UNIVERSITY OF LJUBLJANA
FACULTY OF ECONOMICS

MASTER'S THESIS

SHORT-TERM LOAD FORECASTING FOR INDUSTRIAL AND RESIDENTIAL CONSUMERS

Ljubljana, September 2016 RENATA BLATNIK

AUTHORSHIP STATEMENT

The undersigned Renata Blatnik, a student at the University of Ljubljana, Faculty of Economics, (hereafter: FELU), author of this written final work of studies with the title Short-term load forecasting for industrial and residential consumers, prepared under supervision of associate professor Aleš Berk Skok, PhD

DECLARE

1. this written final work of studies to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies to be language-edited and technically in adherence with the FELU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the FELU's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the FELU in accordance with the relevant FELU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the FELU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, September 28, 2016 Author's signature:

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Problem description and identification of the subject of research. Electricity as a commodity differs from other tradeable goods as it cannot be reliably stored in large quantities. Energy has to be consumed as soon as it is produced, and power demand and supply have to remain essentially balanced in real time in order for the grid frequency to remain stable. The growing portion of *Renewable Energy Sources* (hereinafter RES) in the grid further contributes to the instability of the electrical power system (generation, transmission, consumption).

Energy consumption has to be estimated and announced in advance; however, deviations from these values need to be balanced in real time by either using reserve capacities or buying the missing quantities on the balancing market. Balancing energy is considerably more expensive than energy bought in advance.

To reduce operation costs, *Distribution System Operators* (hereinafter DSOs) started offering different tariffs as early as the 1980s to motivate customers to defer consumption at times of highest demand and highest prices observed on the market. These were the first forms of *Demand Side Management* (hereinafter DSM). The newly liberalized market has set the stage for a new business model: the announced but not consumed energy of a DSO's customers can profitably be offered on the balancing market. The *Demand Response* (hereinafter DR) programs motivate customers to actively defer their consumption, either as a result of price incentives or fair compensation. Customers can also just practice *Energy Efficiency* (hereinafter EE), which is the passive version of DSM. *Virtual Power Plants* (hereinafter VPPs), sophisticated software solutions for DR programs, which connect consumers, prosumers, system operators and the balancing market into a *Smart Grid* (hereinafter SG), have emerged only recently.

With the adoption of a package of laws, called the *Third Energy Package* in 2009 the European Union (hereinafter EU) introduced a legislation that aims gradually to liberalize the electricity market. Its goal is to establish a reliable and competitive supply. Competitiveness is being achieved by separating the ownership of channels of production, sales, and transmission, as well as the introduction of an independent system operator and independent distributors. Furthermore, an internal European electricity market and balancing market are established. Each European country has to establish its own national regulatory authority and they all cooperate within the framework of the *Agency for the Cooperation of Energy Regulators* (hereinafter ACER), codified in the "European Parliament and Council Regulation establishing the Agency for the Cooperation of Energy Regulators" (2009) (hereinafter "2009/713/EC ACER Regulation").

The new legislation incentivizes the development of an integrated European balancing mech-

anism. In this context, in 2011, ACER started developing the Framework Guidelines on Electricity Balancing. Judging from ACER's statements, DR is expected to play a significant role in the future integrated balancing market. The guidelines will enable the emergence of new business models, such as VPPs allowing DR programs and distributed generation resources to compete on an equal footing (Cordis, n.d.).

In the context of VPP, *short-term load forecasting* (hereinafter STLF) plays a central role since, for the VPP to successfully bid in the balancing market, it has to know the precise load of its DR participants (and the quantity they will be able to curtail) for the next 24 hours at least. The goal of STLF is to predict the electrical load for at least one hour and up to a few days ahead. The aim is to ensure a secure supply of electrical energy while minimizing daily operating and distribution costs.

In addition to STLF, there are also

- *mid-term load forecasting*: for which the goal is to predict weekly, monthly and yearly load peaks, which enables efficient operational planning, and
- *long-term load forecasting*: which is used for predicting consumption beyond one year and up to a few years in advance; it plays a vital role in generation, transmission and distribution network planning.

These two forms of forecasting are not the subject of this work. Each of these fields has its specific characteristics that require different approaches/methods to solving problems.

While long-term load forecasting was an integral part of system planning from its beginnings, short-term load forecasting started to gain traction since the emergence of the free market. Maintaining constant grid frequency at all times and thus ensuring the grid's stability is becoming an increasingly challenging task as the traditional power grid is increasingly evolving into a more complex future grid that includes RES, micro grids that are able to decouple from the *RES!see Renewable Energy Sources* main grid and function in island mode, and so on. For these grid forms, fast and accurate load forecasting at all levels is crucial. In addition to the load, the production of RES needs to be forecast. The future will bring even more complex forecasting situations as DSM and DR will have to be considered in the forecasting process. The field of forecasting and STLF has also been gaining importance and is developing the academic realm quickly. The number of articles is increasing rapidly, and this year the 36th International Symposium on Forecasting will be held.

The vast majority of STLF literature is related to forecasting load time series on a *large scale*, that is data that is highly spatially aggregated, usually on the level of a utility, DSO or even a whole country. The present master's thesis is an attempt at short-term load forecasting *indi-*

vidual industrial and residential consumers' loads, to be used in VPPs. It covers forecasting on a *very short scale*, such as a community of a couple dozen houses, as well as on the *individual* level. In a VPP setting, there is a need in the *forecasting module* to predict each consumer's electricity load for the next 24 hours, based on his historical consumption data. The results of this module serve as input data for the *optimization module*, which determines the optimal *activation* (a subset of consumers who are asked to reduce their consumption) at a DR event, based on the predicted load of all consumers, taking into account some limitations while minimizing activation costs.

Purpose of this research. The purpose of this master's thesis is to study the problem of short-term load forecasting for individual residential and industrial consumers, to master statistical methods suitable for addressing the problem, to establish their performance on real data and finally to select a method that on average delivers the best forecast accuracy and that is sufficiently robust to be used in an automated environment. Additional aspects of automation also need to be anticipated and satisfactorily solved, such as missing data imputation, anomaly and outlier detection, the presence of holidays and extraordinary events.

Research questions. The following research questions will be pursued:

- How to predict with satisfactory accuracy the hourly load consumption for the next 24 hours for any *industrial* or *residential* consumer?
- Which method should best be applied for this task?

An outline of research objectives. The fundamental *research objective* is to determine a statistical method or methods, that can predict, with sufficient accuracy for the requirements of a VPP, the load consumption for the next 24 hours of any individual household or industrial consumer. Where that is not possible, why no such method exists should be determined and preferably theoretically explained, and a suitable practical solution should be proposed.

The *core objective* is to determine the type of underlying statistical problem, to select and present each statistical method deemed appropriate for solving this problem, and to test their efficiency on acquired real data and to select among them the one which provides on average the most accurate predictions. Throughout, I will seek not only to specify the results obtained but also to establish root causes and to explain in a theoretical manner what the most probable reasons are for the results obtained.

Identification of research methodologies. The acquired *input data* is primary data, collected by sampling electric energy at automated smart meters, every 15 minutes. The measurements were collected for 123 industrial and 235 residential anonymous consumers between January 2011 and August of 2013 and represent time series of load demand. I will apply an *empirical*

data analysis methodology modified for time series forecasting.

After becoming familiar with the field of electricity and forecasting in general, the relevant scientific literature on STLF will be reviewed. By visually observing data and further exploratory analysis, I will try to understand the characteristics of the data, which will lead the selection of appropriate statistical methods to be applied.

The performance of tested methods will be estimated primarily by forecast accuracy, computational complexity and suitability for use in an automatic environment. If individual forecasting will not be feasible, an attempt to forecast aggregated data of several users will be made. Throughout the process, I will apply all my knowledge of statistics and my experience.

Structure of the master's thesis. The *first chapter* sets the broader context of this work. We briefly present the electric power system and the relatively recent transition of the European electricity market from a traditional and regulated market to a completely liberalized market. The chapter is concluded by a description of the balancing market, which sets the stage for virtual power plants in the following chapter.

The *second chapter* narrows the setting further. It introduces the concept of demand side management and its active version, the demand response. DR programs are also offered by virtual power plants as providers of advanced software solutions. A demand response event, as the central event in DR programs, is described by its phases. In this context, Customer Baseline Load (hereinafter CBL) and its forecasting play a central role. The chapter is concluded by the introduction of the European research project called eBadge which is the framework of this work.

The *third chapter* starts by investigating forecasting, in general, continuing to the specific field of load forecasting. The principles of a forecasting procedure are first outlined, then the concept of a mathematical load model is described, including the stepwise approach to building one. Forecast accuracy measures are compared and the most suitable one is chosen. Finally, major research work in short-term load forecasting is presented.

The *fourth chapter* covers all analytical work done prior to the actual forecasting. Data is prepared for processing. This includes missing data imputation, outlier detection and special events handling. Producing time plots for loads of individual industrial and residential consumers reveals the trend and seasonal patterns contained in a given consumer's load data. This is investigated further by decomposing signals into single patterns.

The *fifth chapter* describes the software used and all methods that were applied. They range

from very simple average-based methods in Section 5.1, followed by the centrally important time series models. These are classified into popular exponential smoothing methods on one side, and the *Autoregressive Integrated Moving Average* (hereinafter ARIMA) models on the other. As these methods are crucial, they are covered in depth. Finally, two alternative approaches are presented for comparison: the contemporary neural network AR models and the U.S. standards baseline, as a straightforward, robust and simple-to-implement method.

Chapters six to eight present results from forecasting the load of individual industrial consumers, individual households and groups of households, respectively. Method performance is compared based on time complexity and the chosen forecast accuracy measure *Mean Absolute Percentage Error* (hereinafter MAPE). *Chapter eight* is central to finding the most appropriate method and covers most of the methods. Their selection is guided by two insights gained from the previous two chapters: the fact that the load of individual households is unpredictable and the decision that the method that is ultimately selected for forecasting the loads of groups of households will also be used in the process of forecasting loads of individual industrial consumers. The chosen method, `tbats()` function, is a very sophisticated state space model, able to handle multiple seasons of the load signal.

Part of this work was carried out at the time of my employment at the XLAB d.o.o. company which gave me the opportunity to study this exciting field.

Throughout this text **bold** will be used to denote the first appearance of a major term, *italic* font is used to introduce a less important term or to emphasize parts of a sentence. `monospace` font is used for annotating `R` functions and packages. `R` functions are given with parentheses, for example `bats()`.

1 ELECTRICITY MARKET

The **electricity market**, as part of the energy market, in the EU has undergone huge changes in the last decade. Since mid-2007 when the last provisions of the second liberalization directives came into force ("Parliament & Council Directive concerning internal market in electricity's common rules", 2003), the EU energy market is, at least on paper, completely *liberalized*. *Production and supply* have essentially become market activities in which various participants compete for market shares, as well as trade in organized markets, such as energy exchanges that offer energy products, as well as derivatives. In truth, liberalization is an ongoing process, not without obstacles, which has so far reached different stages of implementation across EU member states.

1.1 Electric Power System

Electrical power is produced by transforming energy, derived from various energy sources, such as nuclear energy, fossil fuels (coal, gas, and oil), into electricity. Alongside **traditional power generating facilities**, more recently the use of **Renewable Energy Sources** (hereinafter RES), such as bioenergy, hydropower, wind, solar and geothermal energy, is continually on the rise. Electricity is transmitted to the end consumer through the **power grid**.

The first part of the network, called the **transmission grid**, operates at high and extra-high voltage and is dedicated to *long-distance* transmission. It is managed by **Transmission System Operators** (hereinafter TSOs) (EUETS, n.d.).

The second part of the network, called the **distribution grid**, operates partly at high voltage, but mostly at medium and low voltages, and is dedicated to the *regional* distribution of electricity, supplying it to lower-level distribution systems and to directly connected customers. One can think of it as the capillary part of the power grid. It is run by **Distribution System Operators** (hereinafter DSOs). The thresholds, determining the delimitation between a transmission system and a distribution system, are established at the national level.

Both TSOs and DSOs (commonly called system operators) are legally bound to ensure the security and reliability of supply of their part of the system. (EUETS, n.d.) The network of electrical components used to supply, transfer, and use electric power as a whole is called an **electric power system** (hereinafter EPS) also known as a **power grid**.

1.2 Electricity Market

1.2.1 The energy market – traditional model

Traditionally, electrical power was supplied by state-owned companies, which covered all activities (production, transmission, distribution, and retail) along the supply chain and bore responsibility for supplying all custumers of a certain geographic area. They were *vertically integrated organizations*, acting as autonomous EPSs in their own right. These ”traditional versions” of a TSO centrally dispatched production, forecast the cumulative load (on the company level), and balanced the deviations between production and consumption. As these companies had a market monopoly, they were not motivated to reduce production costs or to introduce innovative technologies. All costs were reflected in the end consumer’s price. The prices were also elevated to include an addition for investments into expanding and modernizing the infrastructure and were not competitive.

1.2.2 Liberalization of European electricity market

Starting in 1999, the EU initiated a process of market liberalization. In 2009, the latest round of market legislation was adopted with the codification of the ”Parliament and Council Directive concerning internal market in electricity’s common rules” (hereinafter ”Directive 2009/72/EC”). The directive, also known as the **Third Energy Package**, introduced legislation that aims gradually to liberalize the energy market. This will increase competition, allowing new players to join the sector, and, ultimately, liberalize energy prices. Its ultimate goal is to establish a competitive, reliable and sustainable supply of energy that will be driven by innovation and new technologies. Serena (2014) studied the historical motivation for introducing this legislation, the challenges of its introduction and issues/benefits concerning consumers.

The EU is aiming to reduce the concentration of power in the energy market, which it intends to achieve by separating the ownership of the channels of production, sales, distribution, and transmission (unbundling); the ownership of these companies will be private. There is also a separation within the same activities (vertical separation) where competition in the wholesale market allows for new producers to enter the market and new business models to emerge in the retail market.

Production and sales have become commercial activities as the TSO and DSOs have become independent and regulated. Across EU member states the situation regarding DSOs varies extensively; some member states have hundreds of DSO’s, whereas others have only one or two. The transmission system is accessible to all players under the same conditions. For regulation purposes, every EU member state establishes its own independent National Regulatory Authority. These cooperate within the framework of the **Agency for the Cooperation of Energy Regulators** (hereinafter ACER) (Market legislation, 2009).

The new situation in the market is already enabling new business models, which are competing with traditional ones. Investments into energy infrastructure are being made at an increased pace, with the objective of increasing cross-border trading and access to diverse sources of energy. A **European Internal Electricity Market** is envisioned, but not yet implemented. It will be “implemented bottom-up through regional market coupling projects and top-down through the network codes that ACER, EC and ENTSO-E develop” (ENTSO-E, 2014).

Currently, electrical energy produced, is traded at various regional electricity markets, such as:

- **long-term market**: standardized products of electrical energy are traded for up to three

years in advance. The products offered are weekly, monthly, quarterly and yearly products of base, peak, and off-peak energy. Retailers buy most electrical energy at these markets, in order to avoid price risk.

- **day-ahead market**: insufficiently bought electrical energy, according to a DSO's prediction is bought a day in advance, at elevated prices but in order to minimize loss.
- **intra-day market**: is similar to the day-ahead market, but with the difference that energy is bought at very short notice and delivered one or two hours later. The prices are obviously even higher.
- **balancing market**: a TSO settles imbalances between demand and supply after markets have closed (gate closure). This happens in and near real time.

The day-ahead and intra-day markets are examples of *spot markets*. All the above-listed markets are *physical markets* where the commodity is physically delivered. Apart from these markets, there are *derivatives markets* where prices of the derivative instruments depend on prices of the underlying electricity products. The reasons for trading can be to hedge electricity price risk, or even for speculative reasons. A more detailed description of the above listed markets, with focus on risk management can be found in (Paravan, 2004).

1.3 Balancing Market

Electricity as a commodity is different from other tradeable goods. Since there are still no reliable technologies available for storing large quantities of energy ("bulk storage"), energy has to be consumed as soon as it is produced. Power demand and supply have to remain essentially **balanced in real time** for the grid frequency and the whole power grid to remain stable. (Magliavacca et al., 2015, p. 12). Even the smallest deviations from the prescribed 50 Hz may result in system instability and occasionally even black-outs. Therefore, market players are subject to very strict balancing rules, and inability to fulfill the requirements is heavily penalized.

The inclusion of RES, such as solar and wind generation, into the grid represents new challenges for network operators' daily balancing activities. RES display intermittent behavior, which affects flows within the grid, making it difficult to ensure the security of supply, to guarantee the availability of an adequate reserve margin to cope with unforeseeable events and to ensure that electricity markets function in a proper way (Magliavacca et al., 2015, p. 11).

TSOs have a legal obligation to keep the grid frequency constant at all times. Frequency is **regulated** by means of automatic and manual regulation mechanisms. The TSO holds in store active power capacity made available by producers, which can be activated at any time to bring balance into the grid.

There are several levels of reserves, and they differ by function and response time. In the EU, the *primary reserve* is intended to stop the frequency drift in the case of unpredictable major events (for example a power plant going down). The *secondary reserve* brings the frequency back to its nominal value. The *tertiary reserve* is able to solve longer-term imbalances, which last up to a few hours. The secondary and tertiary reserves are part of a larger network safeguard called **ancillary services**.

Before market liberalization, ancillary services were the only way to **balance** the demand and supply of electrical energy. The overall balancing services' procurement costs were naturally transferred to the price the end users were paying, partly in the form of *imbalance charges* and partly in the form of *network charges*. The new legislation introduced a **balancing market** where demand and supply for balancing electrical energy meet. Trading is done by placing bids to buy and offers to sell energy. The origin of the balancing energy can be a *power generating facility* or *demand facility* (these will be discussed in the next chapter).

Currently, there are only **national** balancing/reserve markets in all EU states. A **pan-European Exchange of Balancing Energy** is envisioned, but, due to the scale and complexity of this task, a phased approach has been adopted. The first phase will be a coordination on a regional basis, followed by a merging of these regional balancing markets. Balancing cost varies by country and is dependent on multiple factors, such as market size, the share of RES in the grid, and so on. It is expected that the pan-European balancing market will benefit all participating countries as they will be able to balance supply and demand more accurately and also reduce their balancing costs.

2 DEMAND RESPONSE PROGRAMS

In this chapter, we will learn about DR Programs, within a VPP environment. These are advanced approaches to stimulate customers to defer their consumption in times of highest market demand. We will discuss phases of a DR event and introduce the central term that will be our focus in the following chapters, Customer Baseline Load. We will also describe the current state of DR programs in the U.S., Europe, and Slovenia.

2.1 Demand Side Management

To ensure stability of the electricity grid, electricity supply and demand must remain balanced in real time. Traditionally, utilities have reached back to peaking power plants, asking them to increase their power generation to meet rising demand. To prevent power outages, **Demand Side Management** was introduced at the beginning of the 1980s. The term covers any activity that helps reduce or conserve energy use. DSM comprises **Energy Efficiency**, **Demand Response**, and even ordinary retail rates (tariffs). DR programs will be discussed in depth in Section 2.2.

More recently, DSM technologies became increasingly feasible due to the integration of ICT and the power system, resulting in a new term: **smart Grid** (Enernoc, 2014). It connects TSOs, DSOs, RES, and end consumers into a new, contemporary type of grid. It has been established that when the share of RES in the power grid surpasses 30%, energy balancing becomes a huge challenge due to the *unpredictable* nature of RES power generation, which disrupts the system's frequency. A smart grid is able to manage that.

In Slovenia, the concept of smart grids is fairly new. A fundamental study was published by Kosmač et al. in 2010 that describes all elements of the smart grids concept in detail and lays out a vision / execution plan of introducing it into the Slovenian electricity market and households.

2.1.1 Virtual power plants

A VPP is a software platform that follows a new business model, which only became possible in the liberalized market. It reaps benefits from the fact than by consuming less than the estimated and announced energy, the curtailed energy can be profitably sold at the balancing market. A VPP, therefore, offers DR programs to interested consumers / prosumers that have a load profile suitable for aggregation.

Prosumer is a term, coined from "consumer" and "producer". These are for example customers, who own a wind or solar power station. By offering curtailed electrical energy to the balancing market, a VPP becomes a *balancing asset*. Frequently, VPPs are owned and run by a distribution utility and maximize value for both the end user and the distribution utility using a sophisticated set of software-based systems. They are dynamic and can react quickly to a change in market demand.

2.2 Demand Response Event

DR programs stimulate consumers to reduce their energy use at times of peak demand as a result of financial incentives or price information. Requests for demand reduction are called when stress on the grid is significant. They are made for a specific and limited period, also referred to as a **DR Event**.

Each DR event has three key measurement components (Enernoc, 2011):

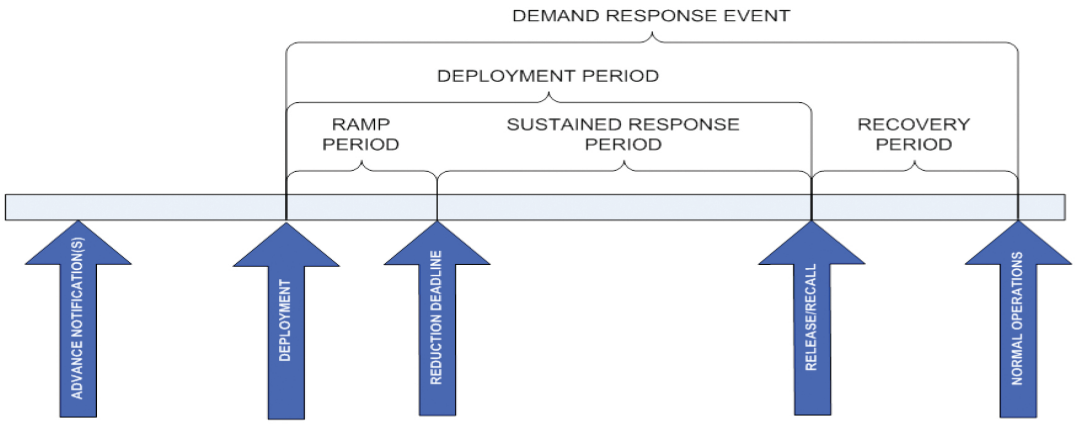
- **baseline** the amount of energy the customer would have consumed in the absence of a DR event,
- **actual use** the amount of energy the customer actually consumed during the DR event period and

- **load reduction** the mathematical difference between the baseline and the actual use:

$$\text{load reduction} = \text{baseline demand} - \text{actual demand}. \quad (1)$$

Demand Response Event Phases. The North American Energy Standards Board (hereinafter NAESB) has developed Figure 1 to clarify phases and standardize the terms of a DR event. The figure is reproduced subject to a limited copyright waiver from NAESB with all rights reserved.

Figure 1. Demand Response Event Phases



Source: E. Winkler et al., Measurement and Verification Standards Wholesale Electric Demand Response Recommendation Summary, 2008, p. 17.

Phases of curtailment in a DR event are (Enernoc, 2011):

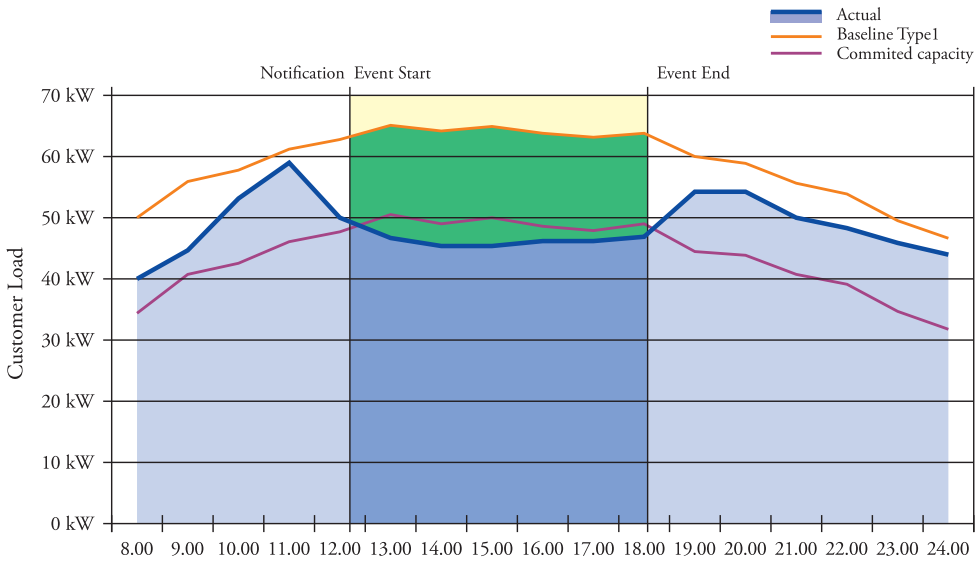
- The **ramp period** is when sites begin to curtail (the beginning of deployment).
- The **sustained response period** is bound by the reduction deadline and the release/recall. In this period, the DR resources are expected to have arrived and to stay at their committed level of curtailment.
- The **recovery period** occurs after customers have been notified that the event has ended and start to resume normal operations.

Notifications are issued to customers/grid operators as an advance notification that the event is likely to take place, as well as when the event is certain to be called and, at the agreed upon schedule, to denote the deployment of the resources.

When a customer enrolls in a DR program, first his **baseline** (yellow line in Figure 2) is generated from his historical load data. Based on the agreed upon curtailment plan, the **committed capacity** line (purple line in Figure 2) is calculated from it. The customer must remain at or below this usage level during the entire DR event.

Customer's curtailment or load reduction is calculated by deducting the actual meter data (blue line in Figure 2) from the baseline during the event period. **Curtailment performance** is tracked by comparing the actual meter data to the committed capacity during the event period. In Figure 2, the deployment occurred at 11:00 and the customer started to decrease energy usage in the preparation of the 12:00 pm reduction deadline. He performed well because the actual meter readings were consistently below the committed capacity during the event. In the following chapters, we will show that such a model for rewarding will not be able to be used for reward individual households.

Figure 2. Baseline and Committed Capacity



Source: Reproduced after Enernoc, The Demand Response Baseline (White paper), 2011, p. 5.

2.3 Customer Baseline Load

Fundamental for operating these programs is the ability to estimate each DR participant's future load. A **Customer Baseline Load** (hereinafter CBL) is by definition an "estimate of the electricity that would have been consumed by a demand resource in the absence of a demand response event. The baseline is compared to the actual metered electricity consumption during the demand response event to determine the demand reduction value" (Grimm, 2008).

In the U.S., in order to unify terminology and procedures for DR programs, efforts for standardization were initiated by the Federal Energy Regulatory Commission (hereinafter FERC). Agreements have been reached as to what constitutes a "good" baseline. It has the following properties (Enernoc, 2010):

- **accuracy**: customers should receive credit for no more and no less than the curtailment they provide,
- **simplicity**: the baseline and the resulting curtailment calculations should be simple enough to calculate for all stakeholders, including end-user customers, during events
- **integrity**: baseline methods should protect against attempts to "game the system" and should not encourage irregular consumption.

Calculations of CBL may differ by type of DR program. Furthermore, baseline design for the aggregation of DR resources *for use as a VPP* needs to consider the following issues: what to do when a customer leaves the aggregation and what to do when more resources join the aggregation.

As we will see in the course of this work, it is *impossible* (for statistical reasons) to predict an individual household's load with sufficient accuracy to be used as a baseline. To reward residential participants in DR programs, new business models have to be developed. Typically, a one-time bill credit is provided to customers who sign up for a direct-load control or other type of load-monitoring and/or control program. Sometimes, rebates are offered to replace older inefficient appliances.

2.4 An Overview of Demand Response Programs

2.4.1 Demand response programs in U.S.

Prior to DR in deregulated markets, regulated utilities had large industrial customers on interruptible load tariffs, since at least the mid-1980s. DR in deregulated markets began in 1999 with the New York Independent System Operator's (hereinafter NYISO) Special Case Resource Program for DR to participate in the installed capacity market.

The FERC defined DR in 2007 as *any* changes in load from either a response to price or an incentive to reduce load. Thus, any retail (regulated) utility tariff or wholesale market DR program, and even energy efficiency, can be considered DR.

DR programs differ from state to state, based on their state regulations and specific environ-

mental objectives and targets to be reached.

In the early days of DR in deregulated markets, participation was primarily from large industrial and commercial customers, in this way the largest reductions for the least effort were achieved. That changed for wholesale markets only within the past year or two, so now there is quite a variety of sizes and types of loads in DR programs.

Smaller, residential customers are targeted by *direct load control programs* for electric hot water heaters and air conditioners. In New York City, for example, there are millions of window air conditioner units, and Consolidated Edison (hereinafter ConEd) has a DR program to cycle those units every 10–15 minutes to help flatten out the load on the system. In July 2016, ConEd also partnered with a solar retailer and a battery supplier to start a VPP program that distributes solar power across a *small grid*. Customers can draw power from solar units during the day or sell surplus energy back to ConEd. The storage batteries will protect them against power outages from major storms. ConEd can tap surplus power in the batteries to meet system demands and keep the lights on if a power line fails. Meanwhile, ConEd will recharge the batteries. This program is an important experiment, following a "Reforming the Energy Vision" initiative, adopted two years ago, that requires the state's utilities to adopt new business models. It is believed that providers will make enough profit from their new roles as energy delivery platforms to offset declines in their monopoly rates of return that support the distribution infrastructure (Behr & Rahim, 2016).

2.4.2 Demand response programs in Europe

A few companies in the Slovenian market (among them CyberGrid, since 2010) have actively been supplying DR and VPP technology solutions to be used by utilities, power traders, DSOs, and large industries. Only a couple of DSOs currently operate a VPP. Their DR programs only include industrial consumers; event notifications are issued by telephone, and active participation is requested on the consumer's side. The few VPPs running are profitable. Outside of the VPP setting, early-stage residential DR programs are offered, unfortunately without broad participation yet.

In Europe, the latest EU market liberalization legislation, codified in "Directive 2009/72/EC" (2009) and ACER's ensuing Framework Guidelines on Electricity Balancing introduced the necessary legal ground for the development of an integrated European balancing market, where DR programs and VPPs will play a significant role. Currently, industrially and commercially targeted VPP programs are on the rise across the EU, but it is a slow process, hindered by the mentality and the resistance of big market players.

Furthermore, a three-year research project, called eBadge, was concluded in November 2015. Its main objective was "to propose an optimal pan-European intelligent balancing mechanism that will include VPPs by means of an integrated communication infrastructure that can assist in the management of the electricity transmission and distribution grids in an optimized, controlled and secure manner" (eBadge, 2014). It has demonstrated the feasibility of the model and confirmed that cross-border balancing will benefit all participants. Results were confirmed by a pilot project, that included such entities as: residential and industrial consumers /prosumers, energy providers, DSOs, TSOs, VPPs and possibly others.

2.5 Role of Load Forecasting in Demand Response Programs

Forecasting DR participants' baseline load is fundamental for the existence of these programs as being able to anticipate near future events and to quantify the capacity, available for offering to the balancing market, is at the core of this business. Moreover, fair participant compensation is the major motivation for participating in these programs in the first place.

In the framework of a VPP, every single participant's baseline should be forecast, as well as the aggregated load of all VPP consumers. Production of participating RES is also predicted. The baseline forecasts serve as input data for the VPP's *Optimization module*, which for every DR event selects an optimal subset of customers whose aggregated curtailment will cover the energy amount the VPP is planning to offer to the market. For this reason, VPPs are also called *aggregators*.

3 LOAD FORECASTING

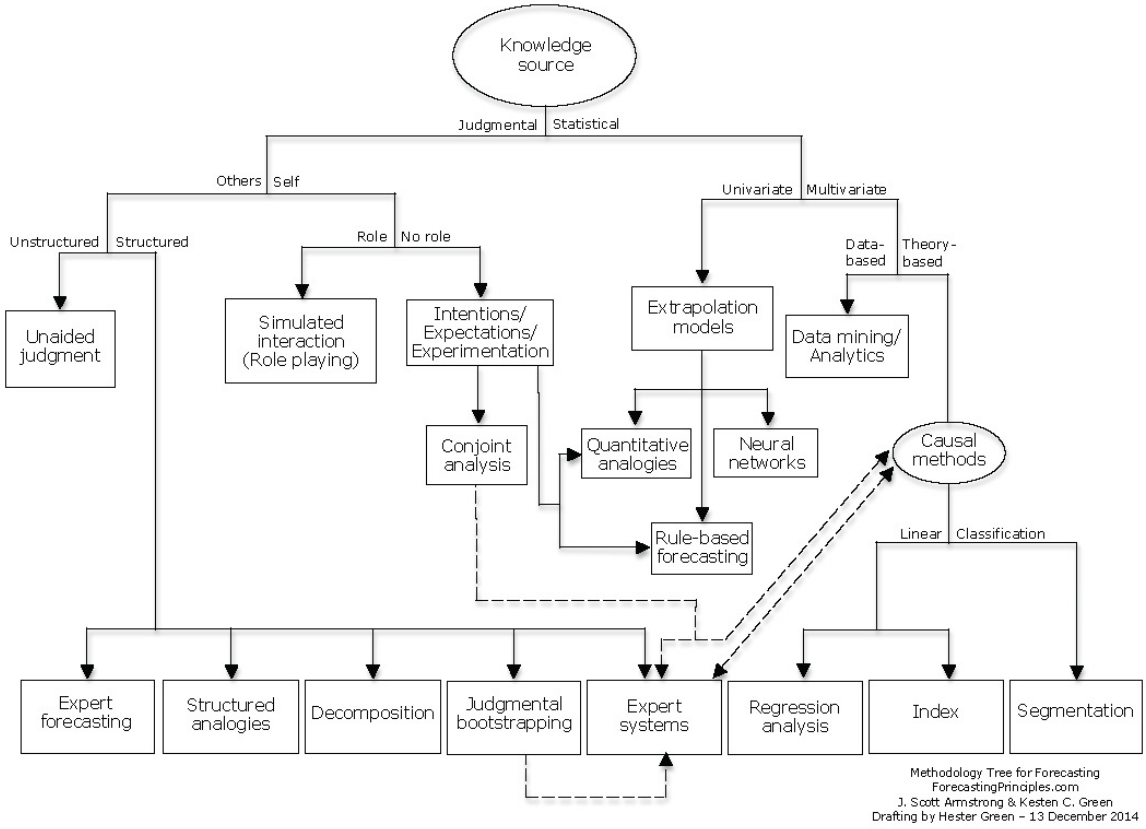
3.1 About Forecasting in General

"Forecasting is about predicting the future as accurately as possible, given all of the information available, including historical data and knowledge of any future events that might impact the forecasts" (Hyndman & Athanasopoulos, 2013, sec.1.2).

The process of forecasting can take on many different approaches that result in different methods. Which approaches are appropriate for a given forecasting situation depends largely on the type of data available. When numerical historical data is available, and we may reasonably assume that some aspects of the past patterns will continue into the future (Hyndman & Athanasopoulos, 2013, sec.1.4), then we will adopt the *quantitative* forecasting approach and use that data for predictions. Otherwise, we will have to take the (less reliable) *qualitative* approach.

Figure 3, reproduced here with the kind permission of J. Scott Armstrong and Kesten C. Green, comprehensively shows the relationships between various groups of methods in a *selection tree for forecasting methods*. In this work, we will only be concerned with the quantitative approach, calling for statistical methods.

Figure 3. Selection tree for forecasting methods



Source: J. S. Armstrong & K. C. Green, Selection Tree for Forecasting Methods, 2014.

Our focus will remain limited to *univariate methods* for which forecasts are based only on present and past values of the quantity being forecast. *Multivariate methods* that additionally include values of one or more external factors (for example, weather) are not included here for two reasons: the satisfactory forecast accuracy of the former methods as well as practical reasons (the care for obtaining weather data and ensuring accurate and timely weather forecasts becomes obsolete). We will show that univariate methods can be a valid alternative to the multivariate methods. Once the historical data has been modeled, the forecasts are computed by *extrapolation*.

Forecasting is a necessary process in every business because it helps to make informed decisions on how to schedule production, transportation, stock and personnel, and provides

a guide to long-term strategic planning (Hyndman & Athanasopoulos, 2013). It also helps track business progress, gives insight into customer behavior, and helps to improve ROI on online advertising by targeting the most promising customer segments.

In today's saturated markets good forecasting is a competitive advantage, so it is no surprise that it has been experiencing a steep increase in the last 15 years. Many scientific journals are exclusively dedicated to this multidisciplinary research field, e.g. Foresight – The International Journal of Applied Forecasting, issued by the International Institute of Forecasters (hereinafter IIF) since 2005, the *International Journal of Forecasting* by Elsevier, and the *Journal of Forecasting*, to name a few. As a research field, forecasting is present in management, behavioral sciences, social sciences, engineering, and other fields.

Good Forecasting Practices. Armstrong, Green, and Graefe (2015) summarized good practices in forecasting into a unifying theory, called *The Golden Rule of Forecasting*. Twenty-eight guidelines were logically derived from the Golden Rule and explained in the paper. Most of the rules have been analyzed regarding their ability to improve forecasting accuracy on 105 scientific papers. It turned out that ignoring any single guideline increased forecast error by more than 40% on average.

Principles of Forecasting (Armstrong, 2001) is the first handbook that summarizes forecasting knowledge, derived from empirical studies and provides guidelines and necessary steps to efficient forecasting in most fields. The book, which is also a comprehensive overview of existing forecasting methods, is a collaborative effort of 40 forecasting researchers and 123 expert reviewers. It contains an extensive forecasting dictionary.

Despite the existence of clear guidelines for good practices, Armstrong et al. (2015, p. 7) conclude that most researchers still ignore cumulative knowledge about forecasting methods and also do not estimate forecast accuracy. This has, in some cases, led to the use of unsuitable methods.

3.2 Load Forecasting

Load forecasting is defined as the science or art of predicting the future load on a given system, for a specified period ahead. It is divided into three different segments (Soliman & Al-Kandari, 2010):

- **Short-term load forecasting**: loads from one hour to seven days ahead are forecast so that daily running and dispatching costs can be minimized.
- **Mid-term load forecasting**: is used to predict weekly, monthly and yearly peak loads up to 10 years ahead so that efficient operational planning can be carried out.

- **Long-term load forecasting**: is used to predict loads from one to several years ahead so that expansion planning can be facilitated.

Each of these three tasks represents a whole different field with its own characteristics and its own set of suitable methods.

3.2.1 Forecasting procedure

First, **historical data|textbf** is gathered by observing the load sequentially over a certain time interval. We will only consider *equidistant* points in time, where data is sampled hourly, daily, weekly, monthly, etc. In mathematical terms, load data is a **time series**, represented by pairs $\{(t_i, y_i)\}_{i=1}^{T}$, where t_i is the i-th point in time and y_i is the electric energy consumed between time t_{i-1} and time t_i. A graphical plot of the load time series of a single consumer/group of consumers is called a **load curve** or **load profile**. Depending on the time interval over which it is plotted, we are speaking about a *daily, weekly, monthly* or *yearly* load curve. Alternatively, we might call it a **load time series**. In forecasting, we want to be able to predict load values beyond the **forecast origin**, which is the time t_T of the last known observation. The forecast values should also refer to *equidistant* points in time, for a specific number of periods, which is called the **forecast horizon** or **lead time**. Only historical data will be used, which is called an *ex-ante forecast|textbf*.

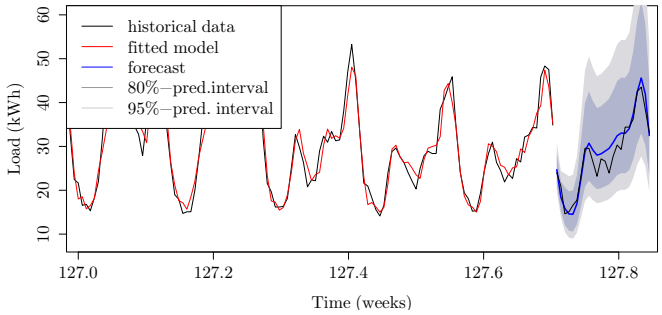
A **forecasting method** is a mathematical procedure for computing forecasts from present and past data. It can simply be expressed in terms of algorithmic rules and need not be based on any underlying stochastic model. Therefore, a "method" should clearly be distinguished from the term "model" (Chatfield, 2000, p. 12).

The general approach to forecasting is to find a **mathematical model** of electricity load that describes the relationship between the load (the *dependent variable*) and the *independent variables*, such as time, weather, economic factors, etc. that influence it. They may also be called *explanatory* or *causal* variables. In load predictions, we will strive to select a most *appropriate load model*, one that takes into account the characteristics of the load curve and the nature of the underlying problem. Most models reflect the underlying stochastic nature of the problem and are **stochastic** models. After a model is selected, the **model parameters** are determined that best fit the relationship between independent and dependent variables. This is usually done with an **optimization technique**. After the mathematical relationship is known, it can be **extrapolated** to the timeframe to be predicted. Any influential variables in that timeframe are either known or predictable, and a **forecast** is made either based on historical data alone or also the predicted values of the explanatory variables.

In Figure 4, all measurements are plotted in black. To the left of the forecast origin, historical

data are displayed, with the fitted model (in red). This model is then used to produce the forecast values (in blue), coming close to the actual values. The two prediction intervals are plotted around the predicted values, which need not lie at the center of the intervals.

Figure 4. Forecasting: Historical and Forecast Data, 80% and 95% Prediction Intervals



3.2.2 The importance of load forecasting

In the energy sector, different forecasts are crucial, load forecasting as well as predicting the production of RES. While long-term load forecasting was an integral part of system planning from the beginnings, short-term load forecasting started to gain traction since the emergence of the free market.

Load research is also a field to which the Association of Edison Illuminating Companies (hereinafter AEIC) has dedicated a separate committee. The company was founded in 1885 by Thomas Edison and his associates and is one of the oldest organizations in the electric energy industry. The AEIC encourages research and the exchange of technical information and best practices, focusing on finding solutions to problems of mutual concern to electric utilities, worldwide. It holds an annual conference dedicated to load research & analytics. (More at: http://aeic.org/.)

3.3 Mathematical Model of Electricity Load

3.3.1 The stochastic process

Statistically speaking, the discrete load time series that we want to predict is unknown at time t_T of forecasting and, as such, can be seen as a sequence of random variables $\{Y_{T+1}, Y_{T+2}, Y_{T+3}, \ldots\}$. The random variable Y_t at time t can take on a range of values, each with a different probability. Usually, its forecast $\hat{Y}_t$ is given as the average of these values. Additionally, a **prediction interval** is specified which will contain the actual value y_t of the random variable Y_t with a specified probability. A 95% prediction interval will contain the actual value with a 95% probability (Hyndman & Athanasopoulos, 2013, sec. 1.7).

The forecast is always based on some observations. Sometimes we will precisely specify which information was used in calculating the forecast. For example, we will write $\hat{Y}_{T+h|T}$ to denote the forecast of random variable Y_{T+h}, taking into account all observations $\{y_1, y_2, \ldots, y_T\}$ up to and including time t_T. The forecasts $\left\{\hat{Y}_{T+1}, \hat{Y}_{T+2}, \hat{Y}_{T+3}, \ldots\right\}$ are based on the outcomes of random variables and can themselves be considered random variables, and so are the forecast errors $\{E_{T+1}, E_{T+2}, E_{T+3}, \ldots\}$, each one being the difference of two random variables: $E_t = Y_t - \hat{Y}_t$.

How should the observed time series be modeled statistically? Every time series represents the evolution of a system of random values over time and, as such, the observations $y_1, y_2, \ldots, y_T$ are inherently dependent. It would be inappropriate to treat the time series as a random sample of independent observations. The *stochastic* (or *random*) time series $\{Y_1, Y_2, \ldots, Y_T\}$ is modeled as the initial part of a **stochastic process** $Y = \{Y_1, Y_2, \ldots, \} = \{Y_i; i = 1, 2, \ldots, \}$. This one, in turn, is a collection of random variables, "connected" by their joint probability distribution. The observed time series $\{y_1, y_2, \ldots, y_T\}$ is seen as a (finite) realization of the stochastic process and is the only realization that we will ever be able to observe (Chatfield, 2000, p. 34). This might sometimes not be enough to estimate the properties of the underlying stochastic process and later use it for forecasting.

A *deterministic process* will only evolve in one way, once the initial conditions $\{y_1, y_2, \ldots, y_T\}$ are known. The *stochastic process* involves *randomness* or *uncertainty* and even when the starting point is known, there are still many directions into which the process may and will evolve. The load time series is real-valued, and there are infinitely many realizations $\{y_{T+1}, y_{T+2}, y_{T+3}, \ldots, \}$ of the future process are possible. The observed future values are only partly determined by past values.

3.3.2 Types of models for time series forecasting

Some models only include information from the past observations of the electrical load. They are generally referred to as **time series models**. Symbolically, we may write (Hyndman & Athanasopoulos, 2013, sec. 1.4):

$$Y_{t+1} = f(Y_t, Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, \text{error}), \tag{2}$$

where t is the current point in time, Y_t is the current value of the load process, $Y_{t-1}, Y_{t-2}, \ldots$ are past values of the load process and Y_{t+1} is its value in the next period. The prediction of the future values is based only on past values of the load process, but no other variable which may affect the load. The effects of any relevant external variables and the random variation of load are contained in the "error" term.

Other models allow for the inclusion of other, possibly relevant information (e.g. strength of economy, effect of holidays, weather variables such as temperature and humidity). Models that predict based on the so-called *predictor* or *explanatory variables* are called **explanatory** or **regression models**. We can describe them by the following formula (Hyndman & Athanasopoulos, 2013):

$$Y_t = f(\text{GDP}_t, \text{temperature}_t, \text{humidity}_t, \text{time of day}_t, \text{day of week}_t, \text{error}). \tag{3}$$

The "error" term accounts for any influential variables not captured in the model, as well as random load fluctuations. The model helps to explain what drives the electricity demand. When using the model for forecasting, the variables on the right side are replaced with their predictions for a future time $t + h$ (e.g. temperature_{t+h}) to predict the future load Y_{t+h}.

Mixed models also exist. They are known under different names, such as dynamic regression models, ARIMA with regression, panel data models, longitudinal models, transfer function models and so on. These models synthesize features of both previous groups. The relation might look like this (Hyndman & Athanasopoulos, 2013):

$$Y_{t+1} = f(Y_t, \text{GDP}_t, \text{temperature}_t, \text{humidity}_t, \text{time of day}_t, \text{day of week}_t, \text{error}). \tag{4}$$

There is a trade-off in using either the *time series model* or the *explanatory* or *mixed models*. The disadvantage of using external predictor variables in explanatory and mixed models is the additional care for regularly obtaining their measurements, as well as ensuring their timely and accurate forecasts, possibly by a robust automated procedure. Due to the propagation of forecasting errors from explanatory variables, a time series model may give more accurate forecasts. In general, our choice of method would depend on the following criteria: the *complexity of the model*, *computational time complexity*, *maintenance complexity*, and *forecast accuracy*. In STLF, it has been observed that *univariate* methods generally outperform regression models in the first 6 hours and, depending on other factors, can be a viable alternative for predictions of up to 24 ahead. For these reasons, we will study univariate methods exclusively.

3.3.3 Time series model building

To **model the time series** means to find a mathematical representation of the process that generated the observed data. A model is just a theoretical concept that tries to *describe* as well as possible the observed data and (potentially) *explain* its relations to factors that influenced it. We want to model the *systematic variation*, related to all identified underlying causes, and the *unexplained variation*.

We will approach the statistical model building in three steps (Chatfield, 2000, pp. 81–88):

1. **Model identification**: this step includes formulating a class of models that suits the perceived data relationships (based on initial analysis of observed data) and reducing the selection to one model family.
   Selecting the proper model family is the most demanding part of model building. Many elements of the correlogram analysis may but do not need to help. For example, the non-seasonal AR and MA processes are characterized by the properties of their ACF and PACF. Sample ACF and PACF are compared to the theoretical ACF and PACF to determine if the model is appropriate for the situation at hand. Also, if the sample ACF "cuts off" at lag q, meaning that it is "effectively zero" for lags $q + 1$ and on, then q will be the degree of the model. (Please, see Sections 5.2.2, 5.4.1 and 5.4.2 for details on ACF, PACF; AR and MA processes.)

2. **Model fitting**: the selected model contains parameters that need to be computed so that the model fits the observed data.
   The fit is usually achieved by minimizing some distance measure, e.g. the *ordinary least squares approximation*. The *fitted model* is just an *approximation* of the data. The goodness of fit depends on the model complexity and the suitability of the selected model and is accessed at step 3.

3. **Model verification**: The fitted model needs to be checked against domain knowledge and the properties of historical data.
   The latter is done by *residual analysis*. Residuals are differences between observed values and their forecasts:
   $$e_t = y_t - \hat{y}_t. \tag{5}$$
   For time series forecasting residuals are based on *one-step forecasts*; that is, the forecast of y_{T+1} is based on observations, available up to that time: $\hat{y}_{T+1} = \hat{y}_{T+1|T}$. The model explains historical data well if the residuals:
   a) have *zero mean*. If they do not, then the forecasting model is *biased*.
   b) are *uncorrelated*. If they are correlated, then there is remaining information contained in the residuals that was not captured in the model.

   Models that do not satisfy the above conditions can be improved. In the case of inconsistencies, the model needs to be either discarded or modified to achieve better performance. Model building is an iterative process.

After we have a valid fitted model, we will use it outside the range of data to which it has been fitted, to *predict* the future values. The systematic variation will be used for computing point forecasts, and the unexplained variation will assist in calculating prediction intervals.

3.4 Evaluating Forecast Accuracy

3.4.1 Forecast accuracy

In statistics, **forecast accuracy** is the degree of closeness of predicted values to the actual (measured) values. For a load time series, this closeness is measured at points in time for which the load is being forecast. The actual values are not known at the time of forecasting, but for the purpose of selecting the best method, the accuracy is calculated by predicting load values for a past time interval, for which historical data is actually known. This is called the *in-sample performance* and differs from the *out-of-sample performance* that can only be evaluated after the fact. The corresponding analysis is called the *ex-ante analysis* and differs from the *a posteriori* or after-the-fact *analysis*.

The accuracy of forecast depends on the following factors:

- the intrinsic **volatility** of the phenomenon being forecast: Greater volatility leads to reduced accuracy;
- the **model uncertainty**: If the *same* data is used to select, fit, and use the model then biases arise. Prediction intervals are usually calculated based on the fitted model and are too narrow;
- the **aggregation level**: Aggregating data is highly correlated to lower volatility, therefore aggregating loads of more consumers (*contemporaneous aggregation*) or sampling load over longer intervals (e.g. 1 hour vs. 15 minutes, weekly vs. daily; *temporal aggregation*) will lead to higher accuracy;
- the **length of the forecast horizon**: The further away we drift from the forecast origin, the more the accuracy decreases, for statistical reasons;
- the **selected load model**: A model that describes the underlying phenomenon better will yield higher accuracy;
- the **accuracy of** (any eventual) **predicted variables**: The prediction errors of input data will add to prediction errors in the result. The same is true for the uncertainty of input data versus uncertainty of the results.

The **parameter estimation algorithms** used in load forecasting in the last five decades have been limited to the *least-error squares minimization criterion*, although the *least absolute value criterion* can be a viable alternative.

The **uncertainty** of a predicted value is expressed in terms of a **prediction interval**. It defines the bounds within which future observed values are expected to fall, with a specified probability. For example, a 95% prediction interval is expected to contain the actual observed value with a 95% probability. The interval is estimated based on statistical methods.

3.4.2 Forecast accuracy measures

Let $\{(t_i, y_i)\}_{i=1}^{T}$ be a time series, representing the load curve to be predicted. We want to predict load at times $t_{T+1}, t_{T+2}, t_{T+3}, \ldots, t_{T+H}$, where t_T is the forecast origin and H is the forecast horizon. Let y_i denote the observation at time t_i and $\hat{y}_i$ denote a forecast of y_i. Then the **forecast error at time t_i** is denoted by $\boldsymbol{e_i = y_i - \hat{y}_i}$.

Forecast accuracy can be determined based on different **accuracy** (or **error**) **measures** that consider forecast errors over the entire forecast horizon, summarized into a single value. A lower accuracy measure (counterintuitively) means better accuracy. Let us look at the most common ones. Most of these are also included in the selected `R` software package that was used throughout this work.

Note that every accuracy measure is "sensitive" to (or detects) a different *type* of difference between the observation and its forecast. When comparing forecast methods on the same dataset, different accuracy measures will therefore produce different performance rankings. The classification of accuracy measures follows Hyndman and Athanasopoulos (2013).

3.4.2.1 Scale-dependent measures

The forecast error $\boldsymbol{e_i}$ is in the same order of magnitude as the forecast values. Any accuracy measure that is based on the forecast error is, therefore, scale dependent and can only be used to compare the performance of different methods on the same dataset. It can not be used to compare the performance of the same method on two time series with different orders of magnitude.

The simplest scale-dependent accuracy measure is **Mean Error**:

$$\text{ME} = \boldsymbol{E}\left[\{\boldsymbol{e_i}\}_{\boldsymbol{i=T+1}}^{\boldsymbol{T+H}}\right] = \boldsymbol{E}\left[\{\boldsymbol{y_i - \hat{y}_i}\}_{\boldsymbol{i=T+1}}^{\boldsymbol{T+H}}\right] = \frac{\boldsymbol{1}}{\boldsymbol{H}} \sum_{\boldsymbol{i=T+1}}^{\boldsymbol{T+H}} (\boldsymbol{y_i - \hat{y}_i}) \qquad (6)$$

It tends to detect *forecast bias* (e.g. systematic overpredicting or underpredicting) but will completely ignore huge individual errors, as long as the forecast is not biased. (That information is lost in the process of averaging.)

Better detection of the magnitude of errors is obtained by measures that use absolute rather than actual values of the forecast errors in the formula. The two most commonly used scale-dependent measures are:

Mean Absolute Error, based on absolute errors:

$$\text{MAE} = \boldsymbol{E}\left[\{|e_i|\}_{i=T+1}^{T+H}\right] = \boldsymbol{E}\left[\{|y_i - \hat{y}_i|\}_{i=T+1}^{T+H}\right] = \frac{\mathbf{1}}{\boldsymbol{H}} \sum_{i=T+1}^{T+H} |y_i - \hat{y}_i| \tag{7}$$

and **Root Mean Square Error**, which is based on squared errors:

$$\begin{aligned} \text{RMSE} &= \sqrt{\boldsymbol{E}\left[\{e_i^2\}_{i=T+1}^{T+H}\right]} = \sqrt{\boldsymbol{E}\left[\{(y_i - \hat{y}_i)^2\}_{i=T+1}^{T+H}\right]} \\ &= \sqrt{\frac{\mathbf{1}}{\boldsymbol{H}} \sum_{i=T+1}^{T+H} (y_i - \hat{y}_i)^2} \end{aligned} \tag{8}$$

3.4.2.2 Normalized measures

We can enhance the usability of RMSE by normalizing it. By comparing it to the greatest difference observed across the forecast horizon, we construct a new accuracy measure, called **Normalized Root Mean Square Error**. It is expressed as a percentage of this difference:

$$\text{NRMSE} = \frac{\text{RMSE}}{\boldsymbol{x_{max} - x_{min}}} \tag{9}$$

This measure is not implemented in the `R` package, but we will need it in Section 3.5.

3.4.2.3 Measures based on percentage error

The **percentage error at time t_i** is defined by $\boldsymbol{p_i = 100 \cdot \frac{e_i}{y_i} = 100 \cdot \frac{y_i - \hat{y}_i}{y_i}}$. It has the advantage of being scale-independent, so all accuracy measures based on it are suitable for comparing forecast performance independently of the dataset.

The simplest measure in this group is **Mean Percentage Error**:

$$\begin{aligned} \text{MPE} &= \boldsymbol{E}\left[\{p_i\}_{i=T+1}^{T+H}\right] = 100 \cdot \boldsymbol{E}\left[\left\{\frac{y_i - \hat{y}_i}{y_i}\right\}_{i=T+1}^{T+H}\right] \\ &= \frac{100}{\boldsymbol{H}} \sum_{i=T+1}^{T+H} \left(\frac{y_i - \hat{y}_i}{y_i}\right) \end{aligned} \tag{10}$$

which suffers from a similar deficit as ME: unless the forecast is biased, single percentage errors will cancel each other out for the most part, so the resulting measure will not be able to detect large percentage errors as such.

This deficiency is overcome by **Mean Absolute Percentage Error**

$$\begin{aligned} \text{MAPE} &= \boldsymbol{E}\left[\{|p_i|\}_{i=T+1}^{T+H}\right] = 100 \cdot \boldsymbol{E}\left[\left\{\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right\}_{i=T+1}^{T+H}\right] \\ &= \frac{100}{H}\sum_{i=T+1}^{T+H}\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right) \end{aligned} \tag{11}$$

which is one of the most useful measures for practical purposes.

Both measures have the disadvantage of being undefined or assuming infinite values whenever $y_i = 0$ applies to some t_i. They also assume extreme values when y_i is close to 0, which is not useful for practical purposes. For these reasons, Anderson (1985, pp. 348) proposed **symmetric MAPE** or **adjusted MAPE**:

$$\text{sMAPE} = 200 \cdot \boldsymbol{E}\left[\left\{\frac{|y_i - \hat{y}_i|}{y_i + \hat{y}_i}\right\}_{i=T+1}^{T+H}\right] \tag{12}$$

Despite the adjustments, whenever y_i is close to zero, $\hat{y}_i$ is also likely to be so, so division by a number close to zero will result in an unstable measure calculation. Furthermore, the value of sMAPE can be negative. For these reasons, Hyndman and Koehler (2006) recommend that the sMAPE not be used.

3.4.2.4 Scaled measures

Mean Absolute Scaled Error was proposed by (Hyndman & Koehler, 2006) as an alternative to using percentage errors when comparing forecast accuracy across series on different scales. The errors are scaled based on the *in-sample* MAE from a simple forecast method (Hyndman & Athanasopoulos, 2013). For *seasonal* time series, like the load time series of an industrial consumer, a **scaled error at time t_j** is defined using *seasonal naïve* forecasts:

$$q_j = \frac{e_j}{\text{MAE}_{\textit{in-sample}}} = \frac{e_j}{\frac{1}{T-m}\sum_{i=m+1}^{T}|y_i - y_{i-m}|}. \tag{13}$$

The numerator is the forecast error $e_j = y_j - \hat{y}_j$ at time t_j, and the denominator is the in-sample mean absolute error of the seasonal naïve forecast method. It uses the last observation from the same season as the forecast: $\hat{y}_{T+h|T} := y_{T+h-km}$, where m is the season's length and $k = \lfloor\frac{h-1}{m}\rfloor + 1$ (Section 5.1). The result is independent of the scale of the data. A scaled error that is less than one indicates that the method performed better than the seasonal naïve method did on average, on historical data. A value greater than one indicates a poorer performance.

MASE is then simply:

$$\text{MASE} = \boldsymbol{E}\left[\{|q_j|\}_{i=T+1}^{T+H}\right]. \tag{14}$$

3.4.2.5 Selected accuracy measure

In the forecasting module, MAPE is the accuracy measure of choice as, due to the absoluteness, it is independent of the direction of the bias and, as it is based on percentage errors, it gives results between $\mathbf{0\%}$ and $\mathbf{100\%}$ which are suitable for ranking purposes.

In contrast, ME has neither of these advantages, RMSE can take on any positive value, MPE is bias-dependent, and so on. The issue with sMAPE is that it is not treating over-forecasting and under-forecasting in the same way. As the following example shows, it penalizes under-forecasting slightly heavier than over-forecasting by the same margin:

- Over-forecasting: $\boldsymbol{y_i = 100, \hat{y}_i = 110}$ gives a sMAPE $= \mathbf{200 \cdot \frac{10}{210} = 9.52\%}$
- Under-forecasting: $\boldsymbol{y_i = 100, \hat{y}_i = 90}$ gives a sMAPE $= \mathbf{200 \cdot \frac{10}{190} = 10.53\%}$

Moreover, sMAPE was not available in R, the software used. MASE was a secondary measure that helped select the best overall forecasting method.

3.5 Review of Literature

The electrical load is affected by external factors, such as *weather* and *social* factors. The principal weather factors are *temperature* and *humidity*. Social factors are captured in *seasonal effects* (hour of the day, day of the week, calendar holiday) and *special events* that result in a significant deviation from the typical load behavior. Marinescu, Harris, Dusparic, Clarke, and Cahill (2013) found the biggest correlation, up to 0.9, between current load and the load of the previous day, followed by temperature (up to 0.8 in the evening and night). Humidity plays a lesser role; nonetheless, its correlation in the evening can reach 0.5. Recent studies (Chen, Cañizares, & Singh, 2001) have shown that when consumers adjust their consumption behaviopricer based on price information, thus are *price elastic*, then including electricity price data as a predictor variable will improve forecasts even further.

The aggregation level of data matters when selecting how to approach STLF. Data that is aggregated more (either *contemporaneous* or *temporal aggregation*, see page 23), demonstrates lower volatility. The level of volatility affects the selection of methods, as well as forecast accuracy. It can mean the difference between being and not being able to forecast a specific future load. In fact, a recent paper by Sevlian and Rajagopal (2014) confirms the observations of the present work by describing the relationship between MAPE and the aggregation

size in the form of an explicit scaling law.

3.5.1 Large scale aggregates

Most scientific literature is based on *large-scale* data, usually a cumulative load of a DSO (for example, Taylor, Menezes, and McSharry (2006) used data of Rio de Janeiro in Brazil and Chen et al. (2001) used historical load data from the Ontario Hydro system) or even an entire country (Taylor (2010) used two load time series; one for Great Britain and one for France). Results from forecasting on a large scale might serve as inspiration for forecasting loads on a *very small scale*, but do not quite address the same problem, even less when predicting load of *individual* consumers. (We consider a very small scale to be an aggregation at transformer level, an equivalent to several dozens of houses.)

We will start by reviewing methods tested on a *large scale*. The simplest methods are based on some *averages* of past consumption, e.g. the seasonal naïve method and the method of typical days. Many *classical methods* have been successfully used for large-scale STLF, the most popular families studied include the multivariate multiple regression methods, the univariate exponential smoothing methods, and seasonal ARIMA models. Contemporary methods are based on *artificial intelligence* (hereinafter AI) and include above all Artificial Neural Network (hereinafter ANN), fuzzy logic (hereinafter FL), neuro-fuzzy method and Support Vector Machines (hereinafter SVM). Two useful STLF literature surveys can be found in (Alfares & Nazeeruddin, 2002) and (Taylor & McSharry, 2007).

A frequent approach to STLF is to use *univariate* methods for lead times of up to 6h ahead and *multivariate* methods beyond that horizon. Over such a short horizon, weather changes gradually and its influence is implicitly captured in the recent historical data. When weather data is difficult to obtain, univariate methods are sometimes used for lead times of up to 24 hours.

When comparing six different univariate methods, Taylor et al. (2006) found that for lead times of up to 48 hours the *simpler* and more *robust* methods, such as the double-seasonal exponential smoothing, outperformed the more complex ones, including ANN and regression with principal component analysis (hereinafter PCA). This method consistently reached MAPE below 3% for the time series of Rio de Janeiro, and below 1.5% for that of England and Wales.

Load profiles on a large scale reveal daily, weekly and yearly seasonality (we will get familiar with the term in Section 4.5). While double-seasonal methods have been known to be competitive for some time, Taylor (2010) extended three of them to include a third seasonality. All double-seasonal methods showed increases in performance over their single seasonal ver-

sions, and all *triple seasonal methods* outperformed their double-seasonal versions, as well as a univariate neural network approach, at all lead times. MAPE increased with lead time and reached around 1.5% for all three triple seasonal methods for a lead time of 24 hours. Unfortunately, these methods require a long historical time series. While this amount of data is provided at the utility or country level, it is frequently not the case for individual consumers, especially single households or small communities.

Artificial neural networks have been increasingly used for STLF in the past two decades and have proven to be successful in the more advanced variants, in particular as hybrid models, which are combinations of neural networks with stochastic learning techniques, such as genetic algorithms (GA), particle swarm optimization (PSO) and similar. For a review of recent research in this field consult (Baliyan, Gaurav, & Mishra, 2015). A critical review of papers covering ANN approaches can be found in (Hippert, Pedreira, & Souza, 2001).
In their *univariate* form, used as an autoregressive time series models, Faraway and Chatfield (1998) have found that due to their flexibility ANN models frequently failed to converge. When they did converge, they failed to find the global minimum of the objective function. In some cases, neural networks fit the in-sample data well but performed poorly at predicting the future (Thielbar & Dickey, 2011). Furthermore, the NN3 competition showed that standard, less compute-intensive forecasting methods, such as ARIMA, outperform neural networks (Crone, Nikolopoulos, & Hibon, 2008). This was also the case in the study by Taylor et al. (2006). Recent work by Thielbar and Dickey (2011) on using neural networks for time series forecasting produced mixed results.

3.5.2 Individual households

There are just a handful of very recent studies covering STLF for individual households. They all use different approaches, mostly sophisticated AI. Prior to the large-scale smart meter installations, hourly residential data was also hardly available. In the near future, much more data will be available as the EU is committed to having installed smart meters for at least 80% of consumers by 2020, with full deployment by 2022 (”Directive 2009/72/EC”, 2009).

Ghofrani, Hassanzadeh Etezadi-Amoli, and Fadali (2011) applied Kalman filtering to individual residential data and observed a MAPE of 30.42% for a sampling period of one hour and forecast horizon of one hour, which is considered very short time load forecasting. In their case accuracy increased with a shorter sampling period. Longer lead times were not studied.

Edwards, New, and Parker (2012) applied several machine learning techniques, which were proven to be successful in commercial buildings, to forecast the next hour’s load for three homes with simulated occupancy. In that case, the best MAPE values obtained on single

houses were between 16.11% and 21.33%. It can be speculated that the results would be worse if the houses had been occupied by people, as they would produce even more complex load patterns.

Gajowniczek and Ząbkowski (2014) studied STLF for an individual household and forecast its load demand for the next 24 hours. They tested seven ANN models and a SVM. Multilayer perceptrons performed best in the ANN category. Forecast accuracy was measured in terms of percentage of forecasts that were up to 10% off from the actual values. ANN's accuracy was between 57% and 69%, depending on the hour of the day, while SVM's was between 52% and 68%. The drawback of this study is that they disposed of one single time series, so additional information on the unit (demographics and a list of appliances) was of no use. Similar data on a larger number of units might potentially yield richer insight.

3.5.3 Small communities/residential buildings

Studies on STLF for residential buildings and aggregations at the level of a single transformer, VPP or microgrid, are still rare, but they play a major role in the management of microgrids and VPPs. Two studies are worth mentioning.

The work by Marinescu et al. (2013) used data from a smart-meter trial on a community of houses in Ireland in 2009–2010, including weather information. Several approaches were used to forecast load for a day and week ahead: ANN, FL, AR, ARMA and ARIMA models and wavelet neural networks (hereinafter WNN). The data was aggregated to investigate two different scenarios: one that covered 90 houses and reached peaks of 140 kWh and another one of 230 houses, with peaks up to 340 kWh. No method outperformed all the others at all times. ANN and WNN performed best at the evening peak, from 5–8 p.m., while FL was best in the first part of the day. ARIMA had best average results across 24 hours. NRMSE, the accuracy measure of choice (see page 25) was between 3.61 and 4.28 for the group of 90 houses. Accuracy increased for the group of 230 houses, due to the effects of irregular human behavior canceling out. By combining four of these methods at their best performance intervals, an improvement in performance of 11–28% was accomplished (Marinescu, Harris, Dusparic, Cahill, & Clarke, 2014).

4 DATA

The first step to selecting methods appropriate for the underlying problem is to visualise data and perform an initial exploratory analysis.

4.1 Software Used

The open source statistical `R` language package (hereinafter `R`) was used throughout this work; from data preparation, analysis, visualization, to performing all experiments, calculating forecast accuracy and visualizing the results.

The main package used was `forecast`, developed by Hyndman (for package documentation, see Hyndman (2016)). The package contains all relevant forecasting methods, especially for time series, including some functions that are the result of the author's research work. It also makes use of all the presented forecast accuracy measures, including ME, RMSE, MAE, MPE, MAPE, MASE, and ACF1.

4.2 Input Data

The historical data used for the development of the forecast model was provided by one of the major Slovenian DSOs, Elektro Ljubljana. It comprises anonymized metering data of electric energy consumption for 235 residential consumers and 123 industrial consumers for the period between January 2011 and August 2013. The data was collected from smart meters, sampled at 15-min intervals and had local timestamps. It contains large blocks of *missing data in all sources* at random positions. The origins of missing data are reading errors in the concentrator. The concentrator also has limited memory, so the data is lost forever.

Later, data from single-chip devices was made available by Telekom Slovenije. These devices, called Home Energy Hubs and Business Energy Hubs, were providing measurements of total power consumption of individual consumers and that of up to 5 devices, at 1-min sampling rate. They were installed at the sites of around 120 residential and industrial participants in a pilot project. While data sampled at a rate this small is able to give customers more insight into their power consumption, it does not aid in better predictions. Results will show that measurements *aggregated to 1 hour are optimal* for forecasting load of individual consumers. Any data that is less aggregated is *too volatile* and results in unacceptably bad forecasting accuracy. (See Table 1 for a comparison of accuracy).

The historical data was used to determine the best methods for the forecasting module, while the field measurements from home/business energy hubs were used as real-time input for the forecasting module.

4.3 Handling Data Anomalies

4.3.1 Missing data

Except for the simplest ones, most forecasting methods will not work when there are missing values in the historical data. Missing values need to be **imputed**, that is replaced by their estimated values, before forecasts can be made. Several techniques can be used (*case deletion, single imputation, multiple imputation*); when using them, the forecaster needs to *prevent introducing bias*, a non-zero average error.

A comprehensive work on the estimation of missing values in time series is the doctoral dissertation by Fung (2006). Sorjamaa (2010) is focusing on the latest and most contemporary methods, contributing some new and improved methodologies, like the deterministic Empirical Orthogonal Functions (EOF) and Self-Organizing Maps (SOM), based on supervised learning.

Imputation Method Used. Missing data estimation procedure was developed *a-posteriori* and *independently*. The idea was to develop a forecasting procedure first, using data with no missing values and later develop a procedure that will estimate the missing values, prior to forecasting.

The fact that the forecasting procedure will run *on-line* and in *real time* played a role in selecting an estimation procedure that is fast and reliable. A simple method was developed that predicted the missing load, corresponding to a timestamp, as the average load of the same hour and same day of the week from data available in the last 28 days. In case this data was missing for all 4 weeks, a longer time interval was considered.

Alternatively, we could let the `forecast` software package handle missing data in an automated manner. The function `na.interp()` is able to impute missing data in non-seasonal, as well as seasonal series. It first fits a seasonal model to the data, and then *interpolates* the seasonally adjusted series, before re-seasonalizing.

4.3.2 Outliers

By **outliers**, we mean observations that differ substantially from the expected value given a model of the situation. An outlier can be identified judgmentally or if it deviates from the mean by a distance of more than two standard deviations. Outliers can be handled by `tsoutliers()`: it identifies them and suggests reasonable replacements.

4.3.3 Holidays and special events

The calendar variables, implemented in the `forecast` package are not suitable for 1-h measurements, so we will have to develop a custom solution, perhaps inspired by the already implemented functions. Holidays and neighboring days could be predicted as a pointwise average of past weekends.

4.4 Data Preparation

For the scope of determining the proper statistical model for the forecasting procedure, data with no missing values was used, for the sake of simplicity. However, in the process of forecasting, we have to use the data available, so the first steps of the procedure always need to include replacing missing data and outliers, as described in the previous section.

4.4.1 Selecting data without missing values

First, the length of the historical period needs to be determined. If then accuracy is insufficient, a longer period should be tested. The optimal length for best forecast accuracy should ultimately be determined empirically. We will start by forecasting from *28 days of historical data* for the following reasons: 1. load values primarily depend on the recent past, 2. the period should be long enough to capture a few weekly seasonal cycles, 3. it will implicitly capture the influence of recent past weather and exclude the irrelevant less recent past weather. This period will hopefully be long enough to capture all major features of this time series. Our goal is to find a long enough period where the most sources have no missing data.

First, all sources containing more than **10%** *overall missing values* were immediately discarded. Then we were looking for the longest period in which most sources had no missing data. June 2011 was selected as the sufficient period (4 weeks of historical data plus additional 2 weeks for the maximum forecast period) where none of the industrial sources had missing values. For residential sources, May and June 2013 were selected. All models were tested on the selected periods.

4.4.2 Data transformation

Data was imported from `txt` files into `R`'s local database structures. The timestamps were converted to Coordinated Universal Time (hereinafter UTC) timestamps. No duplicates were found. Time series for 2011–2013 were concatenated for each source. Sources with too many missing values were discarded. After this step, we were left with 94 industrial and 171 residential sources. Their corresponding time series were labeled as Ind1, Ind2, …, Ind94

for industrial and Dom1, Dom2, …, Dom171 for residential sources. We will alternately be using the terms consumer and customer.

After testing features of several different `R` data objects, we decided to store all consumers' time series individually as multi-seasonal time series (or `msts`) objects with *seasonal periods* (seasonal cycle lengths) of one day and one week, and the primary *frequency* (number of observations per seasonal cycle) of $\mathbf{24 * 7 = 168}$. The initial time series retained the original 15-min sampling.

For predicting groups of residential consumers, the time series of all sources were first aggregated into the SUM171 time series, retaining the 15-min sampling. It contained **46%** overall missing values (time points at which at least one source had a missing value). The longest period without missing values was barely 2 days long and therefore unusable for our purposes. Instead, a smaller group of 69 residential consumers was constructed by gradually adding sources. **SUM69** contained **19%** overall missing values in a 3-year period between 2011 and 2013. The longest continuous period without missing values ranged from April 2013 to September 2013.

4.4.3 Data aggregation

Models were initially tested directly on the 15-min measurements provided and, after these time series turned out to be *too volatile* for good forecasting accuracy, they were aggregated into 1-h measurements and the models were retested. (Compare test set MAPE values for Holt-Winters method between 15-min and 1-h sampled data in Table 2.) In this way, we were able to obtain much better forecasting accuracy. The difference in "smoothness" between a signal sampled at 15 minutes and the same one sampled at 1 hour is clearly visible by comparing Figures 5 and 6.

Figure 5. Time Series of Dom15 Consumer, Week1 of May 2013, 15-min Sampling

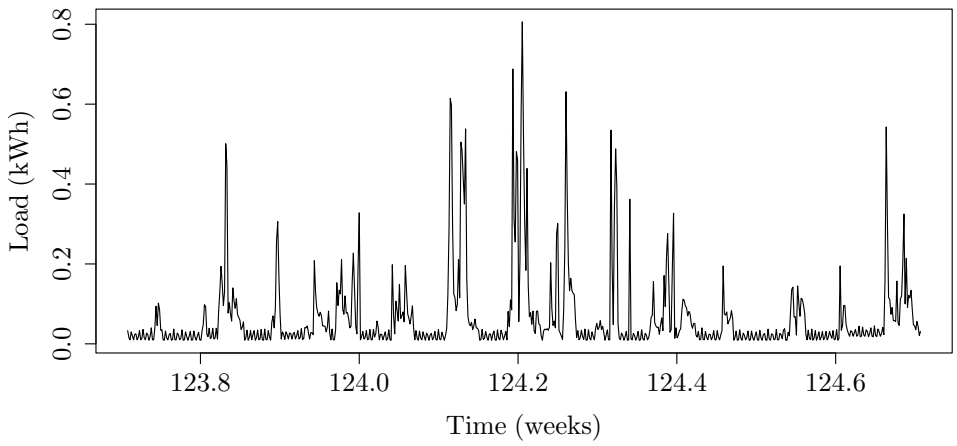
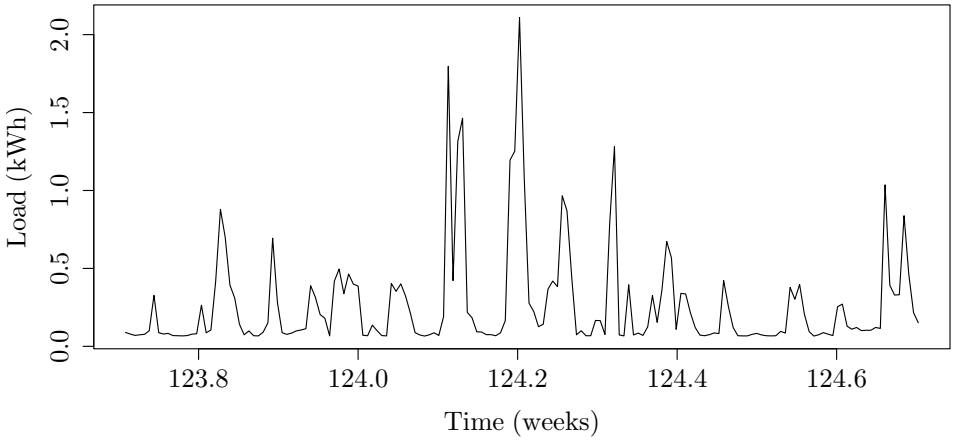


Figure 6. Time Series of Dom15 Consumer, Week1 of May 2013, 1-h Sampling



4.5 Data Visualisation

The first step to obtaining an impression of the signal to be forecast is to visualise the data. This will enable us to see all its major *characteristics*, such as whether the time series is stationary, whether there is a trend, whether its pattern depends on seasons, whether it displays patterns related to business cycles, whether the pattern changes on holidays, whether there are any outliers or noticeable "special events" for which the pattern changes visibly, etc.

The answers to these questions will guide the selection of appropriate models. Any successful forecasting model needs to consider and incorporate all features discovered during *data visualisation* or ones that are discovered during *statistical analysis*.

Let us first explain what is meant by the above questions. First, we will be looking for **patterns** that a time series may or may not display, and that will affect the selection of forecasting methods:

- A **trend** exists when there is a long-term change in the underlying mean level per unit time. The term "long-term" refers to the length of the historical period. A trend can be *linear* or *non-linear*.
- A **seasonal pattern** occurs when a time series fluctuates subject to seasonal factors, such as the time of day, the day of the week or the season of the year. A time series can have multiple seasonal patterns, like a daily, weekly or yearly seasonal pattern and each seasonal cycle's length is a fixed and known period. A "season" is a part of the seasonal pattern.
- A **cycle** occurs when the pattern of the fluctuations in the signal is not of a fixed period and the fluctuations are not related to a specific time in a year but are rather related to "business cycles" and other economic factors. The cyclic pattern has a variable and unknown length

which is usually longer than the seasonal patterns.

- **Irregular fluctuations** are the "left-over" signal after trend, seasonal pattern and cycle have been removed from the original time series. It is desirable that for it to be completely random, in which case it is called *white noise* and is not predicted. In a contrary case, it includes some "left-over" pattern, so the model was not completely successful in explaining the sources of signal fluctuations.

A time series is said to be **stationary** when its statistics (for example mean and variance) do not change over time. A time series exhibiting a trend is not stationary.

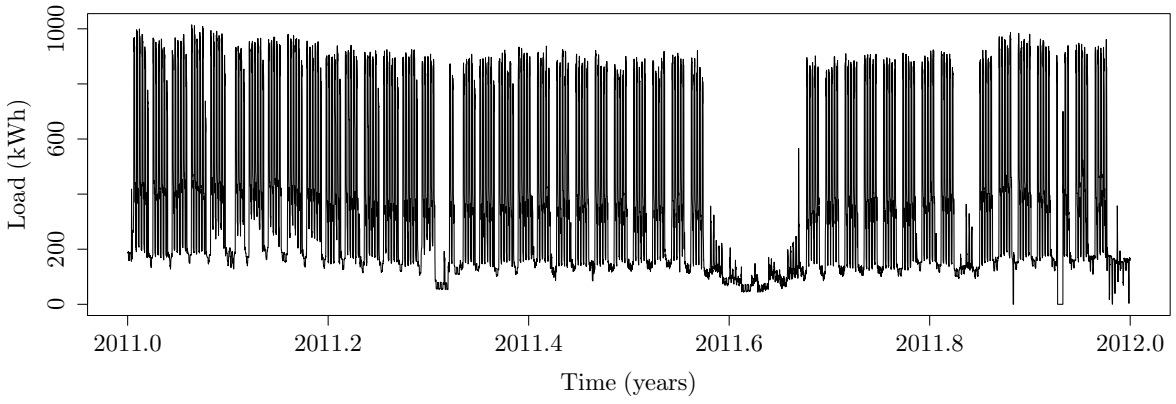
We will consider the following types of plots, which are suitable for a time series:

- **Time plot**: observations are plotted against the time of observation, with consecutive observations joined by straight lines. The time plot will be reveal many features of a time series, but some might require the application of some analytic techniques (for example, the classical seasonal decomposition in Section 4.6.1.)
- **Seasonal plot**: observations are plotted against the individual seasons in which they were observed. Constructing a seasonal plot of a time series for which the period spans over several seasonal cycles is equal to "slicing" the series into periods of the length of a seasonal cycle and plotting them on the same graph, in an overlapping manner. It gives a visual confirmation of the seasonal pattern's existence and enables comparison between the years. It might also reveal a trend.
  Occasionally, we will plot these "chunks" in separate graphs, side by side, to be able to see the details more clearly.

4.5.1 Industrial consumers

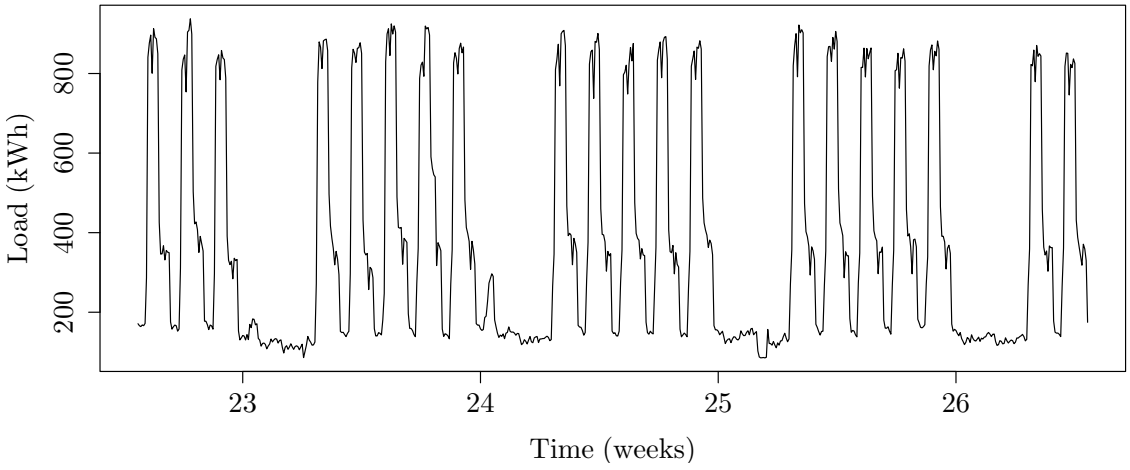
Let us plot the *yearly load time series* of the first industrial consumer, labeled by Ind1:

Figure 7. Time Series of Ind1 Consumer for 2011, 1-h Sampling



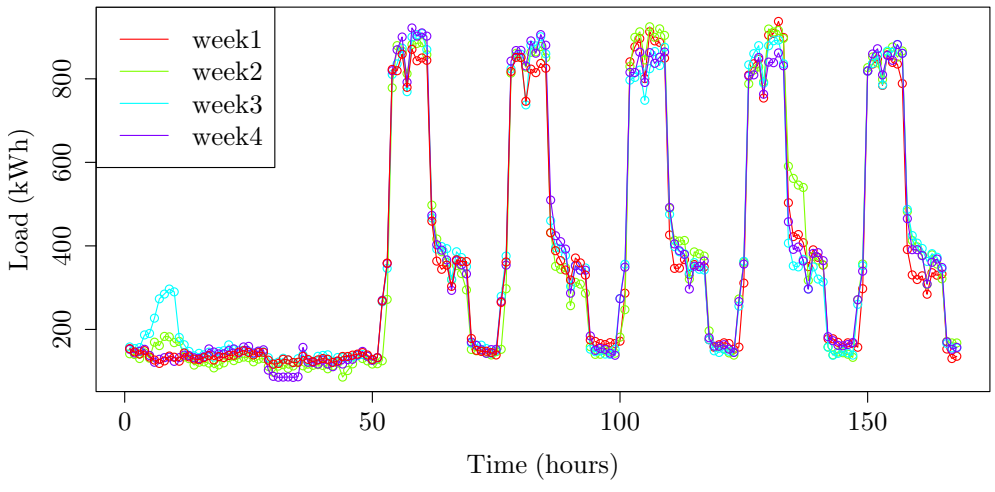
The graph is too compressed to see much, the only thing that stands out are the reduced loads at weeks 17–18, which denote the 1st of May holidays, then again roughly around weeks 32–36, the usual summer vacation period, week 45, indicative of the autumn holidays, and the last week of December, denoting the Christmas season. Plotting a *monthly Time Series* of the same consumer, see Figure 8, reveals a weekly and a daily seasonal pattern with no visible trend.

Figure 8. Time Series of Ind1 Consumer for the First 28 Days of June 2011, 1-h Sampling



Layering these four consecutive weeks on top of each other in a *seasonal plot* (Figure 9) only confirms that the seasonal patterns almost entirely overlap. The differences between observations at the same season (time of the day and day of the week) seem to be more random than anything else. We predict that any statistical method, based solely on historical data, should be fairly successful and conclude that the load of individual industrial consumers should be *predictable*.

Figure 9. Seasonal Plot of Ind1 Consumer, June 2011, 1-h Sampling



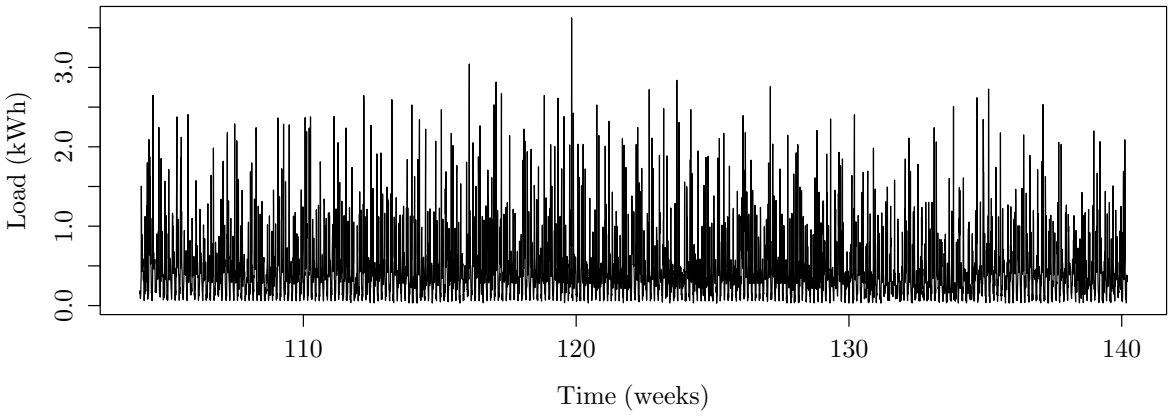
Graphs of many more industrial sources were visualised. They all displayed similar behavior. Every industrial consumer's load profile follows a *different curve*, but they all display the following similarities:

- all workdays follow a similar pattern,
- weekends usually display a different pattern, with much lower values,
- the seasonal plot shows good overlapping (only small variations from week to week).

4.5.2 Residential consumers

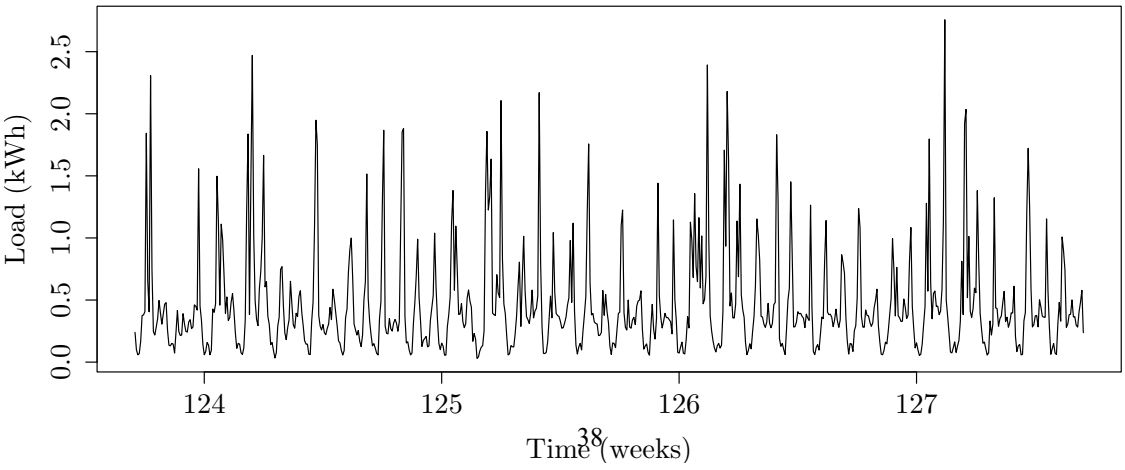
Now, let us look at an 8-month time plot (from January to August of 2013) of the first residential consumer, denoted by Dom1 (Figure 10).

Figure 10. Time Series of Dom1 Consumer for 2013, 1-h Sampling



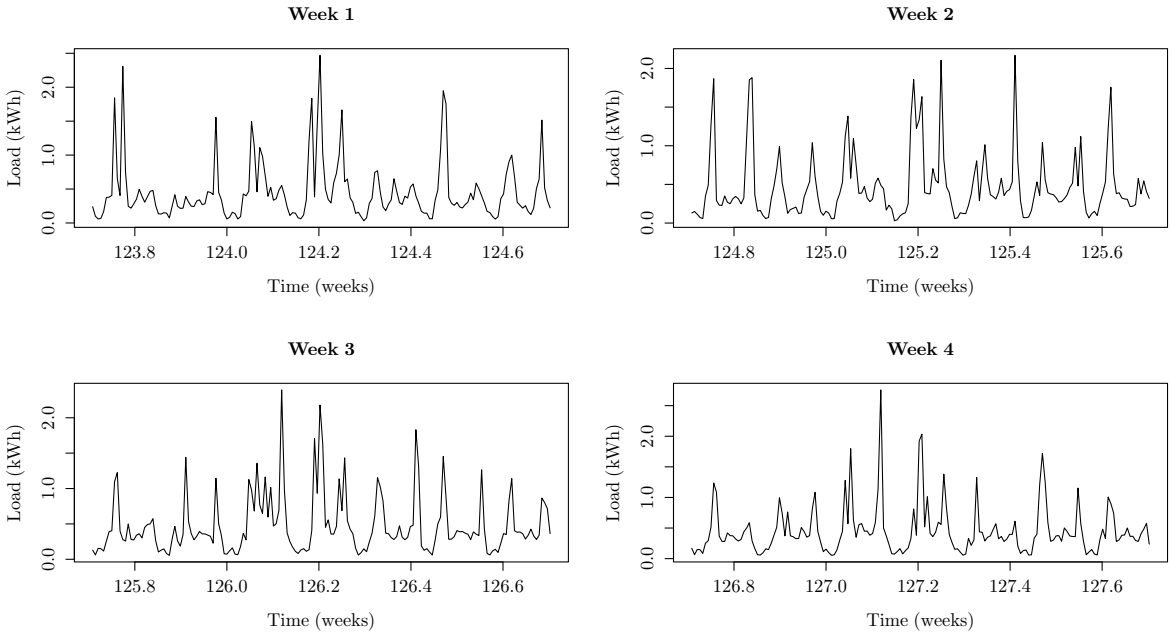
Note that the weeks are labeled from 105 onward, since this time series is part of a 3-year time series, starting in 2011. No holiday effects are visible here. Let us examine the *monthly* load time series of the same consumer (Figure 11):

Figure 11. Time Series of Dom1 Consumer for 4 Weeks in May 2013, 1-h Sampling



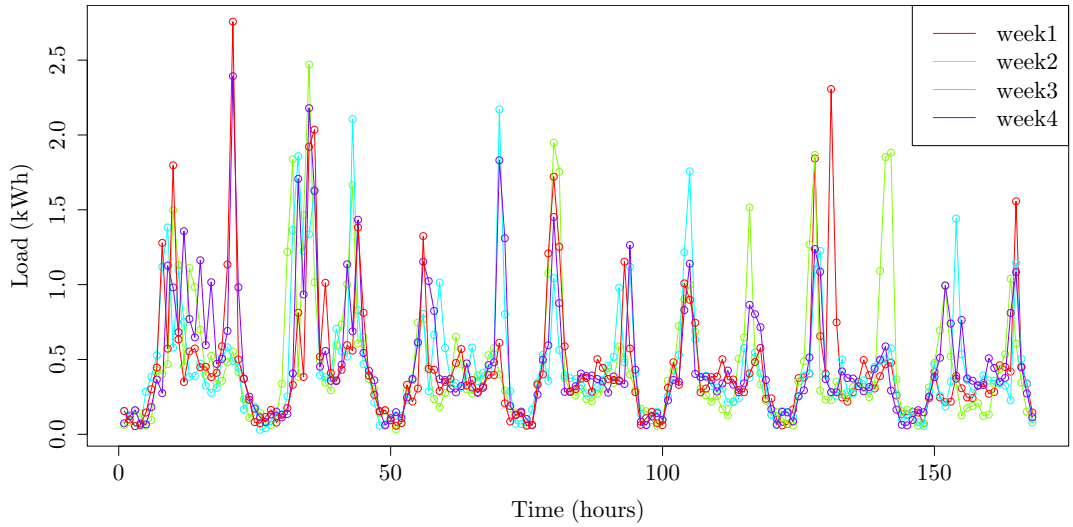
The plot seems to be much more *volatile* (or "spiky") than the industrial one observed in the previous section. A daily fluctuation is visible as an elevated load during daytime and reduced load at nighttime but no two days are exactly the same. Plotting these four weeks for a side-by-side comparison (Figure 12) only confirms this.

Figure 12. Time Series of Dom1 Consumer for 4 Weeks in May 2013, 1-h Sampling



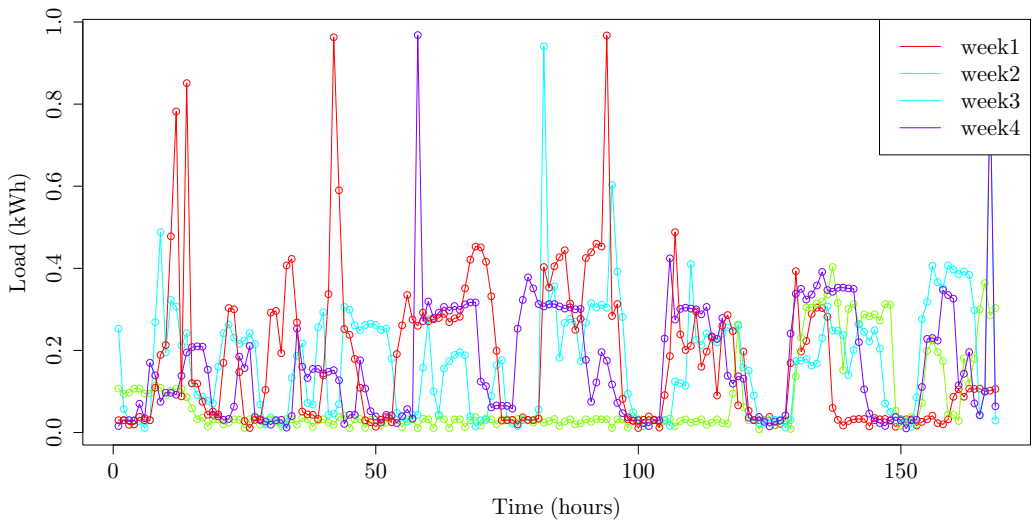
A *seasonal plot* (Figure 13) of these four weeks shows that load peaks are reached at different times every week. The previous week's behavior is not repeated in the following week.

Figure 13. Seasonal Plot of Dom1 Consumer for 4 Weeks in May 2013, 1-h Sampling



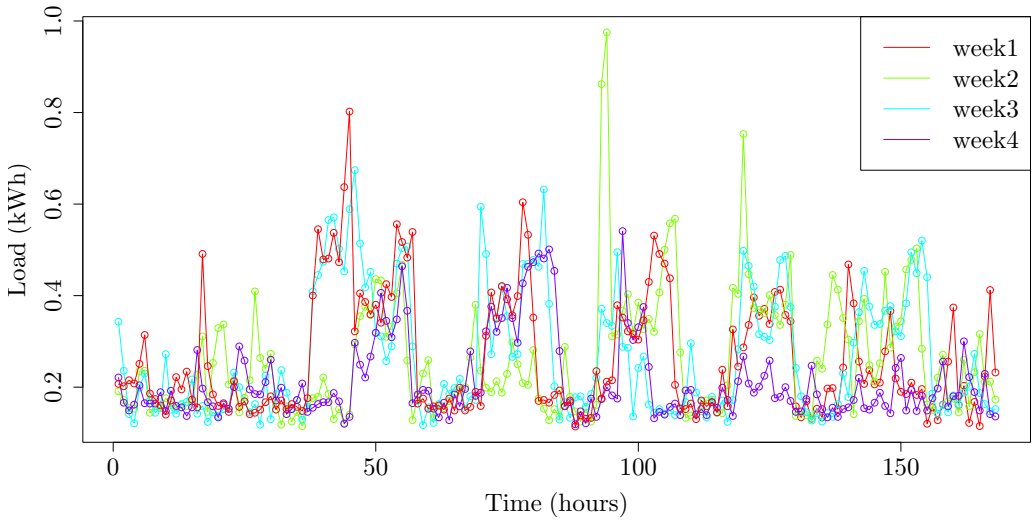
From the seasonal plot, one can see that at a specific season (say, time t_i) in a week *very different values will be observed from week to week* (imagine a vertical line as a visual aid). Therefore, predicting the load of a household for a specific hour will be a hard and unreliable task. Looking at seasonal plots of some other households makes this only clearer:

Figure 14. Seasonal Plot of Dom5 Consumer for 4 Weeks in May 2013, 1-h Sampling



The load of the Dom5 consumer (Figure 14) is showing a different pattern every week. The peaks are reached at random times and the flat profile in week 2 indicates the a temporary absence of the inhabitants. This kind of profile is by no means an exception among residential consumers.

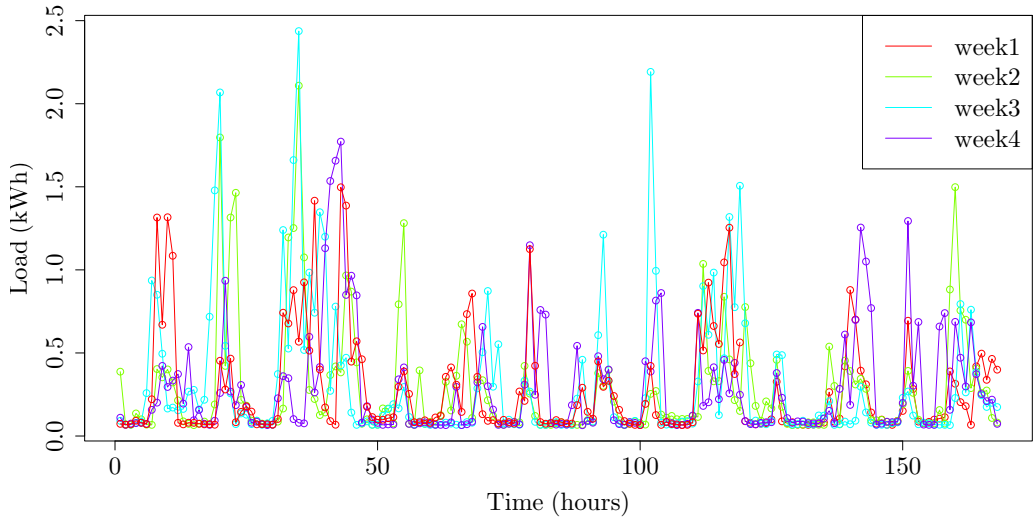
Figure 15. Seasonal Plot of Dom12 Consumer for 4 Weeks in May 2013, 1-h Sampling



The load profiles of Dom12 and Dom15 consumers (Figures 15 and 16) seem less irregular, but the reliability of predictions will not improve considerably.

The more "erratic" behavior related to load consumption in households can be explained with *social and psychological factors*. Regardless of whether the behavior is the cumulative result of several people sharing a household or just one, human behavior is not mechanic and as such *hardly predictable* without considering factors that affect it. The situation is very different from the industrial environment that follows schedules and working hours.

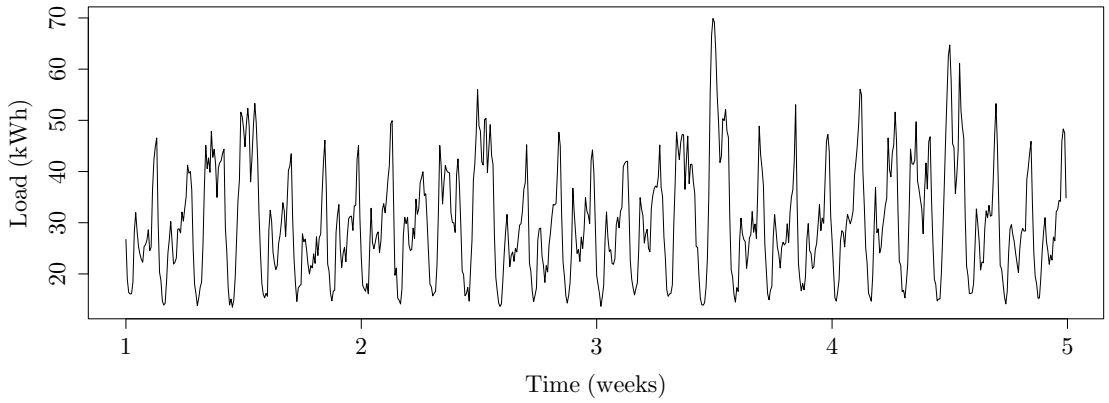
Figure 16. Seasonal Plot of Dom15 Consumer for 4 Weeks in May 2013, 1-h Sampling



4.5.3 Group of residential consumers

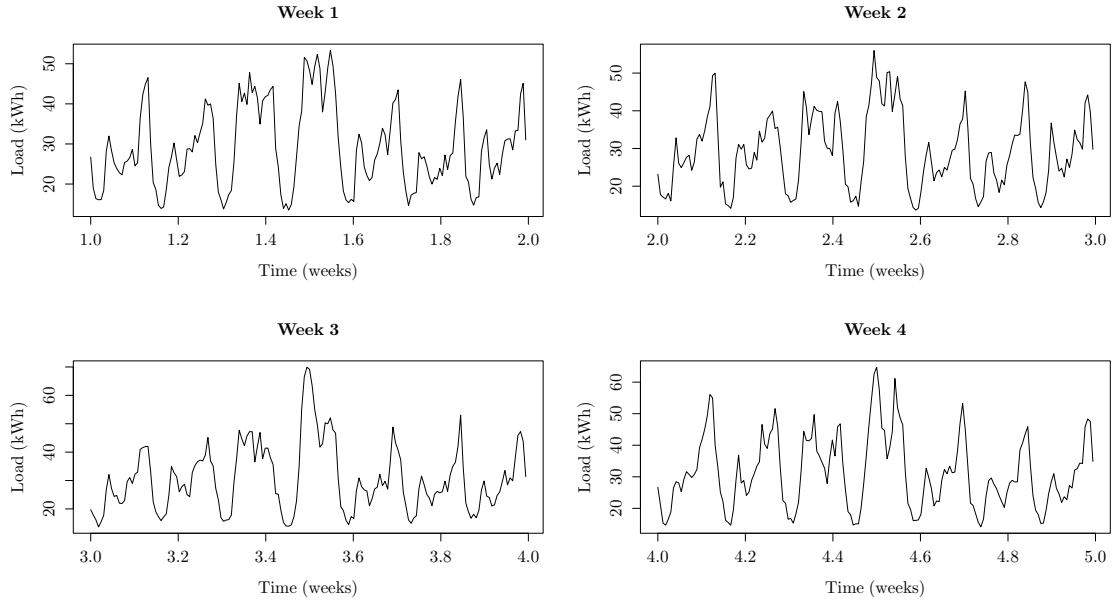
Displaying the aggregated load profile of **69** residential consumers shows that it is less volatile than those of all the individual households and is starting to show some *pattern*. This is due to purely statistical reasons and was expected. The more consumers are aggregated, the smoother the profile will become. Figure 17 is displaying the load profile of SUM69, as we labeled the group of residential consumers, over a period of four weeks.

Figure 17. Time Series of SUM69 Residential Group for May 2013, 1-h Sampling



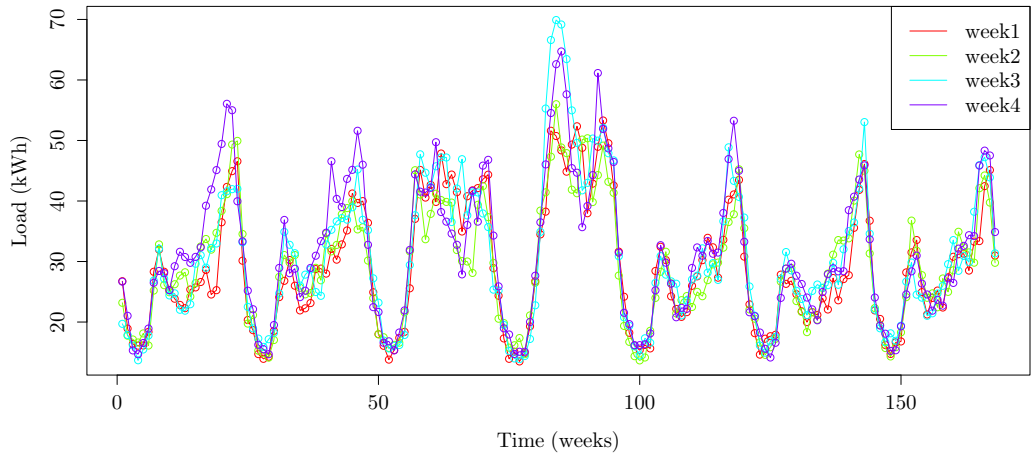
All four weeks display similarities, as we can see in Figure 18.

Figure 18. Time Series of SUM69 Residential Group, 4 Weeks in May 2013, 1-h Sampling



The extent of similarity of the load profiles becomes more evident by examining the seasonal plot in Figure 19. The profile is not only smoother but also displays more consistent behaviour.

Figure 19. Seasonal Plot of SUM69 Residential Group, 4 Weeks in May 2013, 1-h Sampling



We are hopeful that groups of residential consumers will be **predictable**. The fact that observations for the four consecutive weeks at any specific time of week do not differ much from each other indicates that even predicting load as the average of those four measurements would result in a forecast with fairly good accuracy.

The lowest number where one might start seeing this effect is probably **30 households**.

This is a consequence of statistical laws. The aggregated load profile of several households becomes *more predictable* the more households are aggregated together. Loads of individual households are *by a magnitude of 1000 smaller* than loads of individual industrial consumers. Due to the above facts and since we only had 120 household sources for testing in the pilot project, we recommended aggregating household data into *2 groups of 60 households*, at random for the purpose of forecasting in the pilot project.For the "real-life" situation, it would be better to aggregate into groups that have a summary load of around 1000 kW, comparable to that of an industrial consumer.

4.6 Exploratory Data Analysis

In this section, we will analyze the features of load time series beyond what is revealed by looking at time plots and seasonal plots.

4.6.1 Industrial consumers

Since the situation for groups of residential consumers is similar to that of individual industrial consumers, it is important to understand the latter in depth.

We can think of a time series as being composed of several components, each representing one or more of the underlying patterns. **Decomposing the time series** will aid in better understanding. It is also a good start for forecasting, as components are forecast separately and later assembled, and this *bottom-up* forecasting approach usually benefits the forecasting accuracy.

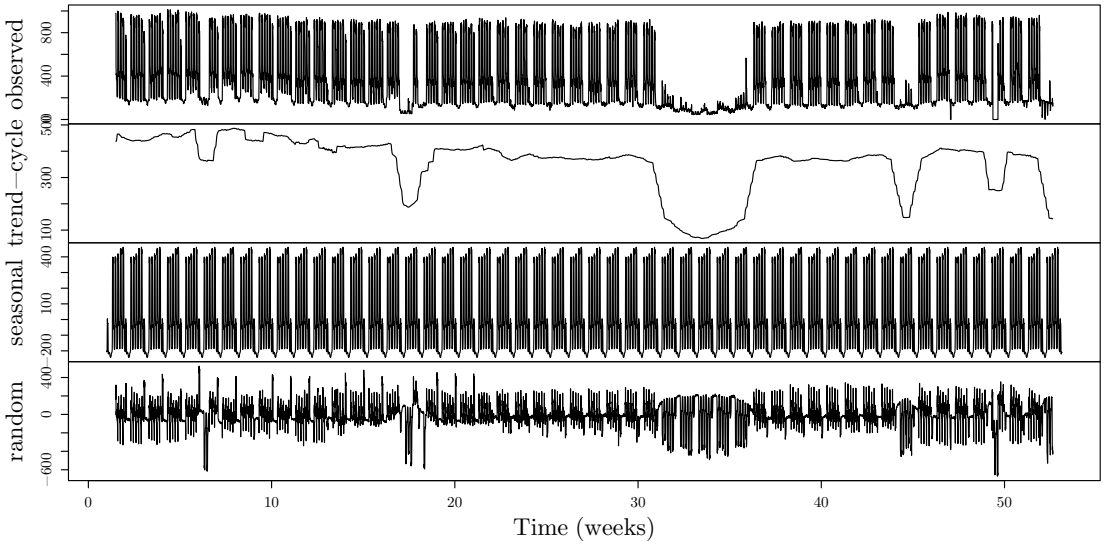
We will split the series into the three components: the **seasonal** component, the **random**, irregular or error component, and the remaining **trend-cycle** component. The last component is not the usual trend component, as it also contains business cycle variations. The decomposition can be obtained by several different methods. We will apply the *classical seasonal decomposition by moving averages* method. (For details on this decomposition method consult Section 5.3.) Since the magnitude of seasonal fluctuations does not change over time, we will select the *additive model*, where components are added, rather than multiplied:

$$\boldsymbol{y_i = T_i + S_i + E_i,} \tag{15}$$

whereby $\boldsymbol{y_i}$ denotes the signal value at time $\boldsymbol{t_i}$, $\boldsymbol{T_i}$ denotes the trend component at time $\boldsymbol{t_i}$, $\boldsymbol{S_i}$ denotes the seasonal component at time $\boldsymbol{t_i}$ and $\boldsymbol{E_i}$ the error component at time $\boldsymbol{t_i}$. All times $\boldsymbol{\{t_1, t_2, t_3, \ldots, t_T\}}$ belong to the historical period.

Figure 20 displays the seasonal decomposition for Ind1 load time series for 2011. We set the frequency of the seasonal component to 168 hours (which is equivalent to 7 days).

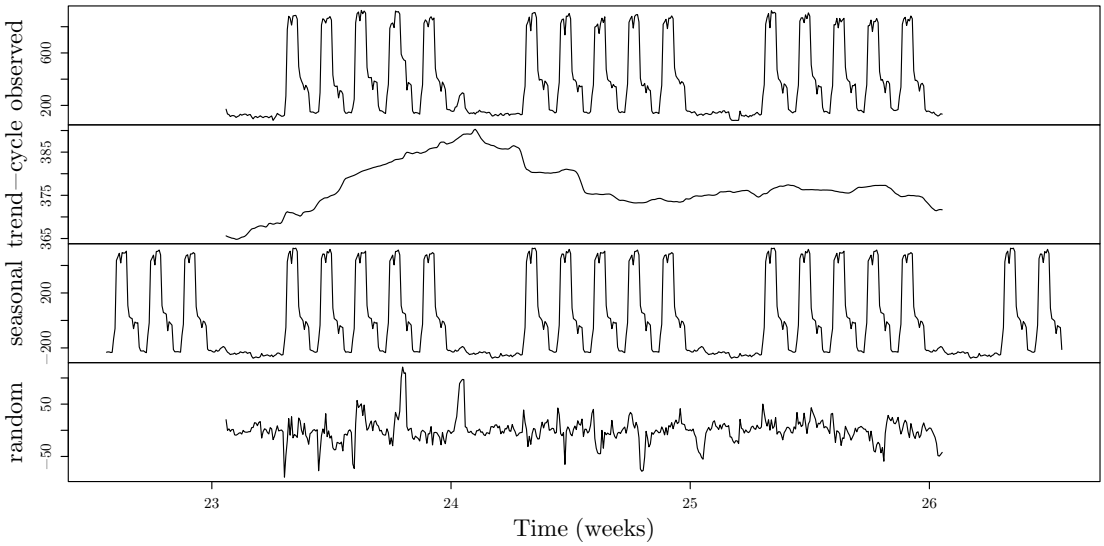
Figure 20. Classical Seasonal Decomposition by Moving Averages of Ind1 Time Series for 2011, 1-h Sampling



The *trend-cycle* component, in our case, does not represent a long-term increase or decrease of signal values but is rather indicative of a business cycle. It clearly shows a consumption drop during the $\mathbf{1}^{st}$-of-May period which is a national holiday in Slovenia, a longer reduction during the summer months of July and August and another drop around November $\mathbf{1}^{st}$ (the so-called "potato holidays"). The *seasonal* component looks regular. The *random* component has a mean of $\mathbf{-0.0138}$ which means that this component is close to being random (so-called *white noise*) and does not contain any more pattern. The decomposition was successful.

In Figure 21, we can observe the same decomposition of Ind1-signal on a monthly scale, for June 2011 data. In subsection 5.3 we will learn an alternative decomposition method, while some forecasting methods that use *two seasons* rather than one will be presented in Sections 5.5.4, 5.5.6, 5.5.7.

Figure 21. Classical Seasonal Decomposition by Moving Averages of Ind1 Time Series for June 2011, 1-h Sampling



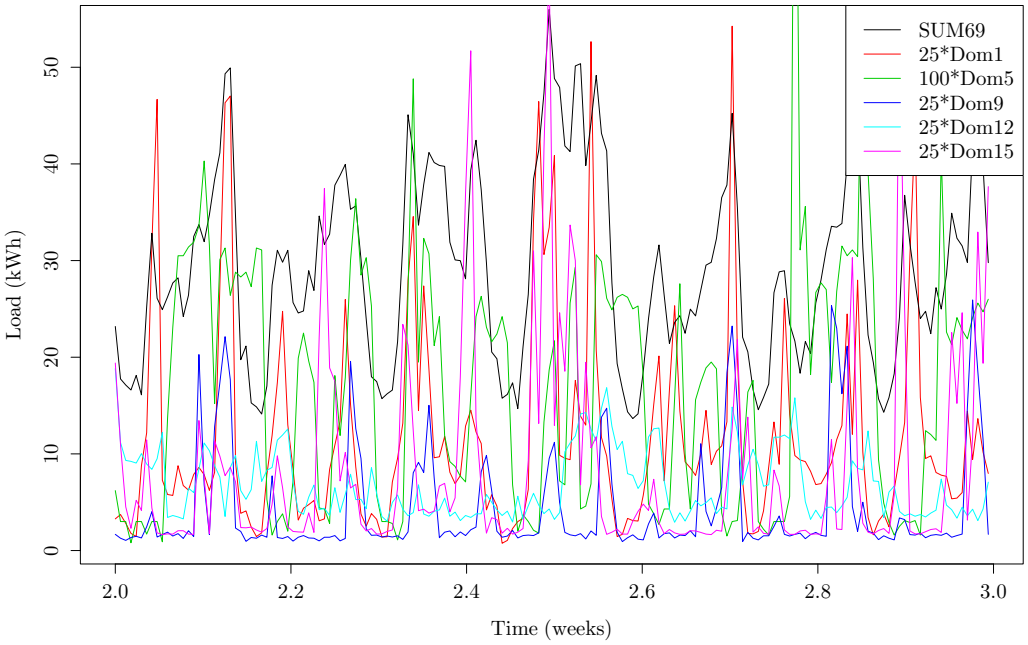
4.6.2 Residential consumers

As was observed in Section 4.5, the load signal of SUM69 is less volatile than any single household's signal. Let us *compare volatility visually*. We will plot the load signals of single households against the plot of SUM69, in the hope of seeing some correlations that would enable splitting the group reward in a DR event at least proportionally. Load signals of single households will be magnified by a corresponding factor to have the same magnitude as SUM69.

How are the load profiles of single households related to the load profile of the entire SUM69

group? We observe in Figure 22 that load signals of single households move in opposite directions at most points and are also not "in tune" with SUM69. This would make sharing the DR reward proportionally between households participating jointly in a DR event unfair. Also, if we observed a spike for a particular household at the time of the DR event, it would not automatically mean it *is not* curtailing. Moreover, vice versa: if we observed a valley, it would not mean that it *is* curtailing. There is just no way to tell which household contributed to the curtailment of the group if we can not predict the behavior of a single household. *New business models* to reward domestic DR participants will have to be invented.

Figure 22. A Plot of SUM69 Residential Group Against Magnified Single Household Plots, Week 2 of May 2013, 1-h Sampling



The reward model will have to be more *qualitative* than quantitative in nature. Since the consumption that would have occurred in the absence of a DR event will not be known at the level of any single household, there is also no possible fair way to reward single households.

Households will have to be *rewarded as a group*, and the reward will possibly be *split* based on a combination of the following factors: the average yearly consumption of each household, the load of each household during the DR event compared to its average consumption at the same timeframe in the past 4 weeks. (The fact that a household would be below the average consumption could be an indicator of curtailment. Only curtailing households would split the prize.) The *fairness* of this model remains *to be tested*.

4.6.3 Anticipation of forecasting efficiency

The *industrial example* showed that a typical load profile of an individual industrial consumer decomposes nicely into the trend, seasonal and random components, and we are hopeful we will be able to successfully predict its short-term future load.

Data visualisation and analysis for residential consumers has shown that:

- Individual households will *not* be predictable.
- If groups of households will be forecast jointly the question of how to *fairly reward* single DR participants remains to be answered. (This might result in many new business models. Some of them might differ considerably from the established U.S. DR models).

5 METHODS

There is a vast pool of methods that aim to solve the short-term load forecasting task. Very different approaches have been tested in academic circles. They range from very simple approaches (such as predicting a season's load as the average of a few past seasons) to the more advanced and popular ARIMA models and exponential smoothing methods (e.g. Holt-Winters), to more contemporary methods (e.g. artificial neural networks) and various hybrid models.

Models that include seasonal factors (e.g. time of day, day of week) are usually more successful in predicting electrical load, which is by its nature seasonal.

Different methods, available in the `forecast` package in `R`, developed by Professor Rob Hyndman will be tested. (Refer to package documentation at (Hyndman, 2016)). The method ultimately selected for implementation into a VPP software solution will have to satisfy some specific conditions. The following features are crucial:

- **automated**: it will be able to predict time series that have different features, without human intervention,
- **robust**: it will handle missing data, outliers, etc. autonomously,
- **streaming data**: it will be able to run online and in *real time*, on a large number of different time series (bulk forecasting).

The last requirement means that highly accurate but time-consuming methods will be discarded. In this section, we will review the forecasting methods which will be used in the experiments and describe their main features.

5.1 A Few Simple Methods

The methods presented in this section are too simple to be able to capture the complex nature of the load profile so, except for the seasonal ones, we do not expect them to perform well at our task. They will only be used to either visually compare their sophistication level to that of more advanced methods or to use their performance as a benchmark to be surpassed.

Average Method. The forecast of all future values is set to be the mean of all historical load observations:

$$\widehat{y}_{T+h|T} = \bar{y} = \frac{1}{T}\sum_{i=1}^{T} y_i;\ h = 1, 2, 3, \ldots, \tag{16}$$

where by $\hat{y}_{T+h|T}$ we denoted the estimate of y_{T+h} based on historical data $y_1, \ldots, y_T$. This is the simplest forecast method and merely a rough estimate of the actual values. It does not consider any specific features of the time series to be forecast, like seasonality, etc. It will be inaccurate most of the time.

Naïve Method. Naïve forecasting models assume that things will remain as they have in the past. In time series, the naïve model only uses the last known observation as the prediction for all future values:

$$\widehat{y}_{T+h|T} = y_T;\ h = 1, 2, 3, \ldots \tag{17}$$

Note that within the `R` software `naïve()` function is simply a convenient wrapper for the ARIMA function that returns forecasts and prediction intervals of ARIMA$(0, 1, 0)$ random walk. It obviously does not take seasonality into account.

Drift Method. This method predicts that the future values will linearly increase or decrease with the rate of the average change between two consecutive historical observations, called *drift*:

$$\widehat{y}_{T+h|T} = y_T + h \cdot \frac{\sum_{t=2}^{T}(y_t - y_{t-1})}{T-1} = y_T + h\left(\frac{y_T - y_1}{T-1}\right). \tag{18}$$

Note that this equals the overall increase/decrease rate over the entire historical period. Again, this method is not appropriate for forecasting a seasonal time series.

Seasonal Naïve Method. This method already considers a simple form of seasonality: each forecast is set to be equal to the last observed value from the same season of the year (e.g., the same day in the previous week).

$$\widehat{y}_{T+h|T} = y_{T+h-km};\ h = 1, 2, 3, \ldots, \tag{19}$$

where $m =$ is the seasonal period, and $k = \lfloor (h-1)/m \rfloor + 1$ is the number of seasons necessary to "reach" historical data. The last observed seasonal cycle is merely repeated as necessary beyond the forecast origin to the forecast horizon.

U.S. Standards Baseline Type I Method. This method is founded in the requirements for a "good baseline" (see Section 2.3). Several variations of this method exist, such as high X of Y, regression, comparable day and rolling average (see Enernoc (2011) for details). We used a simple high 10 of 10 method, which calculates hourly averages, using data from 10 most recent days preceding the DR event, and used this as a prediction. In this way, the method considers the form of the signal (if the signal is seasonal, so will its prediction be). It is straightforward, well-defined and has no parameters, which is desirable whenever a baseline calculation has significant legal or economical consequences.

5.2 Time Series Models – Basic Concepts

Time series models belong to the group of *univariate* methods. They use only information available on the variable to be forecast and do not try to uncover any other factors that might influence its behavior. They will extract and predict the *trend* and *seasonal* patterns but will completely ignore any other factor, that might considerably alter the predicted variable's behavior, such as weather conditions, as well as market and business changes.

All time series models are **stochastic models**. The forecasts are given as the average of all values that the series will likely assume at a specific time. The uncertainty is assessed and specified in a *prediction interval*. (For a definition, consult page 19.)

The basic concepts explained in this section are fundamental for the understanding of time series models that are covered in the Sections 5.3–5.5. These are: *forecasting by time series decomposition, ARIMA models*, and *exponential smoothing methods*.

5.2.1 Stationary stochastic processes

Definition 1. *A stochastic process* $Y = \{Y_i, i = 1, 2, \ldots\}$ *is* **strongly stationary** *if its joint probability distribution does not change when shifted in time.*

As a consequence, all of its statistical properties will remain unchanged in time. A strongly stationary process can also be called a *strictly stationary* or a *first-order stationary* process.

Definition 2. *A stochastic process is* **weakly stationary** *if its first and second moments are finite and constant over time.*

A weakly stationary process is also called *covariance stationary* or *second-order stationary* or simply *stationary*. More precisely, if $\boldsymbol{Y} = \{\boldsymbol{Y_i}, \boldsymbol{i} = \boldsymbol{1, 2, \ldots}\}$ is a stationary stochastic process, then at every time $\boldsymbol{t_i}$ the following conditions hold:

- $\boldsymbol{E(Y_t)} = \boldsymbol{\mu}$ and
- $\boldsymbol{Cov(Y_t, Y_{t-h})} = \gamma_h$,

where $\boldsymbol{\mu}$ is a finite constant and γ_h is independent of time $\boldsymbol{t}$.

Note that $\boldsymbol{E(Y_t)}$ is the expected value of random variable $\boldsymbol{Y_t}$ and calculated by employing its probability density function $\boldsymbol{f_t}$. A similar statement could be made about the calculation of covariance γ_h. As we will see later (e.g. equation (21)), the mean, autocovariance and autocorrelation of the *observed* time series is calculated in a *deterministic* manner, from sample data.

Note that any strongly stationary process is also automatically weakly stationary, due to more relaxed requirements. A time plot of a stationary time series will be roughly horizontal and have a constant variance. There will be no visual patterns, neither trend nor seasonal.

The simplest example of a stationary process is white noise.

Definition 3. *A stochastic process* $\{\epsilon_t\}$ *is called* **white noise** *if it has zero mean, a constant variance* $\boldsymbol{\sigma^2}$*, independent of time and zero covariance.*

When a time series model explains the historical data well then the residual time series will be a white noise process.

5.2.2 ACF, PACF and unit root tests

An **autocovariance coefficient** (hereinafter ACV coefficient) defines the covariance between a random variable and its lag: $\gamma_h = \boldsymbol{Cov(Y_t, Y_{t-h})}$. An **autocovariance function** (hereinafter ACVF) is the set $\{\gamma_h; \boldsymbol{h = 1, 2, 3, \ldots}\}$. Note that $\gamma_0 = \boldsymbol{\sigma^2}$.

More frequently we will use the **autocorrelation coefficients**:

$$\boldsymbol{\rho_h} = \frac{\gamma_h}{\gamma_0} = \frac{\boldsymbol{E[(Y_t - \mu)(Y_{t-h} - \mu)]}}{\boldsymbol{E[(Y_t - \mu)^2]}} = \frac{\boldsymbol{E[Y_t Y_{t-h}]} - \boldsymbol{\mu^2}}{\boldsymbol{E[Y_t^2]} - \boldsymbol{\mu^2}}, \tag{20}$$

where $\boldsymbol{\rho_h}$ is the autocorrelation for lag $\boldsymbol{h}$, γ_h is the autocovariance for lag $\boldsymbol{h}$, $\gamma_0 = \boldsymbol{\sigma^2}$ is the time series variance and $\boldsymbol{\mu_{t-h}} = \boldsymbol{\mu_t} = \boldsymbol{\mu}$ is the stationary time series mean. An

autocorrelation function (hereinafter ACF) is the set $\{\rho_h; h = 1, 2, 3, \ldots\}$. An AC coefficient measures the cross-correlation of a process with itself at different lags.

The ACF is decreasing to zero with an increasing lag h. A plot of the ACF coefficients $\rho_0, \rho_1, \rho_2, \ldots$ against the lag h, called an **ACF plot**, is important for two reasons:

- it may be able to determine whether a time series is stationary or not. A stationary time series' ACF will decline to zero relatively fast, while in the case of a non-stationary time series the decrease will be slow;
- it may be able to identify the presence and length of different seasonal cycles in a given time series and determine specific situations that call for a specific model.

An ACF plot of the historical or sample data is called the **correlogram**. For finite sample data, the empirical autocorrelations are *estimated* from sample data and not calculated from random variables Y_t and Y_{t-h} at a specific time t. For example:

$$\hat{\rho}_h = \frac{\sum_{t=h}^{N}(y_t - \bar{y})(y_{t-h} - \bar{y})}{\sum_{t=h}^{N}(y_t - \bar{y})^2}, \tag{21}$$

where $\hat{\rho}_h$ is an autocorrelation for lag h, of the sample ACF, N is the number of all historical observations, and $\bar{y}$ is the sample mean. Examining the correlogram of a time series helps answer the question whether a time series is random or not. A random time series' AC coefficients will be roughly zero: $\hat{\rho}_h \approxeq 0$. We will examine the ACF of the residual time series to prove or discard its randomness.

The **partial autocorrelation function** (hereinafter PACF) specifies the correlation between a variable and a lag of itself that is not explained by correlations at all lower-order lags. For example, the partial autocorrelation at lag 2 is the difference between the actual correlation at lag 2 and the expected correlation due to the "propagation" of correlation at lag 1.

Unit Root Tests. When the ACF plot was unsuccessful in determining whether a time series is stationary or not, we will have to rely on hypothesis testing. Several statistical tests can tell if a time series needs differencing to become stationary or not. The two most popular ones are the Augmented Dickey-Fuller (ADF) test, presented in (Said & Dickey, 1984) and (Banerjee, Dolado, Galbraith, & Hendry, 1993), and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, which was first published in (Kwiatkowski, Phillips, Schmidt, & Shin, 1993).

5.2.3 Time series differencing

Many of the most important time series models apply to stationary time series. However, most time series that we encounter in practice are non-stationary. A non-stationary time series can often be turned into a stationary one by a process called **differencing**. It creates a new time series, made up of differences $\{Y_t - Y_{t-1}\}$ between consecutive observations. If the resulting time series is not stationary, we can repeat the process of differencing by taking first differences as input. The process of differencing removes changes in the *level* of a time series that is the trend and seasonal components and this way "stabilizes" the mean in time.

We will first introduce the notion of backshift operator, as that will help a more compact presentation.

A **backshift operator** B shifts the stochastic process $\{Y_t\}$ back in time by one period: $BY_t = Y_{t-1}$; for every t. Applying the backshift operator h times will shift the data back by h periods: $B^h Y_t = B^{h-1}(BY_t) = Y_{t-h}$; for every t.

Using the backshift notation, a **first difference** is the difference of a process $\{Y_t\}$ and its first-order lag process $\{Y_{t-1}\}$:

$$\Delta Y_t = (1 - B)Y_t = Y_t - BY_t = Y_t - Y_{t-1}. \tag{22}$$

By creating first differences of first differences, we obtain **second-order differences**:

$$\Delta^2 Y_t = (1 - B)^2 Y_t = (1 - 2B + B^2)Y_t = Y_t - 2Y_{t-1} + Y_{t-2} \tag{23}$$

The dth-order difference is denoted by $(1 - B)^d Y_t$. Most non-stationary processes can be rendered stationary by first- or second-order differencing. Note, that the second-order difference operator $\Delta^2 = (1 - B)^2 = (1 - 2B + B^2)$ is different from the second difference operator $(1 - B^2)$. The latter just subtracts a second-order lag from the original process.

5.2.4 Seasonal differencing

When the time series does not include a trend but appears to be seasonal, then seasonal differencing might be all it takes to obtain a stationary process. A **seasonal difference** is the difference between an observation and a same-season past observation from the previous period.

$$Y_t{}' = Y_t - Y_{t-m}; \tag{24}$$

where m is the number of seasons (e.g. the number of months in a monthly-data yearly time series). If the seasonally differenced time series $y_t{}'$ is white noise, then the time series can be modeled as:

$$Y_t = Y_{t-m} + \epsilon_t \tag{25}$$

Forecasts equal the last observation from the relevant season, hence are seasonal naïve forecasts.

5.2.5 Wold's representation theorem

Any weakly stationary stochastic process $\{Y_t\}$ can, according to **Wold's representation theorem** (Wold, 1938), be written in the form of a *general* **linear time series model**:

$$Y_t = \mu + \left(\epsilon_t + \sum_{i=1}^{\infty} \psi_i \epsilon_{t-i} \right). \tag{26}$$

The *deterministic part* $\mu = \mu_t = E[Y_t]$ is the mean of any of the random variables $\{Y_t\}$. The second part of representation is *stochastic*, where $\{\epsilon_t\}$ is a white noise process, and ϵ_t is the difference between the observed value Y_t at time t and the optimal forecast μ_t of that value based on information available prior to time t. As the process $\{\epsilon_t\}$ captures all new information to the process $\{Y_t\}$ it is also called the **innovations process** and the errors ϵ_t are called **innovations**. In backshift notation we define an **infinite degree lag polynomial** $\psi(\cdot)$ of coefficients $\psi_1, \psi_2, \psi_3, \ldots$ as:

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots, \tag{27}$$

we can rewrite Wold's linear representation of the stationary stochastic process (26) as

$$Y_t = \mu + \left(1 + \sum_{i=1}^{\infty} \psi_i B^i \right) \epsilon_t = \mu + \psi(B)\epsilon_t \tag{28}$$

5.2.6 Causal and invertible stochastic processes

A process $\{Y_t\}$ is said to be **causal** or a **causal function of process** $\{\epsilon_t\}$ if it has a representation

$$Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \tag{29}$$

or in other words, when $\boldsymbol{Y_t = \psi(B)\epsilon_t}$ where $\boldsymbol{\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \ldots}$ and $\sum_{i=1}^{\infty} |\psi_i| < \infty$. As we will see in Section 5.4.2, this means that a causal process can be represented by a MA(∞) process.

A process $\{\boldsymbol{Y_t}\}$ is said to be **invertible** or an **invertible function of process** $\{\boldsymbol{\epsilon_t}\}$ if it has a representation

$$\boldsymbol{Y_t = \sum_{i=0}^{\infty} \phi_i Y_{t-i} + \epsilon_t,} \tag{30}$$

or in other words, when $\boldsymbol{\epsilon_t = \phi(B)Y_t}$ where $\boldsymbol{\phi(B) = \phi_0 + \phi_1 B + \phi_2 B^2 + \ldots}$ and $\sum_{i=1}^{\infty} |\phi_i| < \infty$. In other words, an invertible process can be represented by an AR(∞) process, which is a suitable for forecasting purposes. (See Section 5.4.1 for description of process).

5.2.7 Characteristic polynomial

Here we will present a result, needed in Section 5.4.1, when covering autoregressive processes. An **autoregressive process of order** $\boldsymbol{p}$ (or AR($\boldsymbol{p}$) process) is a weighted linear combination of the past $\boldsymbol{p}$ values plus a white noise process $\boldsymbol{\epsilon_t}$:

$$\boldsymbol{Y_t = (\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p}) + \epsilon_t.} \tag{31}$$

Using the backshift $\boldsymbol{B}$ and setting $\boldsymbol{\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)}$ we can rewrite it as:

$$\boldsymbol{\phi(B)Y_t = \epsilon_t,} \tag{32}$$

where $\boldsymbol{\phi(\cdot)}$ is a polynomial in $\boldsymbol{B}$ of order $\boldsymbol{p}$, called a $\boldsymbol{p}$-degree **characteristic polynomial**. Note that $\boldsymbol{\phi(B)}$ is actually an operator, acting on process $\boldsymbol{Y}$ in the following way: $\boldsymbol{\phi(B)Y = \epsilon}$. Equation 32 is by its nature a difference equation.

Under which conditions is the AR($\boldsymbol{p}$) process $\boldsymbol{Y}$ weakly stationary?
To be stationary, $\boldsymbol{Y}$ must fulfill two requirements: 1) the finiteness of its expected value and 2) the finiteness of its autocovariances and their independence of time.

We will state the result here without proof:

Theorem 1. *Let $\boldsymbol{Y}$ be an AR($\boldsymbol{p}$) process, defined by (32). Equation (32) has a unique causal stationary solution provided that the roots $\{\boldsymbol{\zeta_i}\}$ of the* **characteristic equation***:*

$$\boldsymbol{\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0}\textit{; for } \boldsymbol{z \in \mathbb{C}.} \tag{33}$$

lie outside the unit circle (or: $|\zeta_i| > 1$; for every i).

While trying to satisfy the second condition for stationarity in the theorem, the AC coefficients ρ_h had to be computed. It turned out that the autocorrelations $\{\rho_h\}$ are expressed in terms of the roots of the characteristic equation. This is the connection between ACF and the roots of the characteristic equation.

Note, that with theorem 1 we merely specified another way to test the stationarity of a process, in this case a AR(p) process. It follows that this process can be represented in the form of a MA(∞) process.

5.2.8 Box-Ljung test

This statistical test by Ljung and Box (1978) is used to test the lack of fit of an autoregressive moving average time series model. The test is applied to the residual time series after the model has been fitted to the data. Instead of testing residual autocorrelation at each distinct lag, as the ACF does, this test examines autocorrelations up to lag m and expresses an "overall" judgement of the randomness of the residuals. It is therefore called a *portmanteau test*. The test is defined as:

$\mathbf{H_0}$: The model does not exhibit lack of fit.
$\mathbf{H_1}$: The model exhibits lack of fit.

For a stochastic process Y with N historical observations the **Ljung-Box statistic** is a function of accumulated sample autocorrelations $\hat{\rho}_h$, up to a specified time lag m:

$$Q(m) = N(N+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{N-h}. \tag{34}$$

We will reject the null hypothesis (indicating that the model has a significant lack of fit) whenever the test statistic has a value that is highly unlikely to be a coincidence. This will be the case when

$$Q(m) > \chi^2_{1-\alpha,h}, \tag{35}$$

where $\chi^2_{1-\alpha,h}$ is the chi-square distribution table value with h degrees of freedom and significance level α. Note that, since we are applying the test to residuals, the degrees of freedom must account for the estimated model parameters of an ARMA(p, q) model, hence: $h = m - p - q$.

The test is a substantial improvement over the original test by Box and Pierce (1970).

5.2.9 Information criteria

An **information criterion** measures the appropriateness of a statistical model for modeling a specific set of historical data. It looks for a tradeoff between *model complexity* and *goodness of fit*. Selecting the model with the lowest information criterion will ensure the best fit while keeping the complexity as low as possible.

The most used information criterion is the **Akaike information criterion** or **AIC**:

$$\boldsymbol{AIC = 2k - 2ln(L),} \tag{36}$$

where $\boldsymbol{k}$ is the number of estimated parameters in the model and $\boldsymbol{L}$ is the maximum value of the *likelihood function* for the model. AIC was first described in (Akaike, 1974). For the case of estimating an ARMA($\boldsymbol{p, q}$) model, the AIC will result in:

$$\boldsymbol{AIC(p, q) = \frac{2}{T}\,(p + q)\log \hat{\sigma}^2(p, q),} \tag{37}$$

where $\boldsymbol{\hat{\sigma}^2(p, q)}$ is the estimated variance of the ARMA($\boldsymbol{p, q}$) model's errors $\boldsymbol{\epsilon_t}$.

5.3 Forecasting with Time Series Decomposition

As mentioned in Subsection 4.6.1, when predicting a time series, decomposing it into single patterns, which we described in Section 4.5, and separately forecasting each will most likely result in higher overall forecast accuracy. Several different decomposition methods have been developed since the 1920s, each with its own advantages and disadvantages.

5.3.1 Classical decomposition method

This method by Kendall, Stuart, and Ord (1983, pp. 410–414) splits the time series into a sum or a product of its cyclical, seasonal, trend, and error components. These components are then analyzed individually. The method, in use from the 1920s to the 1950s, is no longer recommended due to its shortcomings. We will briefly outline its **additive** version here as it forms the bases of all later methods:

Step 1: The estimate $\boldsymbol{\hat{T}_i}$ of the trend-cycle component at time $\boldsymbol{t_i}$ is obtained by using a $\boldsymbol{2 \times m}$ moving averages procedure:

$$\boldsymbol{\hat{T}_i = \frac{2}{m} y_{i-m} + \frac{1}{m} \sum_{j=-m+1}^{m-1} y_{i+j} + \frac{2}{m} y_{i+m},} \tag{38}$$

where m is the seasonal period and $\hat{T}_i$ can be computed for all historical times, except the first and last m ones.

Step 2: The *detrended time series* is calculated by subtracting the trend-cycle component from the original time series: $y_i - \hat{T}_i$.

Step 3: In the case of monthly data the seasonal component is estimated by first computing *seasonal indices* for each month as averages of all detrended values available for that month in all data. The indices are then normalized to add up to $\mathbf{0}$. They define the estimate $\left\{\hat{S}_i\right\}_i$ of the seasonal component.

Step 4: The error component is calculated by subtracting the estimated trend and seasonal components from the original signal:

$$\hat{E}_i = y_i - \hat{T}_i - \hat{S}_i \tag{39}$$

The above procedure is implemented in the `decompose()` function. For details of this method consult the original description in (Kendall et al., 1983, pp. 410–414).

Several alternative decomposition methods were developed by the United States Census Bureau since then. The X-11 decomposition method by Shiskin, Young, and Musgrave (1967) was an improvement of the classical decomposition method. This later evolved into the improved X-11 ARIMA decomposition by Dogum (1988) and was superseded with the X-12 ARIMA decomposition method (see: Findley, Monsell, Bell, Otto, and Chen (1998)). The X-12 ARIMA method has overcome all shortcomings of the classical decomposition method: it is able to estimate the trend and seasonal components on all historical points, it is robust to occasional unusual observations; variations can handle trading day variation, holiday effects and effects of known predictors. Recently, the Time Series Research Staff (2016) added X-13 ARIMA SEATS decomposition method to the census II family. In `R` the package `seasonal` represents an interface to the free software for X-13 ARIMA SEATS, provided by the Census Bureau.

5.3.2 Forecasting with STL decomposition

The **Seasonal Decomposition of Time Series by Loess** (or STL decomposition) is an improvement over the classical and X-12 ARIMA methods. It was developed by R. B. Cleveland, W. S. Cleveland, McRae, and Terpenning (1990). The trend and seasonal components are computed by using the *LOESS smoothing procedure* on the trend and seasonal components iteratively over several steps. The LOESS smoother calculates a weighted average of a few neighbouring values where closer values get a bigger weight. After the components estimations are known the decomposition performance needs to be accessed (The random component needs to be “white noise”).

Forecasting for the *additive* decomposition model is done by splitting the time series into the seasonal component $\hat{\boldsymbol{S}}_i$ and the seasonally adjusted component $\hat{\boldsymbol{A}}_i$:

$$\boldsymbol{y_i} = \hat{\boldsymbol{S}}_i + \hat{\boldsymbol{A}}_i; \text{ where } \hat{\boldsymbol{A}}_i = \hat{\boldsymbol{T}}_i + \hat{\boldsymbol{E}}_i. \tag{40}$$

Both components are then predicted separately. The seasonal component is assumed to change little or not at all over the years, so it is predicted using the seasonal naïve forecasting method: last period's actuals are set to be this period's forecasts, without adjusting them. (e.g., we predict next year/week by taking last week/year as the estimate.) The seasonally adjusted component can then be predicted by using any non-seasonal forecasting method, e.g. the Holt method or the non-seasonal ARIMA method. This method is implemented in `R` as the `stl()` function.

5.4 ARIMA models

An autoregressive integrated moving average (hereinafter ARIMA) model consists of three parts: the autoregressive (hereinafter AR), the integrated (hereinafter I) and the moving average (hereinafter MA) and any of these parts (but not all) can be omitted. This group of models is among the most important ones for *linear* time series modelling. They were first described in (Box & Jenkins, 1970, sec. 1.5, 3.1), now revised in (Box, Jenkins, & Reinsel, 1994, sec. 1.5, 2.3, 3.1, 3.5, 4.2, 5.1–5.3, 5.6, 7.4, 7.5, 7.7, 7.8, 8.2).

The model is usually identified (that is, the orders are established) by comparing the sample's ACF and PACF to the properties of the model ACFs and PACFs. The model is then fitted to the data and the resulting model is used for stepwise forecasting.

5.4.1 Autoregressive (AR) models

An autoregressive model (hereinafter AR model) is similar to a multiple regression model in that it forecasts the future values $\{\boldsymbol{y_t}\}$ of the forecast variable $\boldsymbol{y}$ as a linear combination of predictive variables. However, these are not external variables but past observations $\{\boldsymbol{y_{t-1}}, \boldsymbol{y_{t-2}}, \ldots, \boldsymbol{y_{t-p}}\}$ of the same forecast variable.

An **autoregressive model of order p** or **AR($\boldsymbol{p}$)** model can thus be written as:

$$\boldsymbol{Y_t} = \boldsymbol{\phi_1 Y_{t-1}} + \boldsymbol{\phi_2 Y_{t-2}} + \ldots + \boldsymbol{\phi_p Y_{t-p}} + \boldsymbol{\epsilon_t}, \tag{41}$$

where $\boldsymbol{Y_t}$ is the forecast random variable, its "lags" $\boldsymbol{Y_{t-1}}, \boldsymbol{Y_{t-2}}, \boldsymbol{Y_{t-3}}, \ldots$ are the explanatory random variables, $\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \ldots, \boldsymbol{\phi_p}$, $\boldsymbol{p}$ is the number of lags included are model parameters to be estimated and the error terms $\{\boldsymbol{\epsilon_t}\}$ form a white noise stochastic process. The process

$\boldsymbol{Y} = \{\boldsymbol{Y}_t\}$, constructed from equation 41, is called an *autoregressive process of order* $\boldsymbol{p}$.

After the model is fitted to historical data, the solution can be written as:

$$\hat{\boldsymbol{Y}}_t = \hat{\phi}_1 \boldsymbol{Y}_{t-1} + \hat{\phi}_2 \boldsymbol{Y}_{t-2} + \ldots + \hat{\phi}_p \boldsymbol{Y}_{t-p}, \tag{42}$$

where $\hat{\boldsymbol{Y}}_t$ is the forecast random variable and $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$ are estimations of model parameters. This *forecast equation* is used repeatedly for a stepwise forecast.

An AR($\boldsymbol{p}$) process has a unique weakly stationary causal solution if and only if the roots of the characteristic equation are all situated outside of the unit circle (as we have seen in theorem 1).

To *identify* the AR($\boldsymbol{p}$) model by determining its lag $\boldsymbol{p}$, the following characteristics of the ACf and PACF ca be considered:

- the ACF of an AR($\boldsymbol{p}$) process has a geometrical decline,
- the PACF of an AR($\boldsymbol{p}$) process cuts off at lag $\boldsymbol{p}$.

5.4.2 Moving average (MA) models

A moving average model (hereinafter MA model) is similar to the AR model, but rather than using past values $\{\boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}, \ldots, \boldsymbol{Y}_{t-p}\}$ to predict the future of a stochastic process $\boldsymbol{Y}$, it uses past forecast errors $\boldsymbol{\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-p}}$.

A **moving average model of order $\boldsymbol{q}$** or **MA($\boldsymbol{q}$) model** can be written as:

$$\boldsymbol{Y}_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \ldots + \psi_q \epsilon_{t-q}, \tag{43}$$

where $\boldsymbol{Y}_t$ is the forecast random variable, the forecast errors $\{\boldsymbol{\epsilon_{t-h}}\}$, defined by $\boldsymbol{\epsilon_{t-h}} = \boldsymbol{Y_{t-h}} - \hat{\boldsymbol{Y}}_{t-h}$ constitute a white noise random processes and $\psi_1, \psi_2, \ldots, \psi_p$ are model parameters. Note that the number of lags included $\boldsymbol{q}$ can be arbitrarily large.

MA models are extremely important because of Wold's representation Theorem (see page 53) which states that every weakly stationary stochastic process $\{\boldsymbol{Y}_t\}$ can be approximated to an arbitrary degree of precision by a sum of a stochastic and a deterministic time series. The stochastic part is effectively an MA(∞) process:

$$\boldsymbol{Y}_t = \boldsymbol{\delta}_t + \left(\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \right). \tag{44}$$

The degree of precision of approximation is directly related to the order q of MA.

For a MA(q) process, the *weakly stationarity* follows directly from definitions. An AR(p) process can always be represented in form of a MA(∞) process. To represent a MA(q) process in form of an AR(∞) process certain conditions have to be met.

A weakly stationary process' characteristics are captured in its ACF. For a linear model, like the MA(q), its **characteristic function** or **characteristic polynomial**

$$\psi(z) = 1 + \sum_{i=1}^{\infty} \psi_i z^i \tag{45}$$

will determine all equivalent structures. There are up to 2^q equivalent MA models that could have generated a specific ACF. Therefore, in attempting to identify the model that has generated the *sample ACF*, we are faced with the **identification problem**.

Under certain conditions, however, a unique solution exists: if the characteristic polynomial $\psi(z) = 1 + +\psi_1 z^1 + \psi_2 z^2 + \ldots + \psi_q z^q$ has only roots ζ with $|\zeta| \geq 1$, then the MA(q) model is *identified* and *uniquely defined*. Namely, under this condition, we can interpret the MA(q) process as being the MA(∞) representation in Wold's theorem. Large roots translate into small coefficients $\{\psi_t\}$, so the MA(q) model is *invertible* and has an AR(∞) representation:

$$Y_t = \sum_{i=1}^{\infty} \phi_i Y_{t-i} + \epsilon_t \tag{46}$$

To identify the MA(q) model by determining its lag q, the following characteristics of the ACF and PACF ca be considered:

- the ACF of an MA(q) process cuts off at lag q
- the PACF of an MA(q) process has a geometrical decline

5.4.3 Autoregressive moving average (ARMA) models

The process Y, defined by the **autoregressive moving average** model:

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=0}^{q} \psi_j \epsilon_{t-j} \tag{47}$$

is called an **ARMA(p, q)** process, provided it is stable (asymptotically stationary) and uniquely defined. The first sum represents an AR(p) process and the second sum represents a MA(q)

process The process $\{\epsilon_t\}$ is a white noise process. The AR and the MA processes have their corresponding characteristic polynomials $\phi(z)$ and $\psi(z)$, respectively. Note, that AR(p)=ARMA(p,0) and MA(q)=ARMA(0,q).

Thus, not every ARMA model will define an ARMA process. (An unstable one will not.) Next, we will state (without proof) the necessary conditions for characteristic polynomials $\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p$ and $\psi(z) = 1 + \psi_1 z + \ldots + \psi_q z^q$:

Theorem 2. *An ARMA process is uniquely defined and stable if and only if the following statements are true:*

1. *all roots ζ of the characteristic AR polynomial ϕ are $|\zeta| > 1$*
2. *all roots ζ of the characteristic MA polynomial ψ are $|\zeta| \geq 1$*
3. *the polynomials ϕ and ψ have no common roots.*

The first condition ensures stability (the AR process is *causal*) and is unaffected by the MA part. The second condition implies the uniqueness of the MA process and the third condition finally ensures the uniqueness of the entire structure. Note, that when all roots of the characteristic MA polynomial are $|\zeta| > 1$, then the MA part is invertible, and hence the complete ARMA(p, q) process has an AR(∞) representation.

Model Identification. To identify the ARMA(p, q) model, we will not be able to proceed by examining the correlogram, as in the case of AR and MA processes. Namely, for an ARMA(p, q) process both ACF and PACF geometrical decline, that starts approximately at p and q. However, this is difficult to recognise.

Researchers will frequently proceed in the following way:

1. Estimate (fit) various ARMA(p, q) models, for a wider range $0 \leq p \leq P$ and $0 \leq q \leq Q$ of lags. The number of models to be estimated is $(P + 1)(Q + 1)$.
2. Calculate ompare all $(P+1)(Q+1)$ models by one of the *information criterions*. Select the model which has the smallest value of the selected information criterion, thus, the least complex model. Note, that many `R` functions use the Akaike information criterion (consult page 56 for details).
3. Test the model specifications.
4. If the selected model turns out to be specified incorrectly, discard it and test another model with larger p or q.

5.4.4 ARIMA models

If the original process $\boldsymbol{Y}$ is not stationary, we should obtain a stationary process by computing the first order difference process:

$$\boldsymbol{X_t = \Delta Y_t = Y_t - Y_{t-1}} \tag{48}$$

or the second order differences:

$$\boldsymbol{X_t = \Delta^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}.} \tag{49}$$

If the differenced process is stationary, then we attempt to fit an ARMA model to it.

A process $\boldsymbol{Y}$ is called an **autoregressive moving average** or **ARIMA**$(\boldsymbol{p, d, q})$ process if for some $\boldsymbol{d}$ there is a differencing process $\boldsymbol{\Delta^d}$ so that the differenced process $\boldsymbol{X_t = \Delta^d Y_t}$ is an ARMA$(\boldsymbol{p, q})$ process.

The ARIMA family is a broad class of time series models, that includes the AR, MA; ARMA and ARIMA models. The key tool in identifying the correct model is the estimation of ACF.

5.4.5 SARIMA models

All ARIMA models addressed thus far were intended for use with non-seasonal data. A **seasonal ARIMA model** is an extension of ARIMA models, constructed by including additional seasonal terms to the familiar non-seasonal ARIMA models. It is denoted as ARIMA$(\boldsymbol{p, d, q})$ $(\boldsymbol{P, D, Q})_m$, where $(\boldsymbol{p, d, q})$ refers to the non-seasonal part of the model, $(\boldsymbol{P, D, Q})_m$ represents the seasonal part and $\boldsymbol{m}$ is the number of periods per season. For reference refer to (Hyndman & Athanasopoulos, 2013, sec.8.9).

Note that the `auto.arima()` function also includes seasonal ARIMA models.

5.4.6 The `auto.arima()` function

The Hyndman-Khandakar algorithm for automatic ARIMA modelling, implemented in the `auto.arima()` function is a variation of the algorithm, described in Hyndman and Khandakar (2008), which uses unit root tests for determining stationarity and minimization of the AIC and maximum likelihood estimation (MLE) to fit a selected ARIMA model.

5.5 Exponential Smoothing Methods

Exponential smoothing (sometimes denoted by ES) is a technique that can be applied to time series data, either to produce *smoothed data* for presentation, or to make *forecasts*. This family of methods has been in use since the 1950s and does not suffer from any impairment in efficiency and popularity. They originate from a numerical technique from the 17th century, which was adopted by the signal processing community in the 1940s to convert FIR filters into IIR filters.

This familiy of models predict future values as weighted averages of past observations, with recent observations weighted more heavily than older observations. In fact, the name "exponential smoothing" reflects the shape of the weighting function. The weights decrease, following an exponential function.

Despite the exponential smoothing methods having been around for decades, it was not until the article of Hyndman, Koehler, Snyder, and Grose (2002) that a framework incorporating stochastic models, likelihood calculation, prediction interval calculations, and procedures for model selection was developed.

There are 15 versions of the method: the *trend* can be excluded, additive, multiplicative or damped and the *seasonality* can be excluded, additive, additive double-seasonal, multiplicative, or multiplicative double-seasonal. There is also a *triple-seasonal* version.

In the following sections, we will describe the exponential smoothing methods from a simple exponential smoothing, to more complex versions that include trend, seasonality, and double seasonality. At the end, we will present two functions in `R`, `bats()` and `tbats()`, which are both seasonal and include all known components.

5.5.1 Simple exponential smoothing

The simplest ES method is called *simple exponential smoothing* or *single exponential smoothing* and is suitable for time series that display no visible trend or seasonal pattern. This data can be forecast by using simple methods, like the average or naïve method from Section 5.1.

When predicting with the *average method*, all available past observations are considered in the prediction and their power in predicting the future is considered equal, and they are thus given *equal* weights:

$$\hat{y}_{T+h|T} = \frac{1}{T}\sum_{t_1}^{T} y_t. \tag{50}$$

The *naïve method* only considers the last observation and gives it all weight. All forecasts are equal to the last observed value of the time series:

$$\hat{y}_{T+h|T} = y_T. \tag{51}$$

With the **simple exponential smoothing**, we want to achieve a middle way between these two concepts. We will consider more recent observations to be more important in predicting the next value, and give them larger weights than to the more distant observations. The weights will decrease *exponentially* with lag h:

$$\hat{y}_{T+1|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + \ldots + \alpha(1-\alpha)^{T-1} y_1, \tag{52}$$

where $0 \leq \alpha \leq 1$ is called the **smoothing parameter**. The one-step ahead forecast is basically a weighted average of all past observations with decreasing weights, the decrease rate being dictated by α. By definition, the forecast is always in *delay* at least by one period, in comparison to the original time series. The model's capability to adapt to the time series fluctuations depends on α. A greater α will be able to follow the series' behaviour well, while a low α will result in a more smoothed signal.

Formula (52) can be expressed in three equivalent forecasting forms. We will only mention one here: the **component form** representation includes a *forecast equation* and a *smoothing equation* for each included component. We will only consider the *level* l_t here (but the methods in Sections 5.5.2 and 5.5.3 will consider trend b_t and seasonal component s_t additionally):

$$\textit{Smoothing equation} \qquad l_t = \alpha y_t + (1-\alpha)\, l_{t-1} \tag{53}$$

$$\textit{Forecast equation} \qquad \hat{y}_{t+1|t} = l_t, \tag{54}$$

The formulation (53) is attributed to Brown (1956, p. 15). The form defines the one-step ahead forecast $\hat{y}_{t+1|t}$. The forecast for several steps ahead follows as:

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T} = l_T \tag{55}$$

The smoothing process needs to be *initialised* by selecting an initial value for the level l_0. There are several different strategies for doing this, but a common approach would be to set $l_0 = y_1$. Alternatively, one can obtain the initial value l_0 and the smoothing parameter α by an *optimization procedure*. We will be minimizing the forecast errors $e_t = y_t - y_{t|t-1}$ by minimizing the *sum of squared errors* (SSE).

Note that simple exponential smoothing with smoothing parameter α is equivalent to using

the ARIMA$(0, 1, 1)$ model, as $Y_t - Y_{t-1} = \epsilon_t - (1 - \alpha)\epsilon_{t-1}$

5.5.2 Holt's linear trend method

Holt's linear trend method is an extension of the simple exponential smoothing method to time series that display a *linear trend*. It was developed by Holt (1957); the paper was republished in (Holt, 2004) for greater accessibility.

It first estimates the Level l_t and the trend (slope) b_t at present time t, then uses them to forecast the time series in the next moment:

$$\textit{Level equation} \qquad l_t = \alpha y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \tag{56}$$

$$\textit{Trend equation} \qquad b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \tag{57}$$

$$\textit{Forecast equation} \qquad \hat{y}_{t+h|t} = l_t + hb_t \tag{58}$$

Here α and β denote the smoothing parameters for the level and trend, respectively.

The *level equation* estimates level l_t as a weighted average of observation y_t at time t and the within-sample one-step-ahead forecast for time t. The *trend equation* estimates trend b_t as a weighted average of the estimated trend $l_t - l_{t-1}$ and the previous estimate of trend b_{t-1}. The forecast is then obtained as a linear function of the last estimated level plus the last estimated trend value. For an h-step ahead forecast, we just add h-times the last estimated trend value.

It is common to initialize the smoothing process by setting $l_0 = y_1$ and $b_0 = y_2 - y_1$. The initial values l_0, b_0 and the values of the smoothing parameters can also be computed by a similar optimization procedure to the one in Section 5.5.1.

Note that there is also a variation of this method with an exponential trend, a result of multiplying, rather than adding the trend and level estimates for the forecast: $\hat{y}_{t+h|t} = l_t b_t^h$.

5.5.3 Single-seasonal Holt-Winters method

Holt's linear trend method was again extended by Holt (1957) and Winters (1960) to additionally capture *seasonality*. All three components (the underlying level, the trend, and the seasonal component) are estimated by smoothing equations.

There are essentially two versions of this method. The *additive* version is appropriate for time series for which the seasonal variations do not change in size in relation to the level of the series. The *multiplicative* version is used when the seasonal variations are proportional to the level of the series.

The Holt-Winters Additive Method. The component form for this method is expressed as:

$$\begin{aligned}
&\textit{Level equation} && l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) && (59)\\
&\textit{Trend equation} && b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} && (60)\\
&\textit{Seasonal euation} && s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} && (61)\\
&\textit{Forecast equation} && \hat{y}_{t+h|t} = l_t + hb_t + s_{t-m+h_m^+}, && (62)
\end{aligned}$$

where m is the length of the seasonal cycle. By $h_m^+ = \lfloor (h-1) \mod m \rfloor + 1$ we make sure to select the index of the same season in the last year of observations. The seasonal component s_t is estimated as a weighted average of the estimated season $y_t - l_{t-1} - b_{t-1}$ and the previous same-season estimate of season s_{t-m}. It is added to the forecast.

The smoothing parameters α, β, γ and the initial values l_0, b_0, s_0 need to be estimated by optimization.

The Holt-Winters Multiplicative Method. The component form for this method is expressed as:

$$\begin{aligned}
&\textit{Level equation} && l_t = \alpha\frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) && (63)\\
&\textit{Trend equation} && b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} && (64)\\
&\textit{Seasonal equation} && s_t = \gamma\frac{y_t}{(l_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m} && (65)\\
&\textit{Forecast equation} && \hat{y}_{t+h|t} = (l_t + hb_t)s_{t-m+h_m^+}. && (66)
\end{aligned}$$

Here, the seasonal component is just multiplied by the previous $l_t + hb_t$ term.

5.5.4 Double-seasonal Holt-Winters method

The double-seasonal Holt-Winters method (hereinafter DSHW) is an adaptation of the Holt-Winters method that is able to cope with two cycles (a *daily* and a *weekly* one), present in the load profile, rather than just the one in the usual form of the Holt-Winters method. It was developed by (Taylor, 2003) who was using it for short-term load forecasting. Adding a second seasonal component results in a model that is more true to the real situation and leads to better forecast accuracy.

The DSHW leaves the equations for level (59) and trend (60) as is, while the seasonal estimation is replaced by two equations, one for daily season d_t and one for the weekly one w_t, and the final forecast changes accordingly:

DSHW Additive Method

$$\begin{aligned}
\textit{Level equation} \quad & l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1}) && (67)\\
\textit{Trend equation} \quad & b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} && (68)\\
\textit{Seasonality 1} \quad & d_t = \delta(y_t - l_{t-1} - w_{t-s_2}) + (1-\delta)d_{t-s_1} && (69)\\
\textit{Seasonality 2} \quad & w_t = \omega(y_t - l_{t-1} - d_{t-s_1}) + (1-\omega)w_{t-s_2} && (70)\\
\textit{Forecast} \quad & \hat{y}_{t+h|t} = l_t + hb_t + d_{t-s_1+h} + w_{t-s_2+h}, && (71)
\end{aligned}$$

The smoothing parameters are $\alpha, \beta, \delta, \omega$, and the initial values that we need to establish are l_0, b_0, d_0 and w_0.

Later, Taylor also developed a *triple-seasonal* version of Holt-Winters method to add the *yearly* seasonality of the load time series.

5.5.5 State space models

The **state space models** are stochastic models that underlie the exponential smoothing methods described in Sections 5.5.1–5.5.4. Those methods generate *point forecasts* while the present models will not only generate the same point forecasts but also prediction intervals. Every model consists of a *measurement equation* that models the historical data and some *transition equations* that describe the transitions of the unobserved components (or *states* level, trend and seasonal) through time; thus the name "state space".

For each known method we can generate two models: one with additive errors and one with multiplicative errors. So from the 15 mentioned methods we arrive at 30 stochastic models. Some *non-linear* state space models suffer from weaknesses, such as instability.

To overcome this and other shortcomings of previously known state space models, De Livera, Hyndman, and Snyder (2011) developed a state space modeling framework that will be discussed in the next section. Two very general models, BATS and TBATS were developed in this framework.

5.5.6 BATS model

The BATS model was modeled in the new **innovations state space modeling framework**, described in detail in (De Livera et al., 2011). The framework incorporates Box-Cox transformations, Fourier representations with time-varying coefficients and ARMA error correction. The term "innovations" refers to innovations as the single source of error. It was constructed to handle time series which exhibit *complex seasonal patterns*. The load time series is an example of such a series. It has the following key features:

- **multiple seasonality**: it contains daily, weekly and yearly seasonal periods,
- **high-frequency seasonality**: the weekly seasonal period has a frequency (length) of 168, which is considered high frequency,
- **non-integer seasonality**: taking the leap years into account, the yearly seasonal period will have an average seasonal frequency of 365.25 days or 52.18 weeks,
- **dual-calendar effects**: when predicting load in e.g. some Islamic countries, the holiday effects of two calendars need to be considered: the Hijri and the Gregorian.

While some time series models could handle some of the above-listed complexities, the mentioned framework is the first concept that covers all of them. By including a new estimation procedure based on maximum likelihood estimation, it is also less compute-intensive and therefore faster.

BATS is an acronym for the key features of the model: Box-Cox transform, ARMA errors, Trend, and Seasonal components. The model is the most obvious generalization (extension) of the traditional seasonal innovations models to allow for multiple seasonal periods. The system of equations reads:

$$\textit{Transformed observations} \qquad \boldsymbol{y}_t^{(\omega)} = \begin{cases} \frac{y_t^{\omega}-1}{\omega} & \omega = 0 \\ \log \boldsymbol{y}_t & \omega \neq 0 \end{cases} \tag{72}$$

$$\textit{Level equation} \qquad \boldsymbol{l}_t = \boldsymbol{l}_{t-1} + \phi \boldsymbol{b}_{t-1} + \alpha \boldsymbol{d}_t \tag{73}$$

$$\textit{Trend equation} \qquad \boldsymbol{b}_t = (1-\phi)\boldsymbol{b} + \phi \boldsymbol{b}_{t-1} + \beta \boldsymbol{d}_t \tag{74}$$

$$\textit{Seasonalities} \qquad \boldsymbol{s}_t^{(i)} = \boldsymbol{s}_{t-m_i}^{(i)} + \gamma_i \boldsymbol{d}_t \text{; for } \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{S} \tag{75}$$

$$\textit{Error} \qquad \boldsymbol{d}_t = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{p}} \boldsymbol{\varphi}_i \boldsymbol{d}_{t-i} + \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{q}} \boldsymbol{\theta}_i \boldsymbol{\epsilon}_{t-i} + \boldsymbol{\epsilon}_t \tag{76}$$

$$\textit{Forecast} \qquad \boldsymbol{y}_t^{(\omega)} = \boldsymbol{l}_{t-1} + \phi \boldsymbol{b}_{t-1} + \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{T}} \boldsymbol{s}_{t-m_i}^{(i)} - \boldsymbol{d}_t^{(i)} \tag{77}$$

The differences from a traditional model are roughly:

1. the observations $\boldsymbol{y}_t$ are first transformed by Box-Cox transformations (Box & Cox, 1964) to obtain $\boldsymbol{y}_t^{(\omega)}$.
2. an ARMA($\boldsymbol{p}, \boldsymbol{q}$) process $\boldsymbol{d}_t$ is added as an additional error term to a Gaussian white noise process $\boldsymbol{\epsilon}_t$.
3. there are two types of trends included: a *long-term trend* or *global trend* $\boldsymbol{b}$ and a *short-term trend* or *local trend* $\boldsymbol{b}_t$. $\boldsymbol{b}$ is a replacement for a damped trend in traditional models and $\boldsymbol{b}_t$ converges to $\boldsymbol{b}$.
4. the seasonal components are denoted by $\boldsymbol{s}_t^{(i)}$.

The parameters of the BATS($\omega, \phi, p, q, m_1, m_2, \ldots, m_J$) model are: the *Box-Cox* parameter ω, the *damping* parameter ϕ, the ARMA parameters p and q, and a list of the seasonal periods $\{m_1, m_2, \ldots, m_J\}$. For example, BATS$(1, 1, 0, 0, m_1)$ represents the underlying model for the well-known Holt-Winters additive single-seasonal method. The model for the additive double-seasonal Holt-Winters method is given by BATS$(1, 1, 0, 0, m_1, m_2)$.

We will use the **`bats()` function**, which is contained in the `forecast` package in `R`. It was developed by Professor Rob Hyndman. (Refer to package documentation at (Hyndman, 2016)). The function fits a BATS model applied to $\boldsymbol{Y}$, as described in (De Livera et al., 2011). Parallel processing is used by default to speed up the computations. The `bats()` function selects the best model out of a group of BATS *double-seasonal* models, based on lowest AIC (page 56).

5.5.7 TBATS model

The **TBATS model** is a Trigonometric exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components. Parallel processing is used by default to accelerate the computations. It is a further extension of the BATS model, allowing for a *trigonometric* representation of seasonal components based on Fourier series:

$$\textit{Seasonality} \qquad s_t^{(i)} = \sum_{j=1}^{k_j} s_{j,t}^{(i)} \tag{78}$$

$$\textit{Level of seasonality} \qquad s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t \tag{79}$$

$$\textit{Growth of seasonality} \qquad s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t, \tag{80}$$

where $\gamma_1^{(i)}$ and $\gamma_2^{(i)}$ are smoothing parameters, $\lambda_j^{(i)} = 2\pi j / m_i$, and m_i is seasonal period of the i-th seasonal component, $s_t^{(i)}$ is the *level* of the seasonal component and $s_t^{*(i)}$ is its *growth*. Note, that by setting the smoothing parameters to zero, we get a *deterministic* representation.

The **`tbats()` function** is also part of the `forecast` package and is based on the work by De Livera et al. (2011). It is an improvement over the `bats()` function and can be used for *seasonal* or *non-seasonal* time series. It works in several steps and attempts to find the TBATS model with the lowest AIC from a group of generated models. Parallel processing is used when possible. For a step-by-step description of `tbats()` function the reader is referred to Appendix B.

The fitted model is designated as TBATS($\omega, \phi, p, q, \{m_1, k_1\}, \{m_2, k_2\}, \ldots, \{m_J, k_J\}$),

where ω is the Box-Cox parameter and ϕ is the damping parameter. The error is modelled as an ARMA(p, q) process and $m_1, \ldots, m_J$ lists the seasonal periods used in the model and $k_1, \ldots, k_J$ are the corresponding initial seasonal values (the Fourier terms) used for each seasonality.

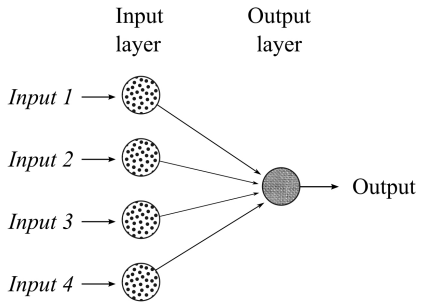
5.6 Autoregressive Artificial Neural Network Models

Artificial Neural Networks (ANN) are forecasting methods that are based on simple mathematical models of the brain. They allow complex nonlinear relationships between the "response variable" and its "predictors". The model is fitted by a "learning algorithm". These models function as *black boxes* and detect relationships without any prior knowledge.

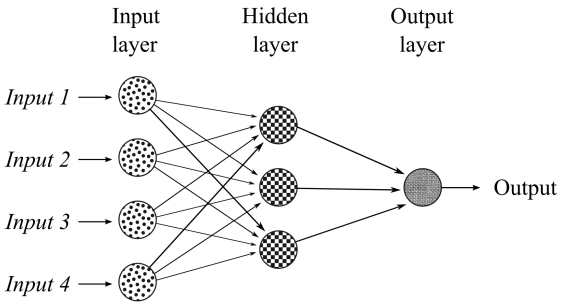
The simplest networks contain no hidden layers and are equivalent to *linear regression* (Figure 23a). Once we add an intermediate layer with hidden neurons, the neural network becomes non-linear. There can be many hidden layers. A simple example of a non-linear neural network is shown in Figure 23b.

Figure 23. An Artificial Neural Network Model, Without and with Regression

(a) A simple neural network equivalent to a linear regression



(b) A neural network with four imputs and one hidden layer with three hidden neurons



Source: R. J. Hyndman & G. Athanasopoulos, Forecasting: Principles and Practice, 2013, sec. 9.3.

The `R` function `nnetar()` is a new experimental function that fits a neural network with lagged values of the time series as inputs. It is a feed-forward neural network with a single hidden layer and lagged inputs for forecasting univariate time series. The notation NNAR(p, k) indicates there are p lagged inputs and k nodes in the hidden layer. For example, a NNAR($9, 5$) model is a neural network with the last nine observations $(y_{t-1}, y_{t-2}, \ldots, y_{t-9})$ used as inputs, an output y_t and five neurons in the hidden layer. A NNAR($p, 0$) is equivalent to the ARIMA($p, 0, 0$) model (Hyndman & Athanasopoulos, 2013, sec. 9.3).

ANNs might be able to capture non-linear relations better than traditional models. In fact, they have been very successfully used for STLF on a large scale. However, results depend on the neural network design, that is number of the layers, size of the hidden layer, number of inputs in the input layer, and so on. They can be used as pure time series models or including other predictive variables, such as weather, day of week, time of day etc. As we are dealing solely with pure time series models, we will also test ANNs in their autoregressive role, predicting load values of the next 24 hours from sufficient historical data.

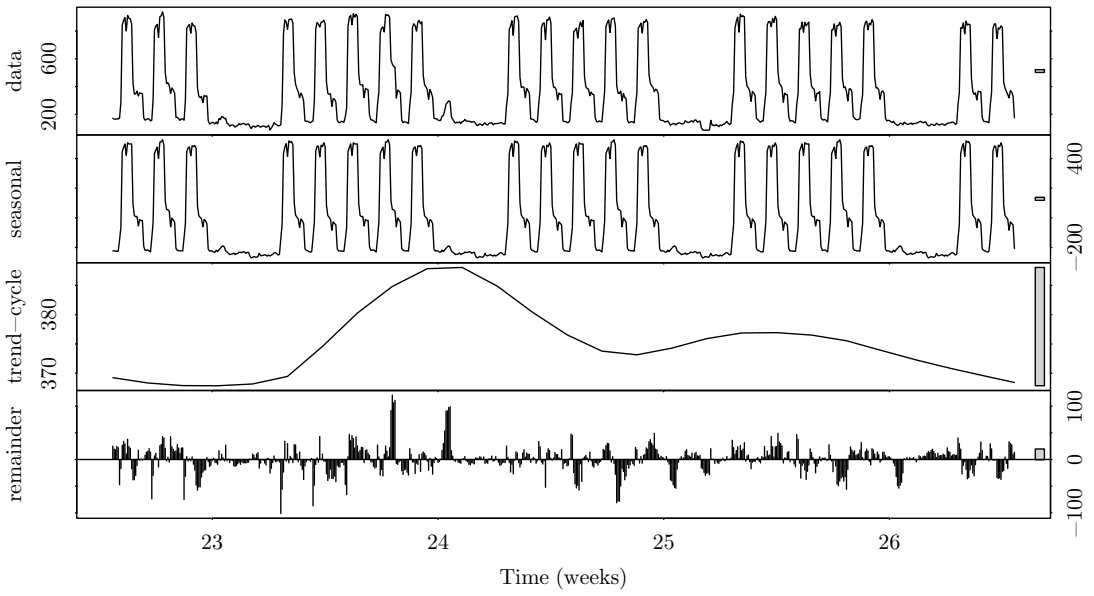
6 RESULTS – INDUSTRIAL CONSUMERS

We will start by testing a few forecasting methods on historical data for industrial consumer *Ind1* and compare forecast accuracy. The selected period ranges from June 1st, 2011 to June 28th, 2011 (*training set*). Data were aggregated to 1-h measurements. The goal was to forecast the electricity consumption for the next 1, 2, 3 and 14 days (*test set*), in 1-h values. A couple of other residential sources were also analyzed and forecast and, since the results were similar, we will only be displaying Ind1 results here.

6.1 Forecasting with STL Decomposition

Experiments. The 4-week historical load time series of *Ind1* consumer was used and STL decomposed, as we can observe in Figure 24.

Figure 24. STL Decomposition of Ind1 Consumer Time Series for June 2011, 1-h Sampling

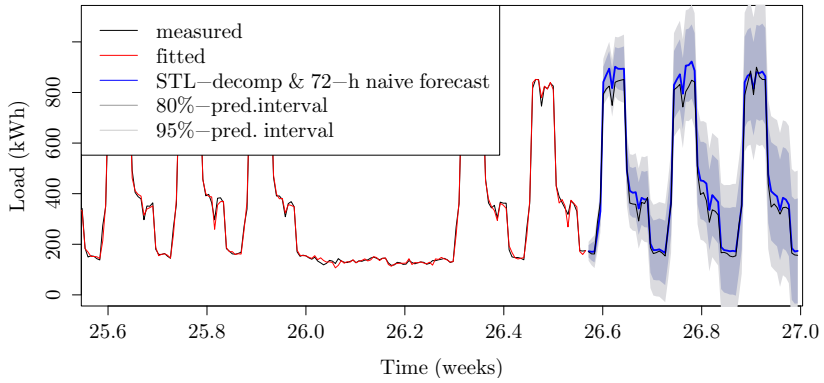


The deseasonalized series was then forecast by using the seasonal naïve method. 1-day, 2-

day, 3-day and 14-day forecasts were visualized and forecast accuracy was calculated by using MAPE. Figure 25 displays the last 7 days of historical data and a 3-day forecast with 80% and 95% prediction intervals.

Results.A MAPE of 4.3% was obtained on the training set and 7.3% on the test set.

Figure 25. 3-Day Naïve Forecast of STL-Decomposed Ind1 Consumer Time Series



Note. Historical Data: 28 Days in June 2011. Sampling: 1 h.

Commentary. The STL decomposition should capture the load signal's nature well. Furthermore, the seasonal naïve forecasting method seems to be performing fairly well. We will consider the achieved MAPE of 7.3% on the test set as a benchmark, against which we will measure more advanced forecasting methods.

6.2 Single-Seasonal Holt-Winters Method

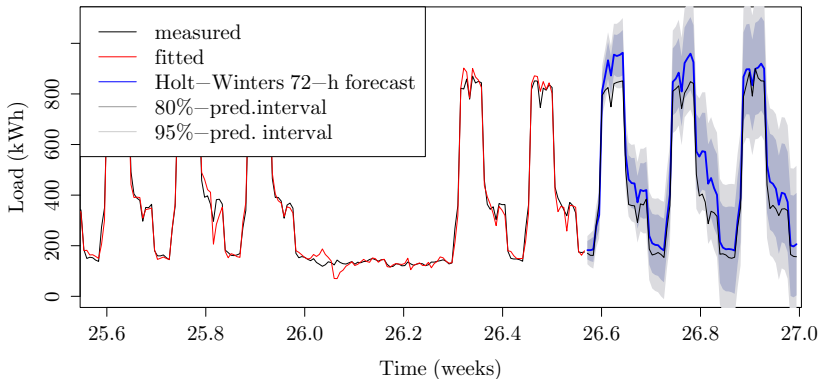
Experiments. The same training and test sets as in section 6.1 were used for industrial consumer Ind1. The future load was forecast by applying the single-seasonal Holt-Winters method. It is an ES method, suitable for time series that display a linear trend and seasonality. Results for a 3-day forecast are visualised in Figure 26.

Forecast accuracy was calculated. Residuals were inspected to assess their eventual autocorrelation and autocovariance. Box-Ljung test for autocorrelations was performed. Residuals were also plotted and their statistics were estimated to check randomness.

Results. MAPE on the training set was not satisfactory: the values for a 1-day, 2-day, 3-day and 14-day forecast were 16%, 19%, 18% and 18%, respectively. Furthermore, the autocorrelation function (ACF) pointed to the autocorrelation of residuals, and the Box-Ljung test confirmed that there are correlations between $\boldsymbol{Y_t}$ and lags $\boldsymbol{Y_{t-1}}$ to $\boldsymbol{Y_{t-20}}$.

Conclusion. The forecast accuracy of 16–19% is insufficient for our task and, due to the autocorrelation of residuals, the Holt-Winters *single-seasonal* method is not the most suitable one for this source, as it can not explain all variation in the data.

Figure 26. 3-Day Single-Seasonal Holt-Winters Forecast of Ind1 Consumer Time Series



Note. Historical Data: 28 Days in June 2011. Sampling: 1 h.

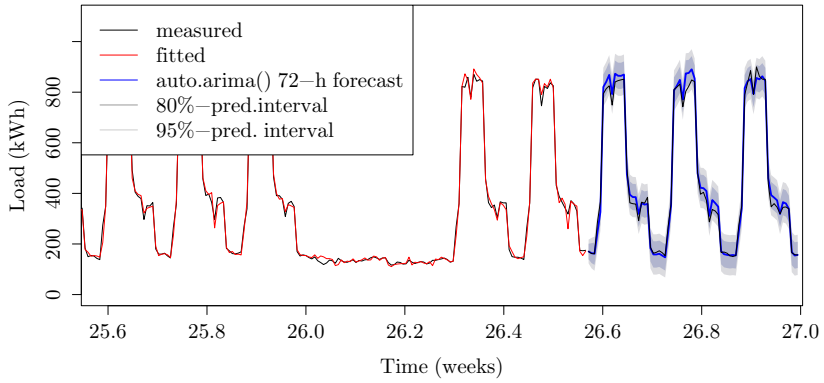
6.3 ARIMA Models

Experiments. We tested the performance of the ARIMA model family for the same source, training and test set as before. The best model was selected by using the `auto.arima()` function, which also fits the model.

Results. The automatic function `auto.arima()` selected the seasonal autoregressive model ARIMA$(1,0,0)(1,0,1)_{168}$, with a week as the lenght of its main seasonal cycle. However, it needed around 5 minutes to automatically determine the most accurate ARIMA model and fit it to the data. MAPE reached 5.2% on the training set and was even better on the test set: it reached 4.5% for the 1-day and 3-day forecast and 5% for the 2-day forecast. The ACF and the Box-Ljung test did not detect any autocorrelation of residuals, so the model is valid. The results of the 3-day forecast are plotted in Figure 27.

Commentary. `auto.arima()` shows a solid improvement in forecast accuracy over the Holt-Winters method and explains the data well. However, baseline calculation time in a VPP environment is crucial. The selection and fitting process is an *optimization problem* and it seems, that the function which fits several models before selecting the best one, takes too much time. Therefore, we will look for a method that is less time consuming.

Figure 27. 3-Day `auto.arima()` Function Forecast of Ind1 Consumer Time Series



Note. Historical Data: 28 Days in June 2011. Sampling: 1 h.

6.4 Discussion – Industrial Consumers

Applying a few very different methods to a single industrial source we showed that forecasting load profiles of individual industrial consumers is feasible. The accuracy achieved varied across the methods tested. Forecasting by STL decomposition and seasonal naïve forecasting method achieved a solid accuracy and can serve as a benchmark to be surpassed. The single-seasonal Holt-Winters method was found to be less suitable, as it is not able to explain all the variations contained in the double-seasonal load signal, resulting in a somewhat poor performance. The `auto.arima()` function achieved the best accuracy but was discarded for use in the VPP environment due to exceedingly long calculation times.

We will now continue by testing the forecasting methods on households and groups of households as they will be harder to predict. If we manage to find a method with good accuracy based on residential data, we will also be able to use it for forecasting industrial consumers.

7 RESULTS – RESIDENTIAL CONSUMERS

We will now extend the pool of methods and aim for methods that give greater consideration to special features of load time series. As fitting different ARIMA models can be too time consuming, we will focus on *Exponential smoothing methods*. All methods will be judged based on MAPE accuracy measure and computational complexity.

Experiments. The historical period ranges from May 9th, 2013 to June 5th, 2013, or 4 weeks. The goal is to forecast the next 24 hourly values of June 6th, 2013. All methods will be tested on several different household data sources (Dom1, Dom5, Dom12). We will also investigate the influence of aggregating measurements to 1-h on forecast accuracy. Time complexity will

be a special concern.

We will plot the results and calculate the forecast accuracy. Residuals will be plotted and their statistics will be calculated to confirm randomness. They will also be checked for autocorrelation and autocovariance, for which the Box-Ljung test will aid in detecting autocorrelations at certain lags.

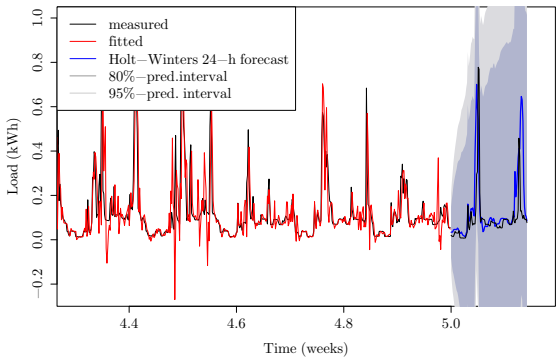
7.1 Single-Seasonal Holt-Winters Method

During the forecasting with this *single-seasonal* method, we will additionally observe the influence of historical data sampling (15 min versus 1 h) on forecast accuracy. All signals that were sampled at 15 min were multiplied by 4 to obtain the same order of magnitude and be visually comparable to signals sampled at 1 h.

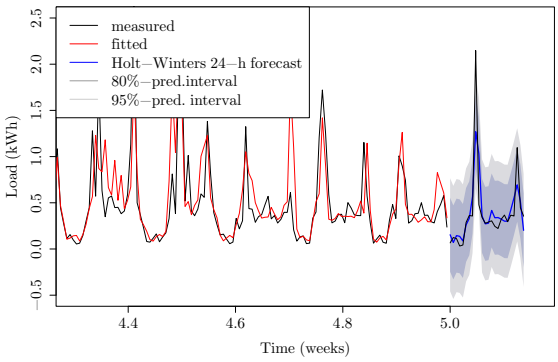
Results. Results for *Dom1* show that MAPE improves by a factor of 2, once we pass from a 15-min sampling of the input signal to the aggregated signal, sampled at 1 h. As we switched to the more aggregated signal MAPE, measured on the test set fell from 81.6% to 45%. Furthermore, the prediction interval shrunk visibly for the aggregated signal, which is also visually smoother, less volatile and therefore easier to predict. Note, that the plot in Figure 28a has been stretched along the y-axes, for better comparability of the signals.

Figure 28. Dom1 Consumer Time Series and 24-h Single-Seasonal Holt-Winters Forecast

(a) Sampling: 15 min



(b) Sampling: 1 h

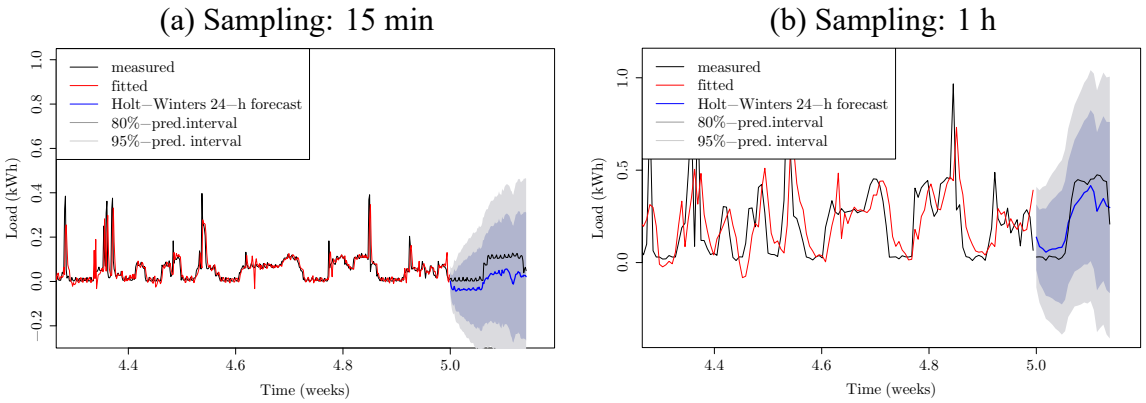


Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

We were focusing on MAPE as our primary selected accuracy measure. Figure 28b seems to indicate a good prediction accuracy, but the value of MAPE, measured on the test set, was 45%. This means that in the 24 forecasted points the forecast value differed from the actual value by 45% on average. In the worst case, the error was most probably even bigger.

The results for *Dom5* are even more extreme: MAPE reached 495% on 15-min sampled signals and 103% on 1-h sampled signals. The forecasting results can be observed in Figure 29.

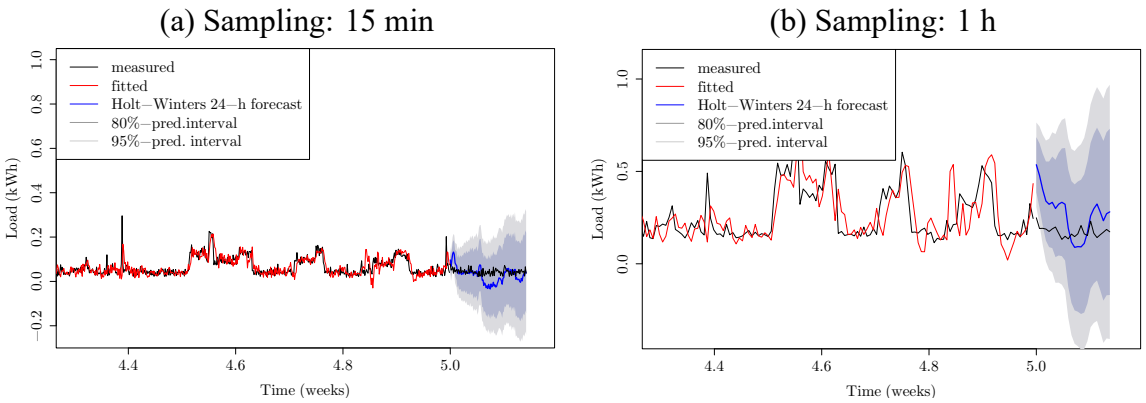
Figure 29. Dom5 Consumer Time Series and 24-h Single-Seasonal Holt-Winters Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

In the *Dom12* case, MAPE is of the same magnitude for both sampling values. The aggregated 1-h sampling gives slightly worse results (MAPE = 71%) compared to the 15-min sampling (MAPE = 68%).

Figure 30. Dom12 Consumer Time Series and 24-h Single-Seasonal Holt-Winters Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

All forecast accuracy measures are summarized in Table 1. ME for the 15-min Dom1 source is negative because the method could not adequately capture the two spikes.

Table 1. Accuracy of Single-Seasonal Holt-Winters Method for Selected Households and Two Different Sampling Periods: 15 min, 1 h

Dataset			ME	RMSE	MAE	MPE in %	MAPE in %	MASE	ACF1
Dom1	15 min*	Training**	-3.2156	0.0896	0.0422	-0.10	38.58	0.5393	NA
		Test***	-4.6170	0.1440	0.0700	-71.77	**81.64**	1.066	0.6968
	1 h	Training**	0.0088	0.3281	0.1770	-13.48	41.25	0.6207	NA
		Test***	0.0188	0.2510	0.1347	-24.55	**44.98**	0.6191	-0.4668
Dom5	15 min*	Training**	-0.0004	0.1380	0.0604	31.07	167.19	0.5439	NA
		Test***	0.2396	0.2598	0.2396	495.42	**495.42**	1.6891	0.7517
	1 h	Training**	-0.0010	0.1388	0.0798	24.18	150.36	0.7464	NA
		Test***	0.0209	0.0889	0.0783	-81.15	**103.65**	0.5936	0.653
Dom12	15 min*	Training**	-0.0003	0.1215	0.0732	5.30	35.45	0.6659	NA
		Test***	0.0569	0.1453	0.1119	32.91	**68.42**	0.9037	0.8167
	1 h	Training**	-0.0006	0.1163	0.0750	1.65	33.57	0.7403	NA
		Test***	-0.0965	0.1441	0.1247	-55.44	**71.83**	1.1785	0.7999

Note. * 4 ×15-min sampling values, ** Training set span over 4 weeks, *** Test set span over 24 h.

Commentary. Running single-seasonal Holt-Winters forecasts on different households for different sampling periods (15 min, 1 h) has provided the following insight: In most cases this method gives considerably more accurate predictions when working with an aggregated time series, sampled at 1 h. However, in terms of accuracy, we begin to see, what was anticipated in Section 4.5.2: the irregular household signals are hard to predict and the accuracy itself is directly related to data's volatility. Even the best observed MAPE of 45% is still insufficient for the precision needed in a VPP environment.

Since the Holt-Winters method only considers one seasonality (the longer one of one week), we hope to obtain better accuracy with methods that respect the *double-seasonal nature of load signals*. We will only work with *aggregated signals* from this point on.

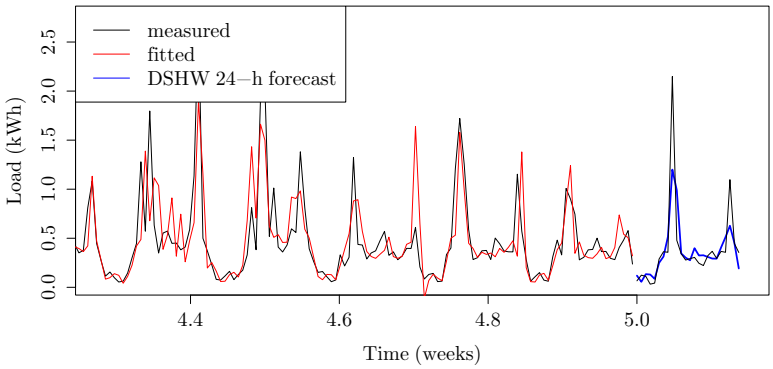
7.2 Double-Seasonal Holt-Winters Method

Experiments. The `dshw()` function was used for forecasting the Dom1, Dom5 and Dom12 signals, sampled at 1 h. Analysis steps are listed at the beginning of Chapter 7. Note, that there are no prediction intervals implemented in this function.

Results. For the *Dom1* signal, MAPE reached 41% on the training set and 45% on the test set. It seems that the DSHW smoothes spikes too much. This is due to the exponential smoothing nature of the method. Plotting the ACF function showed no autocorrelation. The

p-value of the Box-Ljung test statistic was 0.052, proving that the sample auto-correlations of the residuals are not too big. Plotting the residuals showed that the residuals are a random process. The model is valid, thou not very accuarate, as we can observe in Figure 31.

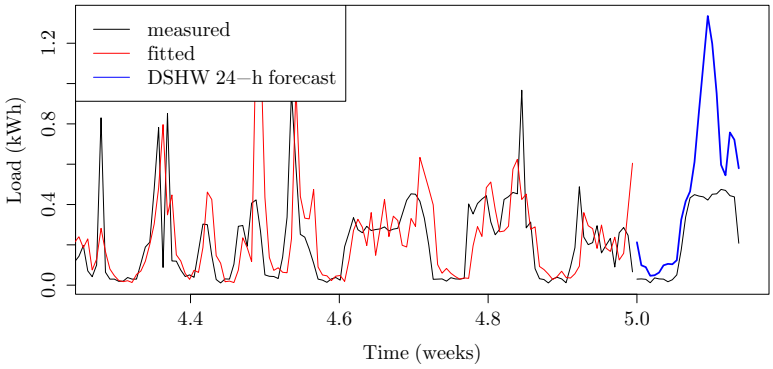
Figure 31. Dom1 Consumer Time Series and 24-h Double-Seasonal Holt-Winters Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

For the *Dom5* source, MAPE reached 87% on the training set and 161% on the test set; the results are thus useless. Furthermore, the ACF function displayed autoregression of residuals at lags 1 and 5, the p-value of 0.0004 for the Box-Ljung statistic proved that the residuals are not "overall" random but are autocorrelated. In other words, the true innovations are not independent and the residuals still contain a periodic component. Hence, the model needs improvement to extract this periodic component. This is also visually verified by plotting the residuals. The model can be observed in Figure 32.

Figure 32. Dom5 Consumer Time Series and 24-h Double-Seasonal Holt-Winters Forecast

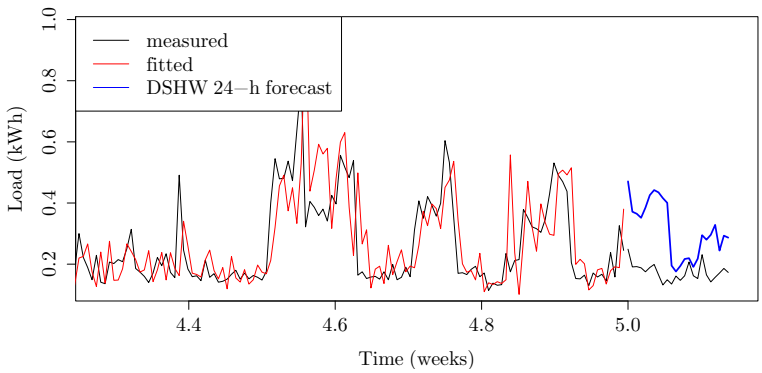


Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

For the *Dom12* signal, the MAPE on the training set reached 31%, but amonuted to 81% on

the test set, which testifies of poor prediction power of this model. The residuals were even more correlated than in the case of Dom5, exposing autocorrelations at lags 1,2,14,15 and 20. The extremely low p-value of $\mathbf{9.68 * 10^{-7}}$ for the Box-Ljung test statistic clearly confirmed this. Thus, the model had to be dismissed. See Figure 33 for results.

Figure 33. Dom12 Consumer Time Series and 24-h Double-Seasonal Holt-Winters Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

Commentary. The DSHW assumes a double-seasonal nature of the signal, but the signals of individual households are very irregular and in fact do not show regular seasonal patterns. Clearly, the DSHW *can not deal with their irregular nature* and *can not predict the individual household signals satisfactorily*. (The MAPE, observed on the test set was between 45% and 161% and renders the results useless and two out of three models failed to explain all the variations, contained in the data).

7.3 BATS Model

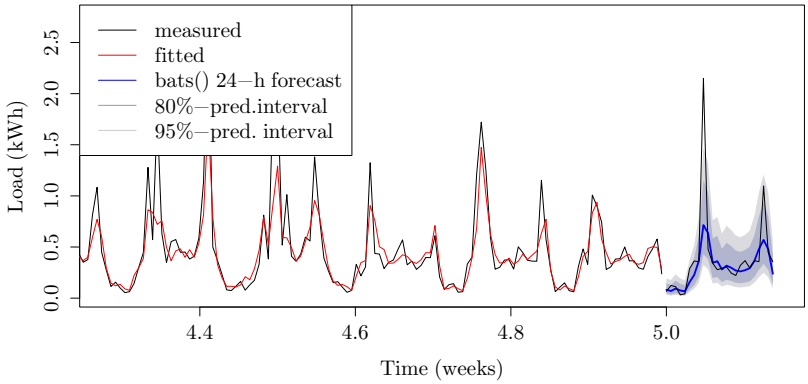
Experiments. The *state space* BATS model was used for forecasting the Dom1, Dom5 and Dom12 load signals, following the described analytic steps. The applied `bats()` function selects the best model out of a group of BATS models, based on the lowest AIC.

Results and Commentary. When forecasting the *Dom1* signal the `bats()` function selected BATS$(\mathbf{0.128}, \{\mathbf{1}, \mathbf{1}\}, -, \{\mathbf{24}, \mathbf{168}\})$ model. The first parameter is the Box-Cox parameter, followed by the two parameters for ARMA errors; no damping parameter was used and we are looking at a double-seasonal model, that successfully uncovered the two seasonal periods, the daily one (24) and the weekly one (168).

MAPE reached 28.7% on the training set and 29.6% on the test set. The ACF of residuals is showing very little autocorrelation, at the more distant lags 15 and 19. The p-value of 0.18 for

the Box-Ljung statistic confirms that the innovations are indeed independent and the model explains the variation well, which can be observed in Figure 34. The residual plot seems normally distributed with zero mean. Forecasting errors are standard normally distributed. The AIC value was 1990.173.

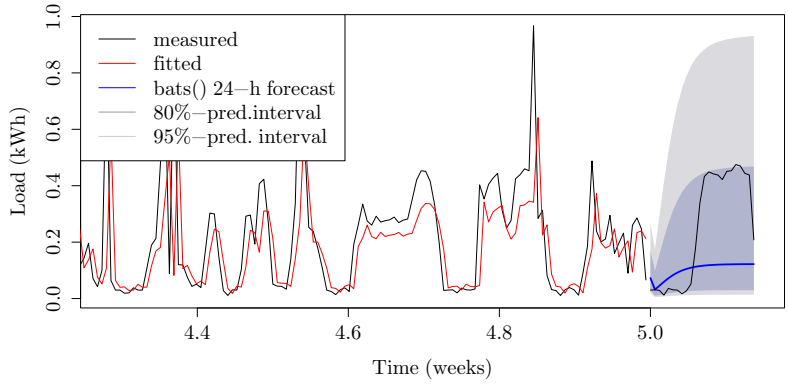
Figure 34. Dom1 Consumer Time Series and 24-h `bats()` Function Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

Forecasting *Dom5* selected BATS$(\mathbf{0.036}, \{\mathbf{2}, \mathbf{0}\}$, -, -) as the best model. It does not include any damping of the trend and is a non-seasonal model with AR errors. MAPE reached 59% on training set and 120% on test set. The residual ACF displayed a strong autocorrelation at lags 3,5,7,8,13 and 14. This was confirmed by the very low p-value of $\mathbf{7.255 * 10^{-5}}$ of the Box-Ljung statistic. The residual plot seems normally distributed with mean close to zero. Forecasting errors are standard normally distributed. The AIC value was 556.86. The model is not able to explain the variation of the data well, which can clearly be seen in Figure 35.

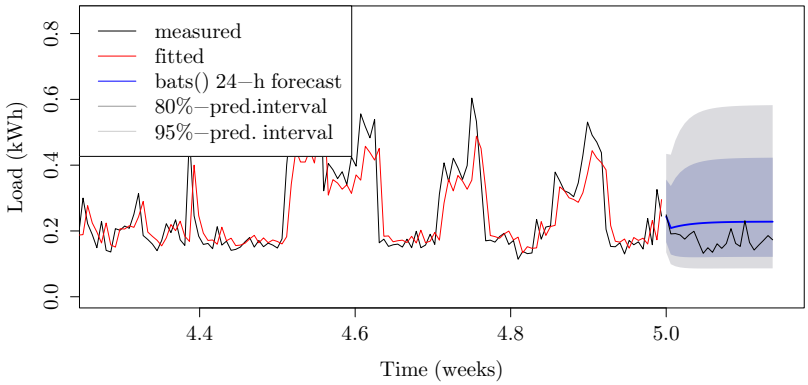
Figure 35. Dom5 Consumer Time Series and 24-h `bats()` Function Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

BATS$(0.041, \{0, 0\}, 0.8, -)$ was selected as the best model for forecasting the *Dom12* signal. It contains Box-Cox transformations and a damped trend, but no seasonality. MAPE amounted to 23% and 32% on the training and test set, respectively. The residual ACF showed negligible autocorrelation at lag 15, confirmed by the high p-value of 0.158 of the Box-Ljung statistic. The residual plot seems normally distributed with the mean close to zero. Forecasting errors are standard normally distributed. The AIC value was 763.28.

Figure 36. Dom12 Consumer Time Series and 24-h `bats()` Function Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

Commentary. Figures show that the BATS model can not "follow" the shape of the "erratic" load curve of individual households and is therefore unsuccessful in predicting their future. In two cases it failed to detect seasonality and in one case strong autocorrelation was proven, confirming that that *model should be further improved* by extracting another periodic component from the residuals.

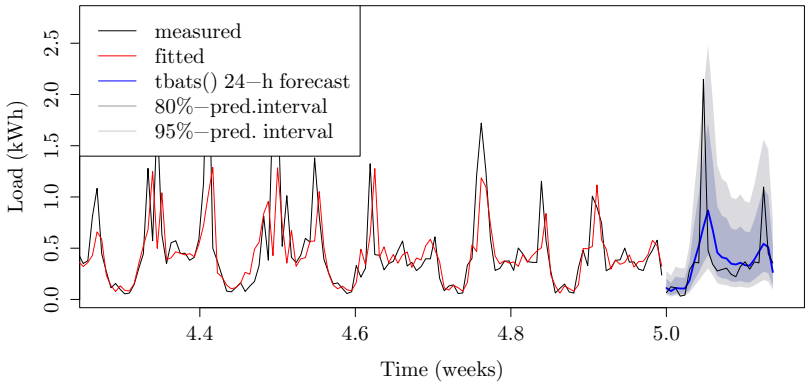
7.4 TBATS Model

Experiments. We hope that the more sophisticated TBATS *state space model* will be able to better explain the load curve of individual households. We will test it on the same sources. The `tbats()` function selects the best model out of a group of TBATS seasonal and BATS non-seasonal models, based on lowest AIC.

Results and Commentary. *Dom1*: Figure 37 shows that for this signal the `tbats()` is slightly less successful in predicting the future volatility than the `bats()` function, resulting in poorer accuracy. In fact MAPE reached 43% on the training set and 45% on the test set. The ACF plot did not indicate any autocorrelation of residuals but the Box-Ljung test showed some lag-dependence. The residual plot seems normally distributed with mean around zero. Forecasting errors are standard normally distributed. The AIC value was 2047.291. The

model selected for having the lowest AIC is TBATS$(0, \{0, 0\}, -, \{24, 6\}, \{168, 6\})$, which captured both the daily and weekly seasonalities.

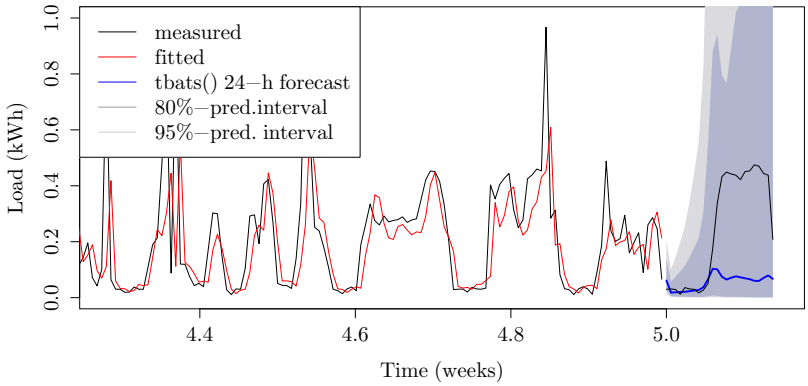
Figure 37. Dom1 Consumer Time Series and 24-h `tbats()` Function Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

Forecasting the *Dom5* signal selected the TBATS$(0.083, \{0, 0\}, 0.8, \{24, 6\}, \{168, 4\})$ model, with mediocre accuracy: MAPE reached 56% on the training set and 65% on the test set. The prediction intervals in Figure 38 are huge and the forecast could not predict the shape of the curve well, which is confirmed by the p-value of 0.0017. The ACF function also showed autocorrelation of the residuals, at lags 5, 8, 13 and 14.

Figure 38. Dom5 Consumer Time Series and 24-h `tbats()` Function Forecast

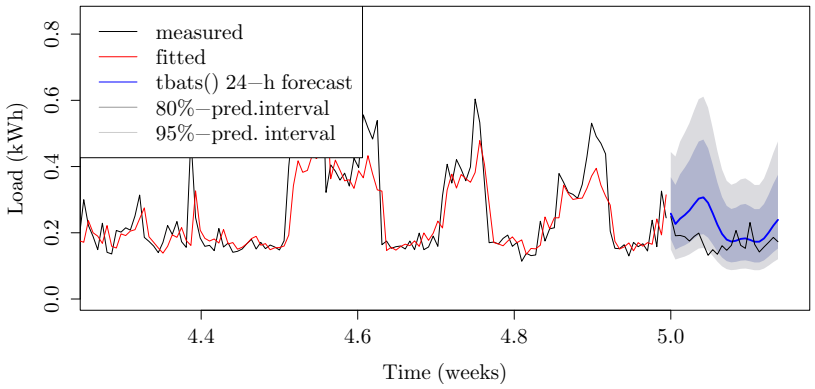


Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

Forecasting *Dom12* also resulted in the selection of a proper TBATS$(0, \{1, 1\}, -, \{24, 5\}, \ldots \{168, 6\})$ model. As Figure 39 shows, with its ARMA errors and double seasonality it managed to produce some prediction, with MAPE reaching 21% and 34% on the training

an test sets, respectively. The model is also valid, with a p-value of 0.47 of the Box-Ljung statistic. It seems that this is the best we will be able to observe.

Figure 39. Dom12 Consumer Time Series and 24-h `tbats()` Function Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

Commentary. Our hopes were not fulfilled. The performance of the `tbats()` function in predicting the load of individual households depends on the specific dataset. Nevertheless, it will produce a mediocre prediction at best and in most cases it will be unable to explain the complete variation in the data. It seems that the very irregular load behaviour of single households can not be captured even by the most sophisticated models.

7.5 Time Complexity

The time complexities of the tested methods are gathered in Table 2:

Table 2. Time Complexity of Exponential Smoothing Methods, Measured on Individual Household Data

Method/Function	**Time complexity (in s)**		
	Dom1	**Dom5**	**Dom12**
Single-Seasonal Holt-Winters	0.027	0.028	0.027
Double-Seasonal Holt-Winters	9.340	10.670	10.060
`bats()`	**160.000**	**96.000**	**91.000**
`tbats()`	15.940	17.570	17.430

The double-seasonal Holt-Winters method is slower than the Holt-Winters method because it considers two seasonalities instead of just one. The more sophisticated `tbats()` function runs considerably faster than the `bats()` function. This is probably because the `tbats()` function is more recent and runs in parallel, by default.

7.6 Discussion – Residential Consumers

The methods that consider *double seasonality* in the load curve perform better than the ones that only consider a *single* seasonality (The double-seasonal Holt-Winters method performs better than the single-seasonal one and `tbats()` function performs better than `bats()` function.) `tbats()` would be the best choice among the tested methods, however it can not cope with the *unpredictability of human behavior* that is reflected in an individual household's load. For the purpose of a VPP and in general we hold that these *can not be predicted with any good reliability*, nor are the functions able to predict the shape of the load curve correctly. We hope to be able to predict at least groups of households and find a solution there.

8 RESULTS – GROUPS OF RESIDENTIAL CONSUMERS

As it is now clear that the load behaviour of individual households is unpredictable, we will attempt to forecast groups of residential users.

Experiments. We will use a vast pool of methods, from simple *average* ones, to the more sophisticated methods of *exponential smoothing*. We will look for the best method based on MAPE but also compare performance between the simpler and the more complex methods. We will use SUM69 historical data as input, aggregated into hourly measurements. The training data period ranged from *May 9th, 2013 to June 5th, 2013* and span four weeks. We will predict 24 hourly values of *June 6th, 2013*.

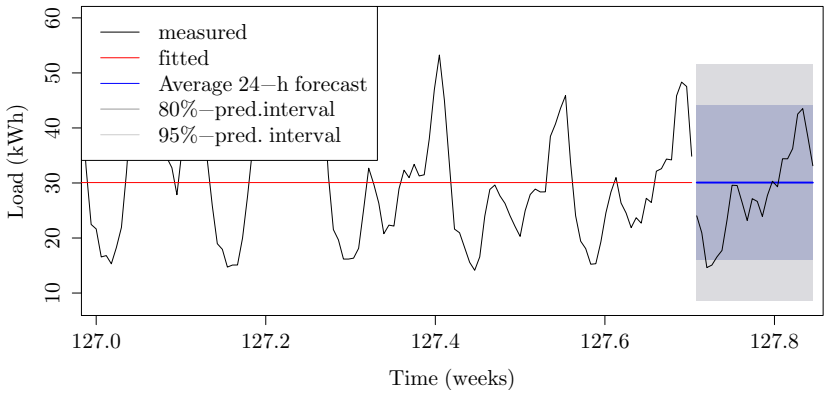
8.1 Simple Methods

Many DSOs use some form of an average method on a daily basis, usually distinguishing between weekdays, weekends and special days, such as holidays. This leads us to first assess the performance of these methods and verify if they could potentially be a good replacement for the more advanced methods, like the `tbats()` function. Due to the simpler nature of these methods we expect the methods to be less accurate; however, we do not know to what extent.

8.1.1 Average method

Results and Commentary. This method takes the average value of all the historical values as the prediction of all future values. The method is only good as a very rough prediction. It does not consider the seasonal nature of the load profile, as can be observed in Figure 40. MAPE reached 34% on the training set and 29% on the test set. The MASE accuracy value of 1.8, a number higher than 1, indicates that the model performed *more poorly than the seasonal naïve method*.

Figure 40. Time Series of SUM69 Group of Households and 24-h Average Forecast

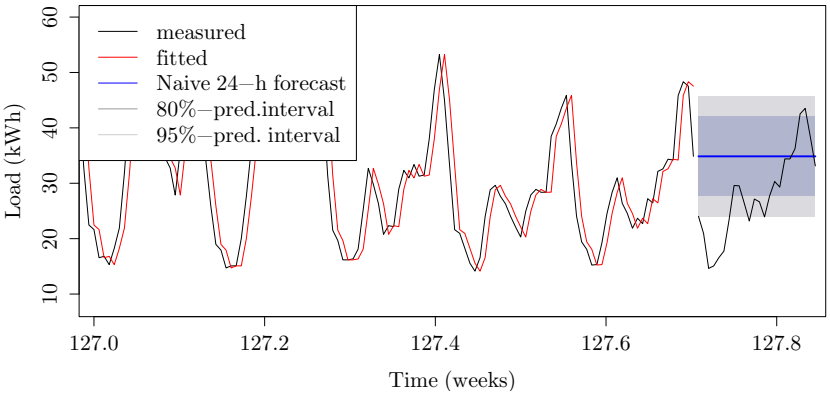


Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

8.1.2 Naïve method

Results and Commentary. The method takes the last observed value as the prediction for all future values (Figure 41). It also does not consider seasonality; thus, the poor performance: MAPE reached 14% on the training set and 41% on the test set, with MASE amounting to 2.4. Seasonal naïve method is to be preferred above this one.

Figure 41. Time Series of SUM69 Group of Households and 24-h Naïve Forecast

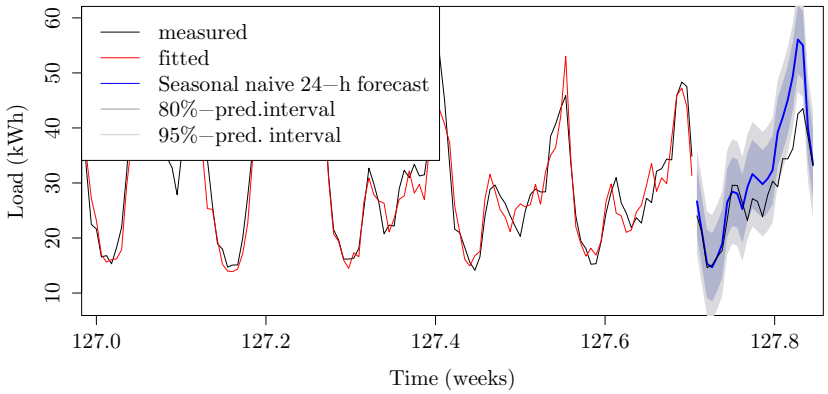


Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

8.1.3 Seasonal naïve method

Results and Commentary. MAPE reached 11% on the training set and 14% on the test set. As expected, introducing the element of seasonality (Figure 42) into a simple forecast is a major improvement for predicting a seasonal time series. The MAPE of this method could serve as a rough benchmark, as it is easily computed.

Figure 42. Time Series of SUM69 Group of Households and 24-h Seasonal Naïve Forecast

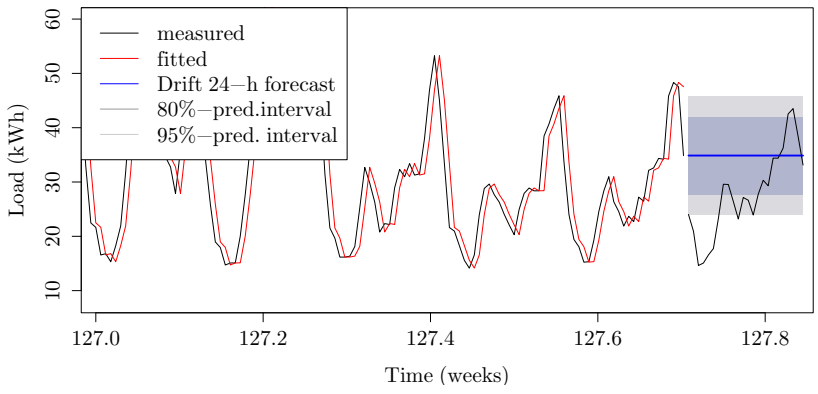


Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

8.1.4 Drift method

Results and Commentary. The drift method increases/decreases with the increase/decrease rate, calculated over the entire historical period. In this specifici case the increase rate is rougly zero, so the drift forecast equals the naïve one (compare Figure 43 to Figure 41). MAPE reached 14% on the training set and 41% on the test set. Drift method is not suitable for a seasonal time series like the load profile.

Figure 43. Time Series of SUM69 Group of Households and 24-h Drift Forecast



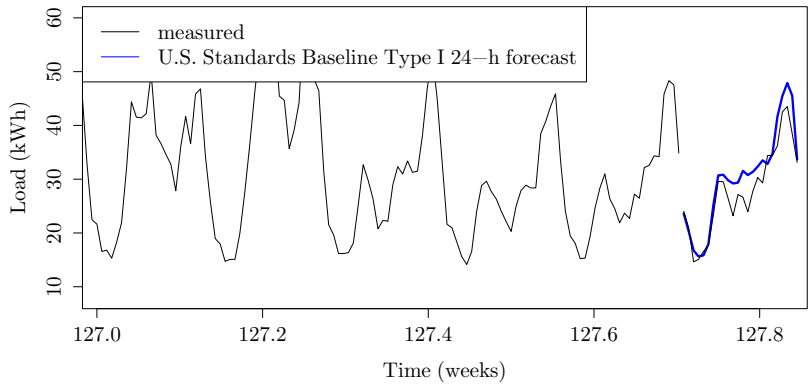
Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

8.1.5 U.S. standards baseline type I

Results. MAPE, achieved by the U.S. standards baseline type I reached 9.5% on the test set, which is better than the similar seasonal naïve method. This is because the former method

forecasts load at a specific time as an average of loads from several different days of the same type, while the later predicts from a single value. In this way, the influence of the random component is somewhat reduced. (Observe, how the forecast in Figure 44 better predicts the observed values than the one in Figure 42.)

Figure 44. Time Series of SUM69 Group of Households and 24-h U.S. Standards Baseline Type I Forecast



Note. Historical Data: 10 Days preceding the forecast. Sampling: 1 h.

Commentary. The U.S. Standards baseline type I, with its MAPE of 9.5% on the test set performed considerably better than the other simple methods and could be acceptably accurate for some purposes. It is still advisable to strive for as good an accuracy as possible, but the baseline can easily be used where other methods do not work or for the purpose of imputing missing data.

8.1.6 Discussion – simple methods

Simple methods should serve as benchmarks. The average, naïve and drift methods are *linear* and do not consider seasonality, and thus do not represent a good prediction method for load profile. The seasonal naïve method and the U.S. standards baseline, in contrast, consider seasonality. U.S. baseline's test set MAPE of 9.5% should serve as a good benchmark to be surpassed by the method selected eventually.

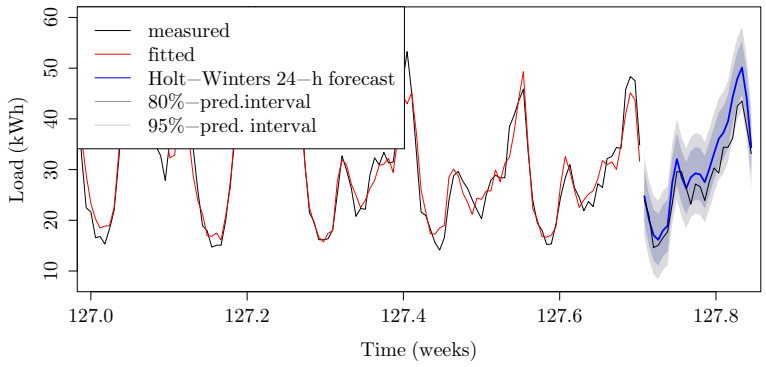
8.2 Exponential Smoothing Methods

8.2.1 Single-seasonal Holt-Winters method

Results. The HW method performs well on SUM69 for the selected period. MAPE reached 9.1% on the training set and 11% on the test set. There is some correlation between lags of

residuals. The residual plot seems normally distributed with mean at 0.1. Forecasting errors are close to being standard normally distributed. Time consumed: 0.02 seconds.

Figure 45. Time Series of SUM69 Group of Households and 24-h Single-Seasonal Holt-Winters Forecast

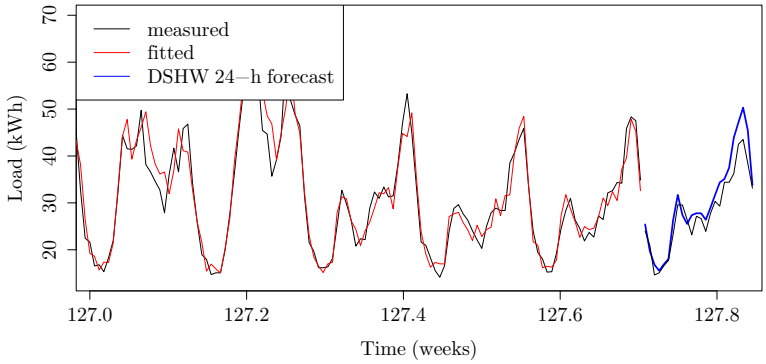


Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

8.2.2 Double-seasonal Holt-Winters method

Results. The DSHW method considers both seasons of the SUM69 load profile and consequentially performs better than HW. MAPE reached 7.6% on the training set and 8.5% on the test set. There is no evidence of autocorrelation of residuals. Residuals seem random and forecasting errors are almost normally distributed, with a slight tendency to over-estimation. The time consumed was 13 seconds.

Figure 46. Time Series of SUM69 Group of Households and 24-h Double-Seasonal Holt-Winters Forecast

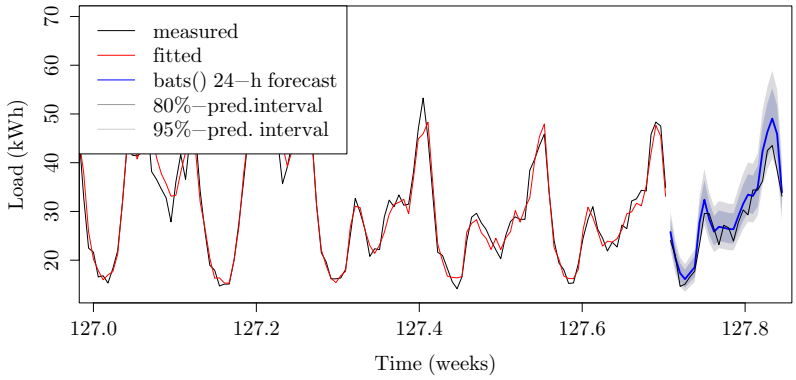


Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

8.2.3 BATS model

Results. The `bats()` function has selected the BATS$(0.002, \{3, 1\}, 0.997, \{24, 168\})$ model. The parameters suggest that it captured both the daily and the weekly seasonalities. MAPE reached 6.2% on the training set and 8.1% on the test set. The good prediction is also visible on Figure 47. There is no evidence of the autocorrelation of residuals. Residuals seem random and forecasting errors are almost normally distributed. Time consumed: 5 minutes and 2 seconds.

Figure 47. Time Series of SUM69 Group of Households and 24-h `bats()` Function Forecast

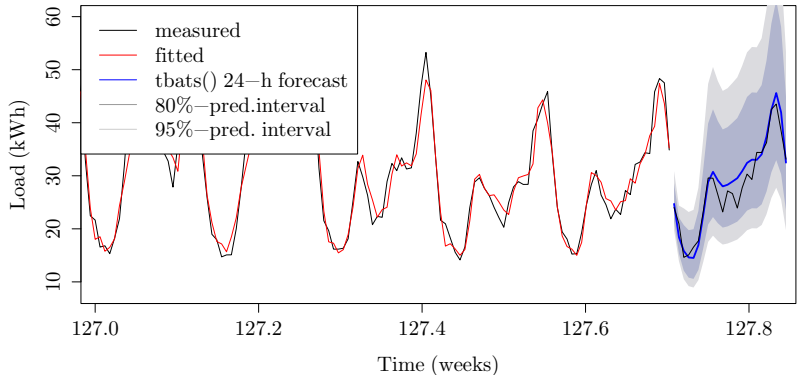


Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

8.2.4 TBATS model

Results. The `tbats()` function has selected the TBATS$(0, \{2, 2\}, -, \{\langle 24, 6\rangle, \langle 168, 3\rangle\})$ model, which also captures both seasonalities. The accuracy has slightly improved over `bats()` but no conclusions may be drawn based on a single case. MAPE reached 8.9% on the training set and 7.3% on the test set. There is autocorrelation within the residuals, which means the seasonal component was not completely extracted from the residuals. Residuals seem random though, with the mean zero and forecasting errors being practically normally distributed. The time consumed was 19 seconds. The results are plotted in Figure 48.

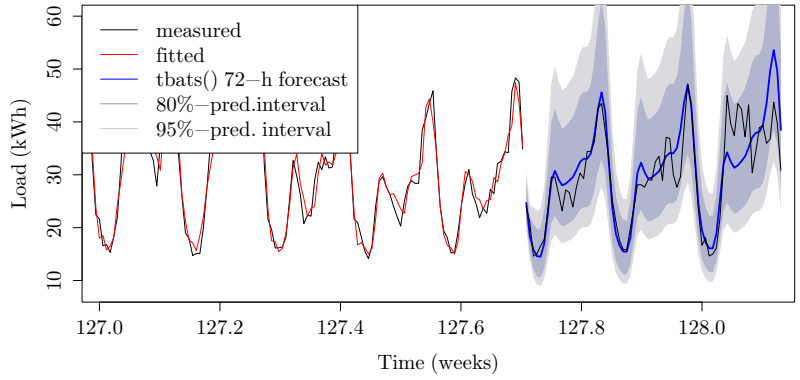
Figure 48. Time Series of SUM69 Group of Households and 24-h `tbats()` Function Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

Using the same fitted TBATS model to forecast 3 days results in MAPE on test set raising to 10.5% (see Figure 49).

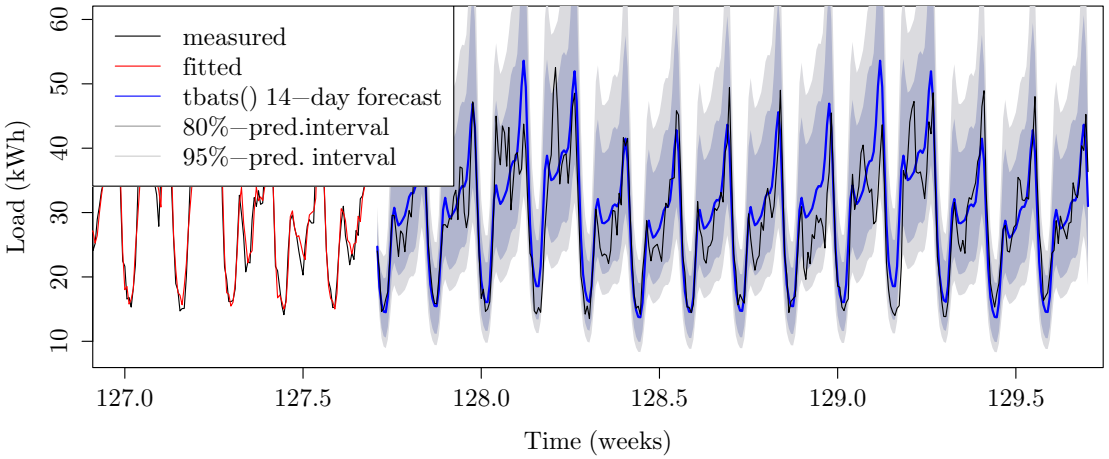
Figure 49. Time Series of SUM69 Group of Households and 3-Day `tbats()` Function Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

When the forecasting horizon is prolonged to 14 days, MAPE rises to 13.1% on the 14-day test set (336 forecast points). Forecast accuracy can be observed in Figure 50.

Figure 50. Time Series of SUM69 Group of Households and 14-Day `tbats()` Function Forecast



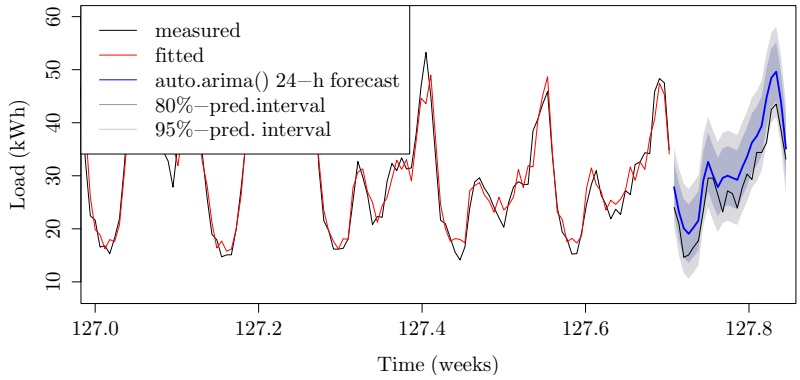
Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

Conclusion. As we can observe, MAPE improves from HW, over DSHW to `bats()` and `tbats()`, based on calculations for the SUM69 time series only. We should not draw conclusions based on one case, but considering experience with forecasting single households, these observations can only be confirmed. Out of these four methods, we would select `tbats()` as it considers two seasonalities and is overall the most sophisticated one. As it includes all models, it would also yield any of the previous models, if they performed better in a specific case. Moreover, it is highly computationally efficient.

8.3 ARIMA Models

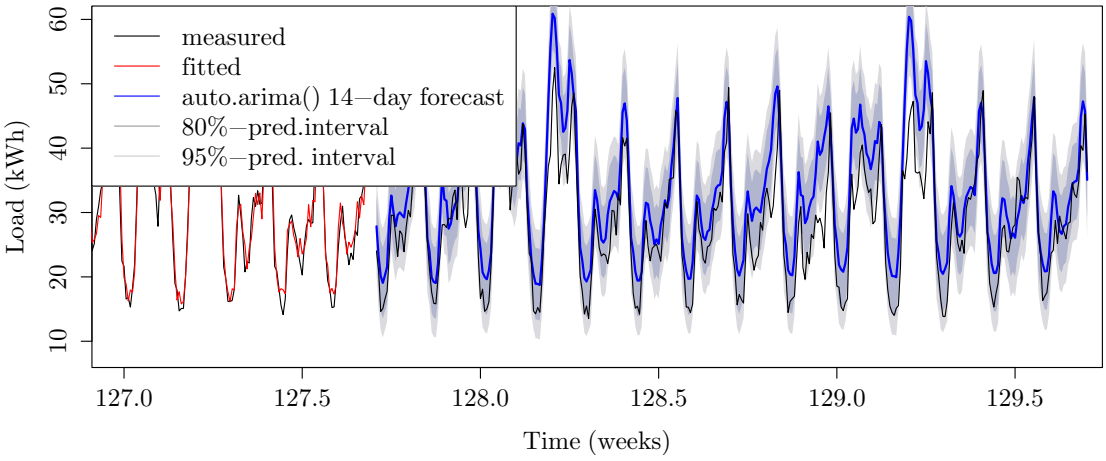
Results. To fit an ARIMA model we used the `auto.arima()` function, which automatically selects the most suitable seasonal or non-seasonal model for the given signal's characteristics and fits it. It chose the ARIMA$(\mathbf{1,0,0})(\mathbf{1,0,1})[\mathbf{168}]$ model with non-zero mean, where (1,0,0) denotes parameters for nonseasonal and (1,0,1) for seasonal component with a weekly seasonality (a frequency of 168 hours). Forecasting based on this model yields test set MAPEs of 16.3%, 15.3%, 14.8% and 18.6% for 1-day, 2-day, 3-day and 14-day forecast, respectively. The 1-day and 14-day forecasts can be observed in Figures 51 and 52, respectively.

Figure 51. Time Series of SUM69 Group of Households and 24-h `auto.arima()` Function Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

Figure 52. Time Series of SUM69 Group of Households and 14-Day `auto.arima()` Function Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

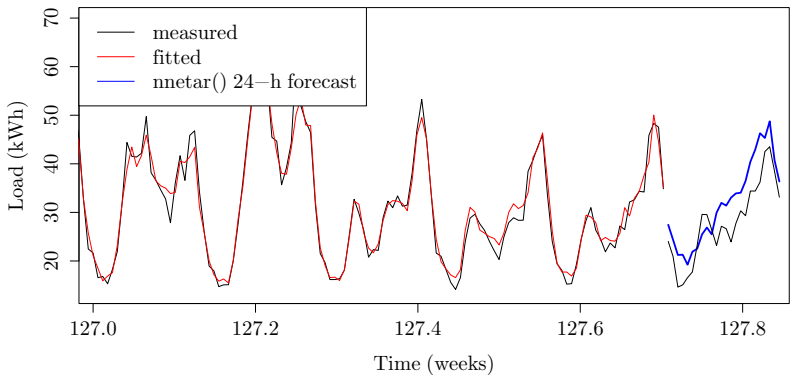
Commentary. One of the prominent methods for STLF on a large scale, ARIMA, seems to perform more poorly on a small scale. It was outperformed by the `tbats()` function; thus on a small scale we will favor `tbats()` over ARIMA. Perhaps we might have been able to improve forecast accuracy if we did not use an automatic algorithm (the type of model, either ARMA, ARIMA or SARIMA would be picked based on results of statistical analyses), but that is what we are looking for.

8.4 Autoregressive Artificial Neural Network Models

This family of methods has been extensively studied in the past decade and was successfully applied for STLF on a large scale. Here we will test how well it performs on a small scale of 69 households and see how it compares to the autoregressive ARIMA.

Results. In our case, we will only consider one "predictor", the historical load data. The NNAR model, selected by the `nnetar()` function, was unsurprisingly the NNAR$(\mathbf{28, 1})$ model, a neural network with the last 28 observations $(\boldsymbol{y_{t-1}, y_{t-2}, \ldots, y_{t-28}})$ used as inputs and output $\boldsymbol{y_t}$ and one neuron in the hidden layer. MAPE achieved 5.8% on the training set and only 18.1% on the test set.

Figure 53. Time Series of SUM69 Group of Households and 24-h `nnetar()` Function Forecast



Note. Historical Data: 28 Days in May 2013. Sampling: 1 h.

Commentary. `nnetar()`'s forecast accuracy is not satisfactory and is slightly worse than that of `auto.arima()`. This is in line with results from a similar study, based on 90 and 230 households in Ireland (Marinescu et al., 2013), where on the 90-household aggregation, ANN performed slightly worse (NRMSE was $\mathbf{3.82\%}$) than ARIMA, which reached a NRMSE of $\mathbf{3.63\%}$. For the rest, the studies are different; ARIMA in the Irish case had a historical interval of 7 days compared to 28 days of the present study. Furthermore, their ANN was not autoregressive but included weather data and day-of-week information. The aggregation level of the 90 households is double of ours, reaching maximum peaks at 140kWh, compared to 70kWh of the SUM69 group of households.

8.5 Summary of Results with Discussion – Groups od Residential Consumers

Forecast accuracy results for all methods, tested on SUM69 time series is summarized in Table 3.

Table 3. Forecast Accuracy for Different Methods for SUM69 Group of Households

Method Group	**Method/Function**	**MAPE (in %)**	
		Training set	Test set
Simple Methods	Average	32.8	28.9
	Naïve	14.5	40.8
	Drift	14.4	41.4
Seasonal Simple Methods	Seasonal Naïve	8.4	14.1
	U.S. Standards Baseline Type I	NA	**9.5**
Exponential Smoothing Methods	Single-Seasonal Holt-Winters	9.1	11.0
	Double-Seasonal Holt-Winters	7.6	8.5
State-Space Models	`bats()`	6.2	8.1
	`tbats()`	8.9	**7.3**
ARIMA Models	`auto.arima()`	6.9	16.3
Artificial Neural Networks	`nnetar()`	5.8	18.1

For the purpose of forecasting a *group of households*, the simplest methods (average, naïve, drift) did not perform well enough. They do not attempt to capture any of the specific features a load profile caries (e.g. seasonality), so their poor performance was expected. Better results were achieved by methods that consider seasonality (seasonal naïve method, U.S. Standards Baseline Type I); they can serve as good benchmarks. Although the ANN models were proven to perform well at STLF on a large scale in a *multivariate* setting including weather, their *univariate* version NNAR performed poorly at STLF on a very small scale.

The single-seasonal Holt-Winters method performed well enough, with MAPE reaching 11.5% on the test set. The double-seasonal Holt-Winters method, as expected, only improved its accuracy. The `auto.arima()` function had an average test set MAPE of 15% and should only be regarded as a *benchmark* to be surpassed. It was also discarded, as fitting several models in some cases resulted in huge time consumption.

`bats()` function performed well, but it only includes single seasonality. The method of choice is the `tbats()` function, as it includes both seasonalities (one day and one week) and also performs well. (A MAPE of 7.3% was reached on the test set for the tested group of 69 households.) The second choice would be the U.S. standards baseline with a MAPE of 9.5%, achieved on the test set. The selected methods can also be used for predicting the loads of industrial consumers, as those are less volatile.

CONCLUSION

Load profiles of *industrial users* display clear seasonal patterns with seasonal cycle lengths of one day, one week and one year. Because of this they can be predicted with good accuracy with most seasonal methods. The test set forecast accuracy measure MAPE of 7.3%, achieved by applying STL-decomposition, should serve as a reference.

Individual households, on the other hand, exhibit very "erratic" behavior with no clear patterns. As a result, they are not reliably predictable by any time series model. This seems to be due to the impact of randomness of human behaviour which cannot be modeled. When aggregating 15-minute data into one hour signals forecast accuracy doubles on average. The best test set MAPE achieved at this sampling rate was 29.6% but can get much worse if the signal is very volatile. (This result is in line with literature.)

In *groups of households*, erratic behavior is reduced for statistical reasons. Individual differences in human behaviour even out in a group setting and the share of this influence is minimized. Load profiles of such groups reveal patterns that are similar to the ones seen in individual industrial consumers. Therefore, it suffices to study the performance of forecasting methods for groups of households only.

Simple average linear methods (e.g., average, naïve and drift) do not capture the *seasonal* nature of the load signal and their poor accuracy of 29–41% was expected. Other average methods (e.g., seasonal naïve method and U.S. standards baseline) manage to capture seasonality and their observed accuracy of 14% and 9.5%, respectively can serve as benchmarks to be surpassed. The family of ARIMA models underperformed, contrary to expectations. It was also too computationally intensive. The neural network AR models that seem to be the state-of-the-art in a large scale multivariate STLF setting, fail as simple univariate time series models and their accuracy of 18% is unacceptable. Exponential smoothing methods (HW, DSHW, BATS and TBATS are generally suitable, especially the last three ones, which are *double-seasonal* methods.

The function `tbats()` achieved a MAPE of 7.3% on the test set. It is our method of choice and a sophisticated state space model. The second choice is clearly the U. S. standards baseline, because of its simplicity, robustness and decent performance. The same two methods can be used for forecasting individual industrial consumers, especially since the lower volatility of their input signal will only result in potentially higher accuracy. It is hard to compare results to known literature due to different settings and different methods used. However, literature reports MAPE values between 5.15% at the university level and 13.8% at the village level, respectively. Our result seems to indicate forecast accuracy on the higher end. This function was, to the best of the author's knowledge, not yet used in studies on real data.

Future work can take on many directions:

1. studying the influence of the *length of the historical time interval* on forecast accuracy,
2. studying different ways of selecting households into groups and this way improving forecasts of existing methods by improving the input signal,
3. studying alternative forecasting methods,
4. combining time series forecasts in hybrid models,
5. improving single steps of the forecasting process,
6. explaining why the noise and irregular components of single households cancel each other out when aggregated.

Ad. 1. All forecasts were done using a 28-day historical time interval, for the following reasons: 1. This period is long enough to capture a load's major seasonality with cycle length of seven days, 2. recent weather is implicitly captured in the signal, 3. earlier weather influences are excluded, 4. for STLF historical data's relevance drops quickly with its age 5. the biggest correlation of the load time series is with its lags -1 and -7. Despite having obtained good forecasting results, longer (and in some cases shorter) historical intervals should be studied to determine the optimal length for historical intervals in terms of forecast accuracy.

Ad. 2. To *improve* the forecast accuracy of existing methods for groups of households *clustering techniques* could be applied to construct clusters (groups) of households with similar load profiles. Furthermore, for VPP purposes customers could be aggregated into groups that reach peaks at different times (for example: one group's peak is another group's valley). Thus, a VPP could place bids on the balancing market at different times. The *minimum dimension of a group* in order to be predictable should be studied. The dimension should be expressed in *group load per time unit*, rather than the number of its members. Households could also be *segmented* based on whether their primary source of heating is electric energy or not and whether they use air conditioning for cooling or not. Another way of segmenting would be by *geographic regions*. Load profiles for the same region will exhibit less volatility, since the weather is similar for all customers.

Ad. 3. A natural extension of research in our direction would be to test the *triple-seasonal* methods (triple seasonal ARMA, triple seasonal HWT exponential smoothing, triple seasonal intraday cycle exponential smoothing), suggested by Taylor (2010). Another direction would be to study *mixed models* that include weather and ARIMA with regression that will include holidays or special events (e.g. soccer game) as predictors.

Ad. 4. As a rule, studies that attempted combining forecast results, obtained from various methods, either by using different methods for different times of day or by taking an average of forecasts from several methods, ended up proving that the *hybrid* method on average

performed better than any one of the single methods included. To combine results from this work, the newly developed `forecastHybrid` package could be used.

Ad. 5. And finally, *outlier detection* and *missing value imputation* should be studied carefully.

Ad. 6. The aggregated signal of many households is smoother and reveals seasonal patterns. This observation should be proved by applying statistical laws.

REFERENCE LIST

1. Akaike, H. (1974). A new look at the statistical model identification. *Transactions on Automatic Control*, *19*(6), 716–723.
2. Alfares, H. K., & Nazeeruddin, M. (2002). Electric load forecasting: Literature survey and classification of methods. *International Journal of Systems Science*, *33*(1), 23–34.
3. Armstrong, J. S. (Ed.) (2001). *Principles of forecasting: A handbook for researchers and practitioners*. New York: Springer.
4. Armstrong, J. S., & Green K. C. (2014, December 13). Selection tree for forecasting methods. *ForecastingPrinciples.com*. Retrieved May 21st, 2016 from http://www.forecastingprinciples.com/index.php/selection-tree
5. Armstrong, J. S., Green K. C., & Graefe A. (2015). Golden rule of forecasting: Be conservative. *Journal of Business Research*, *68*(6), 1717–1731.
6. Baliyan, A., Gaurav, K., & Mishra, S. (2015, December). A review of short term load forecasting using artificial neural network models. *Procedia Computer Science*, *48*(2015), 121–125.
7. Behr, P., & Rahim, S. (2016, June 13). ConEd taps N.Y. customers for 'virtual' power plant project. *Environment & Energy Publishing*. Retrieved June 13th, 2016 from http://www.eenews.net/stories/1060038694
8. Banerjee, A., Dolado, J. J., Galbraith, J. W., & Hendry, D. F. (1993). *Cointegration, error correction, and the econometric analysis of non-stationary data*. Oxford: Oxford University Press.
9. Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, *26*(2), 211–252.
10. Box, G. E. P., & Jenkins, G. M. (1970). *Time-series analysis, forecasting and control* (revised ed., 1976). San Francisco, California: Holden-Day.
11. Box, G. E. P., & Pierce, D. A. (1970). Distribution of residual correlations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, *65*(322), 1509–1526.
12. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis, forecasting and control* (3rd ed.). Upper Saddle River, New Jersey: Prentice-Hall.
13. Brown, R. G. (1956). *Exponential smoothing for predicting demand*. Cambridge, Massachusetts: Artur D. Little Inc..
14. Chatfield, C. (2000, November). *Time-series forecasting*. Chapman and Hall/CRC Press.
15. Chen, H., Cañizares, C. A., & Singh, A. (2001). ANN-based short-term load forecasting in electricity markets. *Power Engineering Society Winter Meeting, 2001. IEEE (Vol. 2)* (pp. 411–415). Columbus, Ohio: IEEE.
16. Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on LOESS. *Journal of Official Statistics*, *6*(1), 3–73.

17. Cordis (n.d.). *Development of novel ICT tools for integrated balancing market enabling aggregated demand response and distributed generation capacity*. Retrieved July 10th, 2016 from http://cordis.europa.eu/project/rcn/105542_en.pdf
18. Crone, S. F., Nikolopoulos, K., & Hibon, M. (2008). Automatic modelling and forecasting with artificial neural networks: A forecasting competition evaluation. *Final Report for the IIF/SAS Grant 2005/6*.
19. De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, *106*(496), 1513–1527.
20. Directive 2003/54/EC of the European Parliament and of the Council concerning common rules for the internal market in electricity.
21. Directive 2009/72/EC of the European Parliament and of the Council concerning common rules for the internal market in electricity.
22. Direktiva 2009/72/EC Evropskega Parlamenta in Sveta o skupnih pravilih notranjega trga z električno energijo.
23. Dogum, E. B. (1988). *X-11-ARIMA/88 seasonal adjustment method – Foundations and users' manual*. Ottawa: Statistics Canada.
24. *eBadge, the pan-European solution, a way forward for integrating the European electricity market*. Retrieved September 7th, 2014 from http://www.ebadge-fp7.eu/
25. Edwards, R. E., New, J. R., & Parker, L. E. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. *Energy and Buildings*, 49, 591–603.
26. ENERNOC (2010, September 20). *Analysis of baseline methodologies*. Retrieved March 14th, 2016 from https://www.naesb.org//pdf4/dsmee092810w1.pdf
27. ENERNOC (2011). *The demand response baseline (White paper)*. Retrieved November 17th, 2013 from http://www.enernoc.com/our-resources/white-papers/the-demand-response-baseline
28. ENERNOC (2014). *What is demand side management?*. Retrieved June 30th, 2014 from http://www.enernoc.com/our-resources/term-pages/what-is-demand-side-management
29. ENTSO-E (2014, October 13). *ENTSO-E Overview of the Internal Electricity Market-related project work*. Retrieved September 14th, 2016 from https://www.entsoe.eu/Documents/Events/2014/141013_ENTSO-E_Update-on-IEM-related%20project%20work_final.pdf
30. EUETS – European Union Emissions Trading Scheme (n.d.). *Internal electricity market glossary*. Retrieved May 26th, 2016 from http://www.emissions-euets.com/internal-electricity-market-glossary/
31. Faraway, J., & Chatfield, C. (1998). Time series forecasting with neural networks: A comparative study using the airline data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *47*(2), 231–250.

32. Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., & Chen, B. C. (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program (with discussion and reply). *Journal of Business and Economic Statistics*, *16*(2), 127–177.
33. Fung, D. S. C. (2006). *Methods for the estimation of missing values in time series* (Doctoral dissertation). Perth: Edith Cowan University.
34. Gajowniczek, K., & Ząbkowski, T. (2014). Short term electricity forecasting using individual smart meter data. *Procedia Computer Science*, 35, 589–597.
35. Ghofrani, M., Hassanzadeh, M., Etezadi-Amoli M., & Fadali, M. S. (2011). Smart meter based short-term load forecasting for residential customers. *North American Power Symposium (NAPS), 2011* (pp. 1–5). Boston, Massachusetts: IEEE.
36. Grimm, C. (2008). Evaluating baselines for demand response programs. *AEIC – Association of Edison Illuminating Companies*. Retrieved June 30th, 2014 from http://publications.aeic.org/lrc/12_Evaluating_Baselines_for_Demand_Response_Programs_PAPER.pdf
37. Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, *16*(1), 44–55.
38. Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*(3), 439–454.
39. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688.
40. Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, *27*(3), 1–22.
41. Hyndman, R. J., & Athanasopoulos, G. (2013, October 17). *Forecasting: Principles and practice*. O'Texts.
42. Hyndman, R. J. (2016, April 14). Package 'forecast'. *CRAN – The Comprehensive R Archive Network*. Retrieved June 14th, 2016 from https://cran.r-project.org/web/packages/forecast/forecast.pdf
43. Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. *Office of Naval Research Memorandum **52***. Pittsburgh, Pennsylvania: Carnegie Institute of Technology.
44. Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10.
45. Kandananond, K. (2012). A comparison of various forecasting methods for autocorrelated time series. *International Journal of Engineering Business Management*, *4*(1), 1.
46. Kendall, M. G., Stuart, A., & Ord, J. K. (1983). *The advanced theory of statistics* (Vol. 3). High Wycombe: Charles Griffin.
47. Kosmač, J., Lakota Jeriček, G., Jurše, J., Kernjak Jager, M., Matvoz, D., Omahen, G., Papič, I., Souvent, A., & Zlatarev, G. (2010, March). *Vizija razvoja koncepta SmartGrids*

v Sloveniji: študija št. 2026. Ljubljana: Elektroinštitut Milan Vidmar.

48. Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, *54*(1–3), 159–178.
49. Ljung, G., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303.
50. Magliavacca, G., Zan, A., Esterl, T., Friedl, W., Kathan, J., Brunner, H., Auer, H., Moisl, F., Lettner, G., Prüggler, W., Schwabeneder, D., Nemček, P., Kolenc, M., Andolšek, A., Šterk, M., & Činkelj, J. (2015, December). Guidelines for the creation of a trans-national reserve/balancing market between AT, IT and SI, Deliverable report D2.5. *eBadge Project*. Retrieved May 25th, 2016 from http://www.ebadge-fp7.eu/wp-content/uploads/2016/01/eBADGE-D2.5-Final.pdf
51. Marinescu, A., Harris, C., Dusparic, I., Clarke, S., & Cahill, V. (2013, May 18). Residential electrical demand forecasting in very small scale: An evaluation of forecasting methods. *Software Engineering Challenges for the Smart Grid (SE4SG), 2013 2nd International Workshop on* (pp. 25–32). San Francisco, California: IEEE.
52. Marinescu, A., Harris, C., Dusparic, I., Cahill, V., & Clarke, S. (2014). A hybrid approach to very small scale electrical demand forecasting. *Innovative Smart Grid Technologies Conference (ISGT), 2014 IEEE PES* (pp. 1–5). Washington, D.C.: IEEE.
53. *Market legislation*. Retrieved May 28th, 2016 from https://ec.europa.eu/energy/en/topics/markets-and-consumers/market-legislation
54. Paravan, D. (2004). *Srednjeročno obvladovanje tveganj proizvajalcev na trgu električne energije* (Doctoral dissertation). Ljubljana: Faculty of Electrical Engineering, University of Ljubljana.
55. Regulation 2009/713/EC of the European Parliament and of the Council establishing an Agency for the Cooperation of Energy Regulators.
56. Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, *71*(3), 599–607.
57. Serena, R. (2014). *The European electricity market liberalization. Motives, problems and benefits for the consumers* (Master's thesis). Tilburg: Tilburg University.
58. *Seventh Framework Programme (FP7)*. Retrieved September 7th, 2014 from http://cordis.europa.eu/fp7/understand_en.html
59. Sevlian, R., & Rajagopal, R. (2014, april 7). Short term electricity load forecasting on varying levels of aggregation. *arXiv preprint arXiv:1404.0058v2*. Manuscript submitted for publication.
60. Shiskin, J., Young, A., & Musgrave, J. (1967). *The X-11 Variant of the Census Method II Seasonal Adjustment Program* (Technical Paper 15). Washington, D.C.: Bureau of the Census, U.S. Department of Commerce.
61. Soliman, S. A., & Al-Khandari, A. M. (2010, April 27). *Electrical load forecasting, modelling and model construction*. Oxford: Butterworth-Heinemann.

62. Sorjamaa, A. (2010). *Methodologies for time series prediction and missing value imputation* (Doctoral dissertation). Aalto: Aalto University School of Science and technology, Faculty of Information and Natural Sciences.
63. Taylor, J. W. (2003, August). Short-term electricity demand forecasting using double seasonal exponential smoothing. *The Journal of the Operational Research Society*, *54*(8), 799–805.
64. Taylor, J. W., de Menezes, L. M., & McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, *22*(1), 1–16.
65. Taylor, J. W., & McSharry, P. E. (2007, December). Short-term load forecasting Mmethods: An evaluation based on European data. *IEEE Transactions on Power Systems*, *22*(4), 2213–2219.
66. Taylor, J. W. (2010, July). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, *204*(1), 139–152.
67. Thielbar, M. F., & Dickey, D. A. (2011). *Neural networks for time series forecasting: Practical implications of theoretical results*. Raleigh, North Carolina: North Carolina State University.
68. Time Series Research Staff (2016). *X-11-ARIMA-SEATS Reference manual, Version 1.1*. Washington, D.C.: Center for Statistical Research and Methodology, U.S. Census Bureau.
69. Winkler, E., Wattles, P., & Pratt, D. (2008, October 3). *Recommendations for measurement and verification standards for wholesale electric demand response*. Retrieved June 9th, 2016 from https://www.naesb.org//pdf3/dsmee100308w7.pdf
70. Winters, P. R. (1960, April). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*(3), 324–342.
71. Wold, H. A. (1938). *A study in the analysis of stationary time series*. Uppsala: Almqvist & Wiksell.

APPENDIXES

LIST OF APPENDIXES

APPENDIX A: Povzetek

OPIS PROBLEMATIKE Z OPREDELITVIJO PREDMETA RAZISKAVE

Električna energija kot dobrina se razlikuje od drugega tržnega blaga, saj je ne moremo zanesljivo skladiščiti v večjih količinah. Energijo je treba porabiti takoj, ko je proizvedena. Da bi ohranjali frekvenco omrežja vseskozi stabilno, pa se morata proizvodnja in poraba usklajevati v realnem času. Naraščajoči delež obnovljivih virov energije (angl. *renewable energy sources* – RES) v omrežju samo še dodatno ogroža stabilnost omrežja.

Poraba električne energije mora biti ocenjena in napovedana vnaprej, odstopanja od teh vrednosti pa je treba izravnavati v realnem času bodisi z aktivacijo rezervnih kapacitet ali z nakupom manjkajočih količin na izravnalnem trgu. Izravnalna energija je občutno dražja od energije, ki je bila kupljena vnaprej.

Da bi znižali operativne stroške, so operaterji distribucijskega omrežja že v osemdesetih letih preteklega stoletja pričeli ponujati različne tarife, da bi tako motivirali odjemalce, da odložijo svojo porabo v obdobjih največje porabe oziroma najvišjih cen na trgu. To so bile prve oblike upravljanja s porabo (angl. *demand side management* – DSM).

Novi liberalizirani trg pa je omogočil tudi nastanek novega poslovnega modela: napovedano, vendar ne tudi porabljeno energijo odjemalcev določenega distributerja je moč donosno ponuditi v odkup na izravnalnem trgu. Programi odziva s spremembo porabe (angl. *demand response* – DR) motivirajo odjemalce, da aktivno odložijo svojo porabo bodisi zaradi cenovnih spodbud ali ustreznega nagrajevanja. Odjemalci lahko tudi zgolj izvajajo energetsko učinkovitost (angl. *energy effiicciency* – EE), ki je pasivna verzija upravljanja s porabo. Šele nedavno so se pojavile napredne programske rešitve za programe DR, ki povezujejo odjemalce, porazdeljene vire energije (angl. *distributed energy resources* – DER) ter operaterje prenosnega in distribucijskega omrežja preko konceptov pametnega omrežja (angl. *smart grid*) z izravnalnim trgom. Imenujemo jih virtualne elektrarne (angl. *virtual power plants* – VPP).

S sprejetjem svežnja zakonov z imenom Tretji energetski paket (angl. *Third Energy Package*) leta 2009 je tudi Evropska unija (v nadaljevanju EU) uvedla zakonodajo, ki je trg električne energije postopno liberalizirala. Njen namen je doseči zanesljivo in konkurenčno oskrbo. Konkurenčnost bo dosežena z ločitvijo lastništva proizvodnih, prodajnih in prenosnih kanalov ter uvedbo neodvisnega upravljavca sistema in neodvisnih distributerjev. Vzpostavi se tudi notranji evropski trg električne energije ter izravnalni trg. Vsaka država ustanovi lasten nacionalni regulatorni organ, vsi organi pa med sabo sodelujejo v okviru Agencije za sodelovanje energetskih regulatorjev (angl. *Agency for the Cooperation of Energy Regulators*, v

nadaljevanju ACER, kodificirano v ”2009/713/EC ACER Regulation”).

Omenjena zakonodaja povsem očitno vzpodbuja razvoj integriranega evropskega izravnalnega mehanizma. V tem kontekstu je ACER leta 2011 pričel razvijati smernice za uravnoteženje oskrbe z električno energijo. Na njihovi osnovi lahko pričakujemo, da bodo na prihodnjem integriranem izravnalnem trgu programi DR igrali pomembno vlogo. Smernice omogočajo nastanek novih poslovnih subjektov, kot so virtualne elektrarne, kjer programi DR enakovredno konkurirajo porazdeljenim proizvodnim virom (Cordis, b. l.).

V kontekstu virtualnih elektrarn pa je kratkoročno napovedovanje porabe električne energije ključno, saj mora virtualna elektrarna z namenom uspešnega licitiranja na izravnalnem trgu natančno poznati porabo svojih sodelujočih (ter količino, ki jo bodo uspeli prihraniti) vsaj za prihodnjih 24 ur, prav tako pa morebitne vrhove porabe na trgu in s tem primanjkljaje enegije. Pri kratkoročnem napovedovanju porabe električne energije (angl. *short-term load forecasting* – STLF) gre za napovedovanje porabe od ene ure do enega tedna vnaprej. Namen je zagotoviti nemoteno preskrbo z električno energijo ob sočasnem minimiziranju dnevnih obratovalnih in distribucijskih stroškov.

Poleg kratkoročnega napovedovanja poznamo še:

- srednjeročno napovedovanje (angl. *mid-term load forecasting* – MTLF): gre za napovedovanje tedenskih, mesečnih in letnih vrhov porabe, kar omogoča učinkovito operativno načrtovanje, ter
- dolgoročno napovedovanje (angl. *long-term load forecasting* – LTLF): uporablja se za napovedovanje od enega do nekaj let vnaprej, kar olajša načrtovanje povečevanja kapacitet.

Ti dve obliki napovedovanja nista predmet tega dela. Vsako od omenjenih treh področij ima svoje značilnosti, ki zahtevajo lastne pristope oziroma metode za reševanje problema.

Medtem, ko je bilo dolgoročno napovedovanje sestavni del sistemskega planiranja vse od njegovih začetkov, pa je kratkoročno napovedovanje pridobilo na pomenu šele ob pojavu liberaliziranega trga. Nenehno vzdrževanje predpisane frekvence omrežja ter s tem zagotavljanje njegove stabilnosti postaja vse bolj zahtevna naloga, kajti tradicionalno električno omrežje postopno prehaja v bolj kompleksno obliko, ki vključuje obnovljive vire energije, mikro omrežja in tako dalje. Za te sodobne oblike omrežij so hitre in predvsem točne napovedi porabe ključnega pomena. Pri tem je treba napovedati tudi proizvodnjo obnovljivih virov, pri napovedi porabe pa upoštevati dogodke DR.

Večina obstoječe literature na področju kratkoročnega napovedovanja porabe električne ener-

gije se nanaša na širše prostorsko območje; bodisi na nivoju posamezne elektrarne, distributerja ali celo države. Delo je poskus kratkoročnega napovedovanja porabe za posameznega industrijskega ali gospodinjskega odjemalca za potrebe virtualne elektrarne. Obsega tako napovedovanje na zelo omejenem območju, kot je na primer skupina nekaj deset hiš, kot tudi na nivoju posameznega odjemalca. V okviru virtualne elektrarne napovedni modul na osnovi zgodovinskih podatkov za posameznega odjemalca napove njegovo porabo v prihodnjih 24 urah. Rezultati tega modula so vhodni podatki za optimizacijski modul, ki na osnovi napovedane porabe vseh odjemalcev ter nekaterih omejitev določi optimalno aktivacijo (skupino odjemalcev, ki naj zmanjša porabo) ob dogodku DR.

NAMEN RAZISKAVE

Namen magistrskega dela je raziskati problematiko napovedovanja kratkoročne porabe električne energije, osvojiti statistične metode, primerne za reševanje tega problema, na realnih podatkih ugotoviti njihovo točnost napovedi ter končno izbrati metodo, ki v povprečju dosega največjo točnost napovedi in je primerno robustna za uporabo v avtomatiziranem okolju. Predvideti in zadovoljivo rešiti je treba tudi druge vidike avtomatizacije, kot so: manjkajoči podatki, prisotnost izrednih dogodkov ter praznikov.

TEMELJNA RAZISKOVALNA VPRAŠANJA

Zasledovala bom naslednji dve raziskovalni vprašanji:

1. Kako bi z zadovoljivo točnostjo napovedali urno porabo električne energije za prihodnjih 24 ur za poljubnega gospodinjskega ali industrijskega uporabnika?
2. Katere metode so primerne za uporabo v avtomatiziranem okolju?

ORIS CILJEV RAZISKAVE

Temeljni cilj raziskave je določiti statistično/e metodo/e, s katero/imi lahko z zadostno točnostjo za potrebe virtualne elektrarne napovemo porabo električne energije za prihodnjih 24 ur za poljubno posamezno gospodinjstvo ali industrijskega uporabnika. V primeru, da to ni mogoče, je treba ugotoviti in po možnosti teoretično razložiti, kaj je vzrok, da takšne metode ni, ter predlagati drugo ustrezno praktično rešitev.

Osrednji cilj je opredeliti, za kakšno vrsto statističnega problema gre, določiti in predstaviti posamezne statistične metode, ki so primerne za reševanje tega problema, preizkusiti njihovo učinkovitost na pridobljenih realnih podatkih ter izmed njih določiti tisto, ki v povprečju daje najbolj točne napovedi. Pri tem vseskozi skušamo ne samo navajati dobljene rezultate, ampak tudi poiskati vzroke ter razložiti, kaj je najverjetneje teoretični vzrok za takšne rezultate.

OPREDELITEV METODOLOGIJE RAZISKOVANJA

Pridobljeni podatki so primarni. Gre za 15-minutne anonimizirane meritve porabljene električne energije iz merilnih števcev, ki jih je zagotovilo podjetje Elektro Ljubljana. Podatki se nanašajo na 235 gospodinjstev in 123 industrijskih uporabnikov med januarjem 2011 in koncem avgusta 2013.

Najprej smo se seznanili s področjem elektroenergetike ter splošnimi principi napovedovanja. Nato smo pregledali obstoječo znanstveno literaturo s podočja kratkoročnega napovedovanja porabe električne energije. Profile porabe smo grafično prikazali ter jih razstavili na komponente. Pri tem smo opazili določene značilnosti profilov in na tej osnovi izbrali ustrezne statistične napovedne metode, primerne za napovedovanje časovnih vrst. Te smo v nadaljevanju primerjali.

POVZETEK VSEBINE

Prvo poglavje predstavi širši kontekst dela. Kompaktno je predstavljen elektroenergetski sistem z vsemi svojimi sestavnimi deli. Pri tem se v funkciji generiranja električne energije poleg tradicionalnih elektrarn v zadnjem času v vedno večjem obsegu pojavljajo tudi porazdeljeni proizvodni viri. Pomembna akterja sta še sistemski operater prenosnega omrežja (angl. *Transmission System Operator* – TSO) ter sistemski operater distribucijskega omrežja (angl. *Distribution System Operator* – DSO).

V *razdelku 1.2* je opisan prehod trga električne energije od tradicionalnega modela, kjer so bili akterji vertikalno integrirane organizacije in trg praktično ni obstajal, do popolnoma svobodnega trga. Proces je EU pričela v letu 1999 in še poteka. Leta 2009 je bil vpeljan zadnji sveženj predpisov, imenovan Tretji energetski paket, kodificiran v direktivi ”Direktiva 2009/72/EC Evropskega Parlamenta in Sveta o skupnih pravilih notranjega trga z električno energijo”. Celotna zakonodaja vzpostavlja tudi podlago za obstoj notranjega trga električne energije v EU. Poznamo dolgoročne trge električne energije, kjer se trguje s standardiziranimi produkti električne energije, poleg tega pa trgovanje za dan vnaprej (angl. *day-ahead*), trgovanje znotraj dneva (angl. *intra-day*) ter izravnalni trg (angl. *balancing market*).

V *razdelku 1.3* je predstavljen izravnalni trg električne energije. Ker električne energije ni moč skladiščiti v večjih količinah ter zaradi zagotavljanja konstantne frekvence omrežja je treba proizvodnjo in porabo usklajevati v realnem času. Sistemski operater prenosnega omrežja ima zakonsko obveznost zagotavljanja stabilnosti omrežja. Odstopanja pokriva iz zakupljenih rezerv proizvodnje ali s kratkoročnim nakupom na izravnalnem trgu.

V *drugem poglavju* prizorišče skrčimo. Predstavljen je koncept upravljanja s porabo ter njena

aktivna oblika, odziv s spremembo porabe. Pri tem odjemalec ob napovedanem dogodku DR aktivno zmanjša svojo porabo električne energije kot odziv na cenovno vzpodbudo ali primerno nagrado. Kot ponudnik programov DR se pojavljajo tudi popolnoma novi poslovni modeli, virtualne elektrarne. Gre za napredne programske platforme, ki povezujejo odjemalce, porazdeljene proizvodne vire ter operaterje prenosnega in distribucijskega omrežja preko konceptov pametnega omrežja z izravnalnim trgom.

V *razdelku 2.2* so predstavljene faze dogodka DR, ki je osreden v programih DR. Ključen pojem predstavlja posameznikova izhodiščna, torej napovedana, poraba v času dogodka (angl. *customer baseline load* – CBL). Učinkovitost omejevanja posameznega udeleženca se meri z razliko med njegovo izhodiščno porabo ter dejansko izmerjeno porabo v času dogodka.

V *razdelku 2.3* pojem odjemalčeve izhodiščne porabe natančno definiramo. Gre za količino električne energije, ki bi jo odjemalec (oziroma vir DR) potrošil, če do dogodka DR ne bi prišlo. Glede na soglasnost ameriških regulatorjev ima "dobra" metoda za izračun odjemalčeve izhodiščne porabe naslednje lastnosti: točnost, preprostost in celovitost.

Poglavje se zaključi s pregledom programov odziva s spremembo porabe v Združenih državah Amerike in Evropi v *razdelku 2.4* ter vlogo, ki jo igra napovedovanje izhodiščne porabe v teh programih v *razdelku 2.5*.

Tretje poglavje uvodoma predstavi problem napovedovanja. Metode so razvrščene v drevo izbora metode napovedovanja. V tem delu se omejimo zgolj na kvantitativne statistične **univariatne metode**, ki napovedujejo prihodnost zgolj na osnovi preteklih in sedanje vrednosti napovedovane količine. Pokazali bomo, da je ta skupina metod veljavna alternativa regresijskim modelom, kjer poleg zgodovinskih meritev v napoved vključimo tudi druge napovedne spremenljivke, kot sta zunanja temperatura in vlažnost. Obrazloženi so "zlati principi napovedovanja", ki omogočajo točne napovedi.

V *razdelku 3.2* se posvetimo napovedovanju porabe električne energije. Zanimalo nas bo zgolj kratkoročno napovedovanje, pri katerem gre za napovedovanje od ene ure do nekaj dni vnaprej, ki je namenjeno zmanjšanju operativnih stroškov. Pri procesu napovedovanja skušamo napovedati prihodnjo porabo odjemalca bodisi s pomočjo napovedne metode, pogosteje pa s postavitvijo matematičnega modela, ki obenem tudi razloži opažene zgodovinske podatke. Model pri tem zgodovinske vrednosti porabe, ki je odvisna spremenljivka, razlaga z vplivom neodvisnih (oziroma napovednih) spremenljivk, kot so vreme, čas dneva, dan v tednu in podobno. Ko izberemo primeren model, ga popolnoma določimo z določitvijo njegovih parametrov z uporabo ene od optimizacijskih tehnik. Nato dani model uporabimo onkraj izhodišča napovedovanja za napoved porabe v prihodnosti. Proces imenujemo ekstrapolacija.

V *razdelku 3.3* kvalitativno vpeljemo matematični model napovedovanja porabe električne energije. Gre za stohastični model, torej model, ki upošteva negotovost vrednosti časovne vrste porabe. Časovno vrsto (angl. *time series* – TS) opaženih zgodovinskih vrednosti pri tem interpretiramo kot realizacijo stohastičnega procesa; tak pristop k napovedovanju pa imenujemo model časovne vrste (angl. *time series model*). Poleg tega lahko same časovne vrste napovedujemo še z regresijskimi (oziroma pojasnjevalnimi) modeli ter mešanimi modeli. Model časovne vrste zgradimo v treh korakih: model najprej identificiramo, nato ga priredimo zgodovinskim podatkom ter končno preverimo njegovo skladnost z realno situacijo. Model dobro razlaga zgodovinske podatke, če so napake (oziroma residuali) nekorelirane ter imajo povprečno vrednost nič.

Razdelek 3.4 predstavi mere točnosti napovedi (angl. *forecast accuracy measures*) ter primerja njihove lastnosti. Za dani problem napovedovanja porabe izberemo povprečno absolutno odstotno napako (angl. *mean absolute percentage error*, v nadaljevanju MAPE) zaradi njene nepristranskosti ter neodvisnosti od velikostnega reda meritev.

V *razdelku 3.5* pregledamo dela drugih avtorjev. Raven združevanja podatkov odločilno vpliva tako na izbor metod, kot tudi na točnost napovedi, ki jo lahko dosežemo. Večina obstoječe literature se nanaša na napovedovanje porabe *širšega prostorskega območja*. Tu so bile uspešne tako preproste metode (npr. metoda tipičnih dni, sezonska naivna metoda), kot klasične metode (multipla regresija, eksponentno glajenje, sezonske verzije modelov ARIMA) ter sodobni modeli na temelju umetne inteligence (npr. umetne nevronske mreže). Najboljša dosežena mera točnosti MAPE znaša okoli 1,5% na nivoju države.

Poskusi napovedovanja porabe *posameznega gospodinjstva* so redki in uporabljajo predvsem sodobne metode umetne inteligence ter so neprimerljivi s pristopi v tem delu. Študije napovedovanja porabe *manjše skupnosti ali komercialnih stavb* so zaenkrat prav tako redke. Omenimo dve študiji, ki na skupini 90 ter 230 hiš uspešno primerjata šest skupin metod ter njihovih kombinacij. Točnost teh napovedi je pričakovano nižja kot pri velikih agregacijah.

Četrto poglavje opiše celoten proces priprave in analize podatkov pred samim napovedovanjem. Podatke opišemo v *razdelku 4.1*. Izhodiščne podatke je dobavilo podjetje Elektro Ljubljana. Gre za 15-minutne anonimizirane meritve porabljene električne energije iz merilnih števcev za 235 gospodinjstev in 123 industrijskih uporabnikov med januarjem 2011 in koncem avgusta 2013. Vsebujejo mnogo manjkajočih meritev, ki so nepovratno izgubljene. V nadaljevanju sem imela na voljo tudi meritve iz tako imenovanih hišnih energetskih prikazovalnikov (angl. *home energy hub* – HEH), ki so bili vzorčeni na 1 minuto. Izhodiščni podatki so bili uporabljeni za razvoj modela, meritve iz prikazovalnikov pa za testiranje razvitega napovednega modula. V obeh primerih sem podatke agregirala na eno uro.

V *razdelku 4.2* je opisan postopek nadomeščanja manjkajočih podatkov. Razvit je bil preprost postopek, ki nadomešča manjkajoče vrednosti na osnovi povprečja istoležnih vrednosti (isti dan v tednu ter ura dneva) preteklih štirih tednov. Z odkrivanjem osamelcev (angl. *outliers*) ter obravnavo prazničnih dni se v tem delu nismo posebej ukvarjali, so pa nakazane programske rešitve.

V *razdelku 4.3* za pripravo obdelave podatkov uporabimo programski paket `forecast`, ki je del programskega okolja jezika `R`. Cilj je bil najti vire in najdaljše obdobje, v katerem večina virov ne bo imela manjkajočih meritev. Obdobje naj bo dolgo vsaj 28 dni ter nadaljnjih 14 dni za napovedi. Vnaprej smo izločili vse vire z več kot 10 % manjkajočih vrednosti. Preostalo je 94 industrijskih in 171 gospodinjskih virov, ki smo jih shranili v podatkovne strukture tipa `msts`. Časovne oznake smo pretvorili v univerzalni koordinirani čas (angl. *coordinated universal time* – UTC). Za primer napovedovanja porabe skupine gospodinjstev smo postopno ustvarili skupino 69 gospodinjstev, ki je imela najdaljše strnjeno časovno obdobje dolgo čez štiri mesece. Časovno vrsto skupine označimo s SUM69. Podatke sem agregirala v enourne, ker so poskusi pokazali, da je nihajnost (angl. *volatility*) 15-minutnih podatkov prevelika za uspešno napovedovanje.

V *razdelku 4.4* si podatke ogledamo v obliki grafov in poskušamo ugledati značilne vzorce, ki jih časovne vrste lahko vsebujejo: *trend, sezonski vzorec, ciklični vzorec, slučajno komponento*. Grafi industrijskih odjemalcev kažejo značilen sezonski vzorec, ki hkrati ustreza trem sezonam različnih dolžin: dnevni, tedenski in letni. Tedenski vzorci se skoraj prekrivajo, zato napovedljivost ni vprašljiva. Grafi gospodinjskih odjemalcev kažejo mnogo večjo nihajnost ter odsotnost vzorcev. Očitno je, da bo napovedljivost teh signalov vprašljiva. Graf skupine gospodinjstev spet kaže podobno sliko kot graf posameznega industrijskega odjemalca, torej prisotnost sezonskih vzorcev. Razlogi za ta (intuitivno pričakovan) pojav so statistične narave.

V *razdelku 4.5* pri raziskovalni analizi podatkov signal industrijskega odjemalca s pomočjo STL dekompozicije razstavimo na posamezne komponente (vzorce) in s tem potrdimo opažanja prejšnjega razdelka. Pri gospodinjskih odjemalcih zaradi odsotnosti vzorcev signala ne poskušamo razstaviti. Namesto tega primerjamo obnašanje signala skupine gospodinjstev SUM69 s signali posameznih gospodinjstev, ki ga sestavljajo. Iščemo morebitne korelacije med signali, vendar jih ni – vsako gospodinjstvo ima vrhove porabe ob drugačnih časih. V bodoče upamo, da bomo uspeli napovedati vsaj skupine gospodinjstev. Izbrane metode bi potem lahko uporabili tudi za napovedovanje porabe posameznih industrijskih odjemalcev, saj je njihov signal kvečjemu bolje napovedljiv.

Peto poglavje temeljito opiše vse uporabljene metode oziroma modele. Najprej v *razdelku 5.1* spoznamo nekaj preprostih metod (povprečno, naivno, drift, sezonsko naivno ter izhodiščno

porabo tipa I po ameriškem standardu (angl. *baseline type I*)), ki temeljijo na povprečjih. Zaradi preprostosti pri prvih treh ne pričakujemo dobre kvalitete napovedi. Slednji dve pa že upoštevata sezonskost signala in sta primera učinkovite in robustne metode, ki ju lahko uporabimo za referenčne vrednosti točnosti napovedi.

Razdelek 5.2 uvaja najpomembnejše metode tega dela, modele časovnih vrst. Najprej se temeljito posvetimo potrebni teoriji. Definiramo močno in šibko stacionarnost stohastičnega procesa. Poseben primer je tako imenovani "beli šum" (angl. *white noise*), ki je popolnoma naključen proces. Definiramo avtokorelacijsko funkcijo ter njeno sliko, korelogram. Njen pomen je v odkrivanju stacionarnosti stohastičnega procesa. Kadar proces ni stacionaren, lahko iz njega z diferenciranjem izločimo stacionaren proces. Stacionarnost lahko preverjamo tudi s testi korenov enote. Woldov izrek o reprezentaciji je ključen za določanje pogojev enoličnega modela v nekaterih primerih. V ta namen je treba definirati tudi vzročen ter obrnljiv stohastični proces. V primeru avtoregresijskega procesa definiramo tudi karakteristični polinom. Proces ima enolično rešitev v obliki stacionarnega procesa natanko tedaj, ko vse ničle tega polinoma ležijo izven enotnega kroga. Na osnovi *informacijskih kriterijev* izbiramo najbolj optimalen model.

V *razdelku 5.3* predstavimo *klasično metodo dekompozicije* časovne vrste ter napovedovanje s pomočjo *STL dekompozicije*.

V *razdelku 5.4* predstavljena družina modelov ARIMA (angl. *autoregressive integrated moving average*) združuje tako avtoregresijske modele AR, kot modele drsečih povprečij MA, kot tudi kombinacije ARMA, ARIMA, ter sezonsko obliko SARIMA. Vse modele predstavimo s formulami ter opišemo način identifikacije modela. V programskem okolju `R` uporabimo funkcijo `auto.arima()`, ki izmed vseh modelov ARIMA na osnovi kriterija AIC izbere najbolj primernega ter ga priredi podatkom.

Metode eksponentnega glajenja (angl. *exponential smoothing*, *ES*) v *razdelku 5.5* so druga ključna skupina v okviru modelov časovnih vrst. Namenjene so tako glajenju signala kot napovedovanju. Predstavimo preprosto eksponentno glajenje, linearno Holtovo metodo, sezonsko Holt-Wintersovo metodo ter dvosezonsko Holt-Wintersovo metodo. Nato vpeljemo sodobne stohastične modele stanja-prostora (angl. *state-space models*). Izpostavimo dve pomembni funkciji, implementirani v `R`: `bats()` ter `tbats()`, ki se nanašata na istoimenska modela. Prva vključuje sezonskost, druga pa dvosezonskost.

Za primerjavo si v *razdelku 5.6* ogledamo še sodobne avtoregresijske modele umetne nevronske mreže (angl. *autoregressive artificial neural network models*).

V *šestem do osmem poglavju* so predstavljeni rezultati napovedovanja porabe električne ener-

gije posameznega industrijskega odjemalca, posameznega gospodinjskega odjemalca ter skupine gospodinjstev. Učinkovitost metode določamo na osnovi časovne zahtevnosti modela ter izbrane mere točnosti MAPE. *Osmo poglavje* obravnava največ metod, kajti v predhodnih dveh poglavjih ugotovimo, da je poraba posameznega gospodinjstva nepredvidljiva ter da bomo metodo, ki jo bomo izbrali za napovedovanje porabe skupine gospodinjstev, lahko uporabili tudi v procesu napovedovanja porabe posameznega industrijskega odjemalca. Izbrana metoda, funkcija `tbats()`, je zelo izpopolnjen model stanja-prostora, ki je sposobna obravnavati večsezonskost signala porabe.

REZULTATI

Pri napovedovanju porabe **posameznih industrijskih odjemalcev** smo testirali nekaj zelo različnih metod. Pri napovedovanju s pomočjo STL-dekompozicije je MAPE na učni množici dosegla 4,3 %, na testni množici pa 7,3 %. To vrednost vzamemo za referenco. Od tu dalje nas zanima in navajamo MAPE samo še za testno množico.

Kljub sezonskosti je (enosezonska) Holt-Wintersova metoda dala nezadovoljive rezultate: na testni množici je MAPE pri napovedovanju za 1 dan, 2 dneva, 3 dneve ter 14 dni vnaprej dosegla naslednje vrednosti: 16 %, 19 %, 18 % ter 18 %. To je nezadostno za potrebe napovedovanja v okolju virtualne elektrarne.

Avtomatska funkcija `auto.arima()` je dala izredno zadovoljive rezultate: MAPE je pri eno- oziroma dvodnevni napovedi dosegla 4,5-odstotno oziroma 5-odstotno točnost napovedi na testni množici. Vseeno se moramo funkciji odpovedati. Samodejna izbira modela je verjetno zaradi prilagajanja več različnih modelov ARIMA zgodovinskim podatkom v danem primeru tekla kar 5 minut, kar je nesprejemljivo pri napovedovanju večjega števila signalov v realnem času. Kljub temu smo uspeli dokazati napovedljivost porabe posameznega industrijskega odjemalca.

Pri napovedovanju porabe **posameznega gospodinjstva** smo testirali štiri metode, ki vključujejo sezonskost. Pri tem je prišla do izraza težavnost napovedi teh zelo nihajnih signalov. Testiranje enosezonske Holt-Wintersove metode na več posameznih gospodinjstvih ter pri dveh različnih vzorčenjih (15 min, 1h) je pokazala, da je v večini primerov točnost napovedi na agregiranih enournih podatkih bistveno boljša zaradi večje gladkosti teh signalov. Vendar je tudi najboljša opažena točnost pri MAPE = 45 % nezadostna za potrebe virtualne elektrarne.

Dvosezonska Holt-Wintersova metoda predpostavlja dvosezonsko naravo signalov, vendar pa so le-ti za posamezno gospodinjstvo zelo nepredvidljivi in ne kažejo izrazitih sezonskih vzorcev. Metoda povsem očitno ni kos naravi teh signalov in jih ne more zadovoljivo napo-

vedati. (Zaznana točnost med 45 % in 161 % pomeni, da so rezultati neuporabni; v dveh od treh primerov pa model tudi ni uspel pojasniti celotne variabilnosti v podatkih.)

Funkcija `bats()` samodejno izbere najboljši model iz družine modelov BATS na podlagi vrednosti kriterija AIC. Uspešnost funkcije je bila odvisna od vsakokratnih podatkov. V enem primeru je model uspel slediti obliki signala, ob 30-odstotni vrednosti MAPE; v dveh drugih pa ni uspel odkriti sezonskosti signala, izbrani model pa ni uspel pojasniti celotne variabilnosti. V najslabšem primeru je znašala mera točnosti MAPE nezadovoljivih 120 %.

Funkcija `tbats()` je nadgradnja funkcije `tbats()` in samodejno izbere najboljši model iz družin sezosnkih modelov TBATS ter nesezonskih modelov BATS na osnovi najnižje vrednosti AIC. Uspešnost je bila tudi tu zelo različna. Vsi izbrani modeli so bili dvosezonski. V najboljšem primeru je bila mera točnosti napovedi MAPE 34 %, v najslabšem pa 65 %, pri čemer model ni uspel napovedati oblike signala, niti pojasniti njegove celotne variabilnosti. Funkcija je tudi časovno mnogo manj zahtevna od funkcije `bats()`, domnevno zaradi novejše paralelne implementacije. Očitno so signali posameznih gospodinjstev preveč nepredvidljivi za napovedovanje, saj jim ob odsotnosti pravilnih vzorcev ni kos niti najbolj prefinjena metoda.

Pri napovedovanju porabe **skupine gospodinjstev** je vpliv nepredvidljivega vedenja posameznika zmanjšan, signal pa kaže podobne vzorce kot pri posameznem industrijskem odjemalcu. testirali smo 12 metod na signalu skupine 69 gospodinjstev.

Preproste linearne povprečne metode (povprečna, naivna ter drift) ne uoštevajo sezosnke narave signala, zato je bila slaba točnost z vrednostmi 29 %, 41 % ter 41 % pričakovana. Drugi povprečni metodi (sezonska naivna ter izhodiščna poraba tipa I po ameriškem standardu) sezonskost upoštevata in njuni zaznani točnosti v višini 14 % ter 9,5 % lahko uporabimo kot osnovno merilo uspešnosti. Funkcija `auto.arima()` je uspela poiskati ustrezen sezonski model ter odkriti 7-dnevni sezonski cikel, vendar pa je bila z mero MAPE v višini 16,3 % njena uspešnost pod pričakovanji. Poleg tega je bilo prilagajanje več modelov časovno prezahtevno. Čeprav so modeli umetnih nevronskih mrež pri uporabi na širšem prostorskem območju zelo uspešni, pa se v univariatni obliki za napoved časovnih vrst niso izkazali. Točnost modela, ki ga je izbrala funkcija `nnetar()`, je znašala nesprejemljivih 18,1 %.

Metode eksponentnega glajenja (enosezonska Holt-Wintersova metoda, dvosezonska Holt-Wintersova metoda, `bats()` ter `tbats()`) so bile najbolj uspešne, ob naslednjih vrednostih mere točnosti: 11 %, 8,5 %, 8,1 % ter 7,3 %. Obe slednji sta uspeli poiskati ustrezen enosezonski oziroma dvosezonski model. Pri funkciji `tbats()` se je točnost pri podaljšanju napovednega horizonta na 14 dni poslabšala zgolj na 13 %.

SKLEPI

Za preciznejše napovedovanje skupin gospodinjstev izberemo funkcijo `tbats()`, za primere, ko potrebujemo preprosto in robustno metodo, pa izhodiščno porabo tipa I po ameriškem standardu. Izbrani funkciji uporabimo tudi za napovedovanje porabe posameznih industrijskih odjemalcev, kjer je posledica manjše nihajnosti vhodnega signala kvečjemu večja točnost metode. Vse testirane funkcije so primerne za uporabo v avtomatiziranem okolju.

APPENDIX B: Function `tbats()`

For *non-seasonal time series*:

1. Apply the non-seasonal BATS model and return the forecasting result.

For *seasonal time series*:

1. Compute the *non-seasonal model* by using BATS.
2. Compute a specific seasonal model using the seasonal TBATS model and store it as *best model*:
   a) For *smaller* periods:
      – generate a few specific TBATS models
      – if any of them has a lower or equal AIC than the current best model, store it as the best model.
   b) For *larger* periods:
      – calculate 3 new specific TBATS models (parallel computing may be used)
      – choose the one of the three with the lowest AIC and set it as the best model
      – attempt to iteratively find a better model of the same type and set it as the best model
3. Save the current best model as the *auxiliary model*. It will be used in Item 4.
4. If the model from Item 1 has a lower AIC than the best model from Item 2b, save it as the best model.
5. Generate further specific seasonal TBATS models:
   a) For *parallel* computing:
      – use function `parFilterTBATSSpecifics()` to generate the TBATS models.
      – Select the one with the lowest AIC and save it as the *best seasonal model*
      – If it has a lower AIC as the current best model, select it as *best model*.
   b) For *serialized* computing:
      – Use function `filterTBATSSpecifics()` to generate the new model.
      – If it has a lower AIC than the current best model, select it as *best model*.

APPENDIX C: Acronyms

ACER Agency for the Cooperation of Energy Regulators. 1, 2, 7, 14
ACF Autocorrelation Function. ii, 22, 50, 51, 55, 58–62
ACVF Autocovariance Function. 50
AEIC Association of Edison Illuminating Companies. 19
AIC Akaike Information Criterion. 8, 56, 62, 69
ANN Artificial Neural Network. 28, 70, 71
AR Autoregressive. 22, 54, 58, 59, 61
ARIMA Autoregressive Integrated Moving Average. ii, iii, 5, 21, 29, 47–49, 57, 58, 62, 65, 70, 73, 74, 91, 92
ARMA Autoregressive Moving Average. ii, 55, 56, 60–62, 67–70

BATS Box-Cox transform, ARMA errors, Trend, and Seasonal components. iii, 67–69, 79–81, 89

CBL Customer Baseline Load. 4, 12, 13
ConEd Consolidated Edison. 14

DER Distributed Energy Resources. 1
DR Demand Response. 1–5, 9–15, 45–47
DSHW Double-seasonal Holt-Winters method. 66, 69, 77, 79, 88, 91
DSM Demand Side Management. 1, 2, 9, 10
DSO Distribution System Operator. 1, 2, 4, 6–8, 10, 14, 15, 28

EE Energy Efficiency. 1
EOF Empirical Orthogonal Functions. 32
EPS Electric Power System. 6
EU European Union. 1, 4, 5, 7, 9

FERC Federal Energy Regulatory Commission. 13

HEH Home Energy Hub. 6
HW Holt-Winters method. 87, 88, 91

ICT Information and Communications Technology. 10
IIF International Institute of Forecasters. 17

LOESS Locally weighted smoothing. 57

MA Moving Average. ii, 22, 54–56, 58–62
MAE Mean Absolute Error. 25, 26
MAPE Mean Absolute Percentage Error. 5, 26
MASE Mean Absolute Scaled Error. 26

ME Mean Error. 24
MPE Mean Percentage Error. 25

NAESB North American Energy Standards Board. 11
NNAR Neural Network Autoregression. 70, 93
NRMSE Normalized Root Mean Square Error. 25, 93
NYISO New York Independent System Operator. 13

PACF Partial Autocorrelation Function. 51, 58, 61

RES Renewable Energy Sources. 1, 2, 6, 8–10, 15, 19
RMSE Root Mean Square Error. 25

SG Smart Grid. 1
SOM Self Organizing Maps. 32
STL Seasonal Decomposition of Time Series by LOESS. ii, 7–9, 57, 71, 72, 74
STLF Short Term Load Forecasting. 2, 4, 95
SVM Support Vector Machines. 28, 30

TBATS Trigonometric Box-Cox transform, ARMA errors, Trend, and Seasonal components. iii, 67, 69, 81, 82, 89, 90
TSO Transmission System Operator. 4, 6–8, 10, 15

U.S. United States. 5, 13, 47
UTC Coordinated Universal Time. 33

VPP Virtual Power Plant. 1–3, 10, 13–15, 47, 96

APPENDIX D: Index

Note: **Bold** font is used for page numbers that contain a term's description; plain font indicates the term's mentioning. `Typewriter` font is used for annotating `R` functions and packages. `R` functions are given with parentheses, for example `bats()`.

APPENDIX E: Author Index

Note: A page number in plain font indicates a citation; *italic* font indicates a mentioning.