SCHOOL OF BUSINESS ADMINISTRATION

DEPARTMENT OF ACCOUNTING AND FINANCE

MASTER OF SCIENCE IN ACCOUNTING AND FINANCE

Thesis

Econometric and Machine Learning Techniques for Electricity Load Forecasting

by

Dipl.-Ing. Apostolos Vavouris

Supervisor: Prof. Achilleas Zapranis

A thesis submitted in fulfilment of the requirements for the degree of MSc in Accounting and Finance

November 2020

Econometric and Machine Learning Techniques for Electricity Load Forecasting

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Professor Achilleas Zapranis for their patient guidance, enthusiastic encouragement and useful critiques of the research work. I would also like to express my very great appreciation to Mr. Bert Lutje Barenbroek and Dr. Dimitrios Doukas from the NET2GRID B.V. company for providing me with the necessary data to perform my analysis, without which this Thesis could not be implemented.

ABSTRACT

Load Forecasting is fundamental for the energy planning sector since a time-ahead power market requires demand-scheduling for power generation, transmission, distribution etc. Forecasting can be performed with different methods, the selection of each method relies on several factors including the quality and the relevance of the available historical data. Also, the methodology used is strictly correlated with the forecast horizon and the level of accuracy of the available data. The time horizon is being adopted taking into account the specific applications in power system planning. Id est, distribution and transmission planning need a short-term horizon while financial or power supply planning require a long-term horizon. In addition, short-term horizon is vital for the hour and day ahead market. Finally, energy forecasting is extremely important for the end consumers as they can get informed about their expected energy consumption and avoid unexpected bill shocks at the end of the year.

Having stated the importance of the load forecasting, the methodology that is followed in order to forecast energy consumption is explained. After having preprocessed the data and having chosen the aggregation level and the time-horizon, the adequate model is being used in order to perform the analysis. Examples of these models are: Regression models, ANNs, SARIMAX (ARIMA family models), ANNs, time series analysis etc. For short-term and high-resolution forecasting, time series analysis and ANNs are preferred while long-term and low-resolution forecasting is mostly done with regression models.

The high-resolution forecasting of a single household is extremely challenging due to the stochastic nature of the appliances' usage. On an aggregated basis since the consumption of several households is added up extreme values and/or abnormalities are reduced to a minimum, making it easier to forecast the consumption on a higher resolution with a smaller error.

Keywords: AMR, Power Grid, Load Forecasting, Neural Networks, Machine Learning

ΠΕΡΙΛΗΨΗ

Η πρόβλεψη της ηλεκτρικής κατανάλωσης είναι απαραίτητη για τον τομέα του σχεδιασμού του δικτύου της ηλεκτρικής ενέργειας, καθώς μια αγορά ενέργειας η οποία βασίζεται στο μελλοντικό προγραμματισμό απαιτεί την ύπαρξη σχεδίου αναφορικά με την παραγωγή της ηλεκτρικής ενέργειας, τη μεταφορά από τις μονάδες παραγωγής στα κέντρα κατανάλωσης και τη διανομή στους τελικούς καταναλωτές. Η πρόβλεψη της κατανάλωσης της ηλεκτρικής ενέργειας μπορεί να πραγματοποιηθεί εφαρμόζοντας πολλές διαφορετικές μεθόδους, η επιλογή τη οποίας κάθε φορά αποφασίζεται βάσει πολλών παραγόντων, με τους κυριότερους εξ αυτών να είναι η ποιότητα και η σχετικότητα των διαθεσίμων ιστορικών στοιχείων. Η επιλογή επίσης της μεθόδου είναι στενά συνδεδεμένη με τον ορίζοντα πρόβλεψης και με το βαθμό ακρίβειας των διαθεσίμων δεδομένων. Ο χρονικός ορίζοντας στον οποίον θα πραγματοποιηθεί η πρόβλεψη ορίζεται βάσει της απαιτούμενης εφαρμογής στον προγραμματισμό του συστήματος. Για παράδειγμα για τη διαχείριση της μεταφοράς και της διανομής της ηλεκτρικής ενέργειας είναι απαραίτητος ένας βραχύς χρονικός ορίζοντας καθώς είναι απαραίτητη η διασφάλιση μη ύπαρξης κορεσμού σε τμήματα ή σε κόμβους του δικτύου κατά τη διάρκεια της ημέρας. Αντίθετα, ο οικονομικός προγραμματισμός για τη λειτουργία ή την απόσυρση νέων ηλεκτροπαραγωγικών μονάδων απαιτεί έναν πιο μακροπρόθεσμο σχεδιασμό. Επιπρόσθετα, η βραχεία πρόβλεψη της κατανάλωσης της ηλεκτρικής ενέργειας είναι απαραίτητη για την ωριαία αγορά και την αγορά επόμενης μέρας, καθώς οι συμμετέχοντες θα επωφελούνταν από τη γνώση του μελλοντικού φορτίου. Τέλος, η πρόβλεψη της κατανάλωσης της ηλεκτρικής ενέργειας είναι σημαντική για τους τελικούς καταναλωτές καθώς μπορούν έτσι να ενημερώνονται για την αναμενόμενη κατανάλωση ηλεκτρικής ενέργειας τους και συνεπώς τους δίνεται η δυνατότητα να προσαρμόζουν τις καταναλωτικές τους συνήθειες αποφεύγοντας το σοκ λογαριασμού στο τέλος του έτους. Η ανάγκη για πρόβλεψη του ηλεκτρικού φορτίου έγινε αισθητή κατά τη διάρκεια της απελευθέρωσης της αγοράς ηλεκτρικής ενέργειας. Και πόσο μάλλον με την επικείμενη αλλαγή στη νομοθεσία η οποία θα επιβάλλει τον ενδοημερήσιο προγραμματισμό και θα επιτρέπει την ενδοημερήσια εμπορία της ηλεκτρικής ενέργειας και των παραγώγων αυτής. Η διαδικασία της απελευθέρωσης της αγοράς της ηλεκτρικής ενέργειας έχει τις βάσεις της στη αρχές της δεκαετίας του ’80. Την περίοδο εκείνη θεωρήθηκε ότι μέσω της απελευθέρωσης

της αγοράς της ηλεκτρικής ενέργειας θα επιτευχθεί μείωση του κόστους και αύξηση των επενδύσεων μέσω του ανταγωνισμού. Από το 1982 έως και σήμερα έχουν δρομολογηθεί αρκετές αλλαγές στην αγορά ενέργειας και σήμερα θεωρείται ότι η αγορά της ενέργειας είναι αρκετά ώριμη για να εφαρμοσθεί το μοντέλο των ενδοημερίσιων συναλλαγών. Ένα μοντέλο το οποίο απαιτεί αρκετή εκ των προτέρων πληροφορία για τους συμμετέχοντες με κυριότερη τη γνώση του μελλοντικού φορτίου. Περισσότερες πληροφορίες αναφορικά με την οργάνωση της αγοράς της ηλεκτρικής ενέργειας παρουσιάζονται στο κεφάλαιο 2. Το κυριότερο χαρακτηριστικό το οποίο δυσχεραίνει την πρόβλεψη του ηλεκτρικού φορτίου και κατ' επέκταση της τιμής της ηλεκτρικής ενέργειας είναι η εξαιρετικά υψηλή μεταβλητότητα της χρονοσειράς. Εκτός της υψηλής μεταβλητότητας η κυματομορφή της ζήτησης χαρακτηρίζεται από έντονη εποχικότητα, τόσο ημερήσια, όσο και εβδομαδιαία και μηνιαία. Η μοντελοποίηση της εποχικότητας μπορεί να επιτευχθεί με διάφορους τρόπους με σημαντικότερη την αποσύνθεση του σήματος σε απλούστερα. Οι μεθοδολογίες εξαγωγής των χαρακτηριστικών από την κυματομορφή παρουσιάζονται στο κεφάλαιο 3. Η επιτυχία των χρησιμοποιούμενων μοντέλων μπορεί να υπολογιστεί μέσω διαφορετικών μετρικών ανάλογα με το βαθμό συγκρισιμότητας που απαιτείται. Οι μέθοδοι αυτοί παρουσιάζονται εκτενώς στο κεφάλαιο 4. Στη βιβλιογραφία έχουν παρουσιαστεί διάφορες μέθοδοι πρόβλεψης της μελλοντικής κατανάλωσης της ηλεκτρικής ενέργειας οι οποίες έχουν διαφορετικό βαθμό πολυπλοκότητας και ακρίβειας αλλά και διαφορετικό πεδίο εφαρμογής. Τα μοντέλα αυτά μπορούν να διακριθούν σε διάφορες κατηγορίες βάσει του χώρου στον οποίο επεξεργάζεται το σήμα – δηλαδή αν η επεξεργασία γίνεται στο σήμα το ίδιο ή στο συχνοτικό του περιεχόμενο –, ή βάσει των χρησιμοποιούμενων μεθόδων – π.χ. τεχνητή νοημοσύνη, εξόρυξη δεδομένων – ή μέσω άλλων χαρακτηριστικών. Τα σημαντικότερα μοντέλα τα οποία είναι διαθέσιμα στη βιβλιογραφία όπως τα στατιστικά μοντέλα (αυτοπαλίνδρομα, κινητού μέσου κτλ) και τα μοντέλα τεχνητής νοημοσύνης και μηχανικής εκμάθησης παρουσιάζονται στο κεφάλαιο 5. Τέλος στο κεφάλαιο 6 παρουσιάζεται η διαδικασία η οποία ακολουθήθηκε και το μοντέλο το οποίο κατασκευάστηκε για την πρόβλεψη της ηλεκτρικής ενέργειας. Χρησιμοποιήθηκαν δεδομένα τα οποία εκτείνονται σε περισσότερα από πέντε έτη με σκοπό την πρόβλεψη της κατανάλωσης της ηλεκτρικής ενέργειας της επόμενης ημέρας σε διαστήματα της μισής ώρας. Τα δεδομένα προεπεξεργάστηκαν και στη συνέχεια τροφοδοτήθηκαν σε ένα νευρωνικό δίκτυο με σκοπό τη δημιουργία ενός μοντέλου το οποίο θα προβλέπει τη μελλοντική κατανάλωση. Τα αποτελέσματα και η ακρίβεια του μοντέλου παρουσιάζονται στο τέλος του κεφαλαίου αυτού ενώ οι μελλοντικές επεκτάσεις και τα συμπεράσματα αναγράφονται στο κεφάλαιο 7.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introductory Concepts

1.1 Power Grids

A Power Grid or an Electrical Grid is a network of interconnected facilities that aims to the delivery of electricity from the producers to consumers. The Power Grid consists of generating stations that produce the required electric power, electrical substations for stepping electrical voltage up for transmission, high voltage transmission lines that enable the transfer of power over long distances - i.e. from sources to demand centers -, electrical substations for stepping electrical voltage down for distribution and finally the distribution lines that connect individual customers to the grid. These grids may cover a single building, or a whole country (national grid) or even across continents (transnational grid). These grids are required because power stations are situated near energy resources - e.g. coal - or in places to take advantage of the renewable energy sources - e.g. solar insolation -, whereas consumers are mostly situated in metropolitan areas away from the production.

In order for the Power Grids to operate correctly and avoid Brownouts, Blackouts, Load shedding and Black starts, a great level of synchronization between the production and the consumption of electrical power is required. In other words the energy that consumers are willing to absorb must be fed simultaneously by the producers into the grid. The electrical power removed by consumers from the grid is known as demand or load. This demand is matched by the power suppliers with the implementation of electricity generators that constantly feed a minimum load into the network over any given period, named baseload, and with the use of peaking plants that are generators of last resort to cover peak demands. The plants that generate the baseload are high capital intensity investments which need several days in order to reach their peak production capacity but produce energy at low cost, whereas peaking plants are lower capital intensity investments which are optimized to come on-line quickly but their energy production cost is extremely higher than the baseload equivalent. It is therefore crucial to the power grid operators to be able to cover the power required using the baseload plants and avoiding putting the peaking plants on-line. Because the baseload plants cannot easily adjust their power production, an almost steady power consumption needs to be established. However this is not easily implemented due to the fact that energy consumption is higher early in the morning and late in the evening due to the nature of daily human activities. This abnormality was initially proposed to be covered by the use of the renewable energy sources, however as is going to be explained in the next subsection this is not possible.

1.1.1 The Duck Curve

The Duck Curve is a graph of power production over the course of a day that shows the timing imbalance between peak demand and renewable energy production. In many grids the peak demand occurs after sunset or before sunrise, when there is no solar insolation. Without any form of energy storage, power grid operators must rapidly increase power output around the time of sunset to compensate for the loss of solar generation. Storage may fix this issue but it can not always be implemented, as it is bound with high construction costs.

1.1.2 Smart Grids

In order to enhance the Power Grids, two-way communications and distributed intelligent devices are being implemented. These devices enable advanced information metering, monitoring and management of the grid, aiming to improve energy efficiency, demand profile, utility, cost and emission based on the infrastructure by using optimization, machine learning and game theory.

1.2 Electricity Load Metering

Power grid operators use electric meters that are installed at customers' premises for monitoring and billing purposes. These meters are most commonly calibrated in billing units, i.e. kilowatt hour (kWh) and are usually read once in each billing period which can most commonly be monthly, quarterly or yearly.

1.2.1 Automatic Meter Reading (AMR)

Automatic Meter Reading is a relative new technology that allows the power grid operators to automatically collect consumption, diagnostic and status data from energy metering devices (electric and gas) and transfer them to a central database for billing, troubleshooting and analyzing purposes. This technology which is currently being implemented in a vast majority of countries (e.g. The Netherlands, Denmark, Sweden, Italy etc) enables the utility providers to limit the expenses that are correlated with the periodic trips to each physical location to read each meter. Using the AMR devices, billing can be based on near real-time consumption rather than on estimates based on past or predicted consumption, thus making billing more accurate. Using this almost real-time information with the help of data analysis enables utility providers and customers to have a better control over the use and the prediction of the energy consumption.

1.3 Demand Response and Demand-Side Management

Demand Response is an alteration in power consumption of an electrical utility customer to match the demand with supply. Utilities have traditionally balanced demand and supply by throttling the production rate of their generators, taking power plants on or off line, or importing power from other utilities. However, due to technical limitations these extra power generators may take long to come up to full power, or other plants may be very expensive to operate or it is even possible that the capacity of all the available power plants be lower than the demand. Demand-Side Management (DSM) also known as Load Management (LM) is the process of balancing the supply of electricity on a grid with the consumption by adjusting or controlling the load rather than the output of the generators. This can be performed either real-time, by the use of frequency sensitive relays that trigger circuit breaks and time clocks or by using special tariffs to influence customers to change their consumption behavior. DSM enables utilities to reduce costs that are associated with the peaking power plants.

CHAPTER 2

Complex Energy Markets

Energy markets are the commodity markets that deal specifically with the trade and supply of energy. Most commonly Energy markets refer to an electricity market but other sources of energy are not excluded. Energy market development come as a result of the energy policies that a government implements. Nowadays these policies promote the development of a highly versatile energy industry in a competitive manner, whereas earlier energy markets where characterized by monopoly-based organizational structures with the most widely-known example that of the Seven Sisters (Exxon, Mobil, Chevron, Gulf Oil, Texaco, The British Petroleum and Shell) who controlled the majority of the world's petroleum reserves.

Nowadays, more and more countries are promoting the liberalization and the regulation of the Energy markets. These markets are strictly regulated by national and international authorities to protect consumer rights and avoid oligopolies. Examples of these authorities are the Energy Community in Europe, and the Australian Energy Market Commission in Australia. The target of the regulating authorities is from one hand to maintain prices volatility in a narrow range in order to protect the consumers from speculators, and from the other hand to discourage anti-competitive behavior such as the formation of monopolies or large conglomerates.

2.1 Electricity Markets

An electricity market is an organized market that enables its participants to purchase - through bids to buy -, to sale - through offers to sell - and to trade in a short-term in the form of financial or obligation swaps of electricity. In this context and using economic terms electricity is a commodity that it can be sold, bought and traded. The price is determined based on the supply and demand principle. Apart from short-term trading, counter-parties may agree on bi-lateral long term trade contracts - similar to Power Purchase Agreements -.

The commodities that are traded in an electricity market are Power and Energy. Power is the amount of energy that is transferred or converted per unit of time and is expressed in Watts (Joules per second). Energy is electricity that flows through a cross section for a

given period and is measured in Joules or Watthours (Joules per Second times Second). Bid and offer transactions are cleared and settled by an independent entity charged with this function such as a market operator, however market operators while clearing trades must simultaneously maintain generation and load balance in order to avoid sudden power losses. Most commonly the trade intervals are in increments of 5, 15 and 60 minutes.
Apart from trading market operators in order to ensure redundancy and reliability of grid, ancillary services are implemented such as operating reserves, responsive reserves etc. Lastly, market operators, may be charged with the organization of electricity derivatives trading such as futures and options on power and energy.

2.2 Electricity Market Liberalization

Power market liberalization was first performed by Chile. This procedure began in 1982 and its aim was to separate the the companies that generate and the companies that distribute electricity. A formula was created, which determined the price paid to each part based on a marginal cost pricing and on the customers requirements. In 1986 in Chile a large-scale privatization efforts began which led to the creation of a wholesale power trading mechanism.
After Chile a vast majority of countries (including all the EU states) have started liberalizing their markets, each on with its own characteristics. The reason for this liberalization is the belief that the same success of liberalization in other industries can be replicated in the electricity sector as well as that vertically integrated monopoly structures (i.e. generation, transportation and distribution of the energy) are going to be vanished. Another fact that led to the introduction of competition in power markets is the fact that due to technological improvements in generation and transmission systems, a once natural monopoly can nowadays be replaced by a market that will promote efficiency gains and stimulate technical innovation.
Countries that followed Chile in market liberalization where Britain in 1990 (only in England and Wales until 2005, thereafter Scotland was added to the scheme), Norway with the Nordic market in 1992 (later Sweden, Denmark and Finland participated in this scheme) etc. Currently European countries participate in a short of market liberalization scheme. In Australia two markets were firstly created in Victoria and New South Wales in 1994 which have been replace by the Australian National Electricity Market (NEM) in 1998.

In Americas market liberalization began from the northeastern states in the late 1990s and then continued to the south of the country, but not always with a success (e.g. California market crash of 2000-2001). During the California market crash of 2000-2001 market liberalization was highly criticized because it led to a situation in which the state suffered from multiple large-scale blackouts and from the collapse of the state's largest energy companies. However California crisis was not the result of market liberalization but a coincidence of several factors including a design flaw in the market.

California's electricity crisis is a failure which should not overshadow the positive impact of the market liberalization. An energy market has a great impact in the economy. From the introduction of the energy markets due to the competition energy prices have significantly dropped and the assets of the electricity sector are more efficiently used. In addition, contrary to vertical monopolies - that traditionally tended to create substantial overcapacity - in a liberal market overcapacity is greatly reduced and the efficiency of the used equipment for the generation, transmission and distribution is augmented. However, what must be stated here is that due to the fact that new investments in generation plants are not centrally planned but are programmed individually, technologies that are capital-intensive and/or require a long construction time are avoided and instead quick-built production plants are preferred. Furthermore, in an environment of low energy prices the participants are not willing to invest to expand the transmission network or augment the generation capacity.

This problem was somewhat solved with the method of the capacity payments. A capacity payments is a compensation to each energy producer based on its installed available capacity, regardless the power that feed to the system. However, due to this measure, energy producers instead of investing in new plants, decided to make fewer capacity resources available in order to increase the capacity payments. This behavior lead to energy shortage. In order to bypass this misbehavior markets where transformed to installed capacity markets (ICAP). With the introduction of these markets adequate capacity is committed on a daily or seasonal basis and therefore system load and reserve requirements are reassured. Each energy distributor that sells electricity is obliged to satisfy its capacity obligations i.e. their predicted peak loads plus a margin. Show in case of an electricity shortage system operators are able to recall extra generators to feed the system.

2.3 Electricity Marketplace

2.3.1 Power Pools and Power Exchanges

Wholesale markets may take two different forms, the first is a power pool and the second is a power exchange. These two forms of factors have many things in common and many times the two terms are wrongly used interchangeably. The most common misconception is the Nord Pool, which although it is called a pool it is actually a power exchange.

2.3.1.1 Power Pools

Two types of power pools can be distinguished, the generation pools and the economic pools also called power pools.
Generation pools are used in order to optimize the generation with respect to cost minimization and optimal technical dispatch. In generation pools all plants are ranked on a merit order, based on their estimated costs of production. The plants that are decided to produce energy are these that satisfy the network constraints and the generation costs constrains.
On the other hand economic pools are created in order to make it easier for competitors that produce energy to participate to the market. These pools are created by governments that are willing to liberalize the energy production, and all the producers are obliged to participate in that pool, i.e. trade outside the pool is not permitted. In these pools a supply curve is being made based on the bid prices that the generators are willing to run their plants. The market clearing price (MCP) is determined by intersecting the estimated demand (market clearing volume - MCV). However due to the complexity of the price determination method economic pools have been criticized for providing a low level of transparency.

2.3.1.2 Power Exchange

Power exchanges is the other type of organized markets. These markets are commonly founded on a private initiative with the participation of producers, distributors, large consumers and traders. This type of market is gaining ground especially in the developed European Markets (e.g. Germany, France, Austria etc.). In a power exchange the price is determined by matching the supply and the demand curves. In this way the market clearing price (MCP) is determined and is publicly announced. In order for the exchange to

work, each day suppliers and consumers (end-users, traders etc.) submit offers and bids for each hour of the next day. These values are then plotted on augmenting diagrams and from the intersection between the supply and the demand curve an hourly MCP is set.
Power exchange work with two different models, as a marginal auction and as a discriminatory auction. In a marginal auction all suppliers are paid the same amount of money for each MWh regardless of his initial bid. In contrast, in a discriminatory auction, also known as a pay-as-bid auction, a supplier is paid exactly the amount of his bid. This form of exchange creates a problem of 'extra money' that is paid by buyers, where as in a marginal auction money paid by buyers equals to the money received by the producers. Currently, most mature electricity markets have adopted the uniform-price auction.

2.3.2 Nodal and Zonal Pricing

The MCP is the only price for the whole system when on every node and zone of the system no transmission congestion exists. However, due to the nature of the electricity and the finite transferability of electricity two prices, the locational marginal price (LMP) or the zonal market clearing price (ZMCP) are employed. LMPs take into account the marginal cost of the producers, the congestion costs of the transmission and the marginal losses costs. These costs may be different even in a local area should different nodes exist. Nodal pricing is the most accurate way of pricing electricity because all the costs related from the production plant to the end-user are taken into account. However, this pricing mechanism is extremely complex and as a matter of a fact the transaction costs are highly augmented.
A simpler and widely used pricing model is that of zonal pricing. In that system different zones may have different prices, but within the same zone prices are stable. What should be pointed is that zonal prices contrary to other financial or commodity prices, may become negative. European network, which is characterized of great complexity, is evolving into a zonal market, where each country is a single zone.

2.3.3 Spot and Balancing Market

In contrast to financial assets and most commodities where a spot market is a market for immediate delivery and financial settlement up to two business days later, spot electricity market is a day-ahead market. A spot market for immediate delivery would not be possible for the electricity markets since the transmission and distribution system operators (TSOs

and DSOs) need a prior notice of the amount of the energy that is going to be transferred in order to reassure the ability of the network to withstand that load. However, currently electricity markets are slowly changing in a 15-minutes ahead market.
Apart from the spot market a balancing market is required for the viability of the system. TSOs are able to call in extra production at very short notice, in order to correct deviations between the supply and the demand. This deviation must be corrected in a matter of minutes or even seconds because otherwise brown-outs or even black-outs may occur.

2.3.4 Traded Products

As electricity is a commodity, electricity contracts can either be sold in over-the-counter (OTC) transaction or on organized market as the power exchanges. These contracts can either be contracts for power delivery, or contracts intended to hedge or speculate over electricity price. As with any commodity contract, electricity contracts have: a delivery period, a delivery location, the quantity of the delivered energy and the price of the delivered energy. In contrast to other commodities such as petrol or gas that can be stored, electricity cannot be easily stored, a fact that leads to the need of an always balanced market. This balance is achieved by physical contracts that their maturity span from next quarter of an hour to several years. Short term contracts are most commonly traded in organized exchanges whereas monthly or annual contracts that are used by distributors are typically negotiated on bilateral basis, although even more power exchanges offer futures and options products on electricity. The reason behind this phenomenon is that the long-term electricity derivatives markets offers a relatively low liquidity and hence market demand cannot be covered. Before the liberalization of the markets a type of long term contract that is still used was created. These contracts where the Power Purchase Agreements (PPAs). Apart from the PPAs a long-term contract that is highly correlated with the electricity price is the CO_2 emissions allowance contract.

2.3.4.1 Power Purchase Aggreements

A PPA is a contract between two parties, that agree to produce and buy a certain amount of electricity. Theses parties are the one that generates electricity (seller) and the one that is willing to purchase the produced electricity (buyer). These agreements define all of the commercial terms for this exchange including the time line of the project, the delivery schedule, the penalties for under delivery as well as the terms regarding the payment.

According to the requirements of the two parties different forms of PPAs exist as of today. The rates are agreed between the two contracting parties given the fact that no regulated environment exist. In case of a regulated environment - which can be found in many EU countries - the price is being regulated by an Electricity Regulator.

2.3.4.2 Carbon dioxide emissions allowances

Each allowance gives the holder the right to emit: one tonne of carbon dioxide, which is the most common greenhouse gas, or the equivalent amount of two more powerful greenhouse gases, the nitrous oxide an the perfluorocarbons. These allowances are traded via emissions trading systems (e.g. EU ETS) and are a cornerstones of the EU's policy to combat climate change by reducing the greenhouse gases emissions. The EU ETS, being the first and biggest carbon market, works on the "cap and price" principle. The total amount of certain greenhouse gases that can be emitted by installations covered by the system is capped. The cap is then reduced over time in order to reduce the total emissions. Within the cap, different participants receive or buy emission allowances, which they can trade with on another as needed. After each year, a company must surrender enough allowances to cover all of its emissions, otherwise fines are imposed. If a company has a surplus of allowances it is possible to keep them for future usage or sell them to another company that is short of allowances. These contracts are greatly influencing electricity prices, because the amount of energy as well as the method of the production are affected by these limitations.

CHAPTER 3

Electricity Load and Price Characteristics

In this chapter characteristics of the electricity load are going to be presented. Certain features of electricity loads differ dramatically from those found in the financial or in other commodity markets. By studying these characteristics, the literature survey as well as the proposed methodology that is presented in the next chapters will be easily understandable. These characteristics are observed and measured via specific statistical tools. These tools perform analysis on time series and the main advantage in contrast to financial data sets is the fact that electricity load is sampled 24 hours a day, 7 days a week, 365 days a year.

3.1 Load and Price Spikes

A characteristic that can in a few seconds evaporate the earnings of a company for the whole year are the load and price spikes. Load and as a consequence price of electricity is extremely volatile, far more volatile than any other commodities that is characterized for extreme volatility (table 3.1). Another feature of electricity is that the spikes inten-

Table 3.1: Volatility comparison

	Volatility
Treasury bills and notes	<0.5%
Stock indices	1-1.5%
Commodities (e.g. crude oil, natural gas)	1.5-4%
High volatility stocks	<4%
Electricity load and price	up to 50%

sity is non-homogeneous in time. Load spikes are observed mostly during business days between 09:00 and 18:00. During a load peak the spot price may increase up to 10 or even 100 times. Although these spikes have a very short duration the impact to the system is extreme and if no measures are taken it may even lead to companies' bankruptcies. A very notorious example of a company that underestimated the importance of load and price peaks was the Power Company of America in 1998. In June of 1998, a combination of prolonged hot spell and power outages caused the OTC price of power in the Cinergy region of USA to climb to the astronomical sum of 7,500USD/MWh when energy typically negotiated in the area of 20 to 40USD/MWh. In European markets due to stricter

regulation, although a load spike may be observed price spikes are rarely observed.
As it was stated in the previous chapters power markets work on a simple principle, the supply-demand equilibrium. In figure 3.1 a visualization of the intersection between supply and demand curve is presented. The supply stack curve is the ranking of all generation units in an order of merit, based on their price per MWh. The operator usually firstly dispatch renewable energy sources, hydro plants, nuclear power source and as a least resort coal plants. These resources tend to cover the so-called base load. These resources usually have a slow respond time and therefore must be continuously operated in order to provide system with electricity. These resources as it can be seen have an almost steady price and the gradient is very small. However, when more power is needed into the system,costlier generators are added to the network. These sources may be steam turbines or gas turbines and in the last resort internal combustion engines. These generators however have a significant higher cost of operation and as it can be seen from the figure 3.1 their price gradient is very steep. This means that for a very small change in the demand curve the price of the electricity may skyrocket. Should a grid be correctly designed almost always base load can cover the majority of the energy demand. However, in cases of extreme weather phenomena can cause sudden shocks in the demand and as a matter of a fact significant raises in electricity price. That's why it is very crucial to be able to forecast future energy consumption, in order to be able to create in time new base loads. A price spike may also occur when the so-called cheap generators are withdrawn from the market for maintenance, or due to fluctuations of their fuel prices etc. However, this is not the only reason for the so high price fluctuations. Another reason is that some market participants are willing to pay almost any price in order to secure a sufficient and continuous power supply. I.e. some market participants place extremely high bids in order to reassure that they will be able to get the energy that they require.

3.2 Electricity Load and Price Seasonality

By observing a time-series of electricity load one can clearly see that the electricity demand exhibits seasonal fluctuations. This fluctuation is mostly due to the different climate conditions (temperature, humidity, daylight hours etc.) that exist in each season. Pilipovic [1] proposed that in order to model the seasonality of the electricity a sinusoidal model should be used.

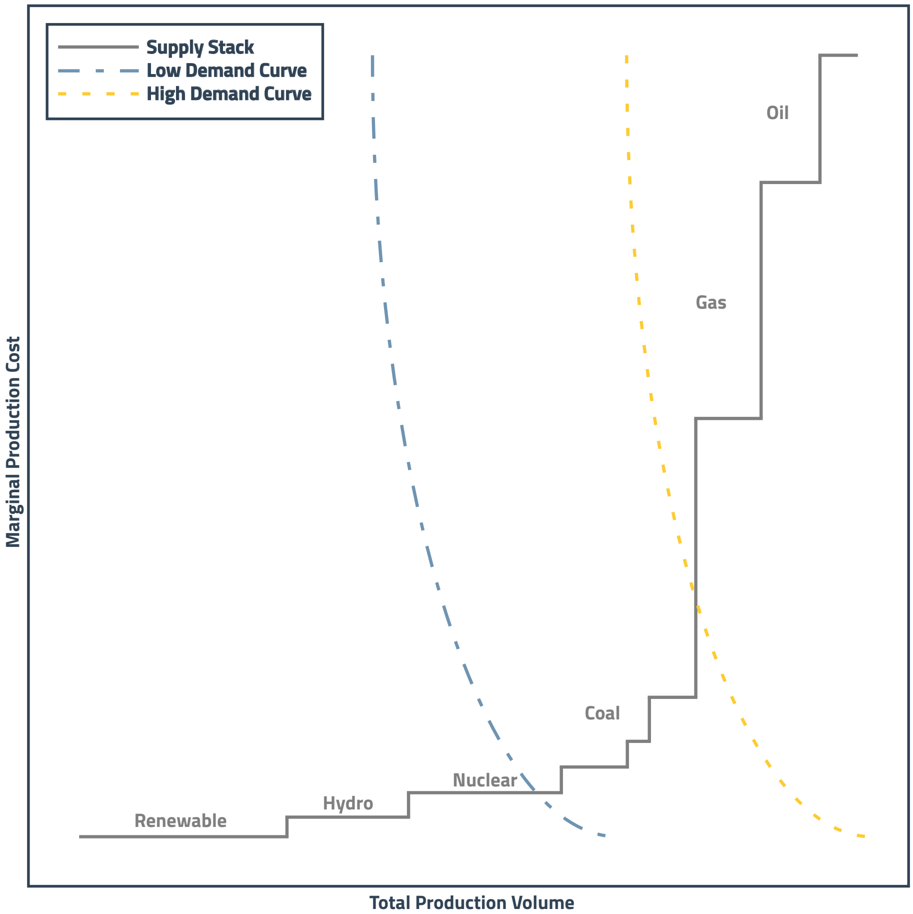

Figure 3.1: Supply stack

However apart from the seasonality that occurs due to the different climatic conditions during a year, electricity load fluctuate also on a daily and weekly basis. This fluctuation in demand and as a matter of a fact in price is due to the human activities. I.e. during the business day and most commonly during the time slots between 09:00 and 18:00 a higher energy demand is observed. Contrary, during weekends or during night the demand is much lower. The same sinusoidal approach can be also used in order to model the intra-week and the intra-day seasonality. What is really important before any further step is to test for a given sample whether a seasonal behavior exists and if exists it should be considered when creating the model. From a mathematical point of view the seasonality can be observed by measuring the serial correlation or by performing a Fourier decomposition that will provide us information for the frequency content of the signal.

3.2.1 Electricity Load Serial Correlation

Given a data-set that consists of the observations: $\{x_1, x_2, \ldots, x_n\}$ the dependency between them can be observed with the use of the sample autocorrelation function plot (ACF).

ACF plot can be given from the following equation:

$$ACF(h) = \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \tag{3.1}$$

where h are the time lags and $\hat{\gamma}$ is the sample autocovariance function (ACVF) which is given by:

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x}) \tag{3.2}$$

where $\bar{x}$ is the mean value of the data-set given by:

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \tag{3.3}$$

using the ACF function and given that a phenomenon is completely random the autocorrelations should be near zero independent of the time-lag. Otherwise, if the phenomenon is not random some autocorrelations will be statistically significant non-zero. It should be stated here that if the first element is closely related to the second element and the second element to the third element then the first element is also related to the third. Regarding electricity price a strong 7-day dependence exist which in contrast to other financial data-sets that the autocorrelation of returns falls into the confidence interval for AWGN after 10-20 days. Apart from using the ACF, the partial autocorrelation function (PACF) is also used. PACF is given by:

$$PACF(h) = \begin{cases} 1 & \text{for } h = 0 \\ \hat{\phi}_{hh} & \text{for } h \geq 1 \end{cases} \tag{3.4}$$

where $\hat{\phi}_{tt}$ derives from the last component of:

$$\hat{\Phi}_h = \hat{\Gamma}_h^{-1}[\hat{\gamma}(1), \hat{\gamma}(2), ..., \hat{\gamma}(h)]' \tag{3.5}$$

where:

$$\hat{\Gamma}_h = [\hat{\gamma}(i-j)]_{i,j=1}^{h} \tag{3.6}$$

The PACF at lag h simple measures the linear association between the $t+h$ and t observation, taking into account all the effects of the middle terms, i.e. $x_{t+1}, ..., x_{t+h-1}$. As with the ACF should there be no dependence between the values then approximately 95% of the ample partial autocorrelations should fall between the bounds $\pm 1.96/\sqrt{n}$. Should value lie outside these bounds then a dependency between the data exists.

3.2.2 Spectrum Domain Analysis

The transformation of a signal from the time domain to the spectrum domain enables us to observe its periodicity. For a data-set with values $x_1, x_2, ..., x_n$ the discrete spectral density, also known as periodogram is given by:

$$I_n(\omega_k) = \frac{1}{n}\left|\sum_{t=1}^{n} x_t e^{-i(t-1)\omega_k}\right|^2 \tag{3.7}$$

where:

$$\omega_k = 2\pi\frac{k}{n} \tag{3.8}$$

are the Fourier frequencies in radians per unit of time, with $k = 1, 2, ..., \lceil n/2 \rceil$. Using the Fourier frequencies the model can then be rewritten as:

$$x_t = \alpha_0 + \sum_{k=1}^{\lceil n/2 \rceil} \{\alpha_k cos(\omega_k t) + \beta_k sin(\omega_k t)\} \tag{3.9}$$

The cosine and sine parameters are the regression coefficients that indicate the degree to which the respective functions are correlated with the data.
Due to the fact that the transformation of a signal from the time domain to the spectral domain is computationally complex, a signal length that is a power of 2 is optimal because then the Fast Fourier Transform (FFT) can be utilized, which can produce results very quickly.

3.3 Modeling Seasonality

After having locate and understand the signal's seasonalities the next step is to decompose them and remove them from the waveform. By performing a seasonal decomposition, using the Census I method, three different components are going to be produced.

- the trend component T_t
- the seasonal component S_t
- and the stochastic component Y_t also known as the error

3.3.1 Differencing

In order to remove trend and seasonal component a time-series can be differentiated. By differencing some autocorrelations are removed were others may become more imminent.

Also, by taking the difference of a time-series the dataset may become stationary which is necessary for further modeling. Different trends can be eliminated taking a different difference. In general an mth-order polynomial trend can be eliminated by differencing m times at lag 1. The transformed series has the following form:

$$y_t = \nabla_h x_t = (1 - B^h)x_t = x_t - x_{t-h} \tag{3.10}$$

where h is the lag. If the time-series consists of hourly data then a common approach is to use differencing at various lags (typically 1, 24 and 168 hours - i.e. 7 days-). The stochastic component will be given by:

$$Y_t = x_t - \left(\frac{1}{N} \sum_{i=1}^{N} x_{t-i\cdot 168} + \frac{1}{7} \sum_{j=1}^{7} x_{t-j\cdot 24} - \frac{1}{7N} \sum_{i=1}^{N} \sum_{j=1}^{7} x_{t-i\cdot 168-j\cdot 24} \right) \tag{3.11}$$

If data is of higher or lower granularity a similar expression can be produced.

3.3.2 Median Method

If a more simple is required then the vector of the medians can be constructed. The advantage of the median compared to the mean is that the median is much less influenced by the outliers. The produced vector can then be subtracted from the original time-series. Using this method the produced vector is assumed to be an estimate of the seasonal component and can lead to the calculation of the stochastic component.

3.3.3 Moving Average Method

Given a time-series of values $\{x_1, x_2, ..., x_n\}$, a moving average filter is applied in order to eliminate the seasonal component. The moving average filter can be calculated from:

$$\hat{m}_t = \frac{1}{h}(x_{t-\lfloor h/2 \rfloor} + ... + x_{t+\lfloor h/2 \rfloor}) \tag{3.12}$$

where h is the length of the moving average filter. Next the average w_k of the deviations is computed. The deviations must comply with the following form:

$$\{(x_{k+h\cdot j} - \hat{m}_{k+h\cdot j}), \lfloor h/2 \rfloor < k + h \cdot j \leq n - \lfloor h/2 \rfloor\} \tag{3.13}$$

After having calculated the average of the deviations the seasonal component can be calculated:

$$\hat{s}_k = w_k - \frac{1}{h} \sum_{i=1}^{h} w_i \tag{3.14}$$

The data after having remove the seasonality component is given by:

$$y_t = x_t - \hat{s}_t \tag{3.15}$$

where $t = 1, 2, ..., n$.

3.3.4 Removing Annual Seasonality

After having removed the daily and the weekly seasonality, annual seasonality must be removed too. As it was stated in the previous sections the annual seasonality can be removed by fitting a sinusoidal signal of an one-year period. The function of the sinusoid is given by:

$$S_t = Asin\left(\frac{2\pi}{365 \cdot i}(t+B)\right) + C_t \tag{3.16}$$

where i is the granularity of the data (e.g. for daily data $i = 1$, for hourly data $i = 24$ etc.) and parameters A, B and C are obtained using a least squares fit [2].

3.3.5 Rolling Volatility Method

Weron et al. in 2001 [3] proposed a method in order to overcome the problem that most energy data-sets covered only a few years and as a matter of a fact the previous techniques could not be applied for the annual seasonality. Rolling volatility can be given by:

$$v_t = \sqrt{\frac{1}{h-1}\sum_{i=0}^{h-1}(R_{t+i} - \bar{R}_t)^2} \tag{3.17}$$

where:

$$\bar{R}_t = \frac{1}{h}\sum_{i=0}^{h-1} R_{t+i} \tag{3.18}$$

and h is the period of rolling volatility calculation. The next step is to calculate the average volatility fo one year:

$$\bar{v} = \frac{v_t^{1styear} + v_t^{2ndyear} + ... + v_t^{nthyear}}{n} \tag{3.19}$$

Then a smoothing is applied taking a h-day moving average over the volatility. The last step is to rescale the output by dividing it with the smoothed annual volatility.

3.3.6 Wavelet Decomposition Method

Wavelet decomposition method is widely used in signal processing as a complement to the Fourier transform. In contrast to the Fourier transform that projects the signal onto an orthonormal set of trigonometric components in wavelet decomposition the signal is projected onto wavelets. Because wavelet decomposition is highly complex, this method would not be used if the Discrete Wavelet Transformation (DWT) was not invented. The family of wavelets that is mostly used is the Daubechies family [4].

The process of wavelet decomposition is going to be explained. The signal is decomposed with the use of a father ϕ wavelet and a sequence of mother ψ wavelets. The signal is expressed:

$$f(t) = S_J + D_j + D_{j-1} + \dots + D_1 \tag{3.20}$$

where:

$$S_J = \sum_k s_{J,k} \phi_{J,k}(t) \tag{3.21}$$

and

$$D_J = \sum_k d_{J,k} \psi_{J,k}(t) \tag{3.22}$$

the wavelet transform coefficients can be calculated as follows:

$$\phi_{J,k}(t) = 2^{-J/2} \phi \left(\frac{t - 2^J k}{2^J} \right) \tag{3.23}$$

and

$$\psi_{J,k}(t) = 2^{-J/2} \psi \left(\frac{t - 2^J k}{2^J} \right) \tag{3.24}$$

These coefficients denote the contribution of the corresponding wavelet function to the sum. After decomposing the signal, the signal can be decomposed only with the parts that are required and therefore have a denoised signal.

CHAPTER 4

Models Evaluation

Model evaluation is mostly done using different rules that are called estimators. Estimators can be grouped in two large categories, point and interval estimators. Point estimators produce a single value for each estimand (i.e. the quantity of interest), or a single vector that can be expressed by using a single function. In contrary interval estimators, produce a range of plausible values, these values may be vectors and/or functions.

4.1 Statistical Estimator

In the field of statistics an Estimator or a Point Estimate, is used to infer the value of a parameter that it is not known. Given the fact that the value that is being estimated is not known the estimator is itself a random variable and as a matter of a fact a function of the data. The evaluation of different estimators can be done by judging their properties, such as their consistency, their asymptotic distribution, their unbiasedness etc. Based on the algebra of random variables, which provides rules for the symbolic manipulation of random variables, a random variable that corresponds to the observed data is denoted as X and the estimator, which is itself treated as a random variable, is denoted as a function of that random variable $\hat{\theta}(X)$. Given these symbols the estimate of a particular observed data $X = x$, denoted as $\hat{\theta}(x)$ is a fixed value. estimate, for simplicity, is sometimes abbreviated and thus denoted as $\hat{\theta}$, which based on the context must not be confused as a random variable.

4.2 Model Evaluation Indexes

The evaluation of the literature's models performance is being done based on certain criteria. Some of these criteria, such as Error, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), the Root Mean Square Error(RMSE) etc., are being presented in the following paragraphs.

4.2.1 Error

Given a sample of a dataset x, the Error of the estimator of the sample $\hat{\theta}$ is defined as:

$$e(x) = \hat{\theta}(x) - \theta \tag{4.1}$$

where θ is the parameter that is being estimated. The drawback of this method is that the error is correlated not only with the procedure that was followed to estimate the model, but also with the sample. That means that the using this metric it is not possible to compare the accuracy of two different models that are used on two different datasets.

4.2.2 Mean Squared Error (MSE)

The Mean Squared Error (MSE) commonly known as Mean Squared Deviation (MSD) is a measurement of the average of the squares of the errors of a procedure that estimates an unobserved quantity. Given the fact that Error is squared, MSE takes always non-negative values, and values that are closer to zero are better. In relation with the shape of the function's graph the MSE is the second moment of the error, making thus possible to incorporate both the variance of the estimator (i.e. how widely are the estimates form one data sample to another spread) and its bias (i.e. the distance between the average estimated value and the true value). The n-th moment of a real-valued continuous function of a real variable about a value c is given by:

$$\mu_n = \int_{-\infty}^{+\infty} (x-c)^n f(x) dx \tag{4.2}$$

By calculating the square root of the MSE one can produce another metric called Root-Mean-Square Deviation that is going to be presented in the next sections. The MSE of an estimator $\hat{\theta}$ with respect to an unknown parameter θ is defined as:

$$MSE(\hat{\theta}) = E_\theta\left[(\hat{\theta} - \theta)^2\right]. \tag{4.3}$$

Another interpretation of the MSE is as the sum of the variance and the bias of the estimator.

$$MSE(\hat{\theta}) = Var_\theta(\hat{\theta}) + Bias(\hat{\theta}, \theta)^2 \tag{4.4}$$

This form of MSE enables as to calculate this metric simply by calculating the variance given the fact that the estimator is unbiased.

4.2.3 Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) or Root Mean Square Deviation (RMSD) is used to determine the deviation between the estimated values and the values observed. The RMSE is simple the square root of the MSE that was described in the previous section. RMSE can be used to measure the prediction accuracy of different models an a specific dataset. Comparison cannot be done between different datasets due to the fact that RMSE is scale-dependent. The values of RMSE are always non-negative, and generally the lower the value of the RMSE the better the fit to the data. Because errors in RMSE are squared, larger errors have a disproportionately larger effect on the metric. RMSE is given by:

$$RMSE(\hat{\theta}) = \sqrt{MSE(\hat{\theta})} = \sqrt{E_{\theta}\left[(\hat{\theta}-\theta)^2\right]}. \tag{4.5}$$

If the estimator is unbiased, the RMSE is simply the root of the variance, i.e. the standard deviation.

4.2.3.1 Normalized Root Mean Square Error

In order to overcome the difficulty of comparing the accuracy of models on different datasets, it is proposed in literature several methods of normalization. The most widely used methods are represented:

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \tag{4.6}$$

where y_{min} and y_{max} represent the range of the dataset.

$$NRMSE = \frac{RMSE}{\bar{y}} \tag{4.7}$$

where $\bar{y}$ is the mean value.

$$RMSEIQR = \frac{RMSE}{IQR} \tag{4.8}$$

where $IQR = Q_3 - Q_1 = CDF^{-1}(0.25) - CDF^{-1}(0.75)$ with CDF^{-1} is the quantile function.

4.2.4 Mean Absolute Percentage Error

The mean absolute percentage error (MAPE), also known as Mean Absolute Percentage Deviation (MAPD), is used to predict forecasting methods' accuracy. MAPE can be given by the following formula:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\theta_t - \hat{\theta}_t}{\theta_t}\right| \tag{4.9}$$

where θ_t is the actual value and $\hat{\theta}_t$ is the forecast value. Most commonly, MAPE is calculated as a percentage by simply multiplying the previous value by 100%.

4.2.5 Symmetric Mean Absolute Percentage Error

The Symmetric Mean Absolute Percentage Error (SMAPE) estimates the accuracy of an estimator based on relative errors. SMAPE is given by:

$$SMAPE = \frac{100\%}{n} \sum_{t=1}^{n} \frac{|\hat{\theta}_t - \theta_t|}{(|\theta_t| + \widehat{|\theta_t|})/2} \tag{4.10}$$

In literature a similar formula was implemented by Armstrong [5].

$$SMAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|\hat{\theta}_t - \theta_t|}{(\theta_t + \widehat{\theta_t})/2} \tag{4.11}$$

This form however is not widely used because it can result in a negative value or even in an undefined value should the denumerator was zero. Also, there is a symmetry problem, i.e. over and under forecasting are not treated equally.

4.2.6 Mean Absolute Error

The Mean Absolute Error (MAE) is an arithmetic average of the absolute errors. The absolute error is given by:

$$|e_i| = |\hat{\theta}_i - \theta_i| \tag{4.12}$$

where $\hat{\theta}_i$ is the estimated value and θ_i is the actual value. Because MAE is scale-dependent (i.e. it uses the same scale as the measured data), this method cannot be used to make comparisons between different datasets. The MAE is given by:

$$MAE = \frac{\sum_{i=1}^{n} |\hat{\theta}_i - \theta_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{4.13}$$

4.2.7 Mean Absolute Deviation

The Mean Absolute Deviation (MAD) or Average Absolute Deviation (AAD) is the average of the absolute deviations from a central point. The Mean Absolute Deviation is given by:

$$MAD = \frac{1}{n} \sum_{i=1}^{n} |x_i - m(X)| \tag{4.14}$$

Due to the fact that the central tendency $m(X)$ has a marked effect on the value of the mean deviation, it should be chosen carefully.

4.2.8 Mean Percentage Error

The Mean Percentage Error (MPE) is the average of percentage errors by which the estimator differs from the estimand. The MPE is given by:

$$MPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{\theta_t - \hat{\theta}_t}{\theta_t} \tag{4.15}$$

MPE is widely used to measure the bias of the estimator, because the actual values are used rather than the absolute values.

CHAPTER 5

Literature Review

In this chapter the most commonly used methods and models in literature for Electric Load Forecasting are going to be discussed. These methods of forecasting are mostly based on statistical approaches (i.e. time series analysis) and on artificial intelligence algorithms (e.g. machine learning). Methods that are used are regression models, neural networks, fuzzy logic systems, chaotic systems, statistical learning approaches, time series models etc. [6–12].

5.1 Methods Categorization

As it is stated in literature [13–15] forecasting methods can be classified in two general categories based mainly on the type of the available data. The first category known as Qualitative or Subjective contains the methods that predict the future load based on structured approaches without using historical data. These methods are widely used when the historic energy consumption is not available and/or when a rough calculation of the future energy consumption is needed. Examples of methods that lie in this category are the Empirical Curve Fitting Method, Delphi Method etc. The second category known as Quantitative or Objective contains the methods that predict the future load based on mathematical and statistical models. These methods require knowledge of the past energy consumption and the assumption that past energy consumption patterns are going to continue to the future even with some small changes. The complexity of these methods is much higher and therefor the implementation cost is higher. Examples of these methods are decomposition methods, exponential smoothing, regression analysis, Artificial Neural Networks (ANN), Box-Jenkins methodology etc.

5.2 Load Forecasting Models

In the next section a description of the most commonly used electricity load forecasting models is going to be presented. Time series models can be classified as follows:

- Time space

  - Simple methods
  - Regression
  - Box-Jenkins models
- Frequency space
  - Fourier Analysis
  - Harmonic Regression
- Phase Space
  - Nonlinear dynamics
  - Chaos theory
- State Space
  - Kalman filters
  - Sequential Monte Carlo filters
- Artificial Intelligence
- Data Mining

5.3 Statistical Load Forecasting Models

Firstly the statistical models that are proposed in literature [16] are going to be presented. The models that are going to be discussed are:

- Box-Jenkins models (SARIMAX family models) [17–19]
- Kalman Filtering Algorithms [20–25]
- Grey models [26–35]
- Exponential Smoothing [36–39]

5.3.1 Box-Jenkins models

These models uses an iterative three-stage modeling approach. The first step is the model identification and the model selection. During this step the stationarity of variables is assured and the identification of the seasonality in the dependent series is done. Using the autocorrelation (ACF) and the partial autocorrelation (PACF) functions of the dependent time series a decision is made on which of the autoregressive and/or moving average components should be used in the model. The second step is the estimation of the parameters with the use of computation algorithms that fit best to the model. The estimation of the fit accuracy is calculated based on the maximum likelihood estimation and/or non-linear least-squares estimation.The last step is the validation of the statistical model. The model that was created must conform to the specifications of a stationary univariate process. During this step the independency of the residuals is checked and if need the model is tuned.

5.3.1.1 Autoregressive (AR) Model

The AR models assume that the future values of a time series can be expressed as a linear combination of past values. These models, due to their simplicity, are widely used in many fields, such as economics, signal processing etc. [14, 40, 41] An AR model is given by:

$$L_t - \sum_{i=1}^{p} \phi_i L_{t-i} = \varepsilon_t \tag{5.1}$$

where $\phi_1, \phi_2, ..., \phi_n$ are the model's coefficients and ε_t is Additive White Gaussian Noise (AWGN) which is the random load disturbance. The more past load values are used the higher the order of the model. In order the forecast to be accurate, a high correlation between the future and the past values of the energy consumption is needed. In order to apply an AR model the time-series ideally should be stationary - i.e. the probability distributions of the process are time invariant -. Because stationarity is difficult to occur a weaker requirement is defined. This requirement is called covariance stationarity and assumes that the mean, variance and autocorrelation structures are invariable over time, i.e. thea mean and variance are constant and the autocovariance is a function of $(t-s)$ only. So the definition of the stationarity is given by reassured tat all roots of the following polynomial lie outside the unit circle. The characteristic polynomial is:

$$1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p = 0 \tag{5.2}$$

If the process has a unit root (i.e. $z = 1$) then the autocovariance varies over time. If a unit root exists then the time-series needs to be differenced in order to become a stationary process.
AR models are used for a long time in fields as economics, digital signal processing and electric load forecasting. Due to the invention of more complex methods AR models are used in literature as benchmarks for more sophisticated approaches.

5.3.1.2 Moving Average (MA) Model

A MA model is a linear regression of the current value of the series against current and previous observed white noise error terms or random shocks. At each point the assumption that the random shocks are mutually independent and that come from the same distribution (most commonly a normal distribution) with location at zero and constant scale is made. Essentially a MA model is a Finite Impulse Response (FIR) filter applied to white noise, with some additional interpretation placed on it. An MA model is given by:

$$L_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{5.3}$$

However, MA models in load forecasting applications can only be used for their filtering properties. If combined with AR models an ARMA model is obtained.

5.3.1.3 Autoregressive Moving Average (ARMA) Model

ARMA models [17, 40] combine the benefits of AR and MA models. In these models the current value of the time-series L_t is expressed as linear combination of its past values and its past values of the noise. An ARMA model [14, 41, 42] of order (p,q) is given by:

$$L_t - \sum_{i=1}^{p} \phi_i L_{t-i} = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{5.4}$$

Due to their simplicity and effectiveness, ARMA models have been widely used in load forecasting. An ARMA model [43] that its parameters were update on-line using the weighted recursive least squares algorithm was presented in 1994. Apart from that model in literature several different models have been proposed for load forecasting [44, 45]. A combination of an ARMA model with the rolling volatility method was proposed by Nowicka-Zagrajek and Weron [46] which outperformed existed techniques.
In order to construct an ARMA model the following steps are followed. Based on the parsimony principle the best model must be the simplest one that describes adequately

the data. In a human driven model estimation the ACF and PACF plots can be used, where as on a non-supervised approach an iterative procedure of trial and error may be used. Final Prediction Error (FPE) proposed by Akaike, Akaike's Information Criterion (AICC) and Schwarz Information Criterion (SIC) are used to determine the goodness-of-fit. Final Prediction Error is given by:

$$FPE = V\frac{n+d}{n-d} \tag{5.5}$$

where:

$$V = \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^2 \tag{5.6}$$

is the variance of model residuals $\hat{\varepsilon}_t = L_t - \hat{L}_t$ and n and d are the sample size and the model size respectively. Akaike's Information Criterion is given by:

$$AICC = -2logL + \frac{2dn}{n-d-1} \tag{5.7}$$

where $logL$ is the log-likelihood function. Schwarz Information Criterion is given by:

$$SIC = -2logL + dlogn \tag{5.8}$$

The next step is to calculate the model's coefficients. AR coefficients can be easily calculated using a least-square regression. However, MA and ARMA coefficients require procedures of higher complexity. To estimate AR models in literature different algorithms have been proposed like Yule-Walker algorithm and Burg algorithm. MA models coefficients are commonly calculated using the Hannan-Rissanen algorithm.

The last step is to check the accuracy of the model. The most important is to check whether the residuals of the models are random. In literature different methods for diagnosing models are proposed. Some of these tests are ACF/PACF test, the AICC AR test, the portmanteau test etc.

5.3.1.4 Autoregressive Integrated Moving Average Model

If the time-series is characterized by non-stationarity then the ARMA models cannot be used. Time-series should first be transformed to a stationary one by diffrencing. Therefore the proposed model has three parameters (p,d,q). An ARIMA model is given by:

$$\phi(B) \cdot \nabla^d \cdot L_t = \theta(B) \cdot \varepsilon(t) \tag{5.9}$$

or by:

$$\left[1 - \sum_{i=1}^{p}\phi_i B^i\right] \cdot [1-B]^d \cdot L_t = \left[1 + \sum_{j=1}^{q}\theta_j B^j\right] \cdot \varepsilon_t \tag{5.10}$$

ARIMA models are non-seasonal models, should a seasonal model needed then a SARIMA model can be implemented. In electricity load forecasting ARIMA and SARIMA models have been widely used [14, 41, 42, 47–49].

5.3.1.5 Autoregresssive Moving Average With Exogenous Inputs Models

In order to further ameliorate SARIMA family models exogenous variables were added to these models (weather conditions, time of the day etc.) creating the ARIMAX and the SARIMAX models [14]. An ARMAX model can be expressed by:

$$\phi(B) \cdot L_t = \theta(B) \cdot \varepsilon_t + \sum_{i=1}^{k} (\psi_0^i + \psi_1^i \cdot B + \ldots + \psi_{ri}^i \cdot B^{ri}) \cdot v_t^i \tag{5.11}$$

5.3.2 Kalman Filtering Algorithm in the State Space

Kalman filter [50] is a set of mathematical equations in the state space that provide an efficient computational means to estimate the state of an observed process. This filter can be used to minimize the error of the model that is produced due to the dependency of the electricity consumption with exogenous factors (e.g. socioeconomic etc.). In literature [25] it is stated that the following factors affect the load behavior:

- Weather conditions, such as temperature, humidity etc.
- Time window
- Socioeconomic factors, such as the load management policy, consumers behavior etc.
- Stochastic processes that don't fall on the previous categories.

These factors are used as input to a Kalman filter, due to the nature of these inputs usually only the weather conditions and the time windows are used [25].

5.3.3 Grey Models

Grey Theory [51] enables to model a system while having partial unknown parameters. Grey models are created based on past values and can be used to forecast future load. Grey models have been used in literature for short-term, medium-term and long-term forecasting [52].

5.3.4 Exponential Smoothing Models

Exponential smoothing is method that the future load values are produced from the exponentially weighted average of the past observations [39, 41, 53]. The weights of the previous values are reduced exponentially, i.e. the first value has the highest weight while the next value has smaller values. These models are widely used due to their computational simplicity, robustness and low implementation costs. The exponential smoothing techniques that are used are Brown's method, Holt's method and Holt-Winters method. Brown's method

5.4 Artificial Intelligence and Machine Learning Models

The models described in the previous section, although simple, their capabilities are often limited and sometimes their computational time is significantly extensive. In order to overcome these difficulties in literature [54] other methods using Machine Learning and Artificial Intelligence have been proposed.

5.4.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) were proposed for forecasting and classification purposes in 1990 by Warren McCulloch and Walter Pitts. In the last decades ANNs become very popular in the field of the electric load forecasting [42]. The main advantage of the ANNs is that due to the non-linear nature of the system, it is capable to perform non-linear curve fitting. ANNs were created by imitating the biologic system of the human brain. An ANN is composed of several interconnected processing elements - neurons -, that are changing their dynamic state response with respect to external inputs [40, 42, 55]. A simple Neural Network is presented in figure 5.1. The system has several inputs and each input has a different weight. The output is produced after transforming the weighted-input data using a hidden process, also known as the hidden layer. Feedback from the output is fed in the input in order to change the weights and reach to the model required. In the literature [24, 26, 40, 41, 55–65] the following categories of Neural Networks have been proposed for load forecasting:

- feed-forward neural networks

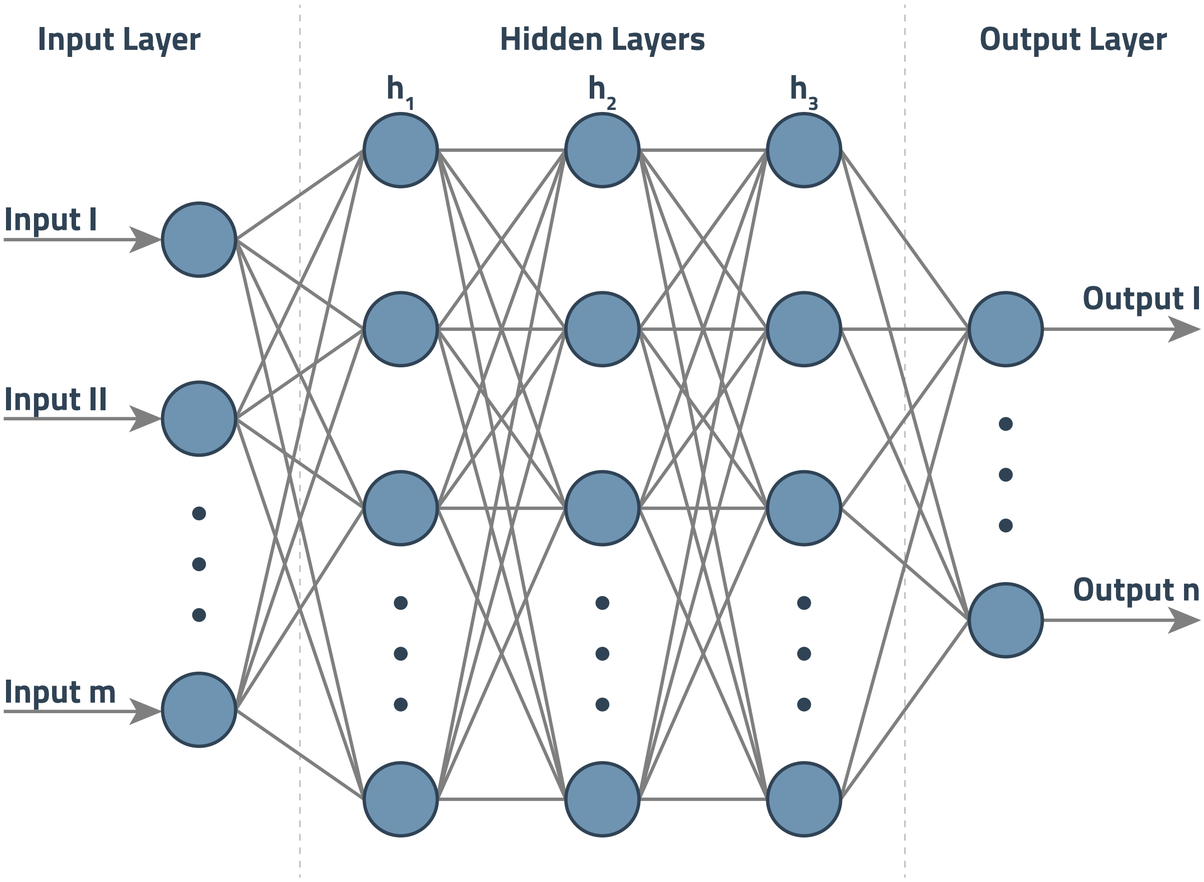

Figure 5.1: A Neural Network

- non-linear autoregressive with exogenous inputs neural networks, which usually outperform the feed-forward neural networks and are relative simple, reliable and accurate [61]
- back-propagation neural networks
- radial basis function neural networks
- the random neural networks
- recurrent neural networks
- self-organizing competitive neural networks

The proposed ANN models can be grouped in two categories [41] based on their output. The first group contains all the methods that have a single output, i.e. the forecasting model only produces the next hour's load, or the next day's load etc. The second group contains the models that have multiple outputs, i.e. the forecasting model produces a vector of future values (e.g. hourly values for the day ahead etc.).

5.4.2 Extreme Learning Machines

Extreme Learning Machines (ELMs) were proposed in 2004 [66] by Huang et al. and are appropriate for regression, classification, clustering, feature learning and sparse approximation. An ELM usually has a hidden FF neural network layer, the input weights of which are randomly selected whereas the output weights are analytically determined based on a least-squares algorithm. The advantage of these network is that the hidden layer can be only once randomly allocated and then never updated. The output weights are settles in a single step, and therefore the computational time needed is much less compared to other ANNs. As it have been proved [66] ELM models can learn with a pace of up to a thousand times faster than other ANNs. These models have been extensively used in forecasting future load values [67–74].

5.4.3 Support Vector Machines

Support Vector Machine (SVM) is a classification and regression tool that is based on Vapnik's 1995 [75] statistical learning theory. ANNs define complex functions of the input space, whereas SVMs perform a non-linear mapping of the data into a high dimensional space. After having created the mapping function, SVMs use simple linear decision boundaries in the multidimensional space. The main advantage of the SVMs is that they converge to the global minimum and so they produce a single solution. Contrary, ANNs may produce multiple solutions as they may converge on local minima. Furthermore SVMs do not rely so heavily on heuristics and therefore have a more flexible structure. One application where SVMs outperformed all other methods, is on medium-term electric load forecasting [76]. In literature different models that use SVMs have been proposed [77–85].

5.4.4 Fuzzy Logic

Fuzzy logic is the theory that generalizes the Boolean theory. Whereas in Boolean theory variables can have only two states - zero and one - in Fuzzy logic variables can have a qualitative range [2, 86–104]. Fuzzy logic is used mostly in combination with other predictive methods in the following cases:

- when a model cannot be defined, or it is difficult to be expressed
- when a model is too complex to be evaluated in a reasonable amount of time

- when a model requires more memory than it is available
- when a model's inputs or definition is uncertain
- when a model is non-linear and other implementation methods won't produce an accurate result

Although Fuzzy Logic systems are widely used in literature [2, 91, 97, 98, 105–113] they should be avoided when other methods can produce adequate results.

5.4.5 Wavelet Neural Networks

In the 1980s Grossman and Morlet proposed the Wavelet theory. Later, in 1992 Zhang suggested the combined use of Wavelets and Neural Networks in order to merge the benefits of these systems. The result was the creation of the Wavelet Neural Netwokrs (WNNs). A WNN can be used instead of a feed-forward neural network in order to approximate the arbitrary non-linear function. By using the wavelet transform theory features can be extracted by calculating the internal product of the wavelets base and the signal vector. Patel et al. [114] proposed a forecasting system based on ELF and wavelet networks.

5.4.6 Genetic Algorithms

Genetic Algorithms are algorithms that try to imitate the genetic variation and the natural selection [115], with the help of operators. These models were first proposed in the 1950s as an optimization tool for solving engineering problems by observing the evolutionary systems. One of the most popular technique of evolutionary computation is the use of Genetic Algorithms [116]. Electric load forecasting methods have highly benefited from the Genetic Algorithms due to the fact that they can indicate the optimal model from a series of forecasting candidates [117]. Also, Genetic Algorithms have been used to estimate the (p,d,q) parameters of the ARIMA models [117]. Lastly, Genetic Algorithms have also been used combined with other Artificial Intelligence and Machine Learning techniques [118–125].

5.5 Hybrid Methods

In order to create an even more accurate load forecasting model different hybrid models have been proposed. These models combine the advantages of different single forecasting techniques in order to improve their prediction performance. These techniques may also combine different optimization algorithms and data pre-processing techniques in order to ameliorate the results. [126, 127].

CHAPTER 6

Proposed Model

In this chapter the proposed model is going to be presented. Firstly, the nature of the data-set and the data pre-processing steps are going to be presented. Then the load forecasting procedure is going to be explained and lastly the results of the proposed method are going to be unveiled.

6.1 Data-set

The Data-Set consists of thousands of end-users from a single area that are served from the same node. The available past load values span more than five consecutive years, thus permitting us to perform our analysis. Apart from past load values, some other values are also used:

- the dry bulb temperature
- the dew point
- the hour of the day
- the day of the week
- a flag that indicates whether a day is a business or off day
- the load from the same hour the previous day
- the load from the same hour and the same day from the previous week
- the average load 24 hours ago

Additionally in order to predict the electricity price in a power grid that the majority of the energy is produced from natural gas plants, the following data are also used:

- Previous day's average price
- Price from the same hour the previous day
- Price from the same hour and same day from the previous week

- Previous day's natural gas price
- Previous week's average natural gas price

Should one need a medium or long term load and price forecasting, then only the hour of the day, day of the week and the type of holiday can be used deterministically. All other values must be specified by a distribution.

6.1.1 Dry-Bulb Temperature

The dry-bulb temperature (DBT) is the temperature of air measured by a thermometer freely exposed to the air, but shielded from radiation and moisture. This temperature is the true thermodynamic temperature and is commonly thought as air temperature. This temperature indicates the amount of heat in the air and is directly proportional to the mean kinetic energy of the air molecules. DBT is one of the most important climate variables for human comfort and therefore plays a significant role in the energy consumption for heating and cooling.

DBT should not be mixed up with wet-bulb temperature (WBT), a temperature that also takes into account the amount of moisture in the air. WBT is the temperature that is read by a thermometer covered in water-soaked cloth over which air is passed. If the relative humidity is 100% then WBT and DBT are identical. If the humidity is lower then WBT is always lower than DBT because of evaporative cooling.

6.1.2 Dew Point

The dew point is the temperature to which air must be cooled to become saturated with water vapor. If the air is cooled further, the airborne water vapor will condense and form liquid water. If the temperature is lower than the freezing point of water, the dew point is called the frost point, as frost is formed via deposition rather than condensation to form dew. Dew point is an indirect method to measure moisture in the air and as a consequence humidity.

Dew point is highly correlated with human comfort. When the air temperature is high, the human body uses the evaporation of sweat to cool down. The cooling effect on a human body is directly related to the velocity of the perspiration evaporation. In an environment with high humidity levels only a small amount of extra moisture can be hold or in extreme situations where the air is saturated no extra moisture can be hold. Therefore perspiration

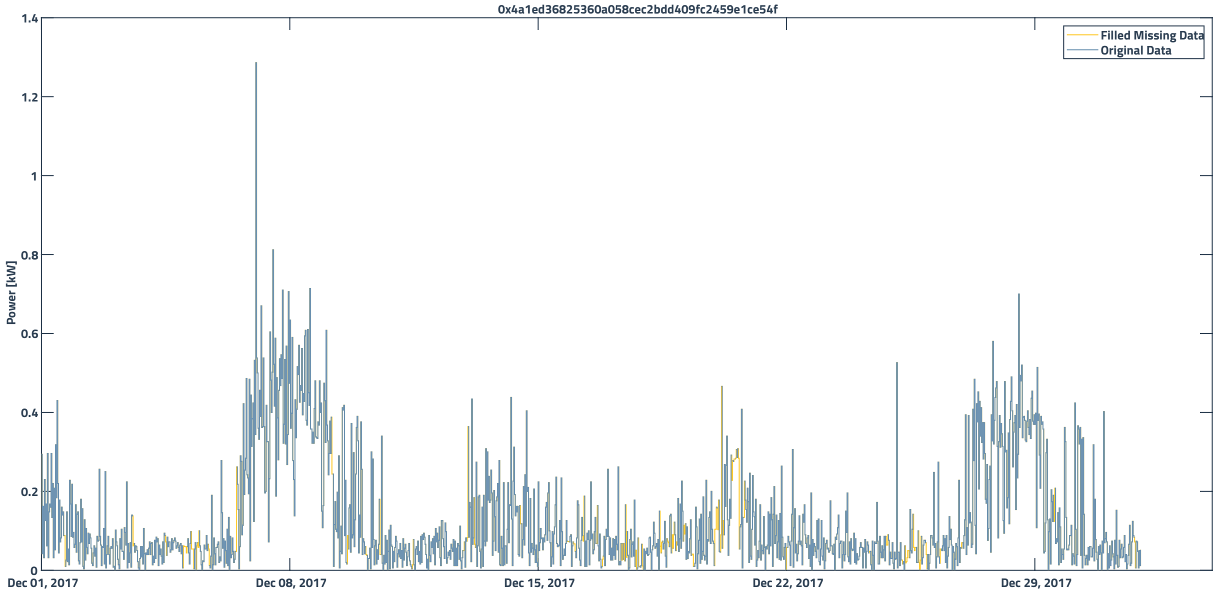

Figure 6.1: Filled Missing Values

will not evaporate and a sensation of discomfort will appear. That's why energy consumption (mostly for heating and cooling) is highly correlated with the dew point.
If no information regarding the dew point is available, it can be easily calculated by the dry bulb temperature and the relative humidity using the Magnus formula:

$$T_{dp} = \frac{c\gamma(T,RH)}{b-\gamma(T,RH)} \tag{6.1}$$

where:

$$\gamma(T,RH) = ln\left(\frac{RH}{100}\right) + \frac{bT}{c+T} \tag{6.2}$$

In the literature other more accurate formulas exist but for the purpose of load forecasting Magnus formula is sufficient.

6.2 Data pre-processing

Before creating the load forecasting model it is important to filter out erroneous data as well as missing values. For each end-consumer past load missing values were filled with the use of a moving median window. Due to the nature of the climate conditions, missing temperature and/or humidity values were filled using a modified Akima cubic Hermite interpolation. In figure 6.1 an example of filled missing load values is presented.
After having filled the missing values, data were aggregated in order to represent a whole area. The aggregation is simply done by adding all the end-consumers' vectors with

respect to the time-stamp. The result was a single vector with granularity of 1 sec. The vector was then under-sampled to a granularity of 30 minutes, and was separated to two parts: the first part, which is the training set, contains data from the first 4 years, while the second part, which is the testing set, contains data from the last year. This separation was done in order to create the model on the training set and then check its performance on the testing set.
During the pre-processing step, the load from the same hour the previous day and the load from the same hour and the same day from the previous week are stored. Lastly, by using an adequate calendar, business and off days are marked accordingly.

6.3 Load Forecasting Procedure

The forecasting procedure is as follows. Firstly a default neural network is initialized, consisting of two layers with 20 neurons. The metric that is used to test the accuracy is Mean Absolute Error (MAE). Then the network is trained using the Levenburg-Marquardt algorithm. This algorithm is widely used in non-linear least squares problems. This algorithm interpolates between the Gauss-Newton algorithm (GNA) and the method of gradient descent, thus making it more robust - as it can reach to a solution even if it starts very far off the final minimum - but it is more time consuming than GNA.

6.3.1 Levenburg-Marquardt Algorithm

Given a set of m empirical pairs (x_i, y_i) of independent and dependent variables the algorithm finds the parameters β of the model curve $f(x, \beta)$ so that the sum of the squares of the deviation $S(\beta)$ is minimized. This problem can be written as:

$$\hat{\beta} \in argmin_{\beta} S(\beta) \equiv argmin_{\beta} \sum_{i=1}^{m} [y_i - f(x_i, \beta)]^2 \tag{6.3}$$

The solution to this problem is found after an iterative procedure. Firstly a uniformed standard guess of the vector β is provided to the system. In each step the vector β is replaced by a new estimation $\beta + \delta$. In order to determine the δ an approximation of the function is taken by converting it to linear:

$$f(x_i, \beta + \delta) \approx f(x_i, \beta) + \mathbf{J_i}\delta \tag{6.4}$$

where the gradient of the function with respect to β is given by:

$$\mathbf{J_i} = \frac{\partial f(x_i, \beta)}{\partial \beta} \tag{6.5}$$

The Sum of square deviations has its minimum at a zero gradient with respect to β. Using the pre-described first-order approximation the following equation can be obtained:

$$S(\beta+\delta) \approx \sum_{i=1}^{m} [y_i - f(x_i,\beta) - \mathbf{J}_i\delta]^2 \tag{6.6}$$

The previous equation can be written in a vector form as:

$$\begin{aligned} S(\beta+\delta) &\approx ||\mathbf{y} - \mathbf{f}(\beta) - \mathbf{J}\delta||^2 \\ &= [\mathbf{y} - \mathbf{f}(\beta)]^T[\mathbf{y} - \mathbf{f}(\beta)] - 2[\mathbf{y} - \mathbf{f}(\beta)]^T\mathbf{J}\delta + \delta^T\mathbf{J}^T\mathbf{J}\delta \end{aligned} \tag{6.7}$$

By calculating the derivative of $S(\beta+\delta)$ with respect to δ and setting the result to zero gives:

$$(\mathbf{J}^T\mathbf{J})\delta = \mathbf{J}^T[\mathbf{y} - \mathbf{f}(\beta)] \tag{6.8}$$

where $\mathbf{j}$ is the Jacobian matrix whose i-th row equals to $\mathbf{J}_i$ and $\mathbf{f}(\beta)$ and $\mathbf{y}$ are vectors with the i-th component being $f(x_i,b)$ and y_i. Levenberg's algorithm replace this equation by a damped version:

$$(\mathbf{J}^T\mathbf{J} + \lambda\mathbf{I})\delta = \mathbf{J}^T[\mathbf{y} - \mathbf{f}(\beta)] \tag{6.9}$$

where $\mathbf{I}$ is the identity matrix. The non-negative damping factor λ is adjusted at each iteration.

6.4 Results

In this section the results and the performance of the proposed algorithm are going to be presented.

6.4.1 Load Forecasting Results

Once the model is built, a forecast is performed on the independent test set. The load forecasting results are presented in figure 6.2 were the blue line presents the actual values and the yellow line presents the forecast values. A more detailed view can be observed in figure 6.3.

The forecast load was contrasted to the actual load and metrics such as the MAE, MAPE and daily peak error were used. The results are presented in the table 6.1.

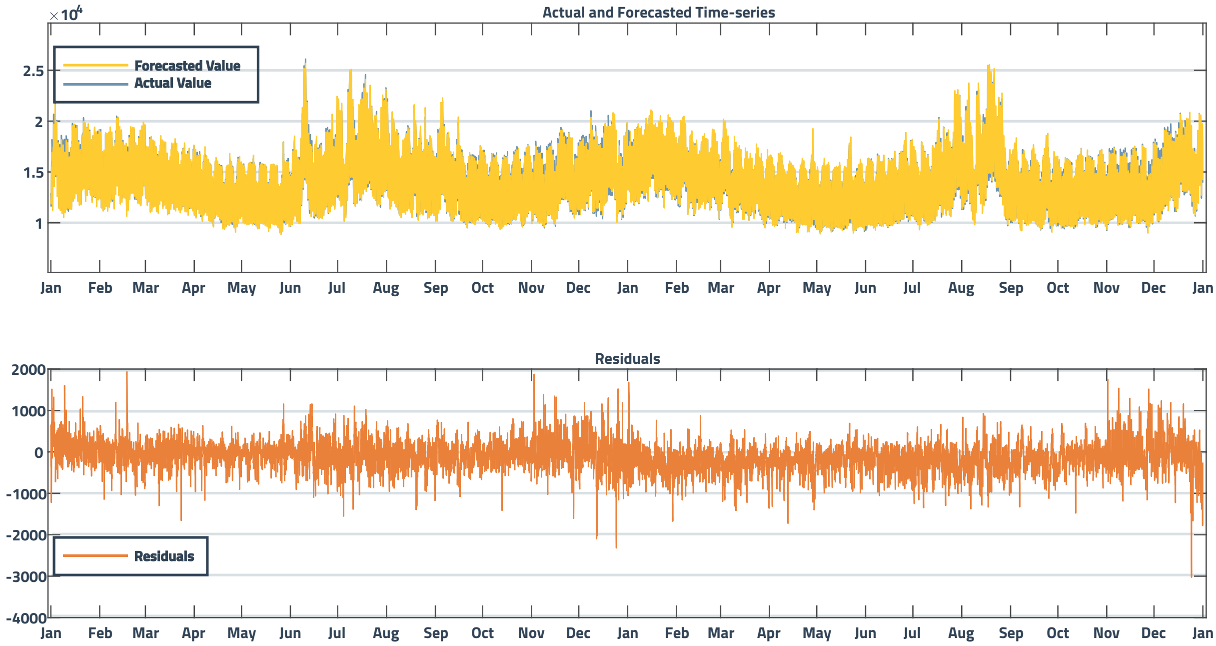

Figure 6.2: Actual and Forecast Load Time-series Residual Plot

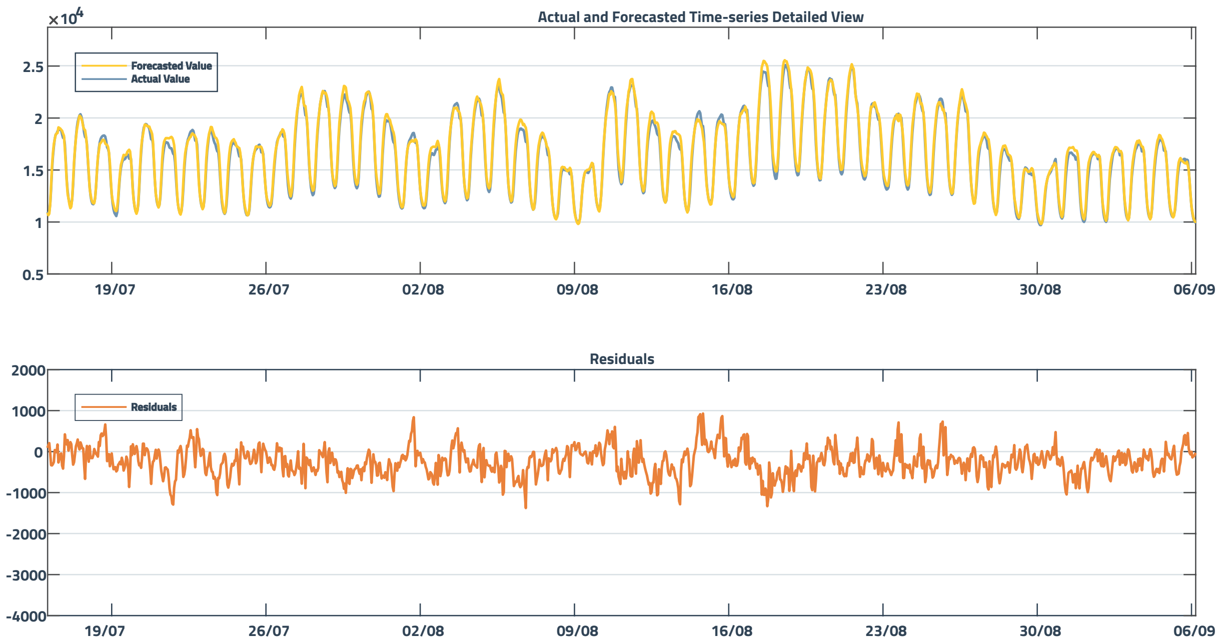

Figure 6.3: Actual and Forecast Load Time-series Residual Plot Detailed View

Table 6.1: Cumulative Load Model Evaluation

	Error
MAPE	1.9%
MAE	277MWh
Daily Peak MAPE	1.8%

6.4.1.1 Monthly Load Forecasting Error

Table 6.2 presents each month's percentage error. As it can be observed the load forecasting error is almost constant. There is an increase to the forecasting error during the winter months mostly due to the unstable weather conditions.

Table 6.2: Monthly Load Forecasting Error

Month	Lower Adj.	25th Percentile	Median	75th Percentile	Upper Adj.
January	0.005%	0.7%	1.5%	2.6%	5.5%
February	0.007%	0.8%	1.6%	2.8%	5.7%
Mars	0.002%	0.8%	1.6%	2.7%	5.5%
April	0.000%	0.6%	1.4%	2.5%	5.2%
May	0.004%	0.6%	1.2%	2.2%	4.6%
June	0.001%	0.7%	1.4%	2.4%	5.1%
July	0.003%	0.7%	1.5%	2.5%	5.4%
August	0.001%	0.7%	1.5%	2.6%	5.3%
September	0.010%	0.8%	1.6%	2.8%	5.8%
October	0.001%	0.6%	1.2%	2.1%	4.4%
November	0.001%	0.9%	1.7%	2.8%	5.6%
December	0.000%	0.8%	1.8%	3.2%	6.6%

6.4.1.2 Daily Load Forecasting Error

Table 6.3 presents each weekday's percentage error. As it can be observed the forecasting

Table 6.3: Daily Load Forecasting Error

Weekday	Lower Adj.	25th Percentile	Median	75th Percentile	Upper Adj.
Monday	0.000%	0.8%	1.7%	2.8%	5.7%
Tuesday	0.001%	0.7%	1.6%	2.8%	5.9%
Wednesday	0.001%	0.6%	1.4%	2.6%	5.4%
Thursday	0.001%	0.7%	1.5%	2.5%	5.1%
Friday	0.001%	0.6%	1.4%	2.5%	5.4%
Saturday	0.003%	0.7%	1.5%	2.5%	5.1%
Sunday	0.000%	0.6%	1.4%	2.5%	5.3%

error is almost non correlated with the day of the week.

6.4.1.3 Hourly Load Forecasting Error

Table 6.4 presents each hour's percentage error. As it can be observed the error is less

Table 6.4: Hourly Load Forecasting Error

Hour	Lower Adj.	25th Percentile	Median	75th Percentile	Upper Adj.
01:00	0.001%	0.5%	1.1%	1.9%	4.0%
02:00	0.001%	0.6%	1.4%	2.2%	4.6%
03:00	0.000%	0.5%	1.0%	1.8%	3.8%
04:00	0.000%	0.5%	1.1%	1.8%	3.7%
05:00	0.005%	0.5%	1.2%	2.0%	4.2%
06:00	0.005%	0.8%	1.8%	2.9%	6.0%
07:00	0.002%	0.7%	1.7%	2.9%	6.1%
08:00	0.003%	0.7%	1.6%	2.9%	6.0%
09:00	0.010%	1.1%	2.1%	3.4%	6.8%
10:00	0.010%	0.9%	2.0%	3.2%	6.0%
11:00	0.011%	0.7%	1.5%	2.4%	4.9%
12:00	0.002%	0.6%	1.3%	2.2%	4.5%
13:00	0.006%	0.7%	1.4%	2.3%	5.0%
14:00	0.001%	0.7%	1.4%	2.5%	5.2%
15:00	0.001%	0.8%	1.7%	2.8%	5.8%
16:00	0.006%	0.8%	1.7%	3.0%	6.2%
17:00	0.010%	0.9%	1.8%	3.0%	6.2%
18:00	0.001%	0.9%	1.9%	3.3%	6.7%
19:00	0.008%	0.9%	1.9%	3.3%	6.9%
20:00	0.016%	1.0%	2.0%	3.6%	7.3%
21:00	0.007%	0.7%	1.4%	2.4%	4.9%
22:00	0.004%	0.6%	1.3%	2.2%	4.6%
23:00	0.000%	0.7%	1.4%	2.3%	4.8%
24:00	0.006%	0.6%	1.2%	2.2%	4.5%

significant during the off-peak hours and higher during the peak hours (06:00 - 10:00 & 15:00 - 20:00), which is expected due to the fact that during these hours more electric devices are online.

6.4.2 Price Forecasting Results

Once the model is built, a forecast is performed on the independent test set. The price forecasting results are presented in figure 6.4 were the blue line presents the actual values and the yellow line presents the forecast values. A more detailed view can be observed in figure 6.5.

The forecast load was contrasted to the actual load and metrics such as the MAE, MAPE and daily peak error were used. The results are presented in the table 6.5. As it can be

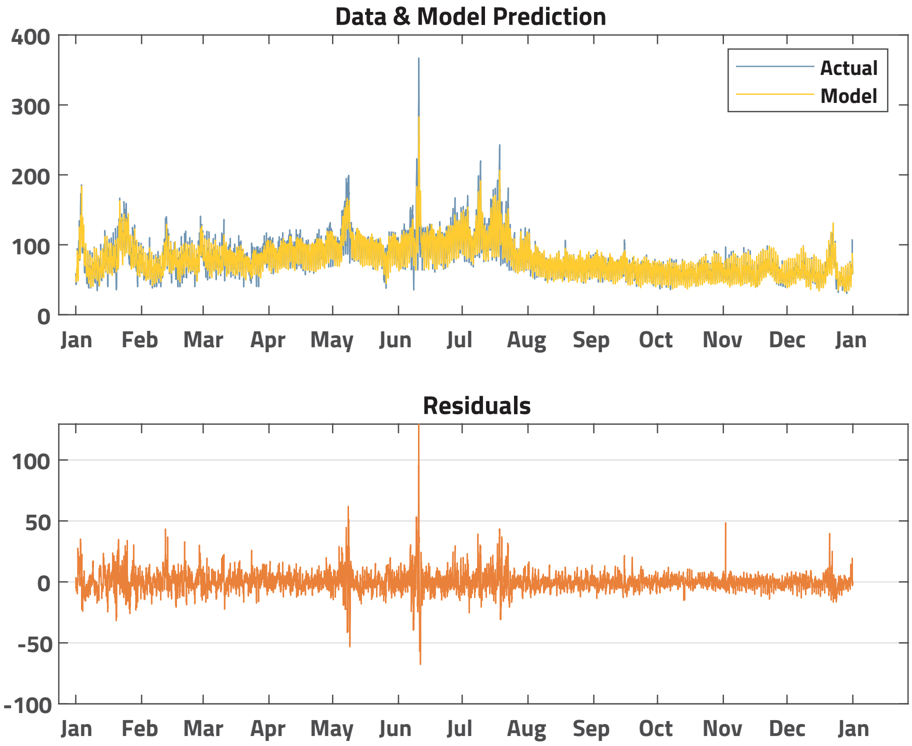

Figure 6.4: Actual and Forecast Price Time-series Residual Plot

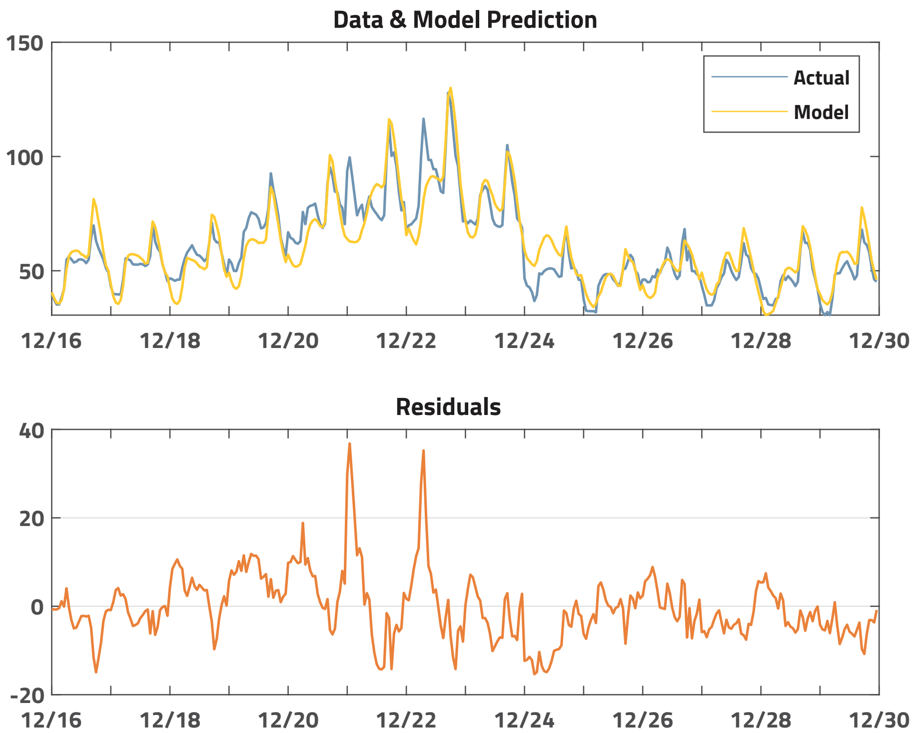

Figure 6.5: Actual and Forecast Price Time-series Residual Plot Detailed View

Table 6.5: Cumulative Price Model Evaluation

	Error
MAPE	6.8%
MAE	5.41 EUR/MWh
Daily Peak MAPE	5.9%

observed, load forecasting procedure leads to more accurate results compared to price forecasting. Price forecasting is by definition a more challenging procedure, due to the fact that price can be influenced by exogenous factors. However, the proposed model and the accuracy of it can be considered sufficient.

CHAPTER 7

Conclusions and Future Research

In this day and age, due to the advances in technology electricity load forecasting has become possible. The importance of the load forecasting is evident both to the end consumers as well as to the producers and the utility providers. End consumers can benefit from load forecasting by adjusting their consumption in advance and avoid energy bill shocks. Producers and utilities providers can benefit by better designing their infrastructures as well as by better planning new power plants. They can also benefit from the advance knowledge of the future energy consumption by better understanding their energy requirements and therefore ensuring the energy required by their customers.

In this Thesis a presentation of the main load characteristics and a brief literature review of the available load forecasting methods was made. Also, a proposed method for load and price forecasting for the day-ahead market on an aggregated basis was presented.

As it can be seen from the results, the proposed method has gave quite satisfying results. However, further fine-tuning can be made in order to ameliorate the accuracy of the produced models. Also, a combination of other techniques (e.g. Genetic Algorithms) can be used in order to improve the robustness and the performance of the proposed model. Lastly, further testing of the model must be performed in order to reassure that by changing the architecture of the sample the same results will be obtained, i.e. that the model is not sensitive to the input sample. However, this testing exceeds the scope of this thesis and is left for future research.

REFERENCES

[1] D. Pilipovic, *Energy Risk: Valuing and Managing Energy Derivatives*. McGraw-Hill Education, 1998.

[2] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. Wiley, 2007.

[3] R. l Weron, “Measuring long-range dependence in electricity prices,” 2001.

[4] I. Daubechies, *Ten Lectures on Wavelets*. USA: Society for Industrial and Applied Mathematics, 1992.

[5] J. Armstrong, “Long-range forecasting, 2nd. ed,” *Contemporary Sociology*, vol. 8, 1985.

[6] S. Fallah, M. Ganjkhani, S. Shamshirban, and K. w. Chau, “Computational Intelligence on Short-Term Load Forecasting: A Methodological Overview,” *Energies*, vol. 12, no. 3, 2019.

[7] W.-C. Hong, M.-W. Li, and G.-f. Fan, *Short-Term Load Forecasting by Artificial Intelligent Technologies*, 08 2019.

[8] I. A. b. W. A. Razak, S. b. Majid, M. S. b. M. Aras, and A. b. Ahmad, *Electricity Load Forecasting Using Data Mining Technique*. IntechOpen, 2012.

[9] F. Su, Y. Xu, and X. Tang, “Short-and mid-term load forecasting using machine learning models,” 10 2017, pp. 406–411.

[10] S. HemaChandra, V. Harish, C. R. Kumar, , and V. Nagarjuna, “A Review of Long Term Electrical Load Forecasting Methods,” *Artificial Intelligent Systems and Machine Learning*, vol. 4, no. 10, pp. 566–569, 2012.

[11] S. Panda, S. Mohanty, and A. Jagadev, “Long Term Electrical Load Forecasting: An Empirical Study across Techniques and Domains,” *Indian Journal of Science and Technology*, vol. 10, pp. 1–16, 06 2017.

[12] M. Jacob, C. Neves, and D. V. Greetham, *Forecasting and Assessing Risk of Individual Electricity Peaks*. Springer Nature, 2019, ch. Short Term Load Forecasting, pp. 15–37.

[13] N. Abu-Shikhah and F. Elkarmi, "Medium-Term Electric Load Forecasting Using Singular Value Decomposition," *Energy Conversion and Management*, vol. 36, no. 7, pp. 4259–4271, 2011.

[14] S. A. h. Soliman and A. M. Al-Kandari, *Electrical Load Forecasting: Modeling and Model Construction*, 1st ed. Butterworth-Heinemann, 2010.

[15] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. OTexts, Melbourne, Australia, 2018.

[16] R. Dawes, *International Encyclopedia of the Social & Behavioral Sciences*, 12 2001, ch. Clinical versus Actuarial Prediction, pp. 2048–2051.

[17] G. Box and G. Jenkins, "Time Series Analysis, Forecasting and Control," *Journal of the American Statistical Association*, vol. 134, 01 1971.

[18] P. S. P. Cowpertwait and A. V. Metcalfe, *Introductory Time Series with R*. Springer New York, 2009.

[19] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power Generation, Operation, and Control*. Wiley, 2013.

[20] J. Casals, A. Garcia-Hiernaux, M. Jerez, S. Sotoca, and A. Trindade, *State-Space Methods for Time Series Analysis: Theory, Applications and Software*, 09 2018.

[21] J. Commandeur and S. J. Koopman, *An Introduction to State Space Time Series Analysis (Practical Econometrics)*, 09 2007.

[22] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods: Second Edition*. Oxford University Press, 2012.

[23] L. Huang, Y. Yang, H. Zhao, X. Wang, and H. Zheng, "Time series modeling and filtering method of electric power load stochastic noise," *Protection and Control of Modern Power Systems*, vol. 2, 12 2017.

[24] S. A. Markoulakis, G. S. Stavrakakis, and T. G. Nikolaou, "Short-term load forecasting based on the Kalman filter and the neural-fuzzy network (ANFIS)," in *2006 IASME/WSEAS International Conference on Energy & Environmental Systems*, 2006, pp. 189–193.

[25] M. Gaur and A. Majumdar, “One-Day-Ahead Load Forecasting using nonlinear Kalman filtering algorithms,” *Energy Procedia*, 06 2016.

[26] J. Cheng, W. Xiong, and L. Ai, *Recent advances in computer science and information engineering*. Springer, Heidelberg, 01 2012, ch. Electric Load Forecasting Based on Improved Grey Neural Network, pp. 651–655.

[27] C. Herui, B. Tao, and L. Yanzi, “Short-term Power Load Forecasting Based on Gray Theory,” *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 11, 11 2013.

[28] M. Jin, X. Zhou, Z. Zhang, and M. Tentzeris, “Short-term power load forecasting using Grey correlation contest modeling,” *Expert Syst. Appl.*, vol. 39, pp. 773–779, 01 2012.

[29] J. A. S. Kelso, P. Érdi, K. J. Friston, H. Haken, J. Kacprzyk, J. Kurths, L. E. Reichl, P. Schuster, F. Schweitzer, D. Sornette, S. Liu, and Y. Lin, *Grey Systems (no. 68)*. Springer, 2011.

[30] S. Liu and J. Y.-L. Forrest, *Grey Systems: Theory and Applications*. Springer Verlag, 01 2010.

[31] S. Liu, Y. Yang, and J. Forrest, *Grey data analysis: Methods, models and applications*. Springer, 2017.

[32] Y. Lu, Y. Teng, and H. Wang, “Load Prediction in Power System with Grey Theory and its Diagnosis of Stabilization,” *Electric Power Components and Systems*, vol. 47, no. 6–7, pp. 619–628, 2019.

[33] J. Mi, L. Fan, X. Duan, and Y. Qiu, “Short-Term Power Load Forecasting Method Based on Improved Exponential Smoothing Grey Model,” *Mathematical Problems in Engineering*, vol. 2018, pp. 1–11, 03 2018.

[34] T. Ozcan, T. Küçükdeniz, and F. H. Sezgin, “Comparative Analysis of Statistical, Machine Learning, and Grey Methods for Short-Term Electricity Load Forecasting,” in *Nature inspired computing*. I. R. Management Association, Ed. Hershey PA: IGI Global, 2017, vol. 1, pp. 1161–1183.

[35] H. Zhao and S. Guo, “An optimized grey model for annual power load forecasting,” *Energy*, vol. 107, 04 2016.

[36] J. Peirong, X. Di, W. Peng, and C. Juan, "A Study on Exponential Smoothing Model for Load Forecasting," in *Power and Energy Engineering Conference (APPEEC), 2012 Asia-Pacific*, 03 2012.

[37] D. Dragan, T. Kramberger, and M. Intihar, "A comparison of Methods for Forecasting the Container Throughput in North Adriatic Ports," in *IAME 2014 Conference Norfolk VA USA*, 07 2014.

[38] D. Dragan, A. Keshavarzsaleh, T. Kramberger, B. Jereb, and M. Rosi, "Forecasting US Tourists' inflow to Slovenia by modified Holt-Winters Damped model: A case in the Tourism industry logistics and supply chains (DeGruyter Sciendo Journal of Logistics and Sustainable Transport)," vol. 10, pp. 11–30, 06 2019.

[39] R. Hyndman, A. Koehler, K. Ord, and R. Snyder, *Forecasting with exponential smoothing. The state space approach*, 01 2008.

[40] C. Kuster, Y. Rezgui, and M. Mourshed, "Electrical load forecasting models: A critical systematic review," *Sustainable Cities and Society*, vol. 35, 08 2017.

[41] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. John Wiley & Sons Ltd, 2006.

[42] R. Adhikari and R. K. Agrawal, *An Introductory Study on Time Series Modeling and Forecasting*. LAP Lambert Academic Publishing, 2013.

[43] J. Y. Fan and J. D. McDonald, "A real-time implementation of short-term load forecasting for distribution power systems," *IEEE Transactions on Power Systems*, vol. 9, no. 2, pp. 988–994, 1994.

[44] J.-F. Chen, W.-M. Wang, and C.-M. Huang, "Analysis of an adaptive time-series autoregressive moving-average (arma) model for short-term load forecasting," *Electric Power Systems Research*, vol. 34, no. 3, pp. 187–196, 1995. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0378779695009771

[45] L. D. Paarmann and M. D. Najar, "Adaptive online load forecasting via time series modeling," *Electric Power Systems Research*, vol. 32, no. 3, pp. 219–225, 1995. [Online]. Available: http://www.sciencedirect.com/science/article/pii/037877969400918T

[46] J. Nowicka-Zagrajek and R. Weron, “Modeling electricity loads in california: Arma models with hyperbolic noise,” Hugo Steinhaus Center, Wroclaw University of Technology, HSC Research Reports HSC/02/02, 2002. [Online]. Available: https://EconPapers.repec.org/RePEc:wuu:wpaper:hsc0202

[47] E. Almeshaiei and H. Soltan, “A Methodology for Electric Power Load Forecasting,” *Alexandria Engineering Journal*, vol. 50, no. 2011, pp. 137–144, 2011.

[48] M. Sarhani and A. E. Afia, “Electric Load Forecasting Using Hybrid Machine Learning Approach Incorporating Feature Selection,” in *International Conference on Big Data Cloud and Application*, Tetuan, Morocco, 2015.

[49] X. Wang, K. Smith-Miles, and R. Hyndman, “Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series,” *Neurocomputing*, vol. 72, pp. 2581–2594, 06 2009.

[50] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[51] E. Kayacan, B. Ulutas, and O. Kaynak, “Grey System Theory-Based Models in Time Series Prediction,” *Expert Systems with Applications*, vol. 37, pp. 1784–1789, 03 2010.

[52] Y. Feng, “Study on medium and long term power load forecasting based on combination forecasting model,” vol. 51, pp. 859–864, 01 2016.

[53] E. Ostertagova and O. Ostertag, “The Simple Exponential Smoothing Model,” in *The Modelling of Mechanical and Mechatronic Systems 2011: The 4th International Conference*, Faculty of Mechanical Engineering, Technical University of Košice, 09 2011.

[54] E. A. Feinberg and D. Genethliou, “Load Forecasting,” in *Applied Mathematics for Restructured Electric Power Systems. Optimization, Control, and Computational Intelligence*, J. H. Chow, F. F. Wu, and J. A. Momoh, Eds. Springer-Verlag New York Inc, 2006, pp. 269–285.

[55] G. Nalcaci, A. Özmen, and G. W. Weber, "Long-term Load Forecasting: Models Based on MARS, ANN and LR methods," *Central European Journal of Operations Research (CEJOR)*, vol. 27, no. 2019, pp. 1033–1049, 2018.

[56] J. Zhang, "Research on Power Load Forecasting Based on the Improved Elman Neural Network," *The Italian Association of Chemical Engineering (AIDIC)*, vol. 51, no. 2016, pp. 589–594, 2016.

[57] R. Wanga, J. Wangb, and Y. Xu, "A Novel Combined Model Based on Bybrid Optimization Algorithm for Electrical Load Forecasting," *Applied Soft Computing Journal*, vol. 82, no. 2019, 2019.

[58] H. Alfares and N. Mohammad, "Electric load forecasting: Literature survey and classification of methods," *International Journal of Systems Science - IJSySc*, vol. 33, pp. 23–34, 01 2002.

[59] A. Singh, I. Nasiruddin, S. Khatoon, M. Muazzam, and D. Chaturvedi, "Load forecasting techniques and methodologies: A review," in *ICPCES 2012 - 2012 2nd International Conference on Power, Control and Embedded Systems*, 12 2012, pp. 1–10.

[60] J. G. Jetcheva, M. Majidpour, and W.-P. Chen, "Neural Network Model Ensembles for Building Level Electricity Load Forecasts," *Energy and Buildings*, vol. 84, no. 2014, pp. 214–223, 2014.

[61] M. Momani, W. Alrousan, and A. Alqudah, "Short-Term Load Forecasting Based on NARX and Radial Basis Neural Networks Approaches for the Jordanian Power Grid," *Jordan Journal of Electrical Engineering*, 10 2015.

[62] L. Andrade, M. Oleskovicz, A. Santos, D. Coury, and R. Fernandes, "Very short-term load forecasting based on NARX recurrent neural networks," *IEEE Power and Energy Society General Meeting*, vol. 2014, 10 2014.

[63] J. Buitrago and S. Asfour, "Short-Term Forecasting of Electric Loads Using Nonlinear Autoregressive Artificial Neural Networks with Exogenous Vector Inputs," *Energies*, vol. 10, p. 40, 01 2017.

[64] W. X. J. Cheng and L. Ai, "Electric Load Forecasting Based on Improved Grey Neural Network," in *Recent Advances in Computer Science and Information En-*

gineering, ser. Lecture Notes in Electrical Engineering, Z. Qian, L. Cao, W. Su, T. Wang, and H. Yang, Eds., vol. 124, pp. 651–655.

[65] H. Li, Y. Zhu, J. Hu, and Z. Li, "A localized NARX Neural Network model for Short-term load forecasting based upon Self-Organizing Mapping," in *2017 IEEE 3rd International Future Energy Electronics Conference and ECCE Asia (IFEEC 2017 - ECCE Asia)*, 06 2017, pp. 749–754.

[66] G.-B. Huang, Q.-Y. Zhu, and C. Siew, "Extreme Learning Machine: Theory and Applications," *Neurocomputing*, vol. 70, 12 2006.

[67] Y. Chen, M. Kloft, Y. Yang, C. Li, and L. Li, "Mixed kernel based extreme learning machine for electric load forecasting," *Neurocomputing*, vol. 312, 05 2018.

[68] S. Dash and D. Patel, "Short-term electric load forecasting using Extreme Learning Machine - a case study of Indian power market," in *2015 IEEE Power, Communication and Information Technology Conference (PCITC)*, 10 2015, pp. 961–966.

[69] O. Ertugrul, "Forecasting electricity load by a novel recurrent extreme learning machines approach," *International Journal of Electrical Power & Energy Systems*, vol. 78, pp. 429–435, 06 2016.

[70] P. García Laencina, "Improving Predictions Using Linear Combination Of Multiple Extreme Learning Machines," *Information technology and control*, vol. 42, pp. 86–93, 03 2013.

[71] M. Albadr and S. Tiuna, "Extreme Learning Machine: A Review," *International Journal of Applied Engineering Research*, vol. 12, pp. 4610–4623, 01 2017.

[72] W. Ting and L. Xueyong, "Research on short-term electric load forecasting based on extreme learning machine," *E3S Web of Conferences*, vol. 53, p. 02009, 01 2018.

[73] Y. Wei, H. Huang, B. Chen, B. Zheng, and Y. Wang, "Application of Extreme Learning Machine for Predicting Chlorophyll-a Concentration Inartificial Upwelling Processes," *Mathematical Problems in Engineering*, vol. 2019, pp. 1–11, 05 2019.

[74] Z. Yang, T. Zhang, J. Lu, Y. Su, D. Zhang, and Y. Duan, "Extreme learning machines for regression based on V-matrix method," *Cognitive Neurodynamics*, vol. 11, pp. 1–13, 06 2017.

[75] V. N. Vapnik, *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.

[76] Bo-Juen Chen, Ming-Wei Chang, and Chih-Jen lin, "Load forecasting using support vector machines: a study on eunite competition 2001," *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1821–1830, 2004.

[77] Y. Fu, Z. Li, H. Zhang, and P. Xu, "Using Support Vector Machine to Predict Next Day Electricity Load of Public Buildings with Sub-metering Devices," *Procedia Engineering*, vol. 121, pp. 1016–1022, 12 2015.

[78] W.-C. Hong, "Electric load forecasting by support vector model," *Applied Mathematical Modelling*, vol. 33, pp. 2444–2454, 05 2009.

[79] Z. Hu, Y. Bao, and T. Xiong, "Electricity Load Forecasting Using Support Vector Regression with Memetic Algorithms," *The Scientific World Journal*, vol. 2013, p. 292575, 12 2013.

[80] S. Qiang and Y. Pu, "Short-term power load forecasting based on support vector machine and particle swarm optimization," *Journal of Algorithms & Computational Technology*, vol. 13, p. 174830181879706, 10 2018.

[81] S. Maldonado, A. González, and S. Crone, "Automatic time series analysis for electric load forecasting via support vector regression," *Applied Soft Computing*, vol. 83, p. 105616, 07 2019.

[82] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*. Cambridge University Press, 03 2000.

[83] N. Deng, Y. Tian, and C. Zhang, *Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions*, ser. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. CRC Press, 2013.

[84] I. Steinwart and A. Christmann, *Support vector machines*, ser. Information science and statistics. Springer, 2008.

[85] J. A. K. Suykens, *Least squares support vector machines*. World Scientific H1 - British Library H2, 2002.

[86] P. L. Anderson, *Business Economics and Finance with MATLAB, GIS, and Simulation Models*. CRC Press, 2004.

[87] M. F. Azeem, *Fuzzy Inference System: Theory and Applications*. IntechOpen, 2012.

[88] J. H. Chow, F. F. Wu, and J. A. Momoh, *Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence*. Springer US, 2004.

[89] S. K. Halgamuge and L. Wang, *Computational Intelligence for Modelling and Prediction*. Springer Berlin Heidelberg, 2005.

[90] R. Jensen and Q. Shen, *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*. Wiley, 2008.

[91] S. Kalogirou, *Artificial Intelligence in Energy and Renewable Energy Systems*. Nova Science Publishers, 2007.

[92] A. Konar and D. Bhattacharya, *Time-Series Prediction and Applications: A Machine Intelligence Approach*. Springer International Publishing, 2017.

[93] E. Ogliari and S. Leva, *Computational Intelligence in Photovoltaic Systems*. MDPI AG, 2019.

[94] A. K. Palit and D. Popovic, *Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications*. Springer London, 2012.

[95] M. Paulescu, E. Paulescu, P. Gravila, and V. Badescu, *Weather Modeling and Forecasting of PV Systems Operation*. Springer London, 2012.

[96] W. Pedrycz and S. M. Chen, *Time Series Analysis, Modeling and Applications: A Computational Intelligence Perspective*. Springer Berlin Heidelberg, 2012.

[97] S. A. Soliman and A. M. Al-Kandari, *Electrical Load Forecasting: Modeling and Model Construction*. Elsevier Science, 2010.

[98] Y. H. Song, *Modern Optimisation Techniques in Power Systems*. Springer Netherlands, 1999.

[99] M. Sudha, *Applied Computational Intelligence*. Educreation Publishing, 2019.

[100] G. Tayfur, *Soft Computing in Water Resources Engineering: Artificial Neural Networks, Fuzzy Logic and Genetic Algorithms*. WIT Press, 2014.

[101] K. E. Voges and N. Pope, *Business Applications and Computational Intelligence*. Idea Group Publishing, 2006.

[102] J. Wang, *Business Intelligence in Economic Forecasting: Technologies and Techniques: Technologies and Techniques*. Information Science Reference, 2010.

[103] L. Wang, C. Singh, and A. Kusiak, *Wind Power Systems: Applications of Computational Intelligence*. Springer Berlin Heidelberg, 2010.

[104] P. P. Wang, *Computational Intelligence in Economics and Finance*. Springer Berlin Heidelberg, 2013.

[105] A. Ali, M. Tayeb, and Z. Shamseldin, "Short Term Electrical Load Forecasting Using Fuzzy Logic," *International Journal Of Advancement In Engineering Technology, Management and Applied Science (IJAETMAS)*, vol. 03, pp. 131–138, 11 2016.

[106] A. Danladi, M. Yohanna, M. Puwu, and B. Garkida, "Long-term load forecast modelling using a fuzzy logic approach," *Pacific Science Review A: Natural Science and Engineering*, 11 2016.

[107] M. Faysal, M. Islam, M. M. Murad, M. Islam, and M. Amin, "Electrical Load Forecasting Using Fuzzy System," *Journal of Computer and Communications*, vol. 07, pp. 27–37, 09 2019.

[108] M. Singla and S. Hans, "Load Forecasting using Fuzzy Logic Tool Box," *Global Research and Development Journal for Engineering*, vol. 38, pp. 12–19, 07 2018.

[109] L. Yao, L. Yu, and J. Xiao, "Short-Term Power Load Forecasting by Interval Type-2 Fuzzy Logic System," in *Information Computing and Applications - Second International Conference, ICICA 2011, Qinhuangdao, China, October 28-31, 2011. Proceedings, Part II*, vol. 244, 10 2011, pp. 575–582.

[110] Z. Ismail and R. Mansor, "Fuzzy Logic Approach for Forecasting Half-hourly Electricity Load Demand," *Fuzzy Logic Approach for Forecasting Half-hourly Electricity Load Demand*, 09 2011.

[111] J. Uddin, D. Hadidjaja, I. Sulistiyowati, E. Suprayitno, I. Anshory, and S. Syahrorini, "Very short term load forecasting peak load time using fuzzy logic," *IOP Conference Series: Materials Science and Engineering*, vol. 403, p. 012070, 10 2018.

[112] J. Kaur and Y. S. Brar, "Short term load forecasting using fuzzy logic of 220KV transmission line," *International Journal of Engineering Research & Technology (IJERT)*, vol. 3, no. 2278, 2014.

[113] A. Laouafi, M. Mordjaoui, and T. E. Boukelia, "An adaptive neuro-fuzzy inference system-based approach for daily load curve prediction," *Journal of Energy Systems*, pp. 115–126, 09 2018.

[114] S. K. Patel and S. Sharma, "A Review of very Short-Term Load Forecasting (STLF) using Wavelet Neural Networks," *International Journal of Science, Engineering and Technology Research (IJSETR)*, vol. 4, no. 2, 2015.

[115] M. Mitchell, *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, 01 1998.

[116] R. L. Haupt and S. E. Haupt, *Practical genetic algorithms*, 2nd ed. Wiley-Interscience, 2004.

[117] Y. Al-Douri, H. Al-Chalabi, and J. Lundberg, "Time Series Forecasting using Genetic Algorithm A case study of maintenance cost data for tunnel fans," in *The Twelfth International Conference on Advanced Engineering Computing and Applications in Sciences*, 08 2018.

[118] J. Carr, "An Introduction to Genetic Algorithms," 2014.

[119] C.-C. Hsu, C.-H. Wu, S.-C. Chen, and K.-L. Peng, "Dynamically Optimizing Parameters in Support Vector Regression: An Application of Electricity Load Forecasting," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, vol. 2, 02 2006, pp. 30c– 30c.

[120] D. Beasley, D. Bull, and R. Martin, "An Overview of Genetic Algorithms : Part 1, Fundamentals," *University Computing*, vol. 15, no. 2, pp. 56–69, 10 1994.

[121] R. Aquino, O. Neto, M. Lira, A. Ferreira, and K. de Farias Santos, "Using Genetic Algorithm to Develop a Neural-Network-Based Load Forecasting," in *Artificial*

Neural Networks - ICANN 2007, 17th International Conference, Porto, Portugal, September 9-13, 2007, Proceedings, Part II, 09 2007, pp. 738–747.

[122] D. P. Sarangi and A. Gupta, "Electrical load forecasting using genetic algorithm based back-propagation method," *ARPN Journal of Engineering and Applied Science*, vol. 7, pp. 1017–1020, 08 2012.

[123] G. M. Khan, F. Zafari, and S. Mahmud, "Very Short Term Load Forecasting Using Cartesian Genetic Programming Evolved Recurrent Neural Networks (CGPRNN)," in *Proceedings of the 2013 12th International Conference on Machine Learning and Applications - Volume 02*, 12 2013, pp. 152–155.

[124] F. Li and X. Zhao, "The application of genetic algorithm in power short-term load forecasting," in *2012 International Conference on Image, Vision and Computing (ICIVC 2012)*, 2012.

[125] J. Silva, B. Cervantes-Bolivar, I. Isaac, G. Valencia, and Y. Cardenas Escorcia, "Demand Energy Forecasting Using Genetic Algorithm to Guarantee Safety on Electrical Transportation System," *Chemical Engineering Transactions*, vol. 67, pp. 787–792, 10 2018.

[126] Y. Lin, H. Luo, D. Wang, H. Guo, and K. Zhu, "An Ensemble Model Based on Machine Learning Methods and Data Preprocessing for Short-Term Electric Load Forecasting," *Energies*, vol. 10, no. 1186, 2017.

[127] X. Zhanga, J. Wanga, and K. Zhang, "Short-Term Electric Load Forecasting Based on Singular Spectrum Analysis and Support Vector Machine Optimized by Cuckoo Search Algorithm," *Electric Power Systems Research*, vol. 146, no. 2017, pp. 270–285, 2017.

APPENDICES

APPENDIX A

Nomenclature

Table A.1 contains a list of the used abbreviations.

Table A.1: Nomenclature

ADD	Average Absolute Deviation
ACF	Autocorrelation Function
ACVF	Autocovariance Function
AICC	Akike's Information Criterion
AMR	Automatic Meter Reading
ANN	Artificial Neural Network
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ARMAX	Autoregressive Moving Average with Exogenous Parameters
AWGN	Additive White Gaussian Noise
CDF	Cumulative Distribution Function
CO_2	Carbon Dioxide
DBT	Dry-Bulb Temperature
DSM	Demand Side Management
DSO	Distribution System Operator
DWT	Discrete Wavelet Transformation
ELM	Extreme Learning Machine
ETS	Emission Trading System
EU	European Union
FF	Feed-Forward
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FPE	Final Prediction Error
GNA	Gauss-Newton Algorithm
ICAP	Installed Capacity Markets
IQR	Interquartile Range
kWh	kilowatt hour

LM	Load Management
LMP	Locational Marginal Price
MA	Moving Average
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MAPE	Mean Absolute Error
MCP	Market Clearing Price
MCV	Market Clearing Volume
MPE	Mean Percentage Error
MSE	Mean Square Error
MWh	Megawatt hour
NEM	National Electricity Market
NRMSE	Normalized Root Mean Square Error
OTC	Over-The-Counter
PACF	Partial Autocorrelation Function
PPA	Power Purchase Agreement
RH	Relative Humidity
RMSD	Root Mean Square Deviation
RMSE	ROot Mean Square Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
SARIMAX	Seasonal Autoregressive Integrated Moving Average with Exogenous Parameters
SIC	Schwarz Information Criterion
SMAPE	Symmetric Mean Absolute Percentage Error
SVM	Suport Vector Machine
TSO	Transmission System Operator
WBT	Wet-Bulb Temperature
WNN	Wavelet Neural Networks
ZMCP	Zonal Market Clearing Price