що моделі та методи, запропоновані в дослідженні, дозволяють підприємствам зменшувати витрати або максимізувати прибуток, оптимізуючи графік споживання або виробництва електроенергії.

4 ЕТАПИ ВИКОНАННЯ РОБІТ

Найменування етапів робіт	**Строки виконання робіт (початок – кінець**
Аналіз теми та постановка задачі	01.09.2020-30.09.2020
Пошук оптимальних шляхів вирішення задачі та побудування моделі	01.10.2020-31.10.2020
Застосування прогностичних моделей	01.11.2020-10.12.2020

Завдання видав ______________ *Сироткіна О.І.*
(підпис) (прізвище, ініціали)

Завдання прийняв до виконання ______________ *Мединський А.Г.*
(підпис) (прізвище, ініціали)

Дата видачі завдання: *01.09.2020 р.*

Термін подання кваліфікаційної роботи до ЕК *15.12.2020*

РЕФЕРАТ

Пояснювальна записка: 69 стор., 35 рис., 2 таблиці, 3 додатка, 64 джерел.

Об'єкт дослідження: процес оптимізації моделей для прогнозування часових рядів.

Предмет дослідження: методи та моделі прогнозування цін на електроенергію.

Мета кваліфікаційної роботи: зниження витрат підприємств на електроенергію за рахунок оптимізації графіків споживання та/або виробництва.

Методи дослідження. Для аналізу даних використані наступні методи: кореляційний аналіз, моделі авторегресії і ковзного середнього, сезонна модель Бокса-Дженкінса, спектральний аналіз. Для побудування прогностичних моделей використані математичні та статистичні методи, нейронні мережі прямого поширення.

Наукова новизна результатів кваліфікаційної роботи полягає в удосконаленні методів прогнозування цін на електроенергію.

Практична цінність полягає в тому, що моделі та методи, запропоновані в дослідженні, дозволяють підприємствам зменшувати витрати або максимізувати прибуток, оптимізуючи графік споживання або виробництва електроенергії.

У розділі «Економіка» проведені розрахунки трудомісткості розробки програмного забезпечення, витрат на створення ПЗ і тривалості його розробки, а також проведені маркетингові дослідження ринку збуту створеного програмного продукту.

Список ключових слів: часовий ряд, прогностична модель, авторегресія, кореляція, гібридна модель, нейронна мережа прямого поширення, спектральний аналіз.

ABSTRACT

Explanatory note: 69 pages, 35 fig., 2 tables, 3 additions, 64 sources.

Object of research: optimization processes for time series forecasting models.

Subject of research: methods and models for forecasting electricity prices.

Purpose of research: reduction of electricity costs of enterprises due to optimization of consumption and / or production schedules.

Research methods. To analyze the data, following methods are used: correlational analysis, models of auto-regression and moving average, Box-Jenkins method, spectral analysis. To create forecasting models, mathematical and statistical methods, feed-forward neural networks are used.

Originality of the results of the qualification work is the improvement of methods for forecasting electricity prices.

Practical value of the results is that the models and methods proposed in the research allow enterprises to reduce costs or maximize profits by optimizing consumption or production schedule.

In the Economics section the calculations of the software development complexity, the costs of software creation and the duration of its development, as well as marketing researches of the market of the created software product were carried out.

Keywords: time series, forecasting model, autoregression, correlation, hybrid model, feed-forward neural network, spectral analysis.

CONTENTS

LIST OF ACRONYMS

ACF	Auto-correlation function
ANN	Artificial Neural Network
AIC	Akaike Information Criteria
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ISO	Independent System Operator
MA	Moving Average
MCP	Market Clearing Price
NNAR	Neural Network Autoregressive
OTC	Over-the-Counter
PSO	Particle Smarm Optimization
PACF	Partial Auto-correlation function
TS	Time Series

INTRODUCTION

Relevance of research. Electricity is a unique commodity on the market that is by nature difficult to manage. Its storage is expensive, non-productive, and requires a tremendous amount of facilities. Consequently, unlike other types of trading products, it is impossible to keep it in stock and have it being available on demand. Demand and supply vary continuously and are not always easily predictable. Such qualities of electricity as a commodity define the essence of electricity market.

The main challenge of electricity market is to ensure that generation and consumption are synchronized and that sufficient capacity is provided even during peaks in demand. Moreover, the recent introduction of smart grids and integration of renewable energy sources increases the complexity of the market and uncertainty of future supply and demand. That`s why electricity price forecasting is increasingly gaining importance for all market participants.

Electricity market has become deregulated and is made up of various competitive submarkets. Electricity price, unlike any other commodity or financial asset, is extremely volatile. These reasons make price forecasts of greater interest for customers, generators, traders, power portfolio managers, etc. Now they are forced to hedge not only volume but price risk. With an accurate forecasting model company can adjust its bidding strategy, production, consumption.

Electricity price forecasting focuses on predicting spot prices and forward prices with different time horizons for wholesale electricity markets.

Purpose of research reduction of electricity costs of enterprises due to optimization of consumption and / or production schedules.

Tasks of research:

• Analyze the basic rules of the electricity market and strategies of its participants.

• Analyze electricity prices for the past period and draw conclusions about its nature.

• Pre-process the dataset for subsequent application of forecasting models.

• Apply ARIMA and ANN models and analyze the results.

• Evaluate and analyze the possibility of creating hybrid prognostic models.

• Draw conclusions about the results obtained.

Object of research is an optimization process of time series forecasting models by combining statistical and non-statistical approaches. Both have been widely used in forecasting separately, however, in case of electricity prices, it was found that combining forecast methods by some means generally results in greater accuracy.

Subject of research. This thesis specifically focuses on developing a two-stage forecasting model, combining Auto-Regressive Integrated Moving Average (ARIMA) and feed-forward neural networks. ARIMA is a linear model that uses lagged values of a variable to predict its future values. It is often used in cases where data show evidence of non-stationarity, i.e. the way time-series changes is not time-dependent. Feed-forward neural networks are used to find a non-linear relationship in time series. Model is applied to day-ahead power market data from the European Power Exchange (EPEX) SPOT over a period of one year till August 2020.

The model presented in this thesis is integrated into a virtual power plant Neckar-Alb and is used to adjust the work timetable of connected distributed energy resources (DERs).

Research methods. To analyze the data, following methods are used: correlational analysis, models of auto-regression and moving average, Box-Jenkins method, spectral analysis. To create forecasting models, mathematical and statistical methods, feed-forward neural networks are used.

Originality of the results of the qualification work is the improvement of methods for forecasting electricity prices.

Practical value of the results is that the models and methods proposed in the research allow enterprises to reduce costs or maximize profits by optimizing consumption or production schedule.

Personal contribution of the author:

- Structure of electricity market and price formation rules were deeply investigated.

- Real prices from the European Power Exchange (EPEX) SPOT were analyzed.
- Knowledge of existing forecasting methods was researched and analyzed.
- Main challenges and tasks for improvement of forecasting processes as well as ideas for future development were described

Structure and scope of work: This paper consists of the introduction, 4 sections, and conclusion.

In the first section analysis of topic and problem statement is made. Structure, history, behavior, and main characteristics of the electricity market are described. Influence of decentralization and deregulation of markets, a transition towards a low-carbon economy, and growth of renewables were discussed. An example of German electricity market was presented. A thorough study of electricity prices was made to find out which determining factors play a vital role in setting the next days` electricity price for a given power market.

Electricity price forecasting as a tool was introduced and its main tasks and challenges were addressed. Previous researches in the area to evaluate the performance of the most efficient forecasting tools were revised.

In the second section methods, models, main approaches as well as the process of variable and model selection to forecasting are described.

In the third section, the analysis of electricity prices from the EPEX Power Exchange is made. This process includes cleaning and pre-processing the data, analysis of behavior, data decomposition. Ideas for future improvement of the models is presented.

In the fourth section, the complexity of software development and important stages are determined. The costs of software creation are calculated. Sales market analysis for developed software was reflected.

Conclusion presents obtained scientific and practical results of the research, recommendations for their scientific and practical use.

SECTION 1
ANALYSIS OF ELECTRICITY DOMAIN

1.1. Electricity market representation

Since the beginning of commercial distribution in the 1880s and until the middle of 20th century electricity was considered a “natural monopoly”. It is an industry in which multi-firm production is more costly than production by a monopoly” [1]. The electric power industry consists of four main processes: electricity generation, electric power transmission, electricity distribution, and electricity retailing. All of these activities were commonly controlled by large utilities and there were no options for smaller electric companies to be partly involved in some of these processes without owning the whole infrastructure.

In the 1990s the traditional vertically integrated electricity market rapidly started to go through a number of significant changes and reforms in many countries [2]. Deregulation and decentralization of the market lead to more opportunities for private companies. And if the functions of transmission and distribution remain monopolized, they are now separated from potentially competitive functions of generation and retail. Thus, wholesale electricity market and retail electricity market were established. The main difference is that the former market is a B-2-B (business-to-business) and the latter is B-2-C (Business-to-consumer). In this paper, wholesale electricity market is addressed.

A wholesale electricity market offers generators to sell their electricity to retailers. Latter then re-price the electricity and take it to end-users [3].

Specific characteristics of electricity as an asset imply certain limitations on power market system. As a regular commodity or financial asset, it can be bought, sold or traded. However, its nature creates following restrictions.

- Electricity cannot be easily stored. One option is to save it in another form of energy by pumping water into storages, however, storage possibilities are limited (there should be enough space for water reservoirs). Another option is to use batteries, which are incredibly expensive and environmentally-unfriendly. This means that electricity must be consumed right after its production.

- Transmission of electricity involves losses;
- Frequency and voltage deviations can lead to major consequences;

Moreover, electricity is grid-based. With a quite recent introduction of smart grids, the structure of the market becomes even more complex [4]. Comparing to a traditional system, smart grids cost-efficiently integrate actions and behaviors of all users. Its most important characteristics are:

- Bi-directional flow of electricity and data (when consumers become prosumers and produce electricity too);
- Systematic communication between producers and consumers, so the supply-demand balance can be balanced better;
- Integration of renewable energy sources;
- A large number of diverse and distributed generation and storage devices complementing the large generating power plants;
- Self-healing capabilities (isolating network failure and protecting the power infrastructure without human intervention);
- Resilience to attacks and natural disasters;

Generally, the main task of electricity market is to ensure the security of supply. Generation and consumption must be synchronized and sufficient capacity needs to be provided during peaks of demand [5].

1.2. Structure of electricity market

Electricity market is made up of various submarkets which can be distinguished based on their time horizons and properties. There are 3 common time horizons: short-term, medium-term, and long-term. Regarding the properties, there are 3 market types: Power Exchanges, OTC markets, and Organized OTC markets [6].

- Power Exchange is a platform where electricity is traded through bids. Market participants submit supply and demand bids. Then market operator collects all submitted bids and clears the market. This type of market is used for anonymous trading and allows direct anonymous contracts. Usually, it is used for short-term and medium-term horizons, but there are some power exchanges for long-term future products. It is a typical example of an auction market system.
- OTC(Over-The-Counter) market – a decentralized market that is used for bilateral trading and excludes a central exchange or broker. Here any type of electricity product with any specific conditions or constraints can be traded. Often most long-term contracts (up to a couple of years) are traded here.
- Organized OTC market – participants submit long-term supply and demand bids to a platform similar to power exchange and then they are matched on a continuous basis.

1.3. Electricity price

The main element to control any kind of electricity market is price [7]. Specific characteristics of electricity as of asset results in incomparably high volatility. Even so every financial commodity experiences a level of volatility, for electricity prices the magnitude it can be up to two orders higher.

Fig. 1.1 demonstrates hourly prices for day-ahead market taken from the European Power Exchange (EPEX) SPOT for a period of one month (December 2019). Prices fluctuate from a local minimum of almost -50 euros up to about 80 euros with a mean value of 38 euros. Weekly and daily patterns can be observed. One of the purposes is to analyze the nature of observable peaks and fluctuations, so that this data could be used as an input to the forecasting models.

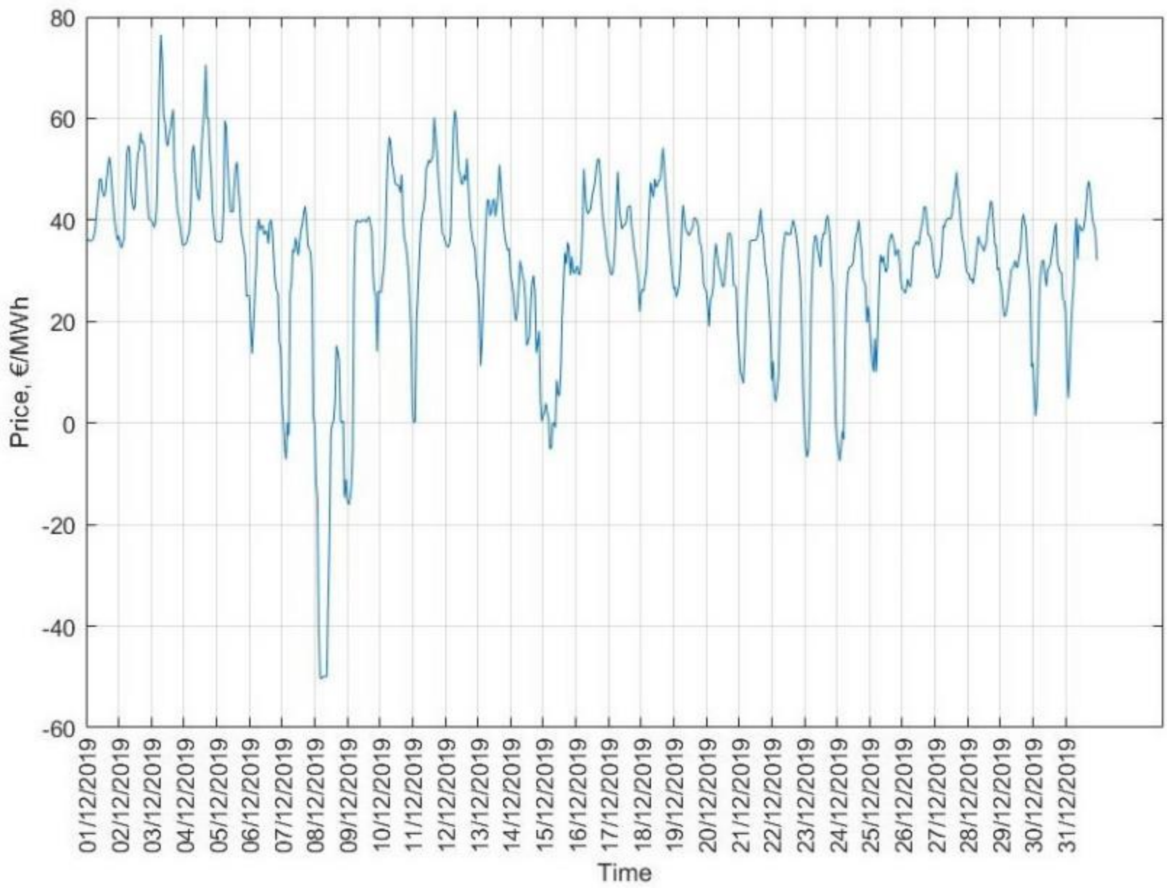

Fig. 1.1. Hourly prices for day-ahead market

Globally, electricity price depends on various factors [8]:

- Government regulations;
- Neighboring competitive markets;
- Global Markets;
- Generation Changes;
- Imports and Exports;
- Financial Speculations;
- Weather conditions;
- Technological innovations;
- Demand levels;
- Supply availability.

All of these factors play a significant role in the process of price determination. Some of them can only be predicted in the short-term, like weather, others in both short- and long-term with different precision (supply, demand). Others either cannot be simplified enough to act as an input to forecasting models or are almost unpredictable by nature.

Wholesale electricity market involves 3 different types of trading with different time horizons: day-ahead, intraday, and continuous [9]. All of them take place in power exchanges, however, they can also take place in OTC markets.

Intraday trading means that buying and selling of electricity happen on the same day when power is delivered. Power is traded in quarter-hour and one-hour intervals. Depending on the exact power exchange, a position can be traded up to 30 minutes prior to delivery. Lead time (time between purchase and delivery) differs in different markets. In Germany, it has been reduced to only five minutes. It can be explained by an increased supply of power from renewable energy sources, which have a high level of fluctuations as they are significantly dependent mostly on weather. It demonstrates the necessity of reliable and precise forecasting models.

Continuous trading happens by the same principles as intraday, but the trading takes place on an ongoing basis. A position is executed without the delay at the best price available.

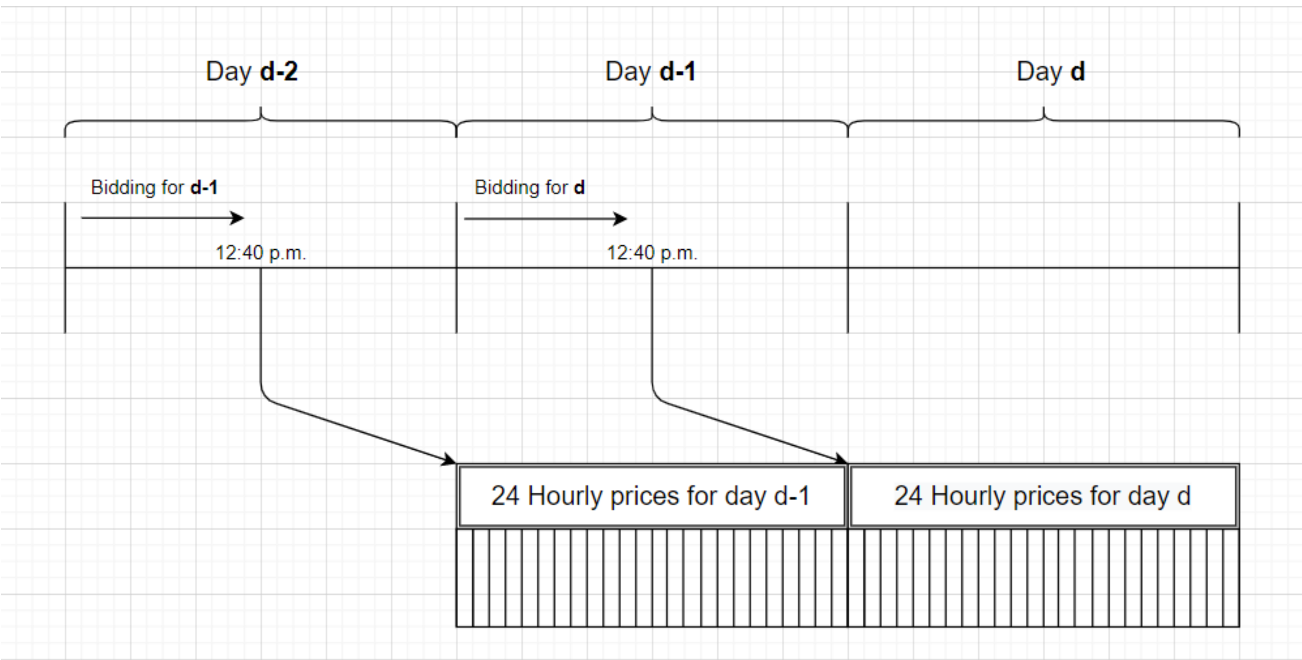

Fig. 1.2. Chronological representation of day-ahead market

Fig. 1.2 demonstrates the schematic chronological representation of day-ahead market. Day-ahead trading refers to the trading of electricity for the following day. It is also presented on power exchanges and OTC markets. Prices are finalized before a particular market closing time. In the markets of Germany bids for auctions for the next day must be submitted by 12 a.m. Resulting prices for the next day are published approximately at 12:40 p.m.

1.4. Electricity Price Forecasting

Electricity price forecasting focuses on predicting spot and forward prices in wholesale electricity markets. All market players use price forecasts as a fundamental input to decision-making mechanisms at the corporate level. As price is extremely volatile, companies have to hedge not only against volume (how much electricity is needed), but also again its price movements [10]. The costs of mistakes in buying or selling strategies and, as a consequence, buying wrong amount of electricity or at the wrong time can lead to huge financial losses or bankruptcy. The more accurate forecast is, the better generator, retailer, or consumer can adjust its bidding strategy or production/consumption schedule in order to reduce the risk or maximize profits.

In the last few decades, a variety of models were applied to the challenging task of electricity price foresting. They can be classified into six groups:

- Multi-agent models [11] – a broad variety of models that simulate the behavior of market and its participants. Market participants called agents are modeled as adaptive software agents who develop their strategies through different ways of learning (usually machine learning). Such model can include different markets, which are interrelated only through the agents` trading strategies. Agents try to find profit-maximizing bidding strategies. Such models are beneficial for companies who try to estimate their market power and market position. They can provide insights on market behavior depending on agents` strategies or how the rise of electricity price can influence the agent`s outcomes. However, these types of models are not recommended for high-precision price forecasts or any kind of quantitative conclusions. This class of models includes cost-based models, game-theoretic approaches (Nash-Cournot framework), equilibrium approaches (supply function equilibrium) [12].

- Fundamental models – class of models that captures fundamental characteristics of physical (transmission lines, generation units, power grids, location, system parameters) and economical (financial speculations, regulatory changes) parts of the market [13]. These approaches are aimed to capture behaviors that cannot be explained by past behaviors in prices. Fundamental methods try to estimate future electricity prices by simulating market clearing process. It is a process of equation of supply to demand, so there is no leftover. Such price when supply and demand are equal is also called equilibrium price. Because of increasing complexity of markets, such models are found to be very difficult to estimate. They strongly rely on the objectivity and completeness of assumptions about economical and physical relationships in the marketplace. Therefore, if these assumptions are violated, model output can be considered as unreliable.
- Reduced-Form models [14 – 16]. Reduced-form models evaluate dependent and independent variables inside market and try to identify their relationships with each other. Dependent values are received from functions of the independent values. Independent values (also called exogenous) are assumed to be determined by outside factors. Their goals are not to provide precise hourly forecasts, but rather reproduce the main characteristics of price, its correlation with other financial assets and commodities. They are often used for derivatives evaluation and risk analytics. Two most famous models are Markov-regime-switching model and jump-diffusion models.
- Statistical models – a variety of models that are often used in econometrics for forecasting future values of the variable by evaluating the relation between previous values of the same variable or values of exogenous factors

[17]. In case of electricity prices, load consumption, weather, demand, and supply-side variables are often used as exogenous variables. Such models are found to work well as long as there are no significant changes in market constraints, rules, and changes. Most of the models use a linear approach to forecasting [18, 19]. Most used models in electricity price forecasting include similar-day and exponential smoothing, regression models, time-series models with and without exogenous variables [20 – 22].

- Computational intelligence models – a wide type of models that uses a non-linear approach to forecasting [23]. They include artificial neural networks, machine learning, information fuzzy networks, support vector machines []. Such models also use previous data to forecast future prices, but unlike statistical models, they can handle and evaluate complex non-linear dependencies.
- Hybrid models – if the model combines two or more approaches it is considered to be hybrid. Often market data is too complex to be handled by only one method. Then a combination of them is used for a more accurate forecast [27 – 30].

In this paper statistical models, computational intelligence models, and hybrid models are further investigated.

1.5. Forecasting Horizons

Depending on how much further in future the price is needed to be forecasted, three main time horizons are determined:

- Short-term forecasting - usually is used for predicting spot and forward prices in intraday and day-ahead markets. Mostly statistical, computational intelligence, and hybrid models are used. Such forecasting plays an important role in daily market operations [31].

- Medium-term forecasting – such models include time horizons from a few days to a few months. They do not focus on accurate forecasts of spot prices and forward prices but are rather useful for estimating the trend and general price changes. Such forecasts are used for adjusting risk management strategies, balance sheet calculations, or derivatives pricing. In many cases, not the actual point forecasts are used, but the price distribution over certain future periods is evaluated for decision making.
- Long-term forecasting – models focus on future months and years and are used for strategic decisions on a corporate level [35]. Such information can also be used for investment profitability analysis and planning, determining the sites of new power plants, or planning on fuel sources usage [36 – 39].

Models developed in this work focus mostly on short-term electricity price forecasting, but can also be applied to medium-term forecasting.

Statistical and computational intelligence models belong to the quantitative forecasting methods. Such methods are applied when numerical data about the past is available and it is assumed that past patterns will continue. Most of the time information about the past is presented in time series datasets or in a form of cross-sectional data.

- Time series is a series of data that is arranged according to the time of its observations. It is a sequence of values of one variable taken at successfully equally spaced time intervals in the past.
- Cross-sectional data includes values of different variables that were taken at a single point in time in the past. In electricity price forecasting, it can be a market-clearing price and load at the same point in time.

1.6. Analysis of common approaches

A lot of researchers dedicated their time to investigate possible solutions for electricity price forecasting. Most important techniques include:

- Statistical time series models;

- Artificial Neural Networks;
- Wavelet transform models;
- Regime-Switching Markov models;
- Fundamental Market models;
- Equilibrium models;
- Ensemble and portfolio decision models.

In [40] authors focus on medium-term probabilistic forecasting of extremely low electricity prices in Spanish market. Paper proposes a novel methodology to simultaneously accomplish punctual and probabilistic hourly forecasts. Model combines Monte Carlo simulation with spatial interpolation techniques. Logistic regression for rare outliers, decision trees, multilayer perceptrons are used. The proposed hybrid models are compared to naïve approach. Author demonstrates empirical results based on real case study of the Spanish electricity market.

In [41] Cheng proposes mid-term electricity clearing price forecasting in a Yunnan electricity market. Author proposes a novel grey prediction model, where the lower and upper bounds are firstly identified to give an interval estimation of the forecasting value. A novel whitenization method is then proposed to determine definite forecasting value from the forecasting interval. Model parameters are identified by an improved particle swarm optimization (PSO). The accuracy of proposed model is compared with multiple linear regression and artificial neural network.

In [42] authors present a hybrid multi-step model for day-ahead electricity price forecasting based on optimization, fuzzy logic, and model selection. Model consists of two stages: particle swarm-optimization with core mapping, self-organizing map and fuzzy set followed by selection rule. Proposed model shows good results in reducing the high volatility of the electricity price.

In [43] author proposes a novel technique to forecast day-ahead electricity prices based on the wavelet transform and ARIMA models. Complex historical price series is decomposed using the wavelet transform in a set of better-behaved constitutive series. Then future values are forecasted using properly fitted ARIMA models. In the end, separate forecasts are reconstructed into a single one. Results from the electricity market of Spain are reported.

In work [44] authors present a well-known ARIMA model to analyze and forecast day-ahead spot price. The model is applied to time series consisting of prices from EPEX power exchange.

1.7. Conclusions and problem statement

This thesis addresses the problem of selecting the effective and sufficient methods for day-ahead electricity market clearing price forecasting. It analyzes complex behavior and extreme volatility of electricity as a financial commodity. Influence of external factors and complexity of power market is taken into account. Historical data from EPEX power exchange is investigated. A combination of statistical and computational methods as an alternative to usage of single approaches is proposed.

SECTION 2
METHODS AND MODELS USED IN FORECASTING

2.1. Statistical methods

Statistical methods forecast future values by mathematical combination either of past values of the same variable and/or values of other (exogenous) variables. All statistical methods can be divided into two main categories: additive and multiplicative. Additive models represent forecasted price as a sum of components, while multiplicative represent it as a product of a number of factors. Below there is a representation of most used statistical approaches in electricity price forecasting.

Average approach is a simple method that suggests that future value is equal to the mean of previous values (eq 2.1).

$$\bar{y} = (y_1 + \cdots + y_T)/T, \tag{2.1}$$

where $\bar{y}$ is forecasted price, $(y_1 + \cdots + y_T)$ are all past values, T is period.

This approach is rarely used separately. More complex version of this approach is called Moving Average (MA). It has been widely used in time series decomposition and forecasting and is a base part of more complex methods.

A moving average of order m can be written as (eq 2.2):

$$\widehat{T_t} = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j}\,, \tag{2.2}$$

where $m = 2k + 1$. The estimation of trend at time t is obtained by averaging historical values within k periods of t. As the process eliminates some randomness in the data, it is also called moving average smoothing.

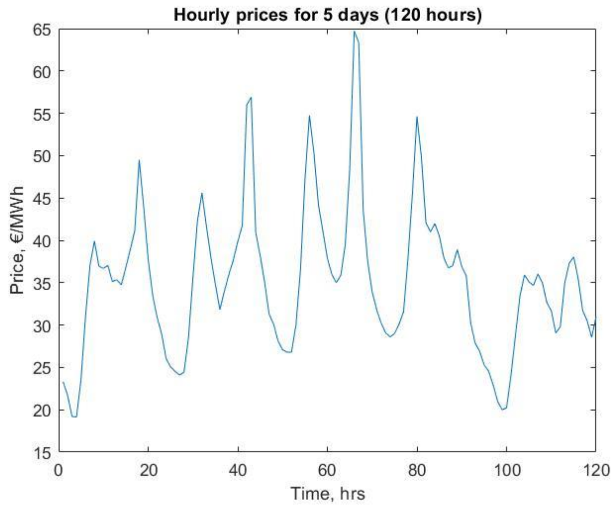

Fig. 2.1. Hourly prices for 5 days

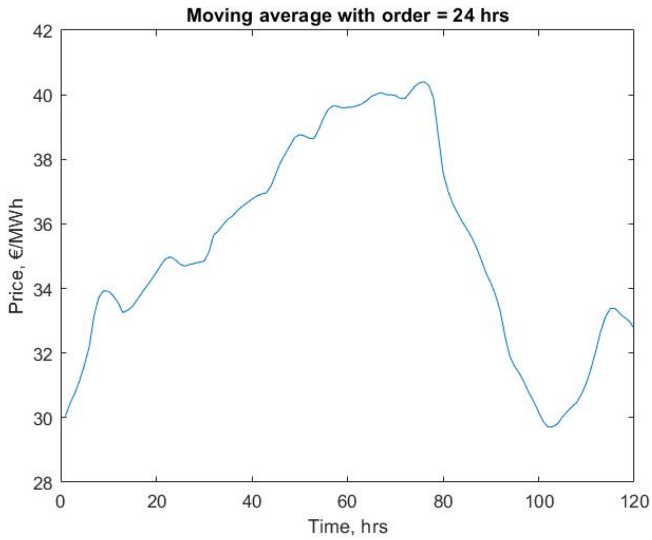

Fig. 2.2. Moving average

Fig. 2.1 demonstrates hourly prices for a period of 5 days, which is equal to 120 hours. Fig. 2.2 shows the moving average of order 24 of the same dataset. Such a technique helps in determining the trend.

It is also possible to apply moving average to moving average. Combination of them results in weighted moving average. In general, it can be represented as follows (eq 2.3):

$$\hat{T}_t = \sum_{j=-k}^{k} a_j y_{t+j} \,, \qquad (2.3)$$

where $k = (m-1)/2$ and the weights are given by $[a_{-k}, \dots, a_k]$.

Main advantage of such approach is smoothing the trend. Weights of observations slowly increase and decrease, resulting in a smoother trend.

Linear regression approach is another basic approach that is present in more sophisticated models. Simply it means that the forecasted value of y has a linear relationship with its past values or with other time series. Forecasted variable is called the regressand or explained. Variable that are used for prediction are called the regressors or explanatory. Relationship between the forecasted variable y and explanatory variable x can be written as (eq 2.4):

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \tag{2.4}$$

where $\beta 0 + \beta 1 xt$ are explained part of the model and εt is the random error.

Linear regression is found to be insufficient in forecasting electricity prices because it doesn`t address the underlying complexity behind historical data.

Exponential smoothing. Forecasted values obtained using this method are weighted averages of past observations with weights decaying exponentially as the observations get older. The further in the past observation is, the smaller is weight that it is associated with (eq 2.5).

$$y_{T+1|T} = \alpha y_T + \alpha\widehat{(1-\alpha) y_{T-1}} + \alpha(1-\alpha)^2 y_{T-2} + \cdots, \tag{2.5}$$

where $0 \leq \alpha \leq 1$ is a smoothing parameter and. $y_1, \dots, y_T$ are averages of time series.

ARIMA approach is one of the most reliable approaches in time series forecasting. To forecast future values, they aim to describe autocorrelations in historical data.

ARIMA stands for Autoregressive Integrated Moving Average. It is a combination of two different models: Autoregressive model and Moving Average model. Autoregressive model uses values from previous time steps to predict values at next time steps. Moving average, for its part, takes arithmetic mean of a set of previous values over the specified number of time steps in the past [45].

For an ARIMA model to be applied, time series must be stationary. A stationary time series is one whose properties such as mean, variance, autocorrelation do not depend on the time at which the series is observed. Thus, if time series has a trend or seasonality, it is not stationary. In simple words, a stationary time series will have no predictable patterns in long term. To get rid of a trend and/or seasonality, differencing can be applied. After that, it is assumed that statistical properties will be the same in the future, as they have been in the past. Differencing means subtracting the previous value from the current value. Practically differencing helps in stabilizing the mean of time series.

An ARIMA model is characterized by 3 terms: p, d, q, where p is the Auto-Regression order, q is Moving-Average order and d is the number of differences required to make the time series stationary. Order in these terms refers to the number of lagged values that should go into the model. ARIMA model has the following mathematical representation (eq 2.6):

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \qquad (2.6)$$

where

y'_t - is the differenced time series;

εt – is a white noise;

ϕ – is AR coefficient for lagged value;

θ – is MA coefficient for lagged value;

To build an accurate ARIMA model, p, d, q parameters have to be determined. As ARIMA model uses only values of a time series, it is necessary to determine the correlation between them. Two functions are used to estimate the correlation: Auto - Correlation Function (ACF) and Partial Auto-Correlation Function (PACF). Autocorrelation function (ACF) is a (complete) correlation function that shows autocorrelation of any value in time series with its lagged values. Lag is a time gap between these values. Mostly, ACF describes how well the present value of the series is related to its previous values and is used to detect non-randomness in data.

Partial Auto-Correlation Function (PACF) is the amount of correlation between a variable and a lag of itself, that is not explained by their mutual correlations with other variables of the same series. For example, let`s assume there is a time series Y. Then, autocorrelation at lag 1 is a coefficient of correlation between Yt and Y_{t-1}, which is presumably a correlation between Y_{t-1} and Y_{t-2}. PACF is needed to find direct correlation between Y_t and Y_{t-2}. The correlation at lag 1 "propagates" to lag 2 and presumably to higher-order lags. The partial autocorrelation at lag 2 is therefore the difference between the actual correlation at lag 2 and the expected correlation due to the propagation of correlation at lag 1.

Order of differencing is the minimal differencing required to get a stationary time series. Fig. 2.3 demonstrates main steps in creating Arima model.

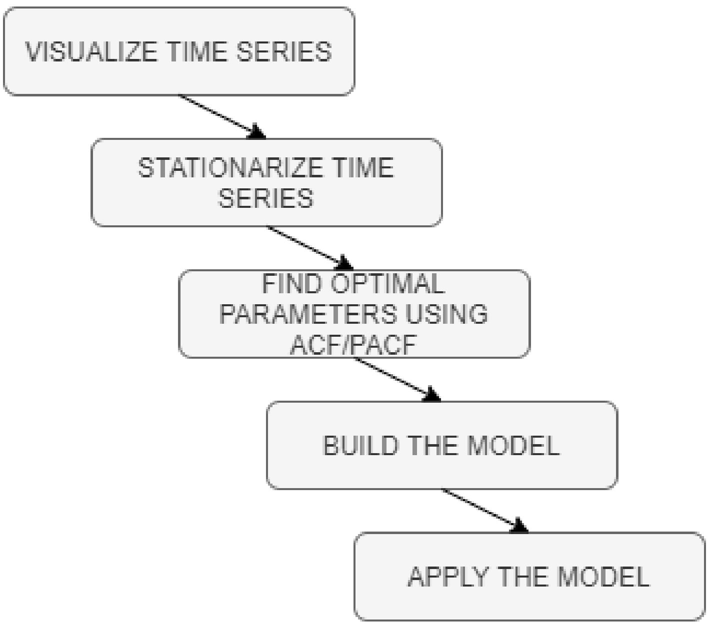

Fig.2.3. ARIMA flowchart

Variations of ARIMA models: ARMA, SARIMA [46], ARIMAX [47].

SARIMA states for Seasonal ARIMA. To the original ARIMA model, seasonal parameters are added. It incorporates both factors in a multiplicative model (eq 2.7):

$$\text{ARIMA}(p, d, q) \times (P, D, Q)s, \tag{2.7}$$

where (p,d,q) are non-seasonal parameters and (P,D,Q)s are seasonal parameters.

Seasonality in a time series is a regular pattern that repeats with a certain periodicity. ARIMAX states for Autoregressive Integrrated Moving Average with Explanatory Variable, where there are one or more autoregressive (AR) terms and/or one or more moving average (MA) terms. It can be roughly seen as a multiple regression model. While ARIMA is used for datasets that are univariate (consists on observations on only one characteristic), ARIMAX is applied to multivariate datasets [48]. Multivariate means there are additional explanatory variables. Its mathematical representation (eq 2.8):

$$\Delta^D y_t = \sum_{i=1}^{p} \phi_i \, \Delta^D y_{t-i} + \sum_{j=1}^{q} \theta_j \, \epsilon_{t-j} + \sum_{m=1}^{M} \beta_m \, Xm, t + \epsilon_t, \tag{2.8}$$

where

y_t – time series data

Xt – exogenous data

2.2. Feed-forward neural networks

Artificial neural networks allow complex non-linear relationship between the response variable and its predictors. In case of time series forecasting, they can be applied both to univariate and multivariate time series. In this paper univariate time series is addressed.

Neural network that can handle univariate time series is called feed-forward neural network (FNN) [49].

The simplest feed-forward neural network contains no hidden layers and is equivalent to linear regression. Forecasts are obtained using linear combination of inputs with weights attached to them. Weights are selected so that cost function such as mean square error (MSE) is minimized. Fig. 2.4 demonstrates the simplest neural network.

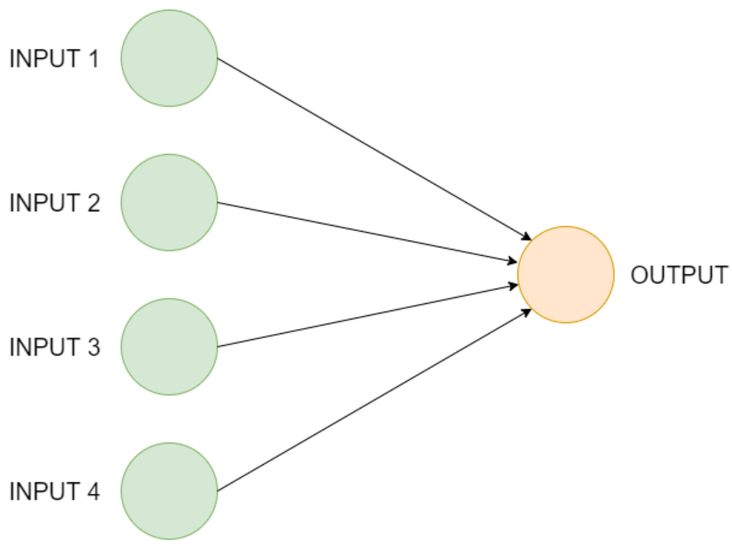

Fig. 2.4. Simplest Neural network

Once a hidden layer with hidden neurons is added, neural networks become non-linear. Fig. 2.5 shows the example of multilayer feed-forward network, where each layer receives inputs from the previous layers. Inputs to each node are combined linearly (eq 2.9):

$$z_j = b_j + \sum_{i=1}^{4} w_{i,j} x_i \, , \tag{2.9}$$

where b_1, b_2, b_3 and $w_{1,1}, \ldots, w_{4,3}$ are "learned" from the data.

In hidden layer, the weighted linear combination of inputs is modified using a non-linear function, such as sigmoid (eq 2.10) to give the input for the next layer. This process reduces the effect of extreme input values [51].

$$s(z) = \frac{1}{1 + e^{-z}} \tag{2.10}$$

Usually, several parameters are used as inputs to the neural network, but in time series forecasting it is possible to take previous values of one parameter and use them as inputs. It can be called a neural network autoregression [52].

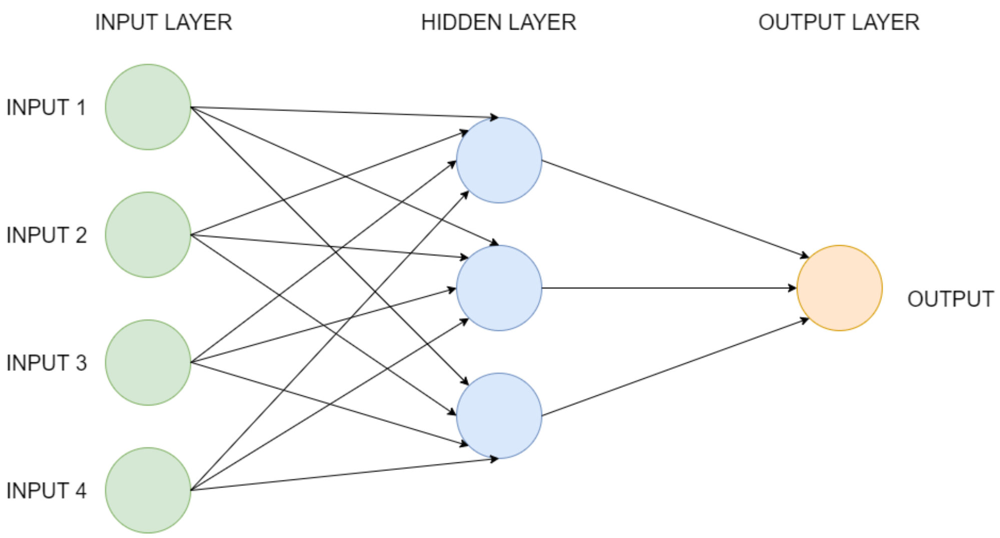

Fig. 2.5. Feed-forward neural network

2.3. Hybrid forecasting models

Usually, time-series data is too complicated to be explained by one or even two models. Thus, hybrid models provide a solution by combining the best of different forecasting techniques. Often statistical and machine learning methods are combined. However, multi-agent or fundamental models can be also taken into consideration when designing a hybrid model. The fundamental idea is that such a combination compensates for the limitations of one approach with the strengths of the other [53]. For instance, statistical methods mostly assume linear relationships in the data and they outperform neural networks in this task. But it is not necessarily the case in real-world data, where datasets contain both linear and non-linear components. On the other hand, machine learning techniques can exploit cross-series information. As they do not have an assumption of linearity, they have an exceptional capability of approximating almost every function [54].

There are 2 main ways to obtain a hybrid forecast (Fig. 2.6).

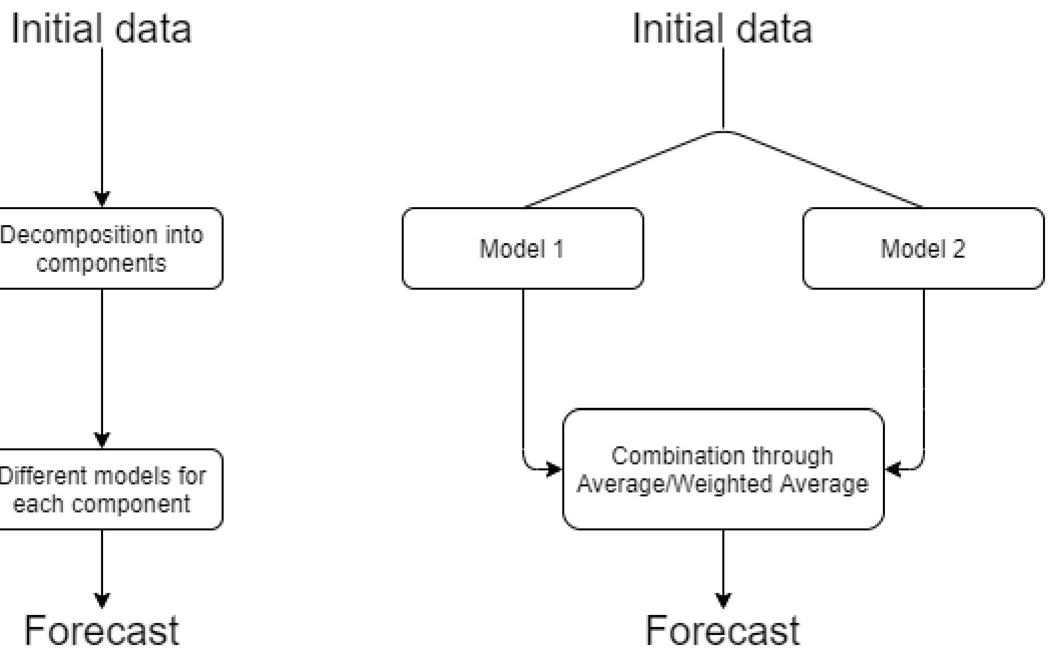

Fig. 2.6. Classification of hybrid approaches

In the first case, models are applied simultaneously with an aim to catch different patterns, model linear or non-linear components etc. It can be understood as a sophisticated combination of different approaches that in one way or another interact with each other. Main challenge of this approach is to find an appropriate way for data decomposition. In electricity price forecasting domain, there is a presupposition that historical data consists of 2 components: linear and non-linear (eq 2.11) [55]

$$y_t = L_t + N_t, \tag{2.11}$$

where $\mathrm{L_t}$ *and* $\mathrm{N_t}$ are linear and non-linear components respectively. Firstly, statistical model such as ARIMA is applied to give a linear forecast. Then residuals from the fitted model are obtained and used as inputs to the non-linear model. If e_t denotes residuals from ARIMA, then (eq 2.12)

$$e_t = y_t - \mathrm{L}_t, \tag{2.12}$$

where $\hat{\mathrm{L}}_t$ is a forecasted value from ARIMA. Then non-linear component $\mathrm{N_t}$ is modelled by neural networks. In the end, results are combined into single forecast. Predicted values of time series can be defined as (eq 2.13):

$$\hat{\mathrm{y}}_\mathrm{t} = \widehat{\mathrm{N}}_\mathrm{t} + \hat{\mathrm{L}}_t. \tag{2.13}$$

First option to split data into subcomponents is to fit ARIMA model first and then obtain residuals from it. Fig. 2.7 demonstrates an example of fitting ARIMA model to electricity prices and obtaining residuals as a difference between forecasted and real value. In a properly fitted ARIMA model, residuals are white noise, which means that they contain no autocorrelation. However, it only means that there is no linear correlation between values. Thus, non-linear methods can be considered to be effective in residuals analysis.

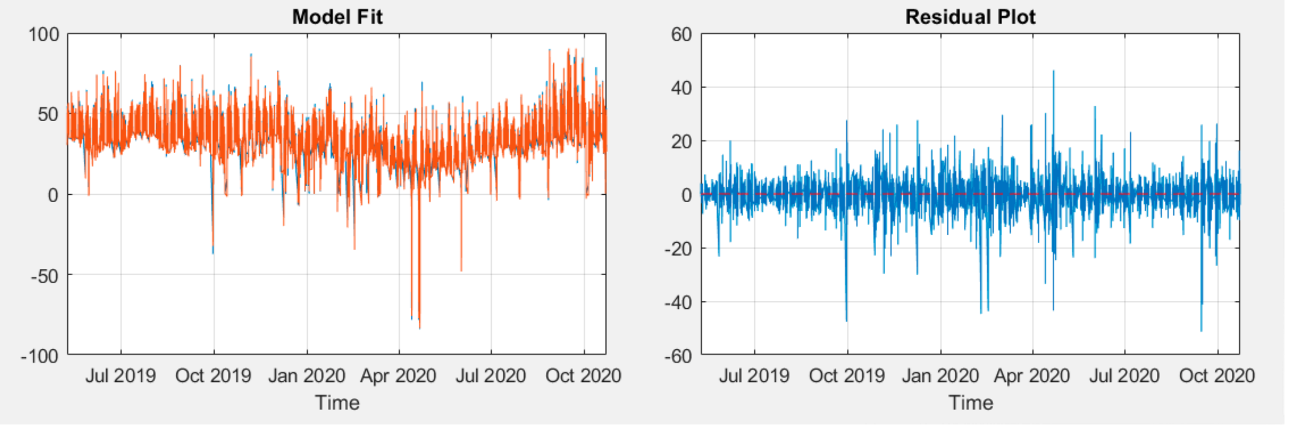

Fig. 2.7. ARIMA and its residuals

As it was said earlier, traditional time series datasets are thought to be an aggregate or combination of the following components: level, trend, seasonality, noise. All series have a level and noise, but trend and seasonality are optional.

In additive model, the components are added together as follows (eq 2.14):

$$y(t) = Level + Trend + Seasonality + Noise. \tag{2.14}$$

An additive model is linear, where changes over time are consistent and are made by the same amount. Trend is a straight line and seasonality has constant frequency and amplitude.

In multiplicative model components are multiplied instead of being added (eq 2.15):

$$y(t) = Level * Trend * Seasonality * Noise. \tag{2.15}$$

A multiplicative model is non-linear, such as quadratic or exponential. Changes increase or decrease over time, Trend is not a curved line, seasonality has changing frequency and amplitude over time.

Datasets can also be transformed through different transformation techniques. Most promising of them are Fourier Transform and Wavelet Transform.

Fourier analysis represents any function by sums of simpler trigonometric functions, usually sines. It is used to map signals from the time domain to the frequency domain. After, Inverse Fourier Transform is used to remap the signals from the frequency domain back to the time domain [57]. The Fourier transform of a time series yt for frequency p cycles per n observations can be written as (eq 2.16):

$$z_p = \sum_{t=0}^{n-1} y_t \exp(-2\pi i\, p\, t/n)\,. \tag{2.16}$$

If applied properly, Fourier Transform can give a powerful insight into the data.

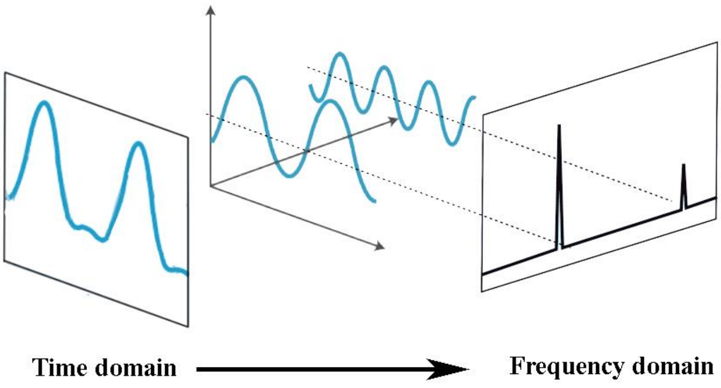

Fig. 2.8. Fourier transform

Fig. 2.8 shows the decomposition of function to 2 sine functions.

However, it cannot cover both time and frequency simultaneously. It is a theoretical limit that is known as the uncertainty principle. The smaller is the size of the window, the more information we get about a location of frequency, but less about the frequency value. In time series analysis it is often important to know not only the frequency of an event, such as price peaks but also the time when it happened. Thus, Fourier transformation implies limitations and has to be used carefully.

As an alternative to Fourier Transform Wavelet Transform is used. It has been used as a preprocessing method since the 2000s [57]. Wavelet transform decomposes time series data into approximation and detail components, and different forecasting methods are applied to each component. Compared to the Fourier Transform, Wavelet has a high resolution in both frequency and time domain. It provides information not only about which frequencies are presented in a signal, but also at which time these frequencies occurred. It is achieved by scaling. Firstly, large features are analyzed, then smaller features are analyzed after shrinking the scale. As Fourier Transform uses a series of sine waves with different frequencies, Wavelet Transform uses functions called wavelets. Sine functions are infinite, while wavelets are localized in time (Fig. 2.9). This localization allows obtaining time information in addition to frequency information.

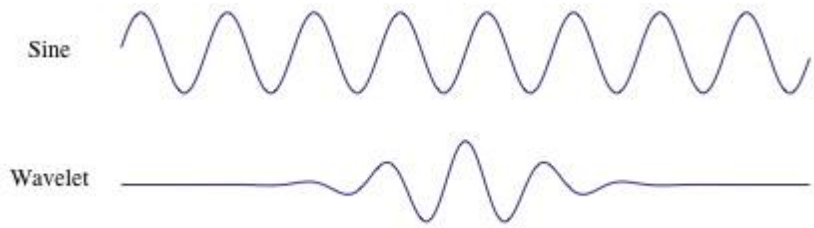

Fig. 2.9. wavelet function

Fig. 2.10 demonstrates the decomposition of differentiated hourly electricity prices. *S* is the original signal, that is decomposed on 7 levels: one approximation level and 6 detailed levels. All levels can be then analyzed separately and with different approaches. As signal is a linear combination of its components, original signal can be obtained through addition [58].

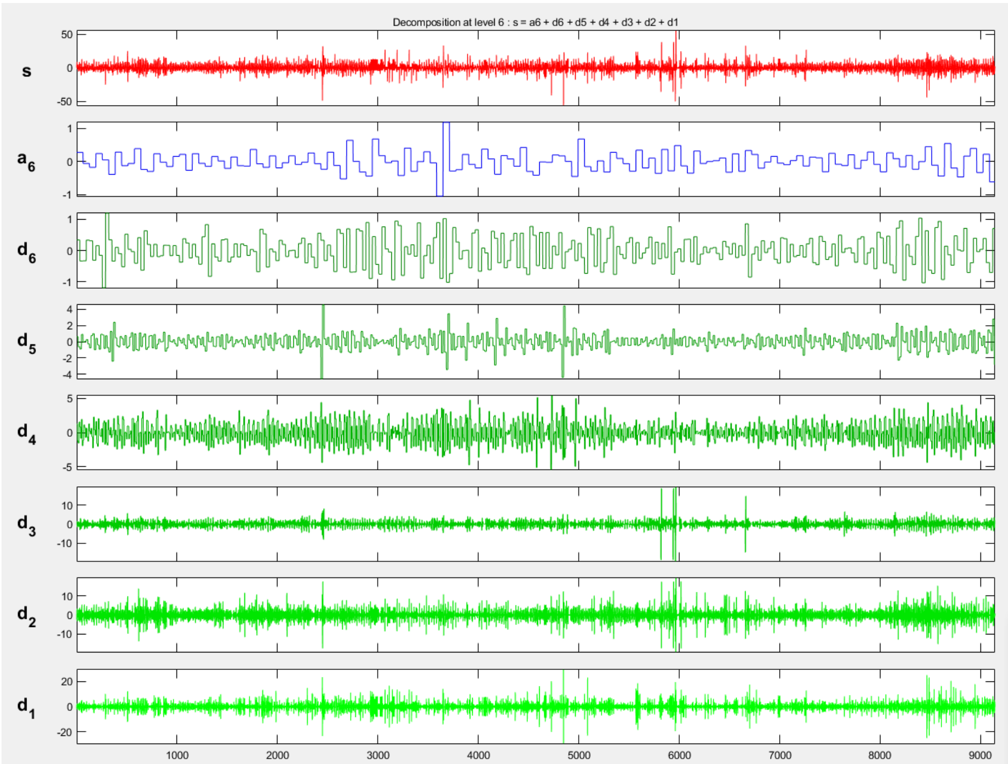

Fig.2.10. Decomposition of differentiated prices through wavelets

Another decomposition approach is splitting the time series into different datasets based on their values. In simple case, electricity prices can be split into three groups: high prices, medium prices, and low-prices. Such process belongs to preprocessing. Core idea of this approach is that high and low prices have different patterns and together create the level of volatility that one model is not able to cover[59]. For this purpose, fuzzy logic can be established:

- **IF** price(i) **IS** High price, **THEN** price(i) equals price(i) * Highweight;
- **IF** price(i) **IS** Medium price, **THEN** price(i) equals price(i) *Mediumweight;
- **IF** price(i) **IS** Low price, **THEN** price(i) equals price(i) * Lowweight.

This method reduces the volatility of price [60]. As an alternative dataset can be divided into 3 different datasets, each containing high, medium, and low prices respectively. Then, they can be analyzed separately and different forecasting techniques can be applied.

2.4. Conclusions

In this section analysis of electricity price formation was made. As predictors, historical values of electricity prices are chosen. Such data gives a deep insight into rules under which electricity price was formed and allows to make generalizations about them. The possibility to add exogenous factors to forecasting models is also described, however, this work does not focus on it.

Main approaches to MCP forecasting are presented. Both ARIMA and ANN are considered to be promising models, however, a deep understanding of underlying data is required to fully utilize their strengths. Volatile nature of electricity makes it difficult to properly fit the model.

There are several ways for models to be combined into a hybrid one. Thus, data-preprocessing techniques, such as Wavelet or Fourier decomposition, as well as fuzzy logic are presented. It allows utilizing strengths of linear and non-linear models while neglecting their weaknesses. Similarities and differences, as well as advantages and disadvantages of decomposition techniques, are reflected.

In the next section, pre-processing, analysis of historical prices, and application of ARIMA and ANN to day-ahead electricity market in Germany is described.

SECTION 3
APPLICATION OF ARIMA, ANN TO DAY-AHEAD MARKET

In this section ARIMA Model, ANN model, and hybrid model are described and compared with each other. Models are trained on dataset containing hourly electricity prices from European Power Exchange (EPEX) Spot.

3.1. Preprocessing the data

Original dataset contains hourly prices for the period from 09 May 2019 to 22 October 2020 that corresponds to 533 days or 12792 hours.

Before analysis of the prices, dataset was preprocessed. Missing entries were filled using linear interpolation (Fig. 3.1).

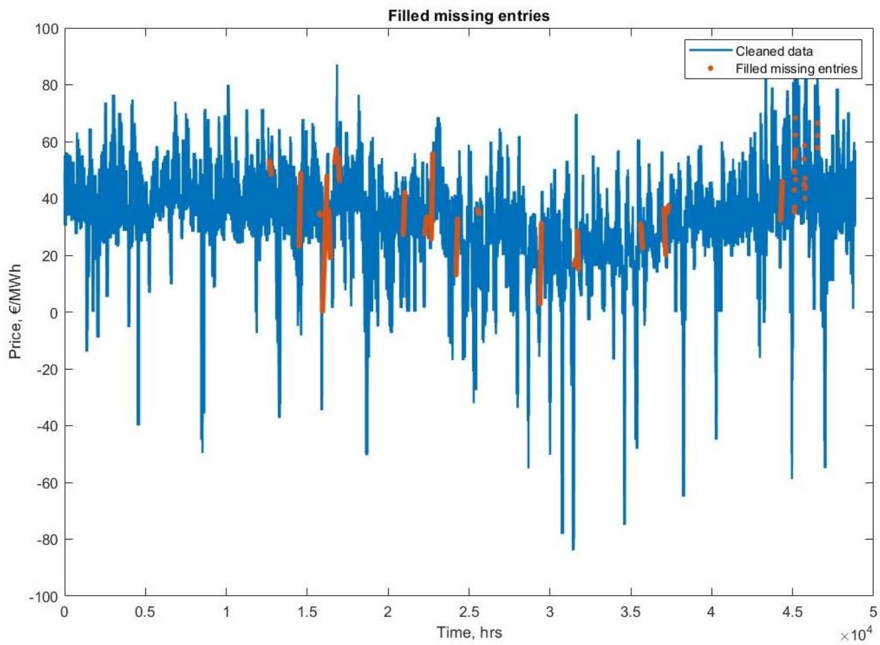

Fig.3.1. Cleaned dataset

Fig. 3.2 demonstrates average price for every weekday. Prices for Monday – Friday stay almost on the same level, while on Saturday and Sunday they are noticeably lower. It was decided to exclude weekends out from the dataset and focus only on working days (Fig. 15).

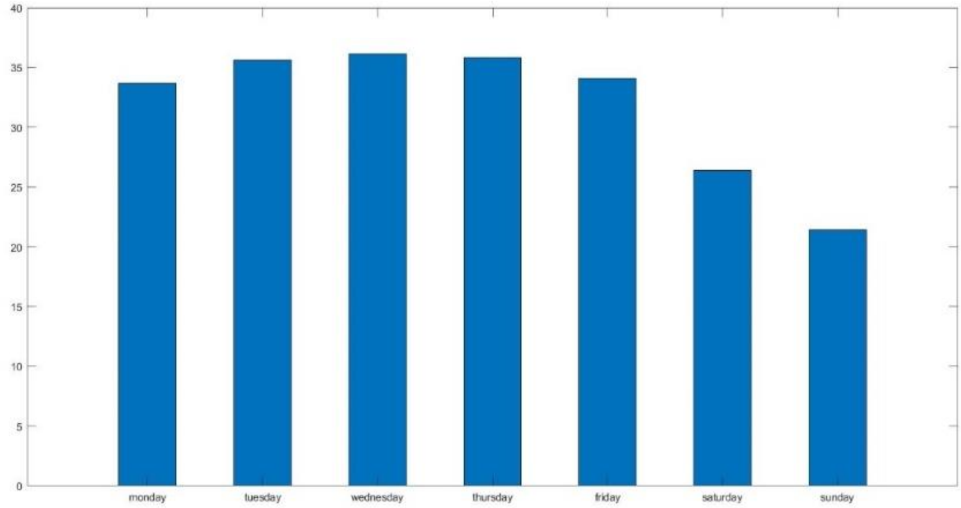

Fig. 3.2. Average prices for every weekday

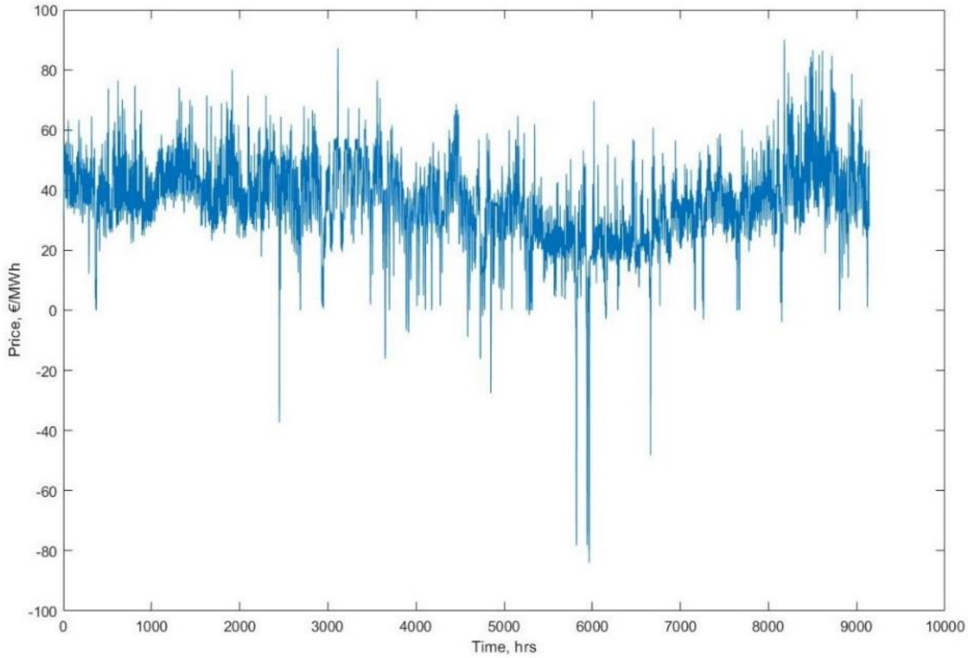

Fig. 3.3. Prices on workdays

Time series dataset consists of systematic and non-systematic components. Systematic are those that have consistency or recurrence and can be described and modeled. Non-systematic components cannot be directly modeled and are called noise.

A given dataset is thought to consist of three systematic components: level, trend, and seasonality. Level is the average value in the series, the trend is the increasing/decreasing value, seasonality is the repeating short-term cycle in the time series and noise is the random variation in the series. Fig. 3.4 demonstrates the overall trend by applying a moving average with a window of twelve months. There is a clear increasing trend since April, 21, 2020.

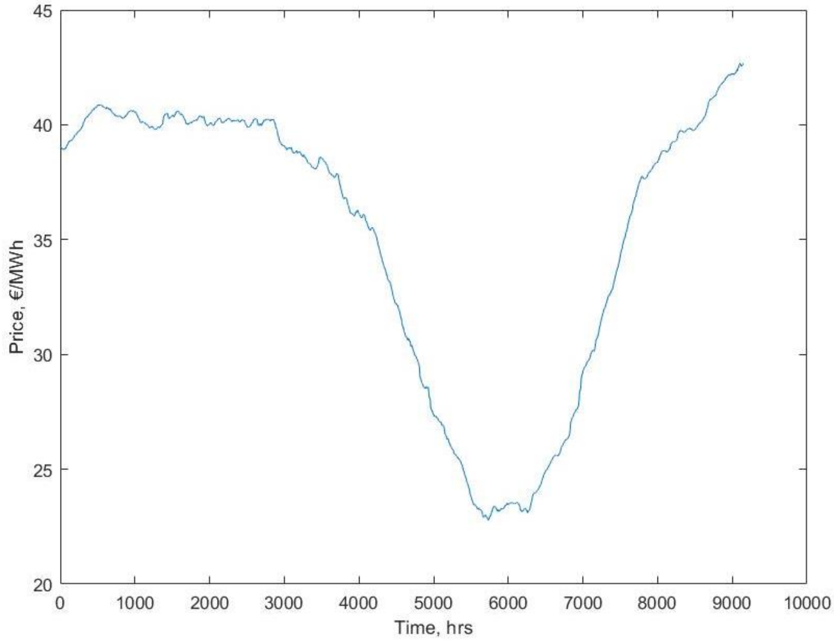

Fig. 3.4. Moving average with window of 12 months

Fig. 3.5 shows averaged daily fluctuations of price. There is a pattern for electricity prices throughout the day. Starting at midnight, price decreases until 02 am, with the local minimum at this point. Then price starts increasing and reaches its first peak at 05:30 am. It is followed by a significant decrease and gets to a local minimum at noon. After that, the pattern repeats, rising rapidly until 05:30 pm, where it reaches day`s maximum value. Since then, there is a massive decrease until midnight, where price gets close to the price at the beginning of a day.

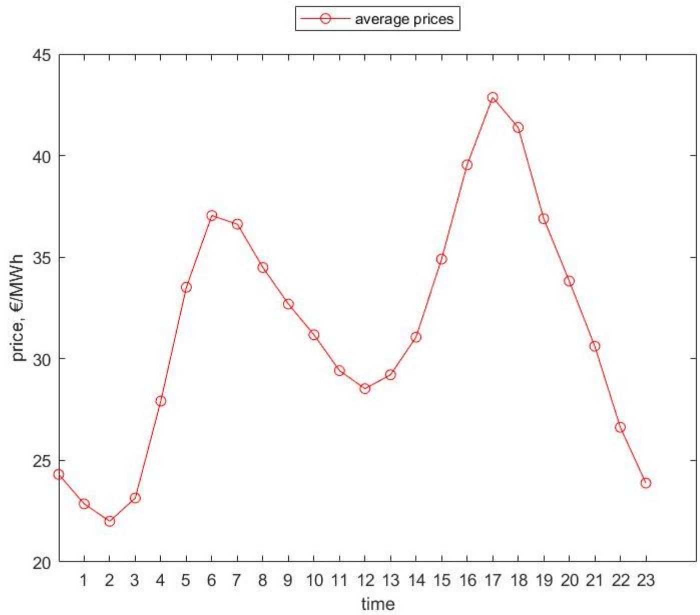

Fig. 3.5. Averaged daily fluctuations

3.2. Estimation of ARIMA model

To estimate an ARIMA model, autoregressive order, differencing order, and moving average order need to be determined.

Looking at Fig. 3.3, it is clear that data, as many macroeconomic time series, is not stationary. There is a downward trend followed by an upward trend and the mean is not constant over the time.

To make data stationary, differencing is applied (Fig. 3.6).

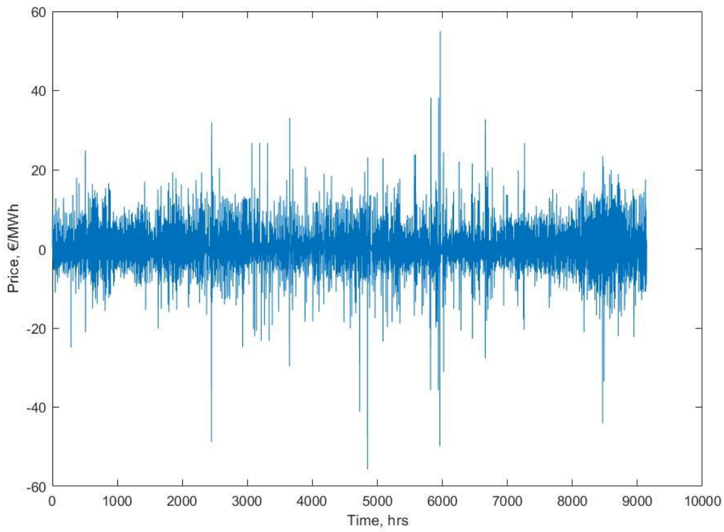

Fig. 3.6. Differenced data

It can be concluded that variance is now constant and differenced prices are distributed around 0. However, besides visual estimation, there are 2 common tests that can be used to test if time series is stationary.

Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test - used for testing a null hypothesis that an observable time series is stationary around a deterministic trend against the alternative of a unit root [62]. Null Hypothesis that the differenced time series is trend stationary is failed to reject (Fig. 3.7), which means that our dataset is trend stationary.

KPSS Test for Stationarity(PriceDiff)

Null Hypothesis: PriceDiff is trend stationary

Results

	Select	Null Rejected	P-Value	Test Statistic	Critical Value	Lags	Include Trend	Significance Level
1	☐	false	0.1000	8.2706e-04	0.1460	0	true	0.0500

Fig. 3.7. KPSS test

Augmented-Dickey-Fuller Test tests the null hypothesis that a unit root is present in a time series. The alternative hypothesis is that time series is stationary. In other words, time series is not stationary, if it contains a unit root. As null hypothesis is rejected, it can be stated that data contains no unit roots and hence is stationary [61] (Fig. 3.8).

Augmented Dickey-Fuller Test(PriceDiff)

Null Hypothesis: PriceDiff contains a unit root

Results

	Select	Null Rejected	P-Value	Test Statistic	Critical Value	Lags	Model	Test Statistic	Significance Level
1	☐	true	1.0000e-03	-64.0518	-1.9416	0	AR	t1	0.0500

Fig.3.8. Augmented-Dickey_Fuller test

First-order differencing is enough to make data stationary. Next step is to determine autoregressive and moving average orders. Figures 3.9 and 3.10 show Autocorrelation and Partial Autocorrelation functions for a given time series.

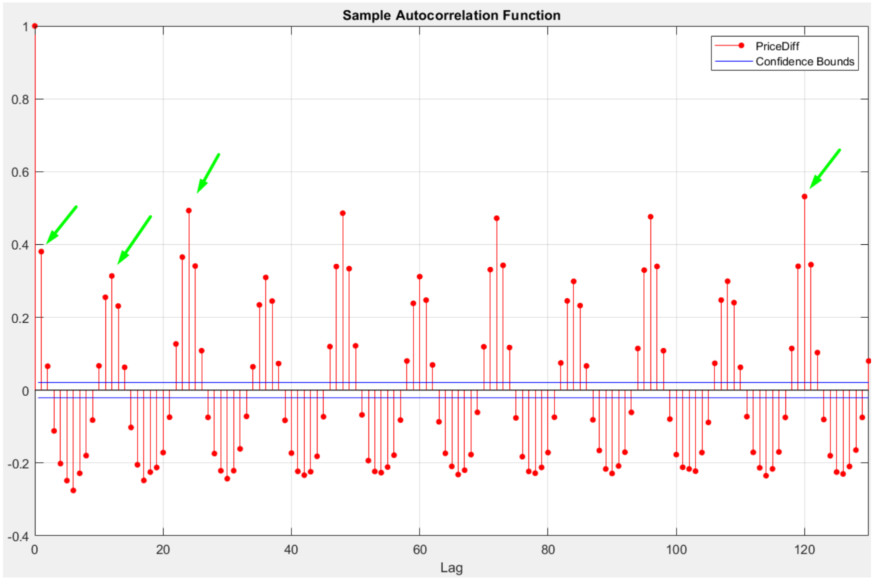

Fig. 3.9. Autocorrelation function

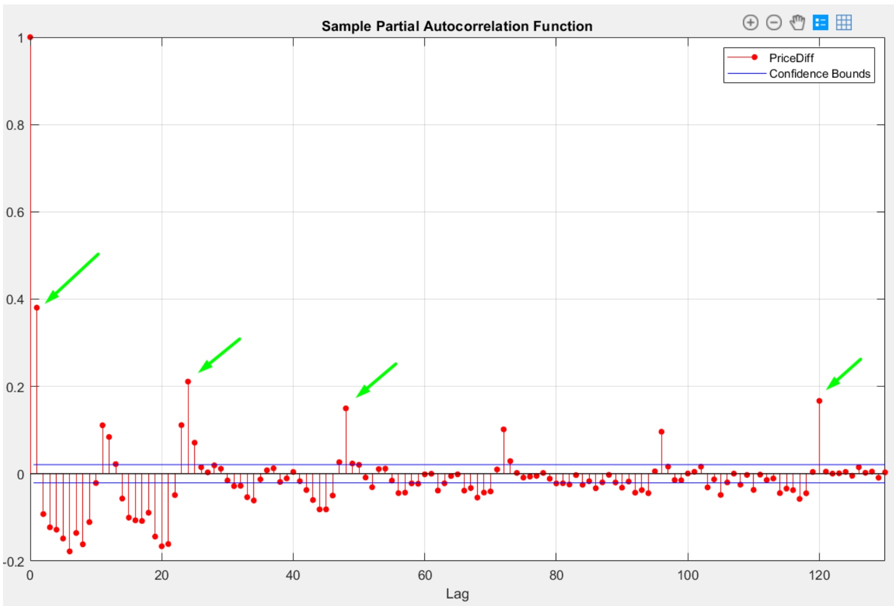

Fig.3.10. Partial autocorrelation function

Partial autocorrelation helps in determining autoregressive order. The most significant partial correlation is in lags 1, 24, 48, 120, 600. Also, partial correlations in the first 12 lags can be considered. Autocorrelation function crosses the x-axis after 1 lag, so moving average of order 1 is considered to be the most likely. However, increasing the moving average order to 2 can also be considered.

Based on these results, the following ARIMA models were estimated and compared with each other.

Table 3.1

Comparison of ARIMA models

Model	AR parameters	Degree of integration	MA parameters	AIC
ARIMA1	1	1	1	5.5253e+04
ARIMA2	1,24	1	1	5.3586e+04
ARIMA3	1,24, 48	1	1	5.2795e+04
ARIMA4	1, 24, 48,120	1	1,2	5.1769e+04
ARIMA5	1, 2	1	1	5.4114e+04
ARIMA6	1,2,24	1	1,2	5.3220e+04

Goodness of fit of the models is compared using the Akaike information criterion (AIC) (table 3.1) [63]. It is estimator of prediction error and thereby shows the relative quality of the model. It can be used to compare only models that were estimated using the same dataset. The lower AIC value is, the better is quality of the model.

Based on the AIC criteria, ARIMA4 model has the best results. Its mathematical representation is (eq 3.1):

$$(1-\phi_1 L-\phi_{24}L^{24}-\phi_{48}L^{48}-\phi_{120}L^{120})(1-L)y_t = c+(1+\theta_1 L+\theta_2 L^2)\varepsilon_t \qquad (3.1)$$

Where

ϕ – autoregressive parameter;

θ – moving average parameter;

L – lagged value;

c – constant;

ε – white noise;

Table 3.2 shows the estimation results of the ARIMA4 model. Standard Error corresponds to the estimate of the standard deviation. T-statistic is the ratio of the coefficient to the standard error. P-value is the probability of obtaining results at least as extreme as observed results.

Table 3.2

ARIMA4 estimation results

Parameter	Value	Standard Error	t Statistic	P-Value
Constant	4.9851e-05	0.03414	0.0014602	0.99883
AR{1}	0.19839	0.010963	18.0961	3.4181e-73
AR{24}	0.19975	0.005291	37.7519	6.99e-312
AR{48}	0.19569	0.0063782	30.6813	1.0103e-206
AR{120}	0.2968	0.0044064	67.3568	0
MA{1}	-0.097981	0.012226	-8.0143	1.1076e-15
MA{2}	-0.13658	0.0067122	-20.3476	4.8741e-92
Variance	16.8095	0.080738	208.199	0

Fig. 3.11 shows the fit of the ARIMA4 model.

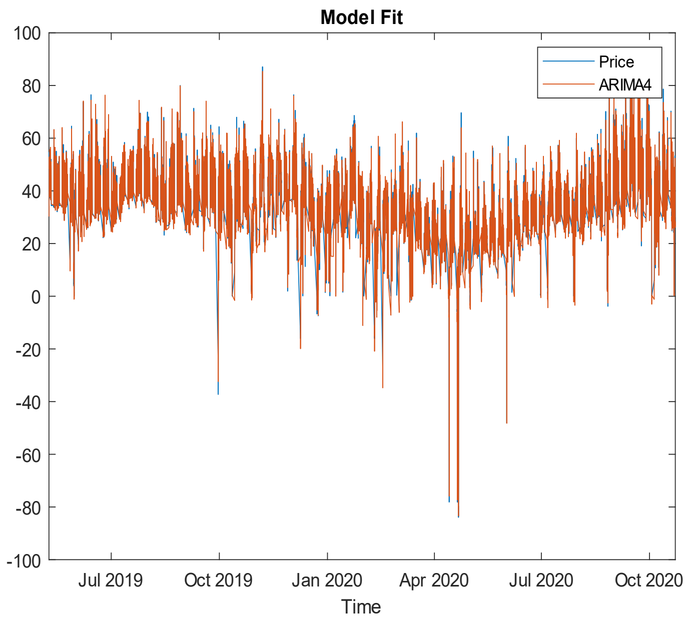

Fig. 3.11. Fit of ARIMA4 model

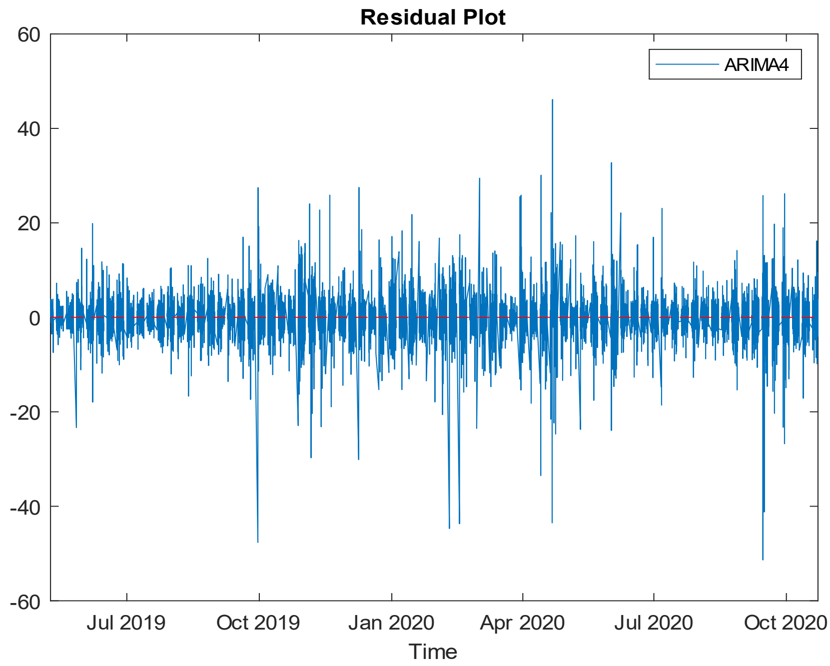

Fig. 3.12. plot of the model residuals

Fig. 3.12 shows the plot of the residuals. Model is considered sufficient if residuals are white noise. It means there is no information left there that can be used for model improvement. Applying Ljung-Box test for residuals autocorrelation shows that there still is some autocorrelation and residuals are not white noise. Visually it can be seen that they often reach up to 10 euros, so this information can be used for significant improvement of the model.

However, improvement of the existing model by entering more autoregressive or moving average parameters can lead to overfitting and lack of generalization. Moreover, such approach may try to explain non-linear dependencies linearly. It can work well on this dataset but will fail badly in a real forecasting scenario. It makes sense to apply other forecasting methods to the information contained in residuals.

Fig. 3.13 – 3.16 show 24 hours ahead forecast made by ARIMA4 model. Daily pattern is repeated properly, however, the deviation of actual prices is higher. First forecasted price almost perfectly matches the actual price

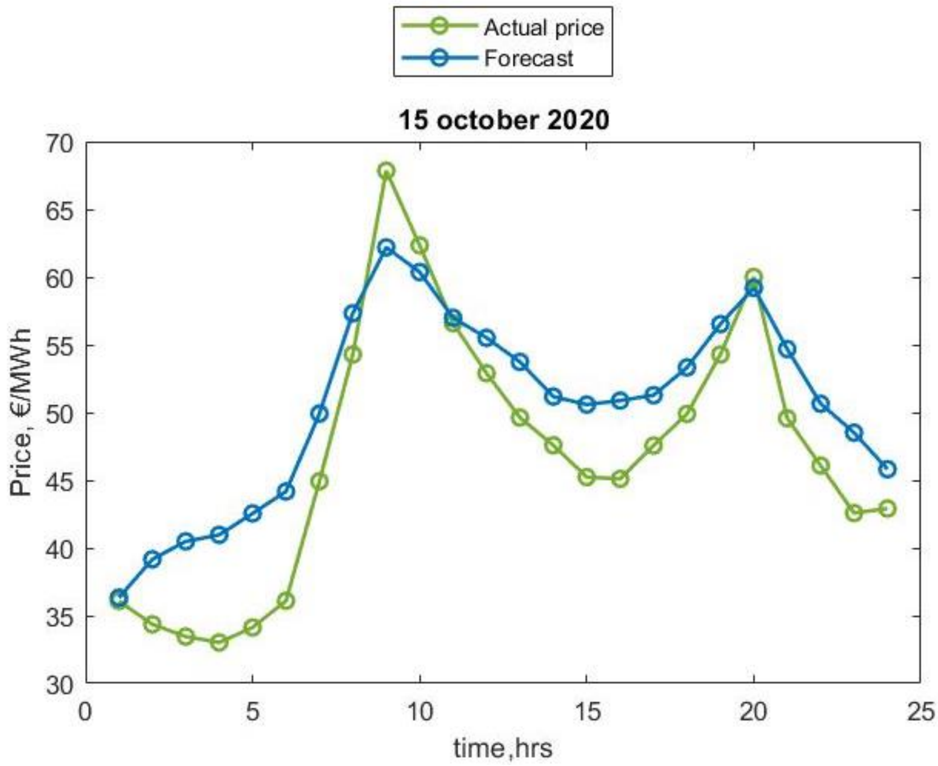

Fig. 3.13. ARIMA forecast 15 October 2020

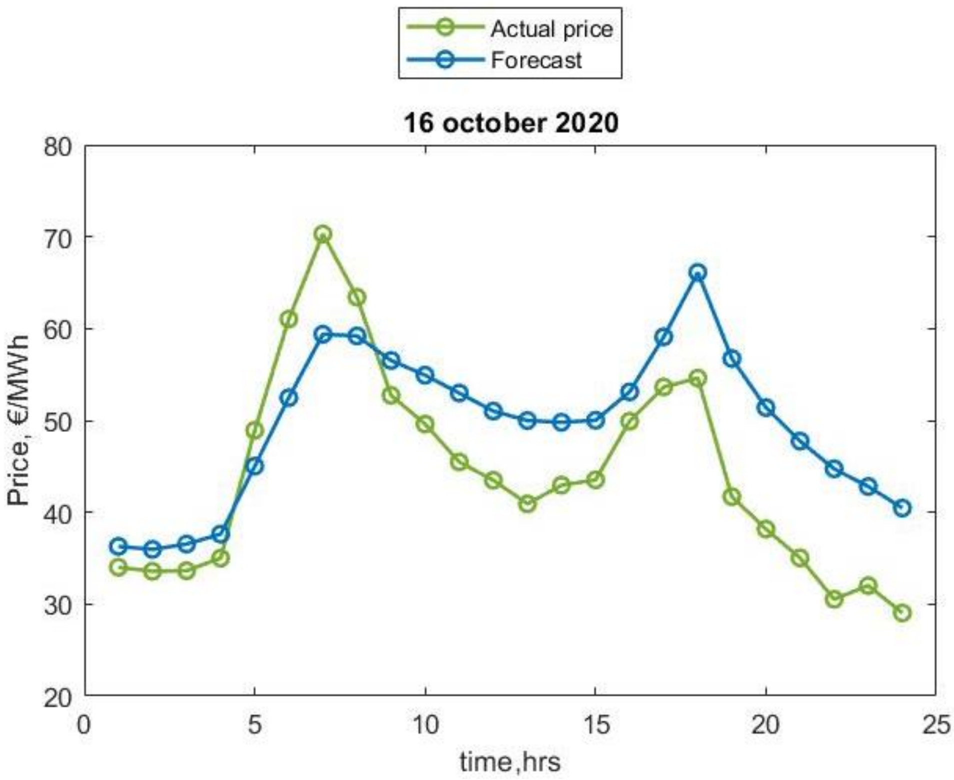

Fig.3.14. ARIMA forecast 16 October 2020

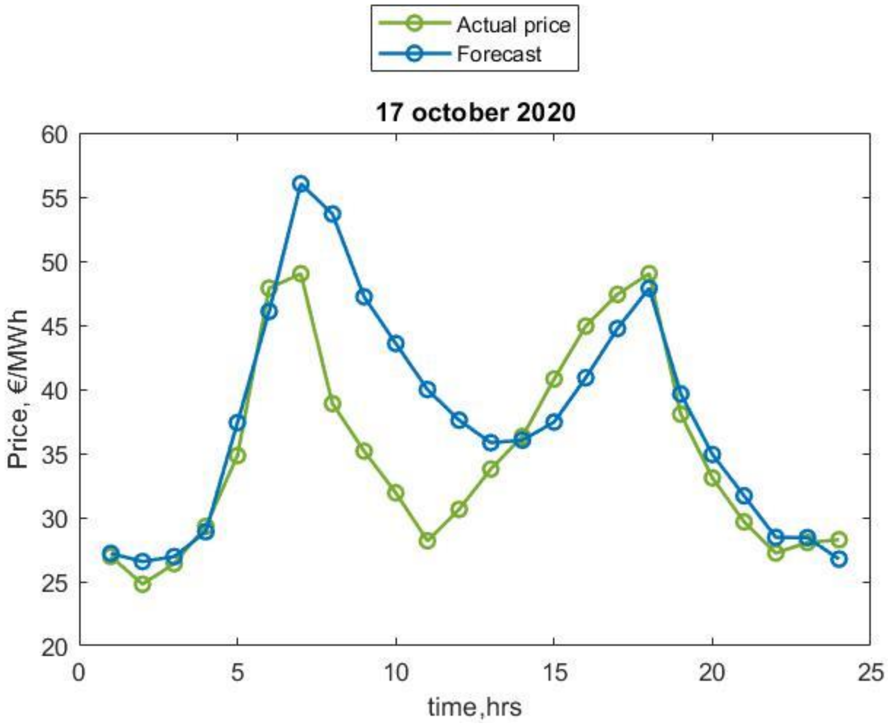

Fig. 3.15. ARIMA forecast 17 October 2020

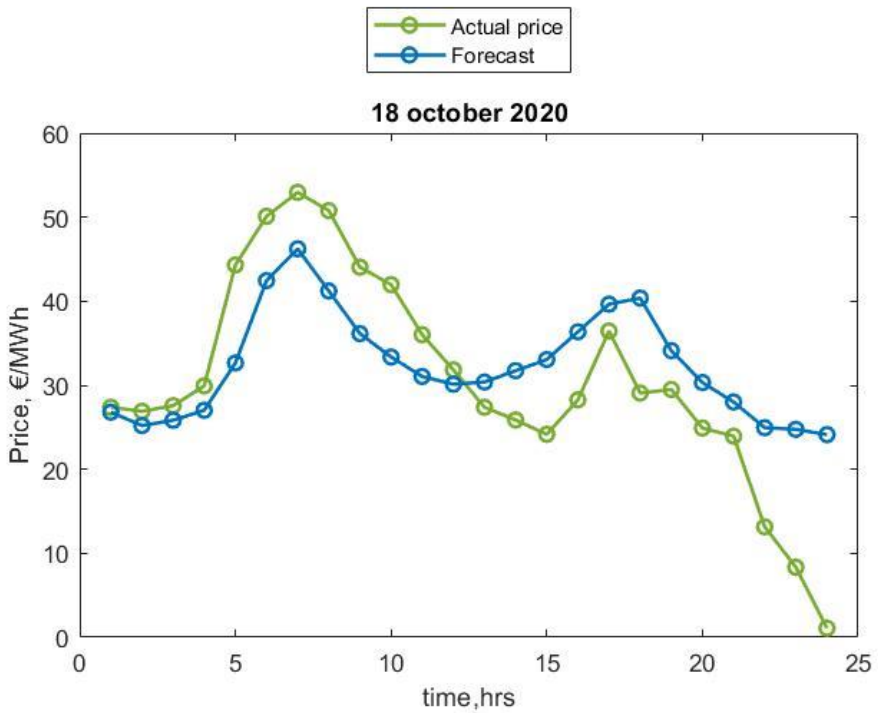

Fig. 3.16. ARIMA forecast 18 October 2020

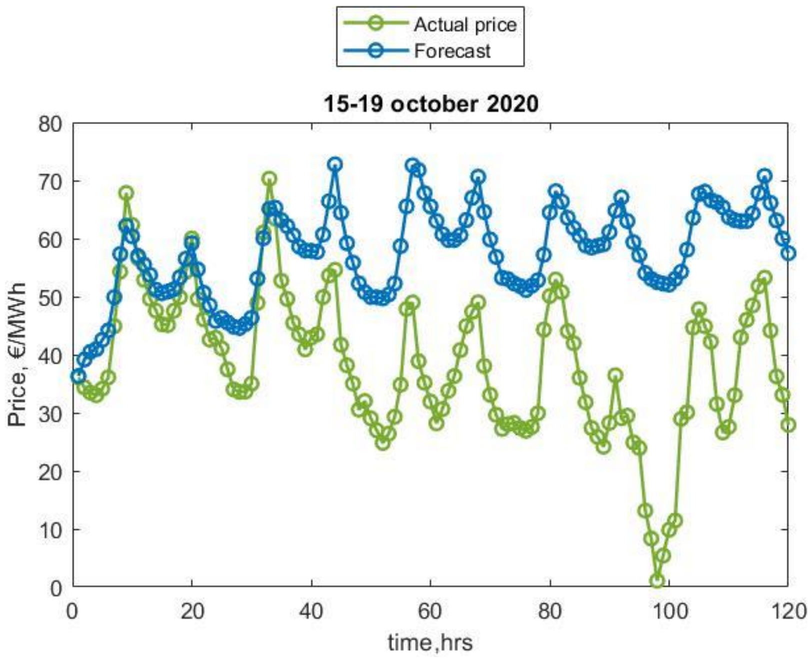

Fig. 3.18. 120 hours ahead forecast

Fig. 3.18 shows 120 hours ahead forecast. The precision of forecast declines as it goes further in the future. It can be concluded that ARIMA model is not suitable for forecasts more than 48 hours ahead.

3.3. Application of feed-forward neural network

Feed-forward neural network was applied to find non-linear dependencies between prices at different lags.

It takes 24 lags as inputs and has 1 hidden layer with 12 neurons (Fig. 3.19).

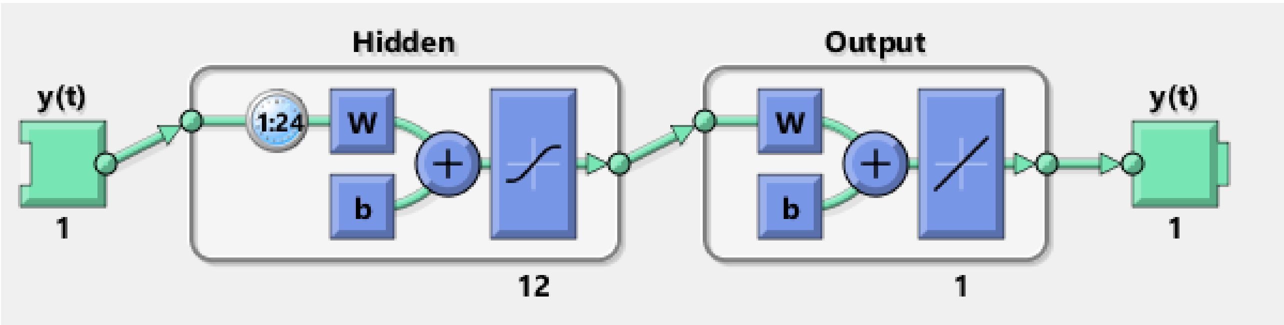

Fig. 3.19. Graphical representation of the neural network

Firstly, model is trained in the open-loop. It means that that it uses target values as inputs to improve performance of the model. Then, it is switched to the closed-loop to obtain a forecast for the time extent that is not included in the dataset. Model is trained using Bayesian Regularization method.

Fig. 3.20 shows error autocorrelation. It can be seen that correlations do not cross the confidence limit and the only significant correlation is at lag 0, that is correlation of the error with itself.

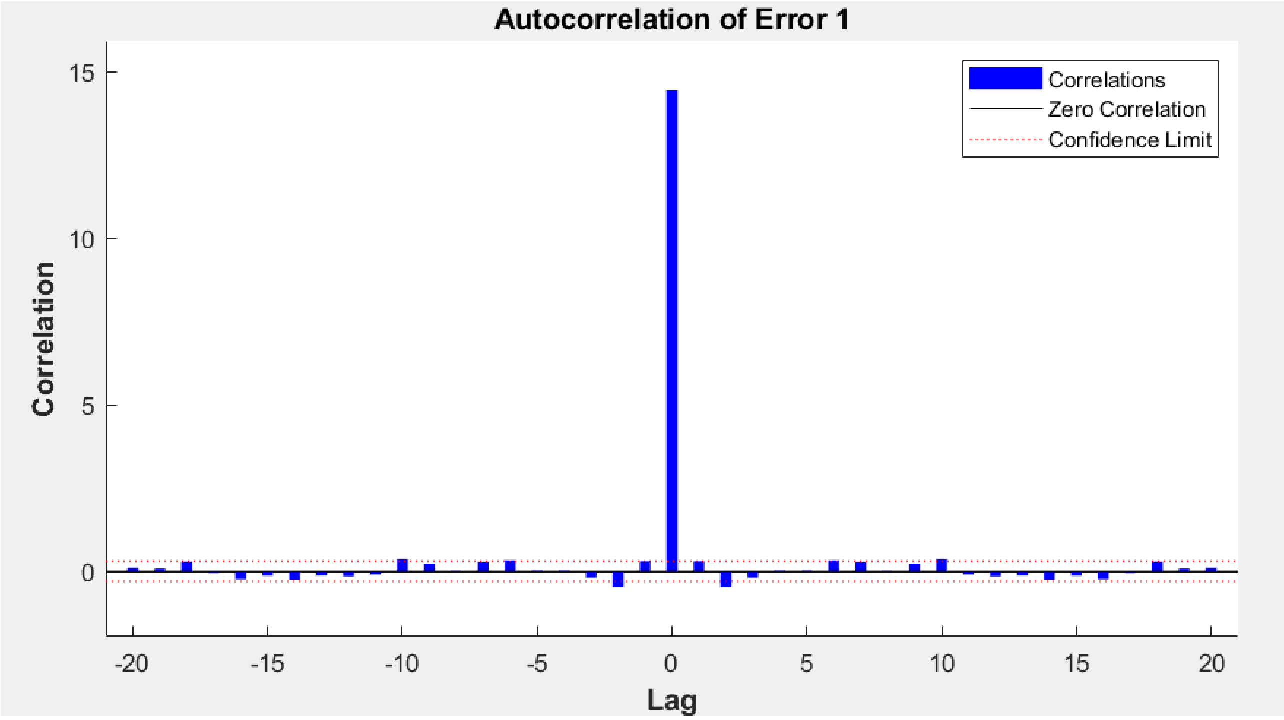

Fig. 3.20. Error autocorrelation

It was found that applying a neural network to stationarized time series gives better results in forecasting. So before training the neural network, dataset was differenced once. After that, forecast was added to initial dataset (concatenated) and the inverse difference was applied.

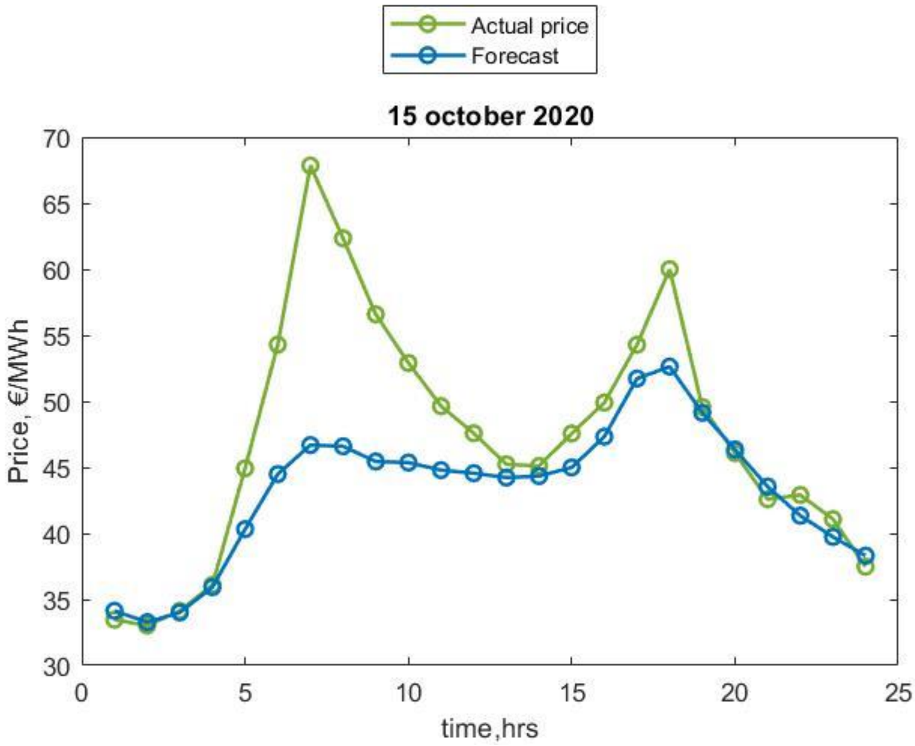

Fig. 3.21. NN forecast 15 October 2020

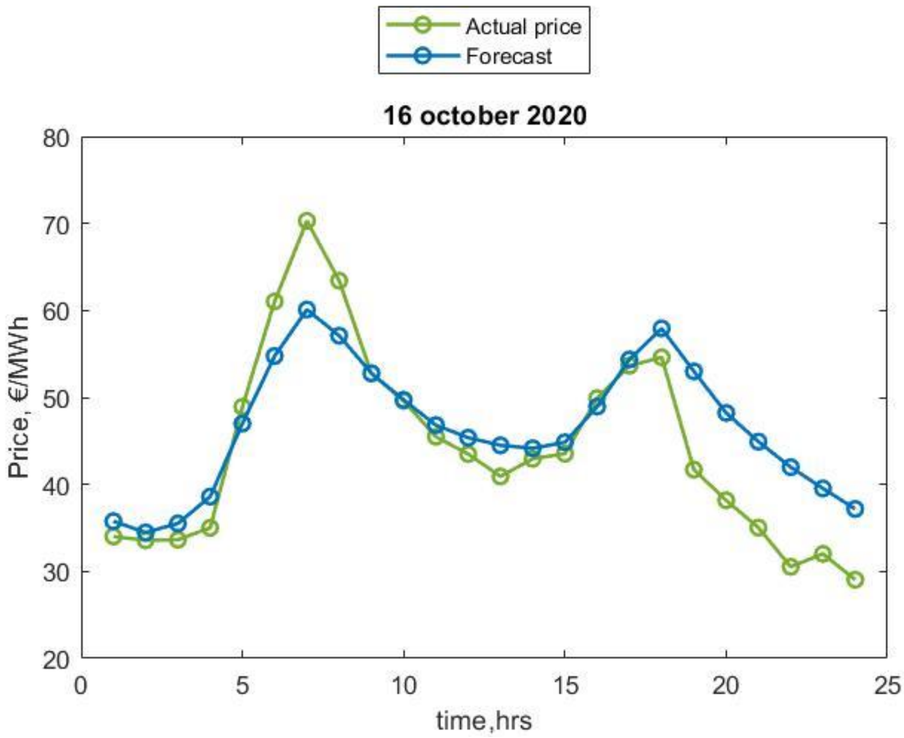

Fig. 3.22. NN forecast 16 October 2020

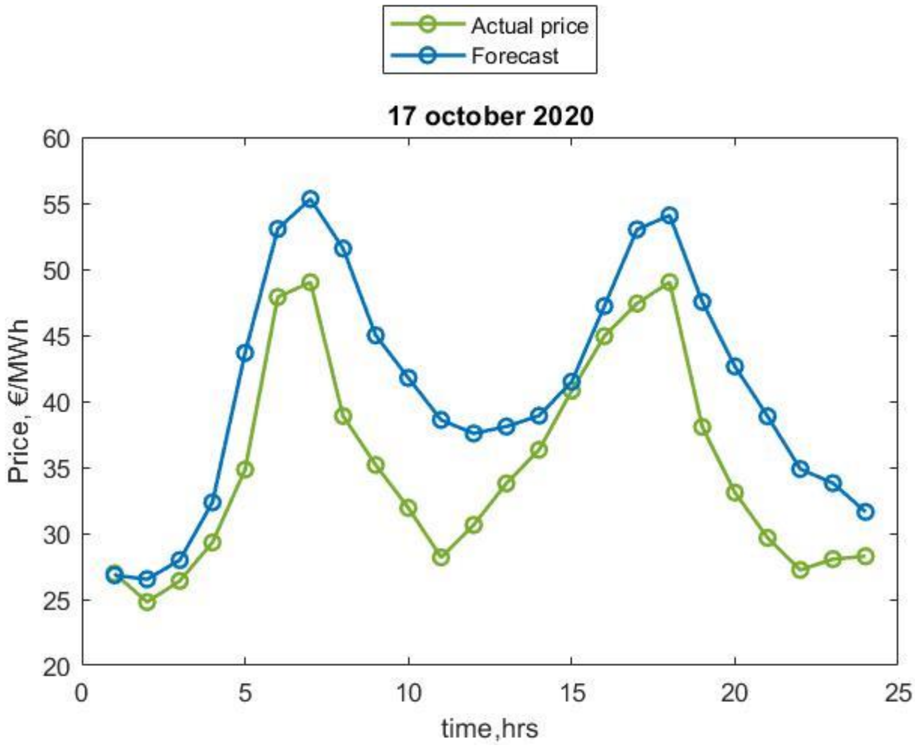

Fig. 3.23. NN forecast 17 October 2020

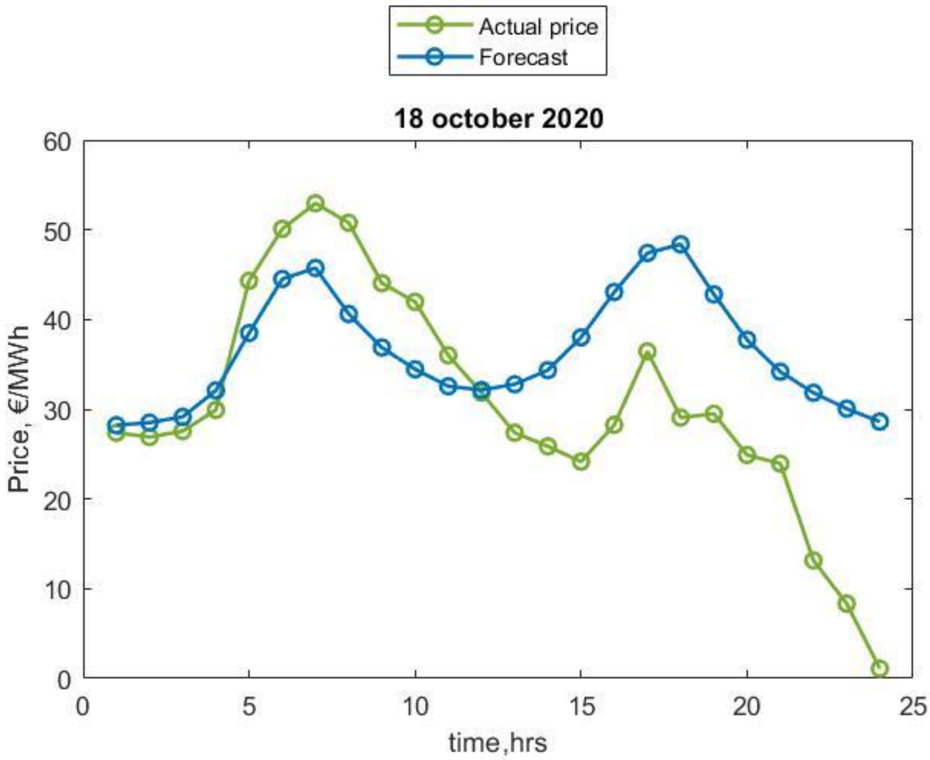

Fig. 3.24. NN forecast 18 October 2020

Figures 3.21 to 3.24 demonstrate the daily forecast made by neural network. After each day, prices of the previous day were added to the dataset, and model was retrained.

3.4. Conclusions

ARIMA and ANN can be considered as reliable models for electricity price forecasting. ARIMA model generally performs better than ANN, however, there are days, when ANN completely outperforms ARIMA. Both models are not reactive to rapid price changes. Daily pattern is repeated properly for both models; however, prediction of outliers needs improvement.

Combination of models through averaging didn`t show the expected results. One of the reasons is that both models were obtained through the same dataset. Hybrid forecasting model that combines ARIMA and ANN was investigated. Residuals from fitting an ARIMA model were obtained and used as an input to the feed-forward neural network. It was supposed, that they contain non-linear autocorrelation and this information can be used for model improvement. Feed-forward network with 24 lags and 12 neurons in the hidden layer was chosen. However, no significant autocorrelation was found. Predicted values were negligible.

In 5 days ahead forecast ANN showed better results than ARIMA. Better performance of ANN in trend estimation can be subjected to further investigation.

Overall, the goodness of forecast does not depend directly on the complexity of the forecasting model. It was found that simpler models can outperform more complex ones in certain situations. In periods when price can be considered as stable, models show similar results.

РОЗДІЛ 4
ЕКОНОМІКА

При розробці програмного забезпечення важливими етапами є визначення трудомісткості розробки ПЗ, розрахунок витрат на створення програмного продукту і аналіз ринку збуту розробленого програмного забезпечення.

4.1. Визначення трудомісткості проведення дослідження та розробки необхідного для його проведення програмного забезпечення

Задані дані:

1. Передбачуване число операторів – 950.
2. Коефіцієнт складності програми – 1,6.
3. Коефіцієнт корекції програми в ході її розробки – 0,3.
4. Годинна заробітна плата програміста, грн/год – 70.
5. Коефіцієнт збільшення витрат праці внаслідок недостатнього опису задачі – 1,1.
6. Коефіцієнт кваліфікації програміста – 1,5.
7. Вартість машино-години ЕОМ, грн/год – 15.

Нормування праці в процесі створення ПЗ істотно ускладнено в силу творчого характеру праці програміста. Тому трудомісткість розробки ПЗ може бути розрахована на основі системи моделей з різною точністю оцінки.

Трудомісткість розробки ПЗ можна розрахувати за формулою:

$$t = t_o + t_и + t_a + t_п + t_{отл} + t_д \text{ ,} людино - годин, \tag{4.1}$$

де t_o - витрати праці на підготовку й опис поставленої задачі (приймається 50);

$t_и$ - витрати праці на дослідження алгоритму рішення задачі;

$t_а$ - витрати праці на розробку блок-схеми алгоритму;

$t_п$ - витрати праці на програмування по готовій блок-схемі;

$t_{отл}$ - витрати праці на налагодження програми на ЕОМ;

$t_д$ - витрати праці на підготовку документації.

Складові витрати праці визначаються через умовне число операторів у ПЗ, яке розробляється.

Умовне число операторів (підпрограм):

$$Q = q * C * (1 + p), \tag{4.2}$$

де q - передбачуване число операторів;

C - коефіцієнт складності програми;

p - коефіцієнт корекції програми в ході її розробки.

$$Q = 950 * 1{,}6 * (1 + 0{,}3) = 1976, людино - годин \tag{4.3}$$

Витрати праці на вивчення опису задачі $t_и$ визначається з урахуванням уточнення опису і кваліфікації програміста:

$$t_и = \frac{Q * B}{(75 \ldots 85) * k}, \quad людино - годин, \tag{4.4}$$

де B - коефіцієнт збільшення витрат праці внаслідок недостатнього опису задачі;

k - коефіцієнт кваліфікації програміста, обумовлений від стажу роботи з даної спеціальності.

$$t_и = \frac{1976 * 1{,}1}{75 * 1{,}5} = \frac{2390{,}96}{112{,}5} = 19{,}32, людино - годин \tag{4.5}$$

Витрати праці на розробку алгоритму рішення задачі:

$$t_a = \frac{Q}{(20 \dots 25) * k}, \qquad людино - годин, \tag{4.6}$$

$$t_a = \frac{1976}{22 * 1,5} = 59,87, людино - годин \tag{4.7}$$

Витрати на складання програми по готовій блок-схемі:

$$t_n = \frac{Q}{(20 \dots 25) * k}, людино - годин \tag{4.8}$$

$$t_n = \frac{1976}{20 * 1,5} = 65,86, людино - годин \tag{4.9}$$

Витрати праці на налагодження програми на ЕОМ:

- за умови автономного налагодження одного завдання:

$$t_{отл} = \frac{Q}{(4 \dots 5) * k}, людино - годин, \tag{4.10}$$

$$t_{отл} = \frac{1976}{4 * 1,5} = 329,3, людино - годин \tag{4.11}$$

- за умови комплексного налагодження завдання:

$$t_{отл}^{k} = 1,5 * t_{отл}\,, людино - годин. \tag{4.12}$$

$$t_{отл}^{k} = 1,5 * 329,3 = 493,95, людино - годин \tag{4.13}$$

Витрати праці на підготовку документації:

$$t_{д} = t_{др} + t_{до}, \qquad людино - годин, \tag{4.14}$$

де $t_{др}$ - трудомісткість підготовки матеріалів і рукопису.

$$t_{др} = \frac{Q}{15..20 * k}, людино - годин. \tag{4.15}$$

$$t_{др} = \frac{1976}{15 * 1,5} = 87,82, людино - годин \tag{4.16}$$

$t_{до}$ - трудомісткість редагування, печатки й оформлення документації

$$t_{до} = 0,75 * t_{др}, людино - годин \tag{4.17}$$

$$t_{до} = 0,75 * 87,82 = 65,86, людино - годин \tag{4.18}$$

$$t_{д} = 87,82 + 65,86 = 153,68, людино - годин \tag{4.19}$$

Тепер розрахуємо трудомісткість ПЗ:

$$t = 195,05 + 164,6 + 329,3 + 153,68 = 842,63, людино - годин. \tag{4.20}$$

4.2. Витрати на створення програмного забезпечення для проведення дослідження

Витрати на створення ПЗ Кпо включають витрати на заробітну плату виконавця програми і витрат машинного часу, необхідного на налагодження програми на ЕОМ:

$$К_{по} = З_{зп} + З_{мв}, грн. \tag{4.21}$$

Заробітна плата виконавців визначається за формулою:

$$З_{зп} = t * C_{пр}, грн, \tag{4.22}$$

де: t - загальна трудомісткість, людино-годин;

$С_{пр}$ - середня годинна заробітна плата програміста, грн/година

$$З_{зп} = 842{,}63 * 70 = 58984{,}1, грн. \tag{4.23}$$

Вартість машинного часу, необхідного для налагодження програми на ЕОМ:

$$З_{мв} = t_{отл} * C_{мч}, грн, \tag{4.24}$$

де $t_{отл}$ - трудомісткість налагодження програми на ЕОМ, год.

$С_{мч}$ - вартість машино-години ЕОМ, грн/год.

$$З_{мв} = 329{,}3 * 15 = 4939{,}5, грн. \tag{4.25}$$

Визначені в такий спосіб витрати на створення програмного забезпечення є частиною одноразових капітальних витрат на створення АСУП.

$$K_{по} = 58984{,}1 + 4939{,}5 = 63923{,}6, грн. \tag{4.26}$$

Очікуваний період створення ПЗ:

$$T = \frac{t}{B_k * F_p}, міс, \tag{4.27}$$

Де

B_k - число виконавців;

F_p - місячний фонд робочого часу (при 40 годинному робочому тижні F_p=176 годин).

$B_k = 1$

$$T = \frac{842{,}63}{1 * 176} = 4{,}7, \quad міс \qquad (4.28)$$

Таким чином, трудомісткість розробки програмного забезпечення становить 4.8 міс.

4.3. Маркетингові дослідження ринку використання результатів дослідження

ЄС перебуває в процесі трансформації своєї економіки з метою мінімізації викидів парникових газів (ПГ). Очікується, що ключову роль в цьому відіграє електроенергія. По-перше, більш ефективне використання електроенергії та зростаюча частка електроенергії з відновлюваних джерел допоможе зменшити викиди парникових газів. По-друге, очікується зростання частки електроенергії в загальному споживанні енергії, особливо в транспортному секторі (електромобілі) та в системах опалення та охолодження (електричні теплові насоси). Виробництво електроенергії розглядається як ключовий компонент для поступового припинення викидів парникових газів з викопного палива до другої половини цього століття. Щоб усі ці зміни відбулись, потрібні значні інвестиції в виробництво електроенергії, її транспортування та розподілу, а також в споживчі електротовари. Більше того, інвестиції та інновації є необхідними для того, щоб забезпечити стабільність постачання електроенергії в умовах підвищення попиту та збільшення кількості електроенергії з відновлюваних джерел, що залежать від сонячного світла, вітру та дощів. У результаті цієї дуже мінливої та невизначеної природи ринку електроенергії та того, що електроенергія - це товар, який споживачі потребують у своєму повсякденному житті в значній мірі, точне прогнозування цін є важливим для всіх гравців ринку: генеруючих компанії, дистриб'юторів, кінцевих споживачів.

Управління фінансовими ризиками часто є головним пріоритетом для учасників дерегульованих ринків електроенергії через значні цінові та об'ємні ризики. Особливі характеристики цього цінового ризику в значній мірі залежать від фізичних основ ринку, таких як поєднання типів генераторних установок та взаємозв'язок між попитом та погодною ситуацією. Ціновий ризик може проявлятися ціновими "стрибками", які важко передбачити, і ціновими "кроками", коли основне паливо або положення заводу змінюються протягом тривалого періоду. Об'ємний ризик часто використовується для позначення явища, коли учасники ринку електроенергії мають невизначені обсяги або кількості споживання чи виробництва. Наприклад, роздрібний торговець не може точно прогнозувати споживчий попит протягом певної години більше, ніж за кілька днів, а виробник не може передбачити точний час, коли буде відключена установка або не вистачатиме палива. Складовим фактором є також загальний взаємозв'язок між екстремальними цінами та обсягом подій. Наприклад, сплески цін часто трапляються, коли деякі виробники перестають працювати на заводах або коли деякі споживачі перебувають у періоді пікового споживання. Роздрібні торговці електроенергією, які в цілому купують на оптовому ринку, та виробники, які в цілому продають на оптовий ринок, піддаються цим ціновим та об'ємним ефектам, а щоб захистити себе від нестабільності, вони укладають між собою "контракти на хеджування". Структура цих контрактів різниться залежно від регіонального ринку через різні конвенції та ринкові структури.

Отже, ринок електроенергії та його учасники все більше залежать від якісних цінових прогнозів.

4.4. Оцінка економічної ефективності впровадження програмного забезпечення

Генератор, комунальне підприємство або великий промисловий споживач, який здатний прогнозувати нестабільні оптові ціни з адекватним рівнем точності, може скорегувати свою стратегію торгів та власний графік виробництва або споживання, щоб зменшити ризик або максимізувати прибуток. Однак, оскільки прогнози навантаження та цін використовуються багатьма підрозділами енергетичної компанії, дуже важко визначити кількісні вигоди від їх покращення. Орієнтовна оцінка економії від зменшення середньої абсолютної процентної помилки (MAPE) на 1% для комунального господарства з піковим навантаженням 1 ГВт становить[64]:

- 500 000 доларів на рік за довгостроковим прогнозуванням навантаження,
- 300 000 доларів на рік від короткострокового прогнозування навантаження,
- 600 000 доларів на рік від короткострокового прогнозу навантаження та ціни.

Оскільки дана робота має аналітичний характер, в ній не розробляється програмне забезпечення. Отже, неможливо розрахувати економічний ефект, обсяг необхідних інвестицій, термін окупності і прибутковість.

CONCLUSIONS

This work investigates the structure and main characteristics of deregulated electricity market. Due to increasing complexity of power markets, electricity price forecasting domain constantly needs improvements. Creation of high-quality MCP forecasts becomes a difficult task, due to increasing complexity of bidding strategies used by participants and interaction of uncertainties in an intricate way. Moreover, in most competitive power markets, hourly prices depend on a number of factors, such as weather, day of the week, load, consumption etc. Changes of these factors due to weather swings or seasonal changes cause MCP to be non-stationary. Other complex stochastic signals like fuel costs and equipment outages also cause nonlinear behavior and sudden unpredictable changes of it.

Many researchers have developed forecasting models. Some of them are described in the first section. Despite that that numerous methods and approaches exist, prediction of market clearing prices still involves large errors.

Two main approaches exist in this field: linear and non-linear. Linear approaches are found to be effective for stable behaviors of MCP, however they have difficulties predicting the non-linear behaviors and rapid changes. Neural are considered to be more efficient in this task. Single neural network cannot reconstruct all parts of the complex mapping function of prices.

Combination of different forecasting methods can lead to a higher accuracy; however, it is highly depended on the market being investigated. While in one market certain forecasting model can show promising results, it can be completely inapplicable to another. Combination of ARIMA and ANN didn’t show the expected performance. Initial suggestion that ANN can model the residual part was not proved.

As was discussed, electricity prices are dependent on various factors. Adding additional parameters to the forecasting models can significantly improve results. Adding exogenous factors to forecasting models can help in modelling non-linear behavior. ARIMAX (Autoregressive integrated moving average with an exogenous variable) or NNARX (Neural Network Autoregressive with exogenous variable) can be used for this purpose.

REFERENCES

1. Baumol W., William J., 1977. "On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry", American Economic Review 67, 809–22.
2. Lucas W., Davis W., 2012. American Economic Journal: Applied Economics Vol. 4, No. 4, 194-225.
3. Milligan M., Frew L., 2016. Wholesale electricity market design with increasing levels of renewable generation: Revenue sufficiency and long-term reliability. The Electricity Journal 29, 26–38.
4. https://www.bmwi.de/Redaktion/EN/Artikel/Energy/smart-grids.html Accessed on 14.12.2020.
5. Leigh H., 2015. Capacity mechanisms in the EU energy markets: law, policy, and economics, Oxford University Press.
6. https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/over-the-counter-otc/ Accessed on 14.12.2020.
7. Herrero I., Rodilla P., Batlle C., 2015. Electricity market-clearing prices and investment incentives: The role of pricing rules. Energy Economics 47, 42–51.
8. https://www.electricchoice.com/blog/8-key-factors-impact-electricity-prices/ Accessed on 14.12.2020.
9. Pape C., Hagemann S., Weber C., 2015. Are fundamentals enough? Explaining price variations in the German day-ahead and intraday power market, EWL Working Paper, No. 02/15.
10. Weron R., “Electricity price forecasting: A review of the state-of-the-art with a look into 742 the future,” Int. J. Forecast., vol. 30, no. 4, pp. 1030–1081.
11. https://energyanalyst.co.uk/electricity-price-forecasting-using-multi-agent-models/ Accessed on 14.12.2020.
12. Beran P., Pape C., Weber C., 2018. "Modelling German electricity wholesale spot prices with a parsimonious fundamental model validation and application," EWL Working Papers 1801.

13. Santos G., Pinto S., Morais H., Sousa T., 2015. Multi-agent simulation of competitive electricity markets: Autonomous systems cooperation for European market modeling, Energy Conversion and Management, Volume 99, 387-399.
14. Derek W., 2004. Modelling Prices in Competitive Electricity Markets.
15. Weron R., 2006. Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach.
16. Carmona R., Coulon M., 2014. A Survey of Commodity Markets and Structural Models for Electricity Prices. Springer New York, 41–83.
17. Jonsson T., Pinson P., Nielsen H., Madsen H., Nielsen T., 2013. "Forecasting Electricity Spot Prices Accounting for Wind Power Predictions". IEEE Transactions on Sustainable Energy. 4 (1), 210–218.
18. Karakatsani Nektaria V., Bunn Derek W., 2008. "Forecasting electricity prices: The impact of fundamentals and time-varying coefficients". International Journal of Forecasting. Energy Forecasting. 24 (4), 764–785.
19. Jonsson T., Pinson P., Nielsen H., 2015. "Forecasting Electricity Spot Prices". IEEE Transactions on Sustainable Energy. 4 (1): 210–218.
20. Antonio J., Contreras J., Espínola R., Plazas A., 2005. "Forecasting electricity prices for a day-ahead pool-based electric energy market". International Journal of Forecasting. 21 (3), 435–462.
21. Weron R., Misiorek A., 2008. "Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models". International Journal of Forecasting. Energy Forecasting. 24 (4). 744–763.
22. Zareipour H., 2008. Price-based energy management in competitive electricity markets. VDM Verlag Dr. Müller.
23. Amjady N., 2006. "Day-ahead price forecasting of electricity markets by a new fuzzy neural network". IEEE Transactions on Power Systems. 21 (2), 887–896.
24. Keles D., Scelle J., Paraschiv F., Fichtner W., 2016. "Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks". Applied Energy. 162, 218–230.

25. Rodriguez C., Anders G., 2004. "Energy price forecasting in the Ontario competitive power system market". IEEE Transactions on Power Systems. 19 (1), 366–374.
26. Yan X., Chowdhury A., 2013. "Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach". International Journal of Electrical Power & Energy Systems. 53, 20–26.
27. Jiang P., Liu F., Song Y., 2016. A Hybrid Multi-Step Model for Forecasting Day-Ahead Electricity Price Based on Optimization, Fuzzy Logic and Model Selection. Energies 9, 618.
28. Conejo A., Plazas M., Espinola R., 2015 "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," in IEEE Transactions on Power Systems, vol. 20, no. 2, 1035-1042.
29. Moazzami M., Khodabakhshian A., Hooshmand R., 2013. A new hybrid day-ahead peak load forecasting method for Iran’s National Grid. Applied Energy. 101(0):489–501.
30. Mellit A., Benghanem M., Kalogirou S., 2013. An adaptive wavelet-network model for forecasting daily total solar-radiation. Applied Energy, 705–22.
31. Terui N., 2002. Combined forecasts from linear and nonlinear time series models. International Journal of Forecasting.18(3):421–38.
32. Nguyen H., Nabney T., 2010. Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. Energy. 35(9):3674–85
33. Luxhoj J., Riis J., Stensballe B., 1996. A hybrid econometric-neural network modeling approach for sales forecasting, Int. J. Prod. Econ.175–192.
34. Wu L., Shahidehpour W., 2014. “A hybrid model for integrated day-ahead electricity price and load forecasting in smart grid”, 1937–1950.
35. Hong T., Wilson J., Xie J., 2014. Long term probabilistic load forecasting and normalization with hourly information. IEEE Trans. Smart Grid 5, 456–462.
36. Kaboli, S., Fallahpour A., Selvaraj J., Rahim N., 2017. Long-term electrical energy consumption formulating and forecasting via optimized gene expression programming. Energy, 126, 144–164.

37. AlRashidi M., El-Naggar K., 2010 Long term electric load forecasting based on particle swarm optimization. Appl. Energy, 87, 320–326.
38. Moral-Carcedo J., Pérez-García J., 2017. Integrating long-term economic scenarios into peak load forecasting: An application to Spain. Energy, 140, 682–695.
39. Khuntia S., Rueda J., Meijden M., 2016. Volatility in electrical load forecasting for long-term horizon—An ARIMA-GARCH approach. In Proceedings of the IEEE PMAPS, 16-20.
40. Bello A., Reneses J., Muñoz A., 2016. Medium-Term Probabilistic Forecasting of Extremely Low Prices in Electricity Markets: Application to the Spanish Case. Energies 9, 193.
41. Cheng C., Chen F., Li G., Tu Q., 2016. Market Equilibrium and Impact of Market Mechanism Parameters on the Electricity Price in Yunnan's Electricity Market. Energies, 9, 463.
42. Jiang, P., Liu, F., Song Y., 2016. Hybrid Multi-Step Model for Forecasting Day-Ahead Electricity Price Based on Optimization, Fuzzy Logic and Model Selection. Energies 2016, 9, 618.
43. Conejo J., Plazas M., Espinola R., Molina B. "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," in IEEE Transactions on Power Systems, vol. 20, no. 2, 1035-1042.
44. Jakaša T., Andročec I., Sprčić P., 2011. "Electricity price forecasting — ARIMA model approach," 2011 8th International Conference on the European Energy Market (EEM), Zagreb, 222-225.
45. Contreras J., Espinola R., Nogales F., Conejo J., ARIMA Models to Predict Next-Day Electricity Prices, IEEE Transactions on Power Systems, vol.18, no.3.
46. Berger J., Yalcinoz T., Rudion K., 2020. "Investigating the Intraday Continuous Electricity Market using Auto Regression Integrated Moving Average Model with Exogenous Inputs," IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and

Commercial Power Systems Europe *(EEEIC / I&CPS Europe)*, Madrid, Spain, 1-6.

47. https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/ Accessed on 14.12.2020.
48. https://365datascience.com/arimax/ Accessed on 14.12.2020.
49. Andreas Z., 1994. Simulation Neuronaler Netze [Simulation of Neural Networks] (in German) (1st ed.). Addison-Wesley, 73.
50. Auer, P., Harald B., Wolfgang M., 2008. "A learning rule for very simple universal approximators consisting of a single layer of perceptrons" Neural Networks, 786–795.
51. https://towardsdatascience.com/feed-forward-neural-networks-c503faa46620 Accessed on 14.12.2020.
52. https://medium.com/max-exchange/an-autoregressive-neural-network-approach-to-forecasting-bitcoin-price-14e3121734ea Accessed on 14.12.2020.
53. https://www.inovex.de/blog/hybrid-time-series-forecasting/ Accessed on 14.12.2020.
54. Skopal R., "Short-term hourly price forward curve prediction using neural network and hybrid ARIMA-NN model,"2015 International Conference on Information and Digital Technologies, Zilina, 335-338.
55. Areekul P., Senjyu T., Toyama H., Yona A., 2009. "Combination of artificial neural network and ARIMA time series models for short term price forecasting in deregulated market," Transmission & Distribution Conference & Exposition: Asia and Pacific, Seoul, 1-4.
56. Shumway R., Stoffer S., 2010. Time Series Analysis and Its Applications. Springer.
57. https://blog.octo.com/en/time-series-features-extraction-using-fourier-and-wavelet-transforms-on-ecg-data/ Accessed on 14.12.2020.
58. Soltani S., 2002. On the use of the wavelet decomposition for time series prediction. Neurocomputing.;48(1–4), 267–77.

59. Song Q., 1993. Fuzzy time series and its models Fuzzy Sets and Systems, vol. 54, 269-277.

60. Jiang P., Liu F., Song Y., 2016. Hybrid Multi-Step Model for Forecasting Day-Ahead Electricity Price Based on Optimization, Fuzzy Logic and Model Selection. Energies 9, 618.

61. Elliott G. Rothenberg T., Stock H., 1996. "Efficient Tests for an Autoregressive Unit Root". Econometrica. 64 (4), 813–836.

62. Bhargava A., 1986. "On the Theory of Testing for Unit Roots in Observed Time Series". The Review of Economic Studies. (3), 369–384.

63. Akaike H., 1973. Information theory and an extension of the maximum likelihood principle, 267 – 281.

64. Hong T., 2015. "Crystal Ball Lessons in Predictive Analytics". *EnergyBiz Magazine*. Spring: 35–37.

APPENDIX A

SOURCE CODE

```
import data
dayahead_auction_result = readtable('Data/dayahead_auction_result_new.csv');


dayahead_auction_result.timestamp_dt = datetime(dayahead_auction_result.timestamp,'ConvertFrom', 'epochtime', 'TicksPerSecond', 1e3, 'Format','dd-MMM-yyyy HH:mm:ss');
dayahead_auction_result.issue_time_dt = datetime(dayahead_auction_result.issue_time,'ConvertFrom', 'epochtime', 'TicksPerSecond', 1e3, 'Format','dd-MMM-yyyy HH:mm:ss');


dayahead_auction_result.timestamp_dt = datetime(dayahead_auction_result.timestamp,'ConvertFrom', 'epochtime', 'TicksPerSecond', 1e3, 'Format','dd-MMM-yyyy HH:mm:ss');
dayahead_auction_result.issue_time_dt = datetime(dayahead_auction_result.issue_time,'ConvertFrom', 'epochtime', 'TicksPerSecond', 1e3, 'Format','dd-MMM-yyyy HH:mm:ss');
Clean Data
Prices = timetable(dayahead_auction_result.timestamp_dt,dayahead_auction_result.price);
Prices.Properties.VariableNames("Var1") = "Price";
msVl =  Prices.Price > 90
Prices(msVl, "Price") = num2cell(NaN);
temp = Prices.Price;
% Fill missing data
[cleanedData,missingIndices] = fillmissing(Prices.Price,'linear');


% Visualize results
clf
plot(cleanedData,'Color',[0 114 189]/255,'LineWidth',1.5,...
    'DisplayName','Cleaned data')
hold on


% Plot filled missing entries
plot(find(missingIndices),cleanedData(missingIndices),'.','MarkerSize',12,...
    'Color',[217 83 25]/255,'DisplayName','Filled missing entries')
title('Filled missing entries');
ylabel("Price, €/MWh");
xlabel("Time, hrs");
hold off
legend
clear missingIndices
Prices.Price = cleanedData;
Prices = retime(Prices,'regular',"fillwithmissing", "TimeStep", minutes(15));
Prices = Prices(1:4:end, :);
Prices(1:2,:) = [];
Prices(12793:end, :) = [];
clear cleanedData temp msVl;
Fill in NaNs
for i = 1:height(Prices)
    try
        if (isnan(Prices.Price(i)))
            Prices.Price(i) = Prices.Price(i-168);
        end
    catch
         Prices.Price(i) = Prices.Price(i+168);
    end
end
clear i;


Prices.weekday = weekday(Prices.Time);
wd = Prices.weekday;
workdays = Prices(wd == 2| wd == 3| wd == 4 | wd == 5 | wd == 6 ,:);
clear wd;


workdays_test = workdays(end-119:end,:);
workdays_training1 = workdays(1:end-120,:);
workdays_training2 = workdays(1:end-96,:);
workdays_training3 = workdays(1:end-72,:);
workdays_training4 = workdays(1:end-48,:);
workdays_training5 = workdays(1:end-24,:);
```

```
ylabel("Price, €/MWh");
xlabel("Time, hrs");



mondayAvg = mean(Prices(Prices.weekday == 2,"Price").Price);
tuesdayAvg = mean(Prices(Prices.weekday == 3,"Price").Price);
wednesdayAvg = mean(Prices(Prices.weekday == 4,"Price").Price);
thursdayAvg = mean(Prices(Prices.weekday == 5,"Price").Price);
fridayAvg = mean(Prices(Prices.weekday == 6,"Price").Price);
saturdayAvg = mean(Prices(Prices.weekday == 7,"Price").Price);
sundayAvg = mean(Prices(Prices.weekday == 1,"Price").Price);


bar([mondayAvg tuesdayAvg wednesdayAvg thursdayAvg fridayAvg saturdayAvg sundayAvg], 0.5,
"stacked");
xticklabels(["monday" "tuesday" "wednesday" "thursday" "friday" "saturday" "sunday"]);
ylabel("Price, €/MWh");
xlabel("Weekdays");


plot(workdays.Price);
ylabel("Price, €/MWh");
xlabel("Time, hrs");


plot(movmean(workdays.Price, 168));
ylabel("Price, €/MWh");
xlabel("Time, hrs");


september = workdays.Price(8233:8760,:);
plot(september);
hold on;
plot(movmean(movmean(september,24),4));
hold off;
ylabel("Price, €/MWh");
xlabel("Time, hrs");


workdays_diff = diff(workdays.Price);
plot(workdays_diff);
ylabel("Price, €/MWh");
xlabel("Time, hrs");


september_diff = diff(september);
plot(movmean(september_diff(end-120:end,:), 120));


Res3 = infer(ARIMA43,workdays_training3.Price);
plot(Res3);


ARIMA41 = arima('Constant',NaN,'ARLags',[1, 24,
48,120],'D',1,'MALags',1:2,'Distribution','Gaussian');
ARIMA42 = arima('Constant',NaN,'ARLags',[1, 24,
48,120],'D',1,'MALags',1:2,'Distribution','Gaussian');
ARIMA43 = arima('Constant',NaN,'ARLags',[1, 24,
48,120],'D',1,'MALags',1:2,'Distribution','Gaussian');
ARIMA44 = arima('Constant',NaN,'ARLags',[1, 24,
48,120],'D',1,'MALags',1:2,'Distribution','Gaussian');
ARIMA45 = arima('Constant',NaN,'ARLags',[1, 24,
48,120],'D',1,'MALags',1:2,'Distribution','Gaussian');


ARIMA41 = estimate(ARIMA41,workdays_training1.Price,'Display','off');
ARIMA42 = estimate(ARIMA42,workdays_training2.Price,'Display','off');
ARIMA43 = estimate(ARIMA43,workdays_training3.Price,'Display','off');
ARIMA44 = estimate(ARIMA44,workdays_training4.Price,'Display','off');
ARIMA45 = estimate(ARIMA45,workdays_training5.Price,'Display','off');
[ARIMA4D1,RMSE1] = forecast(ARIMA41,24,workdays_training1.Price);


[ARIMA4D2,RMSE2] = forecast(ARIMA41,24,workdays_training2.Price);
[ARIMA4D3,RMSE3] = forecast(ARIMA42,24,workdays_training3.Price);
[ARIMA4D4,RMSE4] = forecast(ARIMA43,24,workdays_training4.Price);
```

```
[ARIMA4D5,RMSE5] = forecast(ARIMA44,24,workdays_training5.Price);
[days5for, RMSE] = forecast(ARIMA41,120,workdays_training1.Price);
```

```
plot(workdays_test.Price(1:24),'Color',"#77AC30", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold on;
plot(ARIMA4D1,'Color',"#0072BD", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold off;
legend("Actual price","Forecast", "Location", "northoutside");
title("15 october 2020");
xlabel("time,hrs");
ylabel("Price, €/MWh");


plot(workdays_test.Price(25:48),'Color',"#77AC30", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold on;
plot(ARIMA4D2,'Color',"#0072BD", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold off;
legend("Actual price","Forecast", "Location", "northoutside");
title("16 october 2020");
xlabel("time,hrs");
ylabel("Price, €/MWh");


plot(workdays_test.Price(49:72),'Color',"#77AC30", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold on;
plot(ARIMA4D3,'Color',"#0072BD", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold off;
legend("Actual price","Forecast", "Location", "northoutside");
title("17 october 2020");
xlabel("time,hrs");
ylabel("Price, €/MWh");


plot(workdays_test.Price(73:96),'Color',"#77AC30", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold on;
plot(ARIMA4D4,'Color',"#0072BD", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold off;
legend("Actual price","Forecast", "Location", "northoutside");
title("18 october 2020");
xlabel("time,hrs");
ylabel("Price, €/MWh");


plot(workdays_test.Price(97:120),'Color',"#77AC30", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold on;
plot(ARIMA4D5,'Color',"#0072BD", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold off;
legend("Actual price","Forecast", "Location", "northoutside");
title("19 october 2020");
xlabel("time,hrs");
ylabel("Price, €/MWh");


plot(workdays_test.Price(1:120),'Color',"#77AC30", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold on;
plot(days5for,'Color',"#0072BD", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold off;
legend("Actual price","Forecast", "Location", "northoutside");
title("15-19 october 2020");
xlabel("time,hrs");
ylabel("Price, €/MWh");
```

2 ANN DAY 1

```
wd_training_diff =tonndata(diff(workdays_training1.Price),false, false);
net = narnet(1:24,12);
net.trainFcn = 'trainbr';
net.divideFcn = 'divideblock';
[Xs,Xi,Ai,Ts] = preparets(net,{},{},wd_training_diff);
net = train(net,Xs,Ts,Xi,Ai);
```

```
[Y,Xf,Af] = net(Xs,Xi,Ai);
perf = perform(net,Ts,Y)
[netc,Xic,Aic] = closeloop(net,Xf,Af);
NNfrc1day = netc(cell(0,24),Xic,Aic);

Inverse differencing
temp = [workdays_training1.Price(1)  wd_training_diff NNfrc1day];
temp = cell2mat(temp);
temp2 = cumsum(temp);
NNfrc1day = temp2(:, end-119:end);

plot(workdays_test.Price(1:24),'Color',"#77AC30", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold on;
plot(NNfrc1day(:,1:24),'Color',"#0072BD", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold off;
legend("Actual price","Forecast", "Location", "northoutside");
title("15 october 2020");
xlabel("time,hrs");
ylabel("Price, €/MWh");
```

3 ANN DAY 2

```
wd_training_diff =tonndata(diff(workdays_training2.Price),false, false);
net = narnet(1:24,12);
net.trainFcn = 'trainbr';
[Xs,Xi,Ai,Ts] = preparets(net,{},{},wd_training_diff);
net = train(net,Xs,Ts,Xi,Ai);


[Y,Xf,Af] = net(Xs,Xi,Ai);
perf = perform(net,Ts,Y)
[netc,Xic,Aic] = closeloop(net,Xf,Af);
NNfrc2day = netc(cell(0,24),Xic,Aic);

Inverse differencing
temp = [workdays_training2.Price(1)  wd_training_diff NNfrc2day];
temp = cell2mat(temp);
temp2 = cumsum(temp);
NNfrc2day = temp2(:, end-23:end);

plot(workdays_test.Price(25:48),'Color',"#77AC30", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold on;
plot(NNfrc2day,'Color',"#0072BD", 'LineStyle', "-",'Marker',"o", "LineWidth",1.5);
hold off;
legend("Actual price","Forecast", "Location", "northoutside");
title("16 october 2020");
xlabel("time,hrs");
ylabel("Price, €/MWh");
```

ДОДАТОК Б

ВІДГУК

керівника економічного розділу

на кваліфікаційну роботу магістра

на тему: «Моделі, алгоритми та програмне забезпечення для прогнозування цін на електроенергію на основі статистичних методів та штучних нейронних мереж»

студента групи 121м-19-1 Мединського Антона Геннадійовича

Керівник економічного розділу

доцент каф. ПЕП та ПУ, к.е.н. **Л. В. Касьяненко**

APPENDIX C

LIST OF FILES ON THE DISC

File name	Description
Explanatory documents	
ThesisMedynskyiAnton.doc	Explanatory note to the diploma project. Word document.
ThesisMedynskyiAnton.pdf	Explanatory note to the diploma project in PDF format
Program	
Thesis.zip	Archive. Contains program codes and a program
Presentation	
PresentationMedynskyiAnton.ppt	Presentation of the diploma project