Machine Learning Applications for Load Predictions in Electrical Energy Network

Nils Jakob Johannesen

Machine Learning Applications for Load Predictions in Electrical Energy Network

Doctoral Dissertation for the Degree *Philosophiae Doctor (PhD)* at
the Faculty of Engineering and Science, Specialisation in Renewable Energy

University of Agder
Faculty of Engineering and Science
2022

Doctoral Dissertations at the University of Agder 376
ISSN: 1504-9272
ISBN: 978-82-8427-086-9

Printed by 07 Media
Oslo

Preface

This thesis is submitted in partial fulfillment of the requirements for the Ph.D. degree from the University of Agder (UiA). The research work was conducted during the period from February 2019 to May 2022 at the Department of Engineering and Science, Renewable Energy section.

Acknowledgments

First of all I would like to thank my supervisor, Prof. Dr. Mohan Lal Kolhe and my co-supervisor Prof. Dr. Morten Goodwin. Your knowledge and excellence guided me throughout the best education program I have ever experienced. Your openness and thrive towards genuine insights inspired me throughout the entire process. Thank you for your ongoing encouragement, unconditional support, comments and guidance.

I want to thank the research community at CAIR-entre, who wholeheartedly invited me into the artificial intelligence research centre and accepted my skewed objectives. It was an ascending learning journey for me. In particular i want to thank Thomas Fauskanger, for dedicated conversations on computer programming and artificial intelligence.

I am thankful to everyone I met on my journeys on conferences networking. Presenting the work on conferences gave me valuable feedback and continuously improved the work. I will especially mention the PES general meeting in Beijing 2019, where I had the opportunity to befriend great researchers from all over the world including Imperial College, TU Delft, Carnegie University, ETH Zurich and Tsinghua University.

I am grateful to the scientific community at IEEE SpliTech conferences that I visited several times, and where I met Prof. Dr Petar Sarajcev. Together we started Erasmus program between FESB Split and University of Agder, and I had the opportunity to do partner visit in Split, Croatia in spring 2020. From FESB Split I also want to thank Dr. Petar Solic, co-founder of SpliTech Conference who gave me opportunities to present my work and chair special session within the field of smart energy.

I want to thank Thomas, Immanuel, Bilal for great conversations and keeping the spirit up in our common office. And thank you Alex Ho, whom I shared office with in the beginning of the thesis work, and whom I kept relations with throughout the phd-work, for our ongoing conversations on computer programming and mathematics.

Thanks to my family.

Nils Jakob Johannesen
Grimstad, Norway
May 2022

Abstract

Load forecasting is a required application in Smart-Grid, which provides essential input to other applications such as Demand Response, Topology Optimization and Anomaly Detection, facilitating the integration of intermittent clean energy sources. The big data processing and operation of the energy system will require flexible tools to manage the smart energy system, by using Information and Communication Technologies, Distributed Generation and Artificial Intelligence, together. Machine Learning can provide electrical load demand forecasting, giving information about future loads. In the literature there are many methods on energy prediction, but most of them have used continuous time approach together with complex neural networks which requires huge amount of data. In this work collected operational data of typical urban and rural energy network are analysed for predictions of energy consumption, as well as for selected region of Nordpool electricity markets. The regression techniques are systematically investigated for electrical energy prediction and correlating other impacting parameters. The k-Nearest Neighbour (kNN), Random Forest (RF) and Linear Regression (LR) are analysed and evaluated both by using continuous and vertical time approach. It is observed that for 30 minutes predictions the RF Regression has the best results, shown by a mean absolute percentage error (MAPE) in the range of 1-2 %. kNN show best results for the day-ahead forecasting with MAPE of 2.61 %. The presented vertical time approach outperforms the continuous time approach. To enhance pre-processing stage, refined techniques from the domain of statistics and time series are adopted in the modelling. Reducing the dimensionality through principal components analysis improves the predictive performance of Recurrent Neural Networks (RNN). In the case of Gated Recurrent Units (GRU) networks, the results for all the seasons are improved through principal components analysis (PCA). This work also considers abnormal operation due to various instances (e.g. random effect, intrusion, abnormal operation of smart devices, cyber-threats, etc.). In the results of kNN, iforest and Local Outlier Factor (LOF) on urban area data and from rural region data, it is observed that the anomaly detection for the scenarios are different. For the rural region, most of the anomalies are observed in the latter timeline of the data concentrated in the last year of the collected data. For the urban area data, the anomalies are spread out over the entire timeline. The frequency of detected anomalies where considerably higher for the rural area load demand than for the urban area load demand. Observing from considered case scenarios, the incidents of detected anomalies are more data driven, than exceptions in the algorithms. It is observed that from the domain knowledge of smart energy systems the LOF is able to detect observations that could not have detected by visual inspection alone, in contrast to kNN and iforest. Whereas kNN and iforest excludes

an upper and lower bound, the LOF is density based and separates out anomalies amidst in the data. The capability that LOF has to identify anomalies amidst the data together with the deep domain knowledge is an advantage, when detecting anomalies in smart meter data. This work has shown that the instance based models can compete with models of higher complexity, yet some methods in preprocessing (such as circular coding) does not function for an instance based learner such as k-Nearest Neighbor, and hence kNN can not option for this kind of complexity even in the feature engineering of the model. It will be interesting for the future work of electrical load forecasting to develop solution that combines a high complexity in the feature engineering and have the explainability of instance based models.

Sammendrag

Energiprediksjon er en essensiell applikasjon i driften av smarte elektriske nettverk. Energiprediksjon gir viktig informasjon til nødvendige styringsprogram som forbruksrespons, og til topologioptimalisering og anomalideteksjon. Dette vil hjelpe integreringen av fornybare energikilder i det elektriske nettverket. Stor-data prosessering og operasjon av energi nettverket, vil trenge fleksible verktøy for å kunne administrere smarte energi system, gjennom informasjonsteknologi, distribuerte nettverk og kunstig intelligens. Maskinlæring kan gi informasjon om fremtidig elektrisk forbruk. Det eksisterer mange metoder for energiprediksjon, mange av dem bruker kontinuerlig tidsorganisering kombinert med komplekse nevrale nett, som trenger store mengde data. I dette arbeidet er typiske operasjonsdata fra urbane og rurale energi nettverk, i tillegg til utvalgte regioner fra NordPool spot marked, analysert for å kunne forutsi energiforbruk. Regresjonsteknikker er systematisk gjennomgått for elektrisk energiprediksjon i tillegg til korrelasjonsanalyse av andre influerende faktorer. k-Nearest Neighbour (kNN), Random Forest (RF) og Liniærregresjon (LR) er analysert og evaluert med både kontinuerlig og vertikal tidsorganisering. Det er observert at for halvtimes prediksjoner gir RF best resultat, gjennom en mean absolute percentage error (MAPE) på omkring 1-2 %. kNN gir best resultat for 24 timers prediksjon, med en MAPE på 2,61 %. Den presenterte vertikale tidsorganiseringen presterer bedre enn kontinuerlig tidsorganisering. For å forbedre preprosesseringen av data, er raffinerte teknikker fra statistikk og tidsserieanalyse adoptert i modellene. Ved å redusere dimensjonaliteten på input data gjennom prinsipal komponentanalyse (PCA) prediktive ytelsen til Recurrent Neural Networks (RNN) forbedret. I Gated Recurrent Units (GRU), er alle resultater, uavhengig av sesong, forbedret gjennom bruk av PCA. I dette arbeidet er også unormal operasjon av grunnet ulike variabler (tilfeldige utfall, unormal operasjon av smarte enheter, cyber-trusler, osv.). I resultatene fra kNN, iforest og Local Outlier Factor (LOF) anvendt på urban og rural data, er det observert at anomalideteksjon for disse to scenarioene er forskjellige. For de rurale regionenes data, er de fleste anomaliene observert i den seneste delen av tidslinjen, konsentrert i det siste året av fem. For urban data er de detekterte anomaliene spredt over hele tidslinjen. Antallet detekterte anomalier var betydelig fler for urban enn rural data. Observert i disse brukerscenarioene, er detekterte hendelser mer data drevne, enn unntak i algoritmene. Det er observert, i dette arbeidet, gjennom domene kunnskap om smarte energisystem at LOF kan detektere tilfeller som ikke ville vært synlige med visuell inspeksjon alene, i motsetning til iforest og kNN. Der kNN og iforest ekskluderer et øvre og nedre bånd, er LOF tetthetsbasert og separerer ut anomalier 'midt' i datapunktene. Evnen LOF har til å identifisere anomalier midt i datapunktene vil sammen med inngående domenekunnskap være en fordel, ved anoma-

lideteksjon i data fra smartmålere. Dette arbeidet har vist at instans-baserte modeller kan konkurrere med mer komplekse modeller, selv om noen områder i pre-prosesseringen (slik som circular coding) ikke fungerer for instans-baserte modeller,derfor kan ikke disse modellene nyttiggjøre seg denne graden av kompleksitet. Det vil være interessant for det videre arbeidet i elektrisk energiprediksjon å utvikle løsninger som kombinerer denne graden av høy kompleksitet og som samtidig har forklarbarheten til instans-baserte modeller.

Publications

Appended Papers:

Paper A Relative evaluation of regression tools for urban area electrical energy demand forecasting. Nils Jakob Johannesen, Mohan Lal Kolhe and Morten Goodwin; Journal of Cleaner Production, volume 218, pages 555-564, 2019, doi.org/10.1016/j.jclepro.2019.01.108

Paper B Smart load prediction analysis for distributed power network of Holiday Cabins in Norwegian rural area. Nils Jakob Johannesen, Mohan Lal Kolhe and Morten Goodwin; Journal of Cleaner Production, volume 266, pages 121-423, 2020, doi.org/10.1016/j.jclepro.2020.121423

Paper C Application of regression tools for load prediction in distributed network for flexible analysis. Nils Jakob Johannesen and Mohan Lal Kolhe; Chapter 4 - Flexibility in Electric Power Distribution Networks, pages 67–94, 2021, CRC Press

Paper D Comparing Recurrent Neural Networks using Principal Component Analysis for Electrical Load Predictions. Nils Jakob Johannesen, Mohan Lal Kolhe and Morten Goodwin; IEEE 6th International Conference on Smart and Sustainable Technologies (SpliTech), pages 1-6, 2021, doi=10.23919/SpliTech52315.2021.9566357

Paper E Evaluating Anomaly Detection Algorithms through different Grid scenarios using k-Nearest Neighbor, iforest and Local Outlier Factor. Nils Jakob Johannesen, Mohan Lal Kolhe and Morten Goodwin; IEEE 7th International Conference on Smart and Sustainable Technologies (SpliTech), pages 1-6, 2022, Accepted for publishing.

Thesis related papers, referenced, but not included in the thesis:

Paper F Comparison of Regression Tools for Regional Electric Load Forecasting. Nils Jakob Johannesen, Mohan Lal Kolhe and Morten Goodwin; IEEE 3rd International Conference on Smart and Sustainable Technologies (SpliTech), pages 16, 2018

Paper G Load Demand Analysis of Nordic Rural Area with Holiday Resorts for Network Capacity Planning. Nils Jakob Johannesen, Mohan Lal Kolhe and Morten Goodwin; IEEE 4th International Conference on Smart and Sustainable Technologies (SpliTech), pages 1-7, 2019, doi=10.23919/SpliTech.2019.8783029

Paper H Deregulated Electric Energy Price Forecasting in NordPool Market using Regression Techniques. Nils Jakob Johannesen, Mohan Lal Kolhe and Morten Goodwin; IEEE Sustainable Power and Energy Conference (iSPEC), pages 1932-1938, 2019, doi: 10.1109/iSPEC48194.2019.8975173.

Paper I Load prediction of rural area Nordic holiday resorts for microgrid development. Nils Jakob Johannesen, Mohan Lal Kolhe and Morten Goodwin; Chapter 8 - Residential Microgrids and Rural Electrifications, Academic Press, pages 163-181, 2022. doi.org/10.1016/B978-0-323-90177-2.00005-0

Paper J Recurrent Neural Networks for Electrical Load Forecasting to use in Demand Response. Nils Jakob Johannesen, Mohan Lal Kolhe and Morten Goodwin; Chapter 3 - Industrial Demand Response: Methods, Best practices, Case Studies, and Applications, Institution of Engineering and Technology (IET), pages 41-58, 2022. doi.org/10.1049/PBPO215E_ch3

Paper K Vertical approach Anomaly Detection using Local Outlier Factor. Nils Jakob Johannesen, Mohan Lal Kolhe and Morten Goodwin; Power Systems Cybersecurity, Springer Nature. Abstract accepted. Final submission sent to publisher, 2022.

Contents

List of Figures

List of Tables

Part I
Introduction

Chapter 1

Introduction

1.1 Scope

An analytical approach to understand and manage the increased complexity of energy systems, identifies the D's as key trends to discuss and describe energy system change [1]. What has been known as the 3D's are decarbonisation, decentralisation, and digitalisation [2].

The first D (decarbonisation) has reached considerable research traction across multiple research disciplines, as well as scales of governance. Main challenges is the mitigation of climate change amidst the steadily increasing world supply of energy [3], as shown in Fig. 1.1. It is well documented that the conventional, centralised power system, relying on large-scale fossil-fuel power plants, is causing carbonisation. Untill 2021, the energy consumption and the Greenhouse Gas (GHG) emissions has increased, except from a 6% decline during the Covid-19 [4].

The second D, decentralisation, involves changes in energy system structure, due to decarbonising policies introducing dynamic integration of renewable energy sources (RES). The deployment of RES is facilitated closer to the consumption of electrical loads and is structured in a decentralised power grid. Solar and wind farms are due to their scalability, more easily deployed in different areas and at a different scale, than conventional large-scale fossil-fuelled power plants. This allows for more players, and a transformation of players from passive consumers to prosumers, investing and benefitting from the widespread penetration of RES. Many of the tasks traditionally handled by the Transmission System Operator (TSO) is growing responsibilities for the Distribution System Operators (DSO). Due to the variability in power generation of RES the emerging additional roles of the DSO includes techno-economical challenges involving peak load management through Distributed Energy Resources (DER), network congestion management, providing reactive power support to TSOs, procure voltage support and technical validation for the power market [5].

The third D, digitalisation, has become a key enabler for better, cheaper and faster monitoring to aid network and congestion management through changed consumer behaviour.

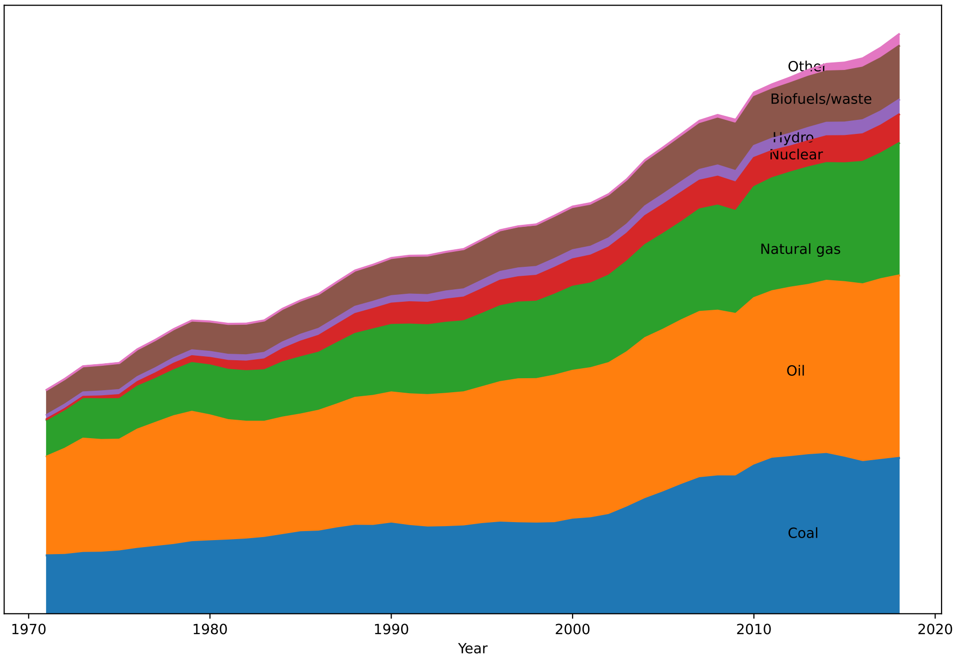


Figure 1.1: The worlds total energy supply by source, from 1971-2018

The deployment of smart meters has served an unprecedented level of data granularity as well as facilitating real-time monitoring of the power grid with enabled two-way communication levelling the playing field for the actors developing a 'smart' energy system [6].

In the so called Business As Usual scenarios (BAUs), where nothing is being done to mitigate the GHG-emissions and control the growing energy demand, electricity demand will be tripled by 2050.

For larger areas, such as a country, region or urban area, the electrical load demand varies typically between 6000 MW (Sydney region), see Fig 1.4 and up to 80 GW (a sample European country), see Fig. 1.2. For larger areas, with extensive data, the patterns in the electrical load consumption is identified through visual inspection, as in Fig 1.2; depicting the electrical load consumption for a week in a big European country. Where load schemes are made based on the quality of consumption. Apparatuses that needs continuous energy demand, is identified as base load. Whereas power that are driven occasionally, represents the peak in the energy consumption and is identified as peak load. In between is the intermediate load. It is necessary for the operators of the power grid to anticipate these load structures through electrical load demand forecasting. Fig. 1.3 shows the development of energy load demand for holiday cabins, in the years from 1993 to 2016.

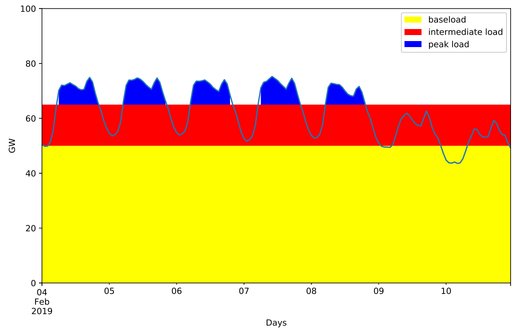


Figure 1.2: The electric load demand curve of a sample European country for one week, indicating level of load curves. Source: ENTSO-E

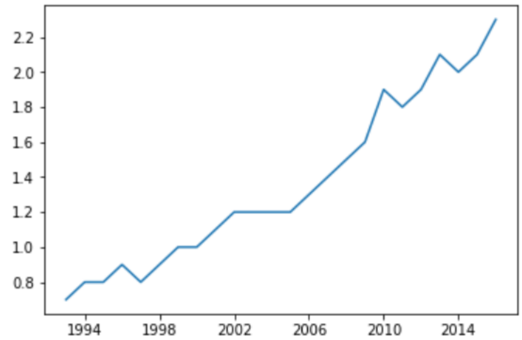


Figure 1.3: Energy demand for cabins and holiday apartments in Norway from 1993-2016, Statistics Norway (2018).

1.2 Background and Motivation

Electricity is regarded a reliable commodity, and all our activities connected to the use of electricity is metered and stored. The need for applicable tools to manage and make use of the complex data gathering achieved through the smart grid, has gained traction in the research community. Big data will play an important role in supplying the world with sufficient energy.

A key solution for a more energy efficient system is smart electrification, making use of a new agent topology introduced in the smart electrical power network, smart grid. The smart grid will manage and operate production, distribution and transmission of electricity by two-ways communication between the producers and end-users. The big data processing and real-time monitoring of the energy system will require flexible tools to manage the smart energy system, by using Information and Communication Technologies (ICT), Distributed Generation (DG) and Artificial Intelligence (AI), together [7][8].

It is needed to have an AI tool based in Machine Learning (ML) to process and correlate the meaningful relation by finding structures and patterns in the considered data. When presented with new data the machine can learn to perform a task without the need of reprogramming [9]. ML can provide electrical load demand forecasting, giving information about future loads. Load forecasting is the most fundamental application in Smart-Grid, which provides essential input to other applications such as Demand Response, Topology Optimization and Anomaly Detection, facilitating the integration of intermittent clean energy sources. Demand side management (DSM) is an umbrella term that describes the utility company efforts to improve energy consumption at customer site, see Fig. 1.5. Demand response (DR) is the customers adaptation to alter their normal electricity usage in response to adjusted electricity prices with grid constraints or other incentives created to decrease energy consumption at times of shortage or when system reliability is at risk.

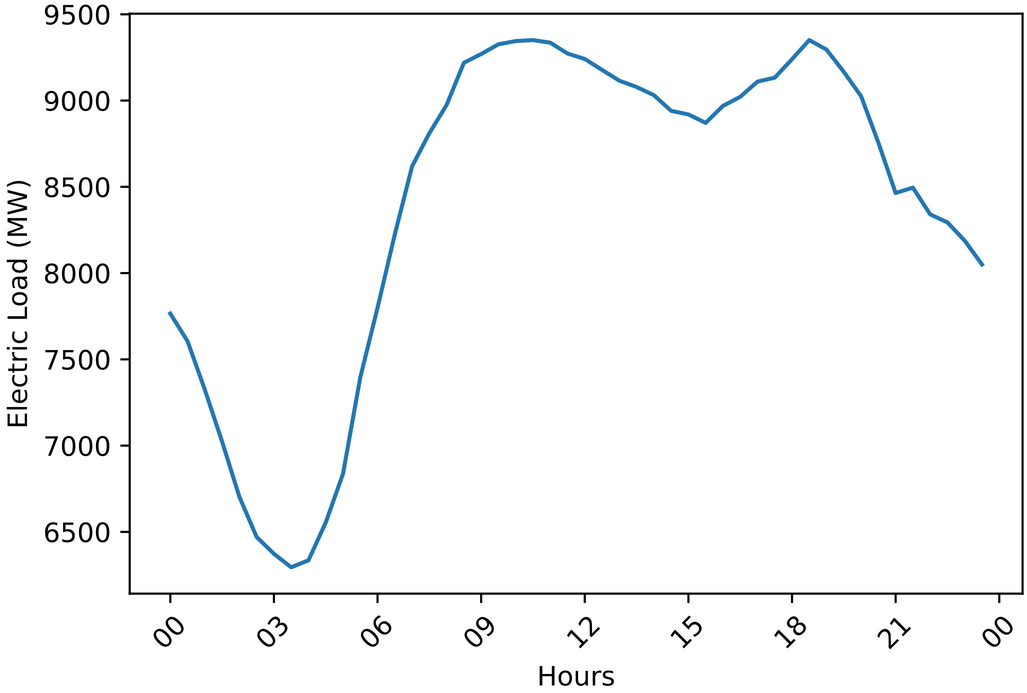


Figure 1.4: Electrical load profile for 24 hours for Sydney Urban Area

Main focus in the field is on load analysis and finding applicable techniques to enhance load forecasting, as well as the various technical solutions for data driven models [10]. To enhance the load monitoring a variety of effective compression techniques of electrical energy data is discussed concerning efficient processing, transmission and storage of data [11]. The amount of such mentioned data is directly related to the choice of algorithms and hence the predictive performance. The accuracy of the algorithms can generally be traced as a function of the system level. For the highest level (national/regional) normally containing big amounts of data Linear Models have been successful, at a more intermediate level, when predicting for cities at Smart City/Smart grid level, hybrid neural network based models are preferred and at the lowest level with least concistency and lowest amount of data at the microgrid level (residential and rural area, islanding cases), hybrid ANN based models are preferred [12]. Traditional modeling can not handle the information flows from all the vectors in the power grid. These tasks includes load forecasting, anomaly detection, stability assessment of the power grid, fault detection and issues with security problems such as cyber attacks [13].

AI is the umbrella term, and ML being a subdivision and deep learning a subdivision of ML. Deep Mining and Statistics appear on the outskirts, both as being used alone or as a part of a hybrid algorithm or in the preprocessing of data. This categorisation is also on the more algorithmic level divided into supervised and unsupervised learning, where the first comprises classification, and regression, and the latter clustering and dimensionality reduction. Deep Learning are both classified as a part of supervised learning, but in other sources as an entity alone alongside supervised/unsupervised learning [12][13].

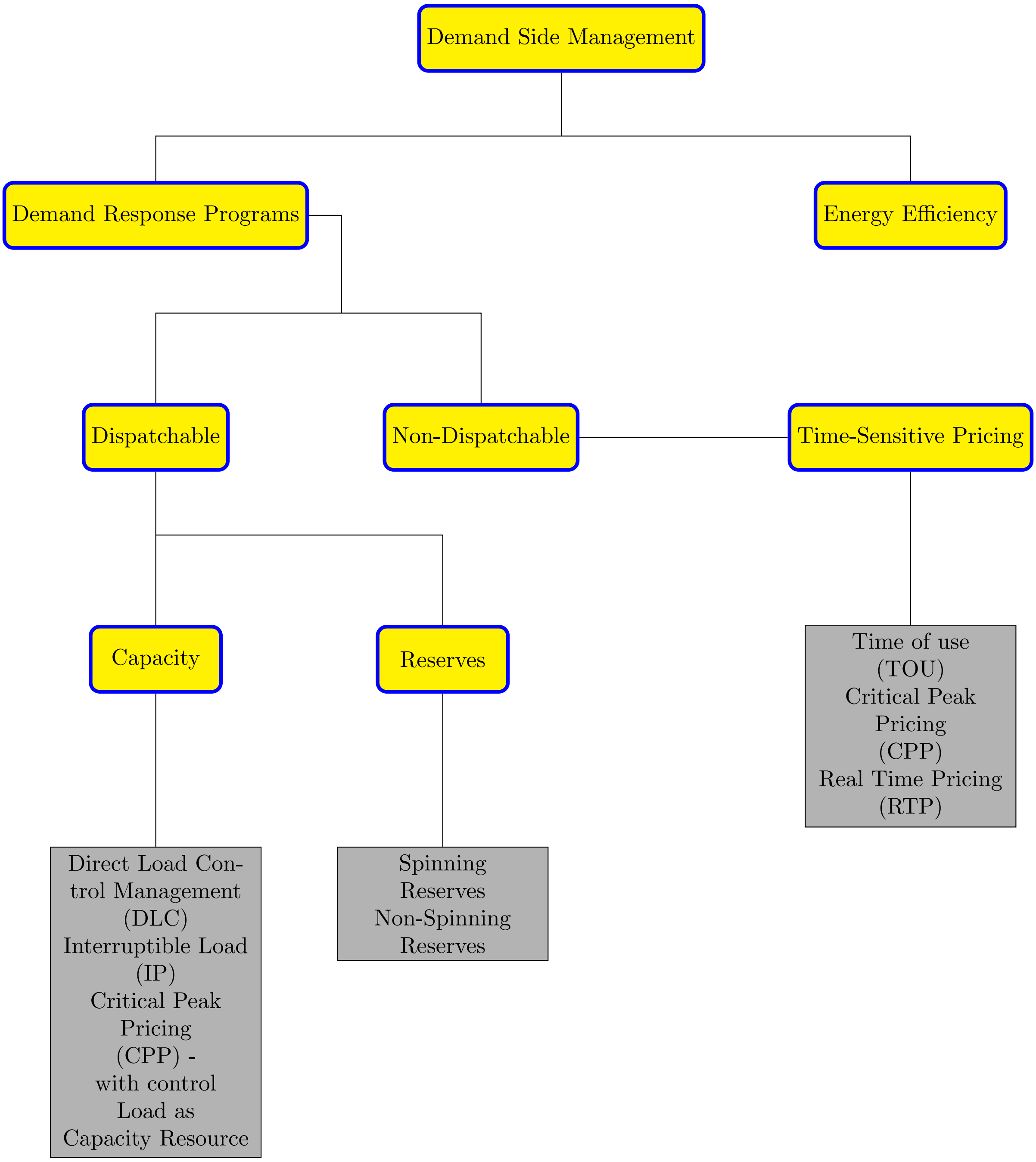


Figure 1.5: Overview of DSM and representative incentives in DR programs, from authors chapter in upcoming book; Industrial Demand Response Methods, best practices, case studies, and applications

The implementation of renewable energy sources is a concern for the operation of the smart grid due to their nature of intermittent behaviour [14]. Remote and rural areas will be needing alternative energy sources and Hybrid Renewable Energy System involves the planning of multiple energy resources (Solar, Wind, Fuel Cell, Electrolyser) in combination with Energy storage (Battery Bank, Hydrogen) [15].

The uncertainty and variability of RES implementation needs a more flexible system, and research on an optimal solution that considers both the supply and demand is adding smartness to the operation of the power grid and yields better results to electricity consumers, providers and generators [16]. In this focus on demand-side and supply-side forecasting, there is also an overlap between the two [17]. The internet of things (IoT) solutions paves way for forecasting on real-time prerequisites, and facilitates equipment that responds automatically and digitally to immediate shift in electrical demand [18].

1.3 Research Gap

The electrical energy distribution network is integrated with intelligent devices for monitoring and controlling the operation of renewable energy sources and energy efficient operation (e.g. demand side management, etc.). The intelligent devices are providing real-time continuous data for making the energy system smart. The collected real-time data needs to be analysed for predictions of energy consumption and renewable energy productions. There is a great research gap related to renewable-based distribution network planning from a flexibility point of view. In the literature there are many methods on energy prediction, but most of them have used continuous time approach together with complex neural networks which requires huge amount of data. These methods have not significantly considered event-based demand prediction with impact of external parameters (e.g meteorological parameters, etc.). When these methods have considered the seasonal impact of the external parameters, it has mainly been done modifying the continuous time algorithm, and not adequately investigated the potential of the pre-processing stage. To bridge the research gap deep domain knowledge needs to be applied together with algorithmic development.

The event-based demand needs more accurate prediction using less amount of available data and the regression techniques need to be systematically investigated for electrical energy prediction and correlating other impacting parameters. There has not been systematic load synthesis and prediction analysis of rural areas seasonal occupancy which requires for network expansion planning and integrated renewable energy sources. Also, the urban area electrical energy demand prediction with weather parameters have not been sufficiently investigated using regression techniques for energy management. Electrical load demand forecasting is useful for energy management operation and can be used for network topology optimisation and anomaly detection. In the literature, electrical demand forecasting is mainly used for operation and planning, it has also the potential for detecting abnormal operation due to various instances (e.g. random effect, intrusion, abnormal operation of smart devices, cyber-threats). The use of load prediction in flexibility of operation of distributed networks has not been investigated sufficiently and there is a research gap on use of flexibility with renewable-based distribution network.

1.4 Research Objectives

Based on identified research gaps through state-of-the-art literature review, following are the key objectives of the thesis:

- Is it possible to improve the predictive performance of regression tools for machine learning algorithms by different time organisation?
- Does the implementation of aspects from Statistics and Time Series Analysis aid the predictive performance of machine learning algorithms for electrical load demand forecasting?
- Will a in general simpler algorithm be able to compete with more complex algorithms?
- How can the use of regression tools bridge the the research gap between statistical models and machine learning models?

1.5 Research Plan and Thesis Outline

The research work considers event-based demand prediction with impact of external parameters (e.g meteorological parameters, etc.). In order to bridge the research gap, deep domain knowledge needs to be applied together with algorithmic development. In this work the event-based demand prediction, using less amount of available data and the regression techniques are systematically investigated. This work focuses on 3 topics: (i) Urban Area Load Forecasting, (ii) Network Capacity Planning for Rural Area Networks applying micro-grid operations, and (iii) Electricity Markets. This work includes case studies with respective data of different types, size and granularity. It is for a rural area and an urban area electric energy load demand as well as Nordic electricity spot price.

This dissertation is composed of two parts. **Part I** summarizes the research carried out throughout the PhD and a presentation of the main findings. **Part II** contains the collection of two journal papers, one academic book chapter, and two conference papers representing the main contribution of this thesis. The organisation of paper contributions is given in Fig. 1.6.

The remaining Chapters in **Part I** are:

- Chapter 2 positions this work within the state of the art literature and identifies the research gaps that this thesis addresses.
- Chapter 3 presents the theory and mathematical modelling.
- Chapter 4 presents this works methodology.
- Chapter 5 presents the data.

- Chapter 6 gives the main results of the evaluation and discusses them.
- Chapter 7 concludes the dissertation, presents the research limitations, and points to future research directions.

The electrical load forecasting has been carried out using conventional mathematical techniques and the traditional forecasting techniques are based on linear regression series. Typical for electric load forecasting is seasonal and diurnal changes. The seasonal behavior is mainly influenced by external parameters, such as weather. An evaluation of 3 different regression techniques and the novel vertical time approach method to deal with seasonal variations, developed on urban area data, are presented in the attached manuscript **Paper A**. The presented and discussed results show that vertical time approach through appropriate data pre-processing considering cross-correlations to external weather parameters can performatively compete with complex neural network architecture.

A time-series is a collected sequence of events. Basically, based on the assumption of an inherent structure, such as autocorrelation, trend, and seasonal behavior. There are many different scenarios of how these sequences of events are collected and described. In the attached manuscript **Paper B** are presented rural area data, which is of different granularity and size than the urban area data. Since the rural area data has fewer end-users the structure and patterns are not as apparent. The method presented investigates classical time series observations and techniques, and uses a persisting autoregression to give finite gradient information to the model, improving its performance.

The electrical load prediction is necessary for distributed network energy management and finding opportunity for flexibility in shifting the operation of non-critical power intensive loads. The flexibility of resources defined by their dynamic capabilities such as ramp time, start-up/shut-down time, operating range (minimum and maximum operating level) as well as minimum up and down times of the energy generation system. Both urban area and rural area data are presented with flexible load analysis in the attached manuscript **Paper C**.

In **Paper D** is analysed how the dimensionality reduction through orthogonal principal components aids the predictive performance of the most used instances of recurrent neural networks (RNN), Vanilla RNN, Gated Recurrent Units (GRU) and Long short-term Memory (LSTM). Further, the 3 RNN's are compared and evaluated for theire applications in electrical load demand forecasting.

An anomaly is defined as a deviation from an established normal pattern. These systems rely on deep domain expertise. Cyber threats could affect the ancillary services that are being delivered from the aggregators, which might lead to stability and security issues resulting in brownout or massive blackouts. In **Paper E** different anomaly detection algorithms are evaluated and analysed.

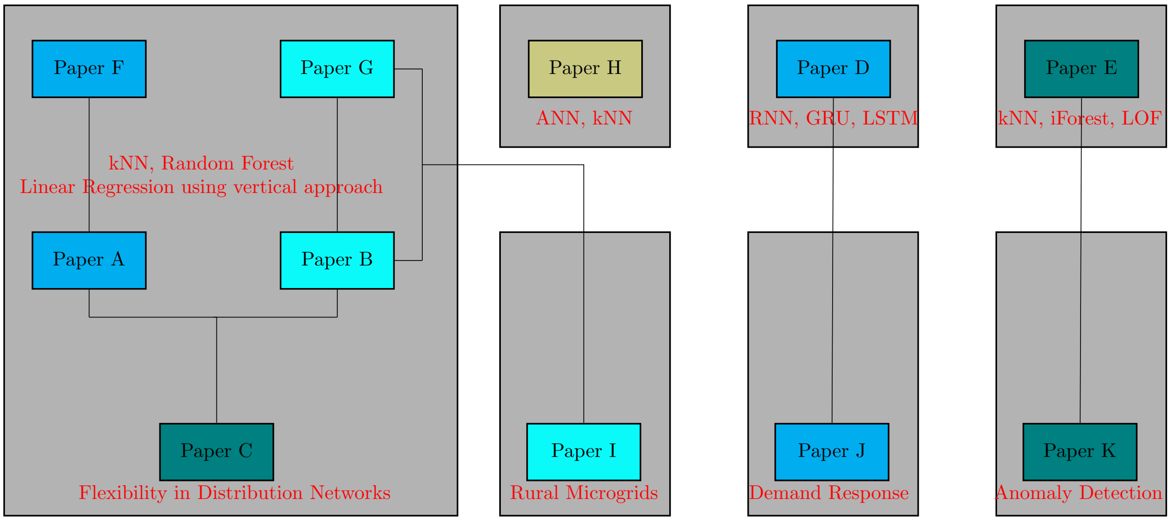


Figure 1.6: Organisations of contributions: The grey boxes defines a research area. Red text the applied techniques. Cyan color are papers (A, D, F and J) based on urban area data. The cyanish color are the papers (B, G and I) based on rural area data. In color teal are the papers (C, E and K) combining urban and rural data. In the color ochre is paper (H) using Nordpool data.

1.6 Description of Publications

Key contributions from research publications are given below. The overview of research publications with key objectives and how they interrelate is illustrated in Fig 1.7

1.6.1 Publication A (Journal Level II)

N. J. Johannesen, M. Kolhe and M. Goodwin, 'Relative Evaluation of Regression Tools for Urban Area Electrical Energy Demand Forecasting', Journal of Cleaner Production, volume 218, pages 555-564, 2019, doi.org/10.1016/j.jclepro.2019.01.108

In Publication A the performance of regression techniques on urban electrical load demand data and the impact of external parameters is evaluated. Load data was collected from Australian Energy Market Operator (AEMO) and weather data from Bureau of Meteorology (BOM). The entire dataset was used to validate findings of k-fold cross-validation, where one fold equals one year. The findings of the paper supported the initial hypothesis that external parameters and vertical time organization improved the predictions both for 30 minutes and day-ahead forecasting. k-Nearest Neighbour (kNN) is found to be the best predictor on day ahead forecasting and Random Forest regression performs best for 30 minutes forecasting. Results of this work is going to be useful for predicting the short term 30 minutes electrical energy using vertical approach and considering Random Forest Regression Tool. For long term prediction of 24 hours kNN Regression Tool can provide better results using vertical approach. It is also important to consider further investigations of the impact of various weather parameters on load prediction.

1.6.2 Publication B (Journal Level II)

N. J. Johannesen, M. Kolhe, M. Goodwin, 'Smart load prediction analysis for distributed power network of Holiday Cabins in Norwegian rural area', Journal of Cleaner Production, volume 266, pages 121-423, 2020, doi.org/10.1016/j.jclepro.2020.121423

The selected rural area power network is used for Holiday Cabins and there is a potential for integrating solar photovoltaic system with energy storage. The selected Bjønntjønn Cabin Area is a typical rural area with low capacity power network in the south-east part of Norway. The demand of Bjønntjønn Cabin Area from 2014 to 2018, shows a peak demand in typically holiday winter seasons, and low load during summer time, where temperature is higher, and evenings are brighter and thus less time for indoor activities. External parameters data (from Norwegian Institute of Bioeconomy Research (NIBIO)) with weather information from the 3 closest meteorological stations to Bjønntjønn Cabin Area (Bø, Gvarv and Gjerpen) are picked for doing correlation analysis. Through correlation analysis, the highest correlating weather station is found. Most of the pattern that constitutes the electric load profile is dependent on individual user behavior. The individual human activities are not enough to make the substantial patterns on its own accord, yet together with the influence of the changing weather, the impact is growing, and it is an important component of feature engineering in load forecasting. To improve feature engineering with the relative smaller amounts of data, than that of an urban area, ideas from time series analysis through autocorrelation where serial dependence, shows how a time-series is related to its own lagged version. Leave-one out method, is compared to crogging, a method aimed at preserving the temporal dependencies of a time series.

1.6.3 Publication C (Academic Book Chapter)

Johannesen, N. J., & Kolhe, M. L. (2021). Application of Regression Tools for Load Prediction in Distributed Network for Flexible Analysis. In Flexibility in Electric Power Distribution Networks (pp. 67-94). CRC Press.

The methodology, used in this chapter, dealing with the problems of irregularities and randomness in the time series considering urban and rural area case studies. Random forest-regressor yields good results on hourly time prediction in load forecasting. The kNN shows precise prediction due to its capability to capture the nearest step in a time series based on the nearest neighbor principle. The presented vertical time approach uses seasonal data for training and inference, as opposed to continuous time approach that utilizes all data in a continuum from the start of the dataset until the time used for inference. The regression tools can handle the low amount of data, and the prediction accuracy matches with other techniques.

1.6.4 Publication D (In proceedings)

N. J. Johannesen, M. L. Kolhe and M. Goodwin, "Comparing Recurrent Neural Networks using Principal Component Analysis for Electrical Load Predictions," 2021 6th International Conference on Smart and Sustainable Technologies (SpliTech), 2021, pp. 1-6, doi: 10.23919/SpliTech52315.2021.9566357.

In the paper principal component analysis (PCA) is used and compared on 3 basic recurrent neural networks; VanillaRNN, Gated Recurrent Unites (GRU) and Long short-term memory networks (LSTM). The Vanilla is the simplest RNN, using hidden states where the output from the previous time step is being fed to the next time step. GRU introduces a higher complexity from the Vanilla, introducing forget gates. LSTM is the highest complexity of the compared recurrent neural networks, using a memory cell together with the hidden state, to evaluate long term state dependencies. PCA reduces and extracts the main components of available data. This work shows that PCA improves the performance of RNNs with use of weather parameters. The historical electrical load dataset from Sydney region is used to test the load forecasting using these techniques considering meteorological parameters. Through load forecasting, it is observed that for the 30 minutes predictions, the simpler RNNs (Vanilla and GRU) has the best overall performance, with GRU trained with a reduced number of principal components performs best for a typical period with a mean absolute percentage error (MAPE) of 0.74%.

1.6.5 Publication E (In proceedings)

N. J. Johannesen, M. L. Kolhe and M. Goodwin, "Evaluating Anomaly Detection Algorithms through different Grid scenarios using k-Nearest Neighbor, iforest and Local Outlier Factor," 2022 7th International Conference on Smart and Sustainable Technologies 2022 (SpliTech 2022), 2022, pp. 1-6, Accepted for publishing.

Detection of anomalies based on smart meter data is crucial to identify potential risks and unusual events at an early stage. The available advanced information and communicating platform and computational capability renders smart grid prone to attacks with extreme social, financial and physical effects. The smart network enables energy management of smart appliances contributing support for ancillary services. Cyber threats could affect operation of smart appliances and hence the ancillary services, which might lead to stability and security issues. In this work is given an overview of different methods used in anomaly detection, and evaluates the performance of 3 models, the k-Nearest Neighbor, local outlier factor and isolated forest on recorded smart meter data from urban area and rural region.

1.6.6 Publication F (In proceedings)

N. J. Johannesen, M. Kolhe and M. Goodwin, "Comparison of Regression Tools for Regional Electric Load Forecasting," 2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech), 2018, pp. 1-6.

Urban area electrical load forecasting is important for power generation capacity planning and also to integrate environment friendly energy sources at district level. Load predictions will help in developing demand side management in coordination with renewable power generation. Urban area load is influenced by meteorological conditions therefore it is important to include weather parameters for load predictions. Machine Learning algorithms can effectively contribute for electrical load predictions. The most commonly used algorithm for load prediction is Artificial Neural Network (ANN), which is a complex predictor that utilizes a big amount of training data. k-Nearest Neighbour (kNN) has proven to be efficient by the introduction of binary dummy variables for categorisation and it can be used for short term (30 min) and long term (24 hours) load forecasting. This paper explores the use of regression tool for regional electric load forecasting by correlating lower distinctive categorical level (season and day of the week) and weather parameters. The historical electrical load datasets with meteorological parameters are available for the Sydney region and they have been used to test the regression tools. Data correlation over seasonal changes have been argued by means of improving Mean Absolute Percentage Error (MAPE). By examining the structure of various regressors they are compared for the lowest MAPE. The regressors show good MAPE for short term (30 min) prediction and Random Forest Regressor scores best in the range of 1-2 % MAPE.

1.6.7 Publication G (In proceedings)

N. J. Johannesen, M. Lal Kolhe and M. Goodwin, ”Load Demand Analysis of Nordic Rural Area with Holiday Resorts for Network Capacity Planning,” 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech), 2019, pp. 1-7, doi: 10.23919/SpliTech.2019.8783029.

In this paper Artificial Neural Networks were first evaluated and benchmarked to K-NN the first journal paper where the vertical approach was introduced, to enhance the performance. In order to deal with the relative smaller amounts of data ideas from time series analysis was used with an alternative version of autoregression to help the algorithm searching the curvature and give a finite gradient based on the latest updates from the targeted vector, in this case the load.

1.6.8 Publication H (In proceedings)

N. J. Johannesen, M. Kolhe and M. Goodwin, ”Deregulated Electric Energy Price Forecasting in NordPool Market using Regression Techniques,” 2019 IEEE Sustainable Power and Energy Conference (iSPEC), 2019, pp. 1932-1938, doi: 10.1109/iSPEC48194.2019.8975173.

In this the electric energy price forecasting of two regions of NordPool market has been compared and analysed with reference to energy generation scenarios two regions, as well as to combine forecast models. Due to the dynamics of the electricity markets and region divisions this work proposes electric energy price forecasting using regression tools based in the k-Nearest Neighbor (kNN) regressor, to capture the small increments in changing price behavior, and an autoregressor on top to capture the finite gradient of the occasional spikes in the price cycles. A hybrid model, combining kNN-regressor and autoregressor is used to improve the prediction accuracy of electric price forecasting. The NordPool market trading behavior and the unanticipated price peaks at daily, weekly and annual level, show improved prediction accuracy, when enhancing the model from a simple kNN-regressor to the hybrid. The kNN-regressor captures the small increments in changing price behavior, and the autoregressor captures spikes in the price cycles, as it is described in the results especially considering the non-controllable Danish electricity generation system, where the MAPE improves from 27.52 without autoregressor to 18.60 MAPE with the autoregressor.

1.6.9 Publication I (Academic Book Chapter)

N. J. Johannesen, M. Kolhe, M. Goodwin, Power and energy management for microgrids in residential systems and rural electrifications. In Residential Microgrids and Rural Electrifications. © Academic Press 2021 Published: 26th November 2021, Paperback ISBN: 9780323901772

Rural electrifications scenarios based on load analysis. The performance of a rural area distribution power network can be improved by operating it as a microgrid with integration of energy storage, RES, and distributed generators. The microgrid is a complex system encompassing various subsystems at various stages of aggregation. It accommodates multidirectional power and information flows among all the vectors (e.g., power generation, transmission, and distribution system operators; distributed intermittent RES; demand response aggregations; end users). The load forecasting within the distributed network requires good insight into user behavior, geographical location, and algorithm assessment. Residential and rural energy networks have significantly lower load demand in comparison to urban systems. This chapter focus on how to appropriately fit a suitable predictor for a typical rural load system prediction and analysis.

1.6.10 Publication J (Academic Book Chapter)

N. J. Johannesen, M. Kolhe, M. Goodwin, Recurrent Neural Networks for Electrical Load Forecasting to use in Demand Response Industrial Demand Response: Methods, Best practices, Case Studies, and Applications. Institution of Engineering and Technology (IET), pages 41-58, 2022. doi.org/10.1049/PBPO215E_ch3

Electric load forecasting is a fundamental technique to understand end-user behaviour and therefore a crucial factor in the design of demand response programs. Load forecasting will also identify the appropriate design of demand response programs. In this Chapter a range of different machine learning applications are covered to represent the influential factors for electrical load demand forecast in a DR context, with a variety of different data scenarios, temporal and technical scenario. This Chapter explores and compares the load prediction analysis through basic Recurrent Neural Networks (RNNs); Vanilla RNN, Gated Recurrent Units (GRU), and Long Short-Term Memory (LSTM), using Principal Component Analysis (PCA). It is found that PCA can be used to reduce the number of principal components for Vanilla RNN, GRU and LSTM networks. Reducing the number of principal components using PCA is one of the techniques that is used in dimensionality reduction. Reduction in dimensionality will relieve the computational burden. In this work the dimensionality reduction improves the predictive output. It is observed that for electric load demand forecasting, the preferred technique is Gated Recurrent Units, trained with a principal components. The performance is evaluated through mean absolute percentage error (MAPE), which is relatively lower than other techniques.

1.6.11 Publication K (Academic Book Chapter)

N. J. Johannesen, M. Kolhe, M. Goodwin, 'Vertical approach Anomaly Detection using Local Outlier Factor', Power Systems Cybersecurity. Springer Nature. Abstract accepted. Final submission sent to publisher.

Detection of anomalies based on smart meter data is crucial to identify potential risks and unusual events at an early stage. In addition anomaly detection can be used as a tool to detect unwanted outliers, caused by operational failures and technical faults, for the pre-processing of data for machine learning, to detect concept drift as well as enhancing cyber-security in smart electrical grid operations. It is known that anomalies are defined through their contextual appearance. Hence, anomalies are divided into point, conceptual and contextual anomalies. In this work the contextual anomaly detection is examined, through a novel type of load forecasting known as vertical approach. This chapter explores the use of anomaly detection in the relevant learning systems for machine learning in smart electrical grid operation and management through data from New South Wales region in Australia. The presented vertical time approach uses seasonal data for training and inference, as opposed to continuous time approach that utilizes all data in a continuum from the start of the dataset until the time used for inference. It is observed that Local Outlier Factor identifies different local outliers given different vertical approaches. In addition, the local outlier factor score vary vertically. An anomaly is defined as a deviation from an established normal pattern. Spotting an anomaly depends on the ability to defy what is normal. Anomaly detection systems aim at finding these anomalies. Anomaly detection systems are in high demand, despite the fact that there is no clear validation approach. These systems rely on deep domain expertise.

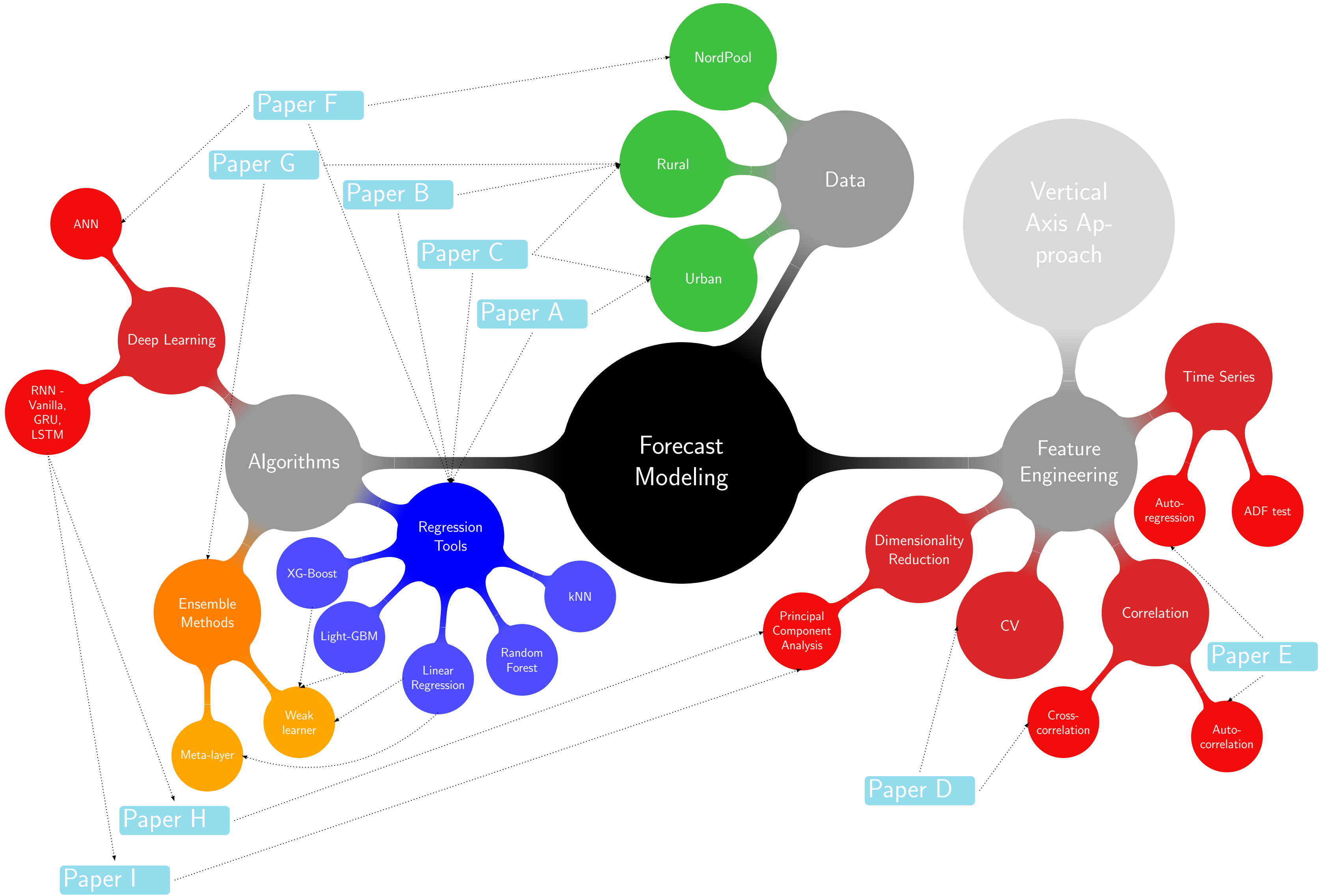


Figure 1.7: Thesis overview

Chapter 2

Literature Review - State of The Art Forecasting Techniques

2.0.1 Introduction

Reliable forecasting is a necessary part of decision identification, information gathering, and assessing solutions in power networks. The principal objectives of forecasting is to provide basic scheduling of power loads. Issues raised in the literature on electrical load demand forecasting and price prediction on electricity markets in the 1980's are still challenges for the industry [19] [20]. Main recent review articles in the field focus on the analysis of load and finding applicable techniques to enhance load forecasting. The authors of review [10] focuses on the various technical solutions for data driven models and aims to categorize and schematically give an overview of the different algorithms, and debates solutions in data driven AI solutions. To enhance the load monitoring a variety of effective compression techniques of electrical energy data is discussed concerning efficient processing, transmission and storage of data [11].

The amount of such mentioned data is directly related to the choice of algorithms and hence the predictive performance. The accuracy of the algorithms can generally be traced as a function of the system level. For the highest level (national/regional) normally containing big amounts of data Linear Models have been successful, at a more intermediate level, when predicting for cities at Smart City/Smart grid level, hybrid neural network based models are preferred and at the lowest level with least consistency and lowest amount of data at the microgrid level (residential and rural area, islanding cases), hybrid ANN based models are preferred [12] Common for several of the review articles that there is a hierarchical categorisation of the the field in general, where Artificial Intelligence is the umbrella term, and machine learning being a subdivision and deep learning a subdivision of ML, and where Deep Mining and Statistics appear on the outskirts both as being used alone or as a part of a hybrided algorithm or in the preprocessing of data. This categorisation is also on the more algorithmic level divided into supervised and unsupervised learning, where the first comprises classification, and regression, and the latter clustering and dimensionality reduction. And on the side with a bit of both is Reinforcement Learning. Deep Learning are both classified as a part of supervised learning but in other as an entity alone alongside supervised/unsupervised learning [12][13].

To enhance the performance of load forecasting models all influential parameters must be considered comprehensively, to search for the influential parameters on the load profile [21]. The small area of Tunis (with only installed capacity of 4425 MW) is considered for analysis of load prediction with seasonal variations [22]. A variation in load due to season is only once a year during heat wave in the summer. For training set they have used horizontal time-series approach, where almost 10 months (more than 14400 datapoints) of training was used for testing on one week. According to Lahouar and Slama (2015) [22], who used random forest for day-ahead load forecast for the Tunisian market with historical data from 2009-2014, they obtained an average MAPE of 2.24% when crediting for the next 24 hours. Presented method of [22] does not improve, when predicting for the heat wave season, as the average MAPE for heat wave period (7-13 July) has increased to 2.6899%. During the Arabic spring in Tunis 2011, Tunis experienced a random effect caused by a much lower power demand during the Tunisian Revolution, the MAPE for some 24 hour intervals of prediction as high as 19.61%. It was even worse during the Blackout of August 2014 where the MAPE rose to 398.09%. This show the machine learning algorithms inabilities in forecasting rare events. [22] also makes a comparison with ANN, and for the testing period of 7-13 of July it scores 2.9140 MAPE. They state that the main advantage with Random Forest over other methods is that there are few hyper-parameters to set and generalize by saying default settings is normally enough to compete with ANN and Support Vector Machine (SVM)/Support Vector Regressor (SVR), which accuracy depends on the tuning of their hyper-parameters. In our work we have used the experiences from Tunis to understand the random effects and their input on electrical energy demand forecasting as well as the understanding of hyper-parameters.

Jinkyu and Sup (2015) [23] recognizes artificial intelligence techniques like ANN or Kalman filter, to show promising results in the load forecasting predictions, although the hidden structures in AI might limit the understanding of the complex spatiotemporal developments in correlation between meteorological conditions and electricity demand. Electrical load demand and the temperature effects have been studied and short term load forecast needs to take temperature effects into consideration for day-ahead predictions. In the very short time load forecasting the time scale is to short for the temperature to have any effect, and in the long run the effect tend to even out [24, 25]. On the load forecasting for the UK electricity demand Al-Qahtani and Crone (2013), proposes a multivariate k-NN approach that, opposed to the univariate model that does not take into account the underlying sub-categories of the calender, create a binary dummy variable where dt = 1 for all nonworking days and dt = 0 for working days. The load forecast MAPE of both univariate and multivariate show improved results by the use of dummy variables. A MAPE of 2.3284 was found using the univariate model, and a 1.8133 was found using the multivariate model. The dataset contained data for more then 7 years (2001-2008). The complete year of 2004 was used for training and 2005 used for validation [26]. Based on their research we developed the relevance of doing multiple correlation analysis with different time factors, where we can observe that meteorological parameters increase their importance on the prediction output as time window increases. In this context we regarded the work of Afkhami and Yazdi who proposes a way to quantisize the day into 3 periods of 8 hours for neural networks to enhance their performance [27]. Local Interpretable Model-Agnostic Explanations (LIME) aims to reflect the behavior in proximity to the predicted outcome, and does so by offering a interpretation that can explain doubts about the model. By explaining here means to provide some mean of qualitative understanding in the relation between a decision making and the predictive outcome. In medical diagnosis LIME highlights what features in the dataset that led to the prediction, and what was evidential against it [28]. ANN studies have shown an MAPE of 1.9, resulting in a Mean Absolute Error (MAE) of 167.91 MW, based on training data for a whole year. The research includes studies of temperature effects and introduces two threshold values where the load and temperature exhibits close correlation, at below 10 degrees Celsius due to heating, and above 23 degrees because of cooling [29].

2.0.2 Forecasting in Distributed Networks

The rural area distribution network performance can be improved by operating it as a micro-grid with integration of energy storage, renewable energy sources and distributed generators. The smart micro-grid (i.e. smart distributed network) is a complex system encompassing of various sub-systems at various stages of aggregation. Smart micro-grid is going to accommodate multi-directional power flow to go together with multi-directional information flows between all the vectors (e.g. power generations, transmission and distribution system operators, distributed intermittent renewable energy sources, demand response aggregations, end-users, etc.). Over the past decade the power system is changing from centralized grid to more decentralized and its operational management is going to be real-time monitored smart and micro-grids [30]. Reference [31] has reviewed energy technologies for application in smart distributed network using IOT technologies, various different types of solar technologies has been reviewed in the same paper and discusses control strategies PV's and hybrid energy systems. For effective operation of micro-grid and demand side management, the load prediction analysis with impact of external parameters is required.

Machine learning algorithms can be electively used for electrical energy demand as well as predicting the output from the renewable energy sources. It is important to do the prediction of future load consumption to balance the electrical energy supply and demand [32]. Existing research into micro-grid electric energy load demand forecasting is scarce. The majority of the existing research selected micro-grids of large power scale with electric energy load demand ranging from 10 MW scale, to larger ones at 1000 MW. The GW-scale which is the size of a medium city and forecasting results from such a large scale micro-grid is comparable to urban area load forecasting. Hence the smaller scale micro-grid is more difficult to predict due to higher load fluctuations and randomness. At a smaller scale the load fluctuations within the same time period may be higher than for bigger more stable load. A comprehensive study compares small and large scale micro-grids in China. The Chinese case study uses five different scale of micro-grid where the two smallest micro-grids have subsequently maximum load of 273 and 463.8 kW. To efficiently predict the electric energy load demand for these micro-grids they propose to use different hybrid forecasting models based on Empirical Mode Decomposition (EMD), Extended Kalman Filter (EKF), Extreme Learning Machine with Kernel (KELM) and Particle Swarm Optimization (PSO). For the small scale micro-grid the hybrid models achieves acceptable MAPE of 7 to 10 % [33]. Existing research on network capacity planning deal with much larger data samples. The term Big Data is a relative concept and not an absolute definition, at best it is ambiguous and to quantify dataset is a difficult task as the capacity and computational power is continuously increasing. Typical Big Data is regarded as that quantification of collected data in different sampling rates is in the Terabyte (TB) area [34] [35].

State of the art research in electrical load demand forecasting focuses on three main aspects in order to make sound predictions. These inputs are from weather parameters, holidays and time of day. The mentioned relations has been found equally important both for simpler instance based machine learning models to the more complex black box neural networks [36]. And the results of this are provided in the research for short [37] [38], mid-term [39], as well as long-term forecasting [40]. The impact of external weather parameters has proven also to be important for forecasting on limited data, such as for households and buildings [41], as well as cabin areas [42]. Hybrid forecast combining neural networks with autoregression has proven to aid in tracing the curvature of the peak in the volatile electricity markets [43].

In short-term electric load demand forecasting, Recurrent Neural Networks (RNN) by Levenberg-Marquardt and Bayesian regularization on 30 minutes predictions had achieved a mean absolute percentage error (MAPE) of an average in one week 1.4792 [44]. One hour ahead prediction, has been performed on hourly power consumption in Toronto Canada using Long Short-Term Memory (LSTM), achieving a MAPE of 2.639, which was an improvement of the Vanilla RNN of 3.712 MAPE [45]. The Resnetplus model for the ISO-NE dataset proposed a day-ahead load forecasting model based on deep residual networks. A basic structure of several fully connected layers to produce preliminary forecasts of 24 hours. A forecast is then made on the residuals of the preliminary forecast provided with a formulation of Monte Carlo dropout for probabilistic forecasting, achieving an average MAPE of 1.447 [46]. Gated Recurrent Unit (GRU) was used to predict the electricity market in Singapore. Multi-features input models of different time structural architecture named Multi-GRU has been used to give 30 minutes predictions [47].

2.0.3 Detection Methods for Abnormal Operation

Anomaly detection is done on any time series data. Various anomalies can be detected in historic time series data, due to human error, false meter measurement, inaccuracies in data processing and failure of delivery due to extreme weather or other failures. A two-stage method is proposed in reference [48] combining two probabilistic anomaly detection approaches for identifying anomalies in time series data of natural gas. Exogenous variables are known to influence the electrical load consumption, and loads are identified accordingly as baseload, intermediate load and peak load. An autoregressive integrated moving average with exogenous inputs (ARIMAX) model is used to extract weather dependency to find the residuals, then through hypothesis testing the extremities, maximum and minimums are found [49]. This procedure was reproduced, with linear regression finding the residuals and a Bayesian maximum likelihood classifier to identify anomalies [48]. A data-mining based framework using DBSCAN was used to detect anomalies in office buildings. The framework is aimed to identify typical electricity load patterns and gain knowledge hidden in the patterns and to potentially be used in an early fault detection of anomalous electricity load profiles [50]. Also to detect anomalies of electricity consumption in office buildings an improved kNN is proposed, ikNN, to automatically classify consumption footprints as normal or abnormal [51]. Dynamic Bayesian Networks and Restricted Boltzman Machine has been proposed for anomaly detection in large-scale smart grids. Simulated on the IEEE 39, 118, and 2848 bus systems the results were verified [52]. Real-Time Mechanism for detecting FDIA analyzed the change of correlation between two phasor measurement units parameters using Pearson correlation coefficient on IEEE 118 and 300-bus sytems [53]. Machine learning techniques have been highlighted for theire ability to differentiate between cyberattacks and natural disturbances. By a simulating a variety of scenarios the ability for One R, Random Forest, Naive Bayes and J-Ripper to recoginze attacks was investigated: Short Circuit faults; location is represented by the percentage range, Line maintenance; identified through remote relay trip command, Remote tripping command injection; the attacker operates the relay remotely that causes a breaker to open, Relay setting change; the attacker misconfigures the relay settings to cause maloperation of relays, FDIA; attacker manipulates measurements sensors. The simulated scenarios was grouped into classes; natural events, attack events, and no events [54]. In machine learning concept drift occurs when the underlying distribution of the data changes over time, making the model unfit to predict for future events [55]. In concept drift, models are inaccurate due to change in the underlying data [56]. Thus the observation can be a result of an improved energy system, and not anomaly [57]. An adaptive sliding window method, that calculates an adoptive window-size on the fly is has been proposed to deal with concept drift [58]. Another adaptation raises the complexity by using multi sliding window detecting growth length over several windows detecting the drift length by adjoining several windows finding the optimal window length useful for online learning [59]. In Recurrent Neural Networks like Long Short-term Memory networks (LSTM) different gates are used to remember and forget time-occurrence over different time windows. To deal with concept drift in LSTMs it has been proposed a novel forgetting mechanism for anomaly detection [60]. It has also been proposed to narrow down the scope, by critical lines detecting distribution change. The first step in the proposed method is to reduced dimensionality through an orthogonal transformation of the data reducing the feature space to its principal components [61]. After reducing the feature space the distributions are compared by two-sample Kolmogorov–Smirnov Test (KS test) [62].

Chapter 3

Pattern Recognition in Machine Learning Techniques

Since ancient times man has observed patterns in nature and aspired to make sense of them. Understanding future conditions to optimize crops was crucial since the rise and fall of ancient civilisations is correlated to changing climate conditions. The survival of ancient civilisations where dependant on regular agriculture and water availability. Mother earth itself keeps timely track of climatic changes through the science of paleoclimatology, and allows for observations that reveals hypothesis.

In a giant databank, the Greenland glacier, the gas trapped in the ice core are measured and used as a time-stratigraphic marker reducing the uncertainty of the time series analysis by correlation the different ice core layers in high precision correlations of ice core records of the past 100 000 years [63].

Nowadays, advanced meters provide data granularity at an unprecedented level. Smart meters, as defined by The Energy Information Administration; "meters that measure and record usage data at a minimum, in hourly intervals and provide usage data at least daily to energy companies and may also provide data to consumers. Data are used for billing and other purposes. Advanced meters include basic hourly interval meters and extend to real-time meters with built-in two-way communication capable of recording and transmitting instantaneous data" [64].

Advances was made in the fields of astronomy, weather, economics and medicine, and recordings of observations within these fields lead to the general field of Time Series Analysis. Questions of causality where treated strictly within the separate fields, yet these fields searched to reveal hypothesis based on how the past influence the future.

Analysis of time series data with both statistical methods and machine learning techniques will be increasingly important with more available data through Internet of things, digitilization and smart systems. This type of continuous monitoring will increase the need for some of the time series specific exploratory methods, that helps identifying stationarity, applying window functions, understanding self-correlations and spurious correlations

[65].

3.1 Time Series Analysis

The origins of time series analysis applied autoregressive methods to understand real time data. Autoregression (AR) is a simple and straightforward regression technique, where past values of the univariate time series are dependent on their own lagged version defined by a parameter weighting of each input, ϕ, and therefore a parametric model. The current value of y(t) is expressed by previous values of time $y_{t-1}, y_{t-2}, ..., y_{t-p}$. The order of an AR process is defined by the number of past values of y(t) it is regressed on. AR(p) is defined by the last y_{t-p}, considered in the process, denoted as:

$$y(t) = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t \tag{3.1}$$

Where the error term ϵ_t, is white noise defined by a constant mean and some unknown fixed variance $\sigma_\epsilon^2(t)$, a stationary process.

The parameters of AR are estimated from the data. In the 1920's experimental physicist Udny Yule, developed a set of equations together with Gabriel Walker, known as the Yule-walker equations, based in the autocovariance and autocorrelation function of a univariate time series. He applied the equations to an autoregressive model to Wolfer's sunspot data [66].

Wolfer's sunspot data depicts an interesting development in the field of data collection, in assuring a level of quality in the data. Ever since the beginning of the 1600's when astronomers had access to advanced equipment sunspot data have been collected. Yet, these observations are in the eye of the beholder, namely that the observations change according to the geographical location of the observer. Wolfram proposed a scaling factor k that varies with location and instrumentation, and developed the relative sunspot number R, based on k, number of sunspot groups (g) and the number of individual spots (s):

$$R = k(10g + s) \tag{3.2}$$

With the relative sunspot number, more observations are considered, to increase the accuracy of the data.

Certain assumptions about the time series is certified through Augmented Dickey-Fuller (ADF) test, such as stationarity. To prove stationarity a search for no trend, constant variance and constant autocorrelation is conducted. Testing for stationarity is done by introducing the null hypothesis H_o: Time series is non-stationary due to trend. By the ADF test, if certain criteria are met the null hypothesis is rejected and the time series is assumed to be stationary. The ADF basically searches for trends in the dataset by evaluating mean and variance over time.

The ACF of a white noise process is zero at all lags other than lag zero where it is unity, to indicate that the nature of its process is completely uncorrelated. Autocorrelation (AC) is the serial dependence of a time-series shows the relation to its own lagged versions [30].

$$\rho_k(t) = \frac{\sum_{i=1}^{n-k}(x_t - \hat{x})\sum_{i=1}^{n-k}(y_{t+k} - \hat{y})}{\sqrt{\sum_{i=1}^{n-k}(x_i - \hat{x})^2}\sqrt{\sum_{i=1}^{n-k}(y_{t+k} - \hat{y})^2}} \tag{3.3}$$

By using backshift operator (B), the previous value of the time series is related to the current value $y_{t-1} = By_t$, and thus; $y_{t-m} = B^m y(t)$, and the error term is explained as:

$$\phi(B)y_t = \epsilon_t \tag{3.4}$$

An AR process p-value is defined by the autocorrelation of residuals of the AR process. If the residuals autocorrelation falls within a confidence interval, normally considered as 95%, the autocorrelation function of the residuals are considered to be white noise. If not, the AR process will still continue to find another parameter, until its residuals satisfy the criteria of white noise. If the current and previous values of a white noise series $\epsilon_t, \epsilon_{t_1}$ are expressed linearly, it is known as moving average process (MA), and an equivalent implementation of backshift operator (B) would be:

$$y(t) = \theta(B)\epsilon(t) \tag{3.5}$$

A combination of the two processes is autoregressive moving average (ARMA). If the mean or covariance of the time series observations change with time, the series is defined as nonstationary. A nonstationary series becomes a stationary series through integration, by taking first difference of each data point with its previous one. The differencing process introduces the ∇ operator, and the AR, MA and ARMA processes are transformed into ARI, IMA or ARIMA process.

$$\nabla Y_t = Y_t - Y_{t-1} \tag{3.6}$$

If the series needs to be differenced multiple times (d) to become stationary, stating the integrated degree of d, shown as $Y_t \to I(d)$ [67].

3.2 Information Criteria

One key aspect in modelling time series is to search for the most parsimonious model. A parsimonious model explains the observations by the the simplest model. In 'Almagest', written by astronomer Ptolemo (cirka 150 CE), he argues to practice the principle of choosing the simplest hypothesis to explain a phenomena [68]. William of Occam, defined parsimony further in his theory known as 'Occams Razor'; when two hypothesis explain a phenomena equally good, it is most likely that the simplest one is the correct hypothesis [69].

In any case a-priori assumptions of observations are likely to be noisy, hence performance metrices are used to numerically compare model performance. The R-squared, R^2, is calculated by dividing the sum of squared residuals ($\sum(y_i - \hat{y}_i)^2$) by total sum of squares ($\sum(y_i - \bar{y})^2$), subtracted by 1:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{3.7}$$

The ajdusted R^2 is found by in the total sample size (N), including the degrees of freedom (N-1) of the estimated variance and, including the number of predictors (p), the degrees of freedom (N-p-1) of the error variance:

$$Adjusted R^2 = 1 - \frac{(1-R^2)(N-1)}{N-p-1} \tag{3.8}$$

To select the best model it is possible to choose the one with the largest adjusted R^2. This is a tedious and computational heavy operation, and might not end up opting for parsimonity.

Various information criterion techniques considers both the model fit as well as introducing a 'penalty' for the number of parameters. The lag length of an autoregressive model can be determined by The Aikike Information Criteria (AIC), Bayesian Information Criteria (BIC) or Hannan-Quinn Criteria (HQC). The mathematical formulation, for AIC and BIC, considers the residual sum of squares of the regression, sample size (n), and the number of parameters (k) in the following equations:

$$AIC = ln\left(\frac{\sum(y_i - \hat{y}_1)}{n-k}\right) + \frac{2}{n}k \tag{3.9}$$

$$BIC = ln\left(\frac{\sum(y_i - \hat{y}_1)}{n-k}\right) + \frac{k}{n}ln(n) \tag{3.10}$$

Whereas HQC considers the maximum likelihood (L_{max}), number of parameters (k) and number of observations (n):

$$HQC = -2L_{max} + 2kln(ln(n)) \tag{3.11}$$

3.3 Linear Regression

Another parametric model is multiple linear regression (MLP) that assumes a linear relationship in the training data and to explanatory variables to explain relationship to the response-vector (y):

$$y(t) = a_0 + \beta_1 x_1(t) + ... + \beta_n x_n(t) + \epsilon(t) \tag{3.12}$$

where $x_1(t), ..., x_n(t)$ are independent explanatory variables correlated with the dependent load variable y(t). The independent variables are found through correlation analysis, and coefficient estimation normally found through least square estimation, or iteratively reweighted least squares (IRWLS). All parameters start at 0 and is step-wise improved using backpropagation through a loss function to find appropriate weights, or through finding the intercept a_0. Each explanatory variable finding its coefficient based on the covariance and standard deviation of dependent and independent variables is defined as:

$$\beta_x = \frac{\sigma_{xy}}{\sqrt{\sigma_x}} \tag{3.13}$$

$$L_{OLS}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 = ||y - X\hat{\beta}||^2 \tag{3.14}$$

3.4 k-Nearest Neighbor Regression

Opposite to the linear regression (LR) is the k-nearest neighbor (kNN) regressor, which is non-paramteric, relying on its own table look-up and mathematical foundation, and highly non-linear.

$$y_{knn}(x) = \frac{1}{K}\sum_{k=1}^{K} y_k \text{ for K nearest neighbours of x} \tag{3.15}$$

The kNN-classifier is illustrated in Fig. 3.1, where the left diagram with a small encirclement options for k = 1, where simply the nearest neighbor decides the class of prediction, whilst in the right diagram in Fig. 3.1, the number of k is increased to more then one [70].

Using k = 1 can lead to false prediction, and a set of kNNs is often used. When classifying the dependent variable is categorical, it can easily be made numerical by regression. The kNN regressor makes a regression based on the number of kNNs to minimize false predictions. The model considers a range of different k-values to find the optimal value. The kNN regressor needs thorough pre-processing and feature engineering to limit the effect of noise caused by irrelevant features, and is, therefore, dependent on finding the appropriate distance model [71]:

3.4.1 Distance

A variety of distances is used in the algorithm. As seen in Equations (3.16, 3.17, 3.18, and 3.19), they are mostly used, since it is easy to intersect by changing the variable q. The variable q is also considered to find the optimal value.

3.4.1.1 Manhattan/City Block Distance

$$d(x,y) = \sum_{i=1}^{k} |x_i - y_i| \tag{3.16}$$

3.4.1.2 Euclidean distance

$$d(x,y) = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{3.17}$$

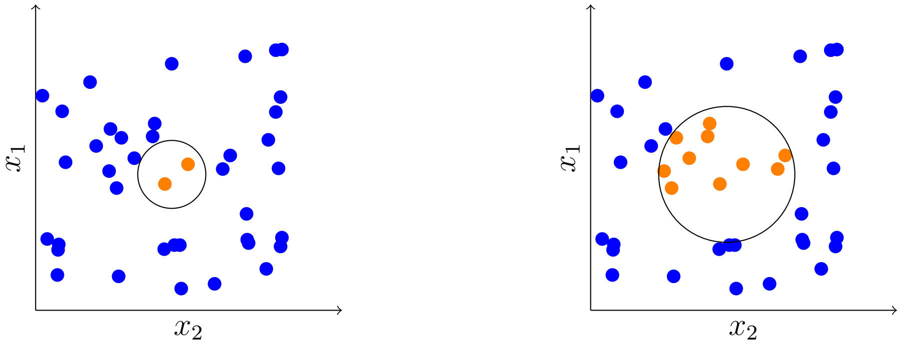


Figure 3.1: k-Nearest Neighbour classifying based on the k'th observation.

3.4.1.3 Minkowski Distance

$$d(x, y) = \left(\sum_{i=1}^{k} (|x_i - y_i|)^q \right)^{\frac{1}{q}} \tag{3.18}$$

3.4.1.4 Chebychev Distance

$$d(x, y) = \lim_{q \to \infty} \left(\sum_{i=1}^{k} (|x_i - y_i|)^q \right)^{\frac{1}{q}} \tag{3.19}$$

3.5 Decision Trees

One of the most practical methods is Decision Trees. It is a popular inductive inference algorithm. The classification and regression trees (CART) algorithm, developed by Leo Breiman. Other tree learning algorithms are ID3, ASSISTANT, and C4.5.

The Learning in a decision tree is represented by a tree, with a root that evolves into different branches, where a leaf node, the final output node, provides a classification of an attribute and in between are internal nodes [72].

The function that are used to make a decision is the splitting criteria. In scikit-learn it is usually gini impurity, or entropy, but other methods are used as well.

The Decision Trees branches uses if else statements and it bases its decisions on a set of rules. The fact that the Decision Tree is a rule based learner makes en attractive model in a machine learning interpretation or explainability aspect. This field is called 'Explainable ML' and 'Interpretable ML' or 'Explainable AI' and 'Interpretable AI'

Overfitting occurs when a tree is growing into a number of branches that makes the accuracy improve while regarding the training data, but subsequently fails to improve, and even reduce accuracy for new data. This issues are dealt with in C4.5, and has to do with the inductive bias and the phenomena Occams Razor. In C4.5 a decision tree is pruned to improve accuracy.

By imposing mathematical functions on the Decision Tree, it can return the tree structure with the highest accuracy:

$$H(s) = -p * log2(pos) - neg * log2(neg) \tag{3.20}$$

To enhance the decision tree a mathematical function is used to decide how well an attribute classifies the training data. The entropy function H(s), see equation [3.20], measures the homogeneity of a classifier, and gives a indication of how pure a classifier is, from 0 to 1, where 0 is the pure classifier and 1 is an impure classifier since it gives you a 50/50 chance of being either which class.

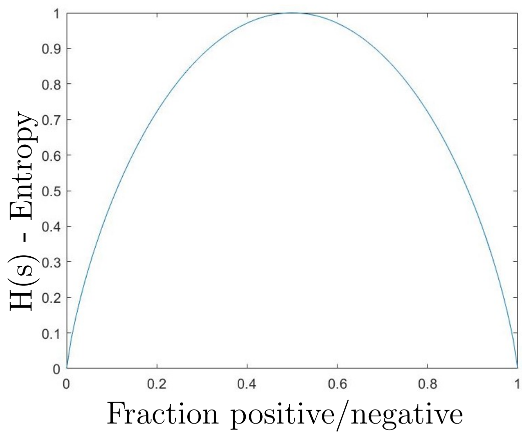


Figure 3.2: The entropy function returns a measure of which attribute makes the best classifier.

Note that a the entropy can also be measured for c number of values, then the function is:

$$Entropy(S) = \sum_{i=1}^{c} -p_i log_2 p_i \tag{3.21}$$

The entropy is then used to evaluate the information gain of an attribute, and thus to know how to choose the highest gaining attribute as the next branch in the decision tree. The equation yields the expected reduction in entropy, by imposing another branch in the decision tree:

$$Gain(S, A) = Entropy(S) - \sum_{v \epsilon Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{3.22}$$

$$\begin{aligned} Gain &= \text{The information gathered} \\ S &= \text{Entropy} \\ A &= \text{Attribute} \\ |S_v| &= \text{the value of the subset} \\ |S| &= \text{Sum of the subsets} \end{aligned}$$

3.6 Random Forest Regression

Random forest (RF) regression is a combination of decision trees, found through recursive partitioning to build a piece-wise linear model. From these tree models, it uses a majority vote for the most popular class. The trees grow dependant on a random vector, and the outputs are numerical scalars [73]. Each leaf on the tree is a linear model constructed for the cases at each node by regression techniques. One sole decision tree encompasses attributes and classes in the data and uses a entropy function to distinguish its structure. Entropy is known from thermodynamics as a measure of disorder, and later adopted by the information theory. In information theory, entropy is a measure of uncertainty of a variable, and defines a pure classifier [74]. In Equation (3.23) p is positive and n is negative:

$$Entropy(S) = -p * log2(p) - n * log2(n) \tag{3.23}$$

The entropy function is then used to evaluate the information gathered (gain) of an attribute, and thus to know how to choose the highest gaining attribute as the next branch in the decision tree. The equation yields the expected reduction in entropy, by imposing another branch in the decision tree.

$$Gain(S, A) = Entropy(S) - \sum_{\nu \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{3.24}$$

In equation (C.15), A are attributes used for splitting the data into subsets (S). S is the sum of subsets, and S_v is the value of subsets. Using prior known input/output relationships, the algorithm searches for a model for the best prediction in the training set. The mathematical equations are structured in the algorithm, see Fig. C.6, based on the past knowledge.

3.7 Ensemble Methods

In ensemble methods, multiple machine learning algorithms is used to perform ensemble learning. A combination of models, in an ensemble, is the basis for ensemble learning. An ensemble learning algorithm is called ensemble models or ensemble methods. The underlying premise for ensemble methods is that an aggregated answer improves predictions when the training data cannot provide sufficient information. When averaging the measurements, the estimate is less pruned to random fluctuations in a single measurement, and hence are a more reliable and stable output. The aggregated averaged answer designed to improve the predictions, is also the same principle as 'Wisdom of the crowds'. The best-known Ensemble Methods include Bagging, Boosting and Stacking. Usually they consists of simple algorithms, like Light Gradient Boosting Machine (LightGBM), includes a 'boosting' paradigm based on gradient descent that adds expansion [75]. Another version is XGBoost that maneuvers the out-of-core computation, by chunking data into manageable sizes for next to load data chunk by chunk, in a way that it allows for fast processing of hundred millions of examples [76]. Decision Trees are ensemble methods, when, as explained in the section above, see Section 3.6, multiple Decision Trees form a Random Forest Regressor.

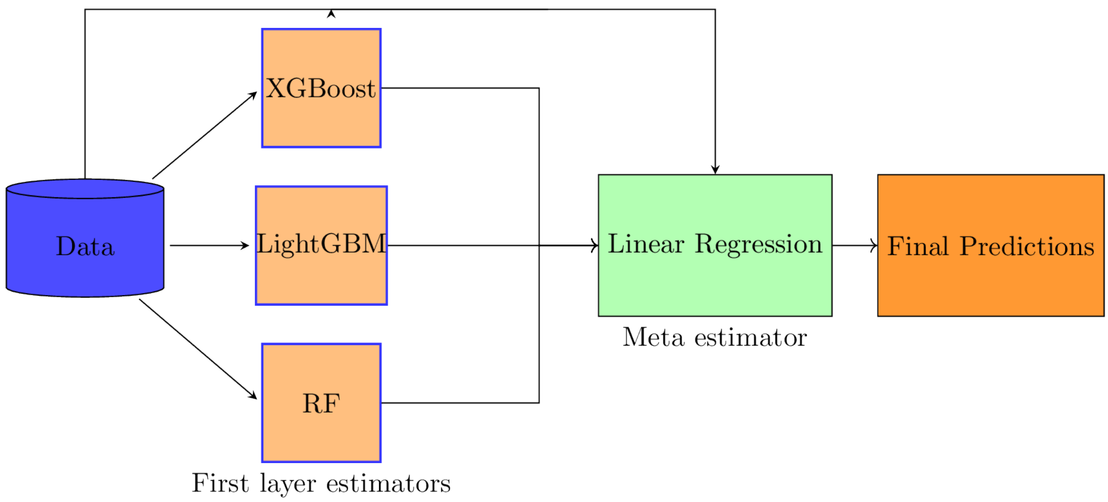


Figure 3.3: Ensemble methods

3.8 Artificial Neural Networks

Artificial Neural Networks (ANN) has been used successfully for complex decision making. Their advantage over basic regression tools is that they encompass a higher degree of complexity, hence can detect structural combinations hidden for the naked eye. Inputs are connected as nodes in hidden layers. Essentially the computed weights of the inputs defines the final model. The complexity of associations made in the networks of hidden layers, and the lack of explainability and interpretability have made ANN-models referred to as black box models. When decision-making is done without being able to explain the logic behind the decisions, it has questioned the ethics and fairness of neural networks and introduced the aspect of fairness in algorithms. The term XAI, refers to explainable AI, aimed to raise fairness through awareness, where focus is aimed at how data is selected and used in algorithms [77], as well as well as advancing machine learning algorithms to be fairness-aware [78].

The traditional Deep Neural Networks (DNN's) learn patterns on the assumption that inputs and outputs are independent of each other. The first DNN's used stacked generalization to developed deep learning based on the concept of a perceptron [9]. A perceptron mimics the behaviour of neurons in the brain for decision making. When the sum of weights and inputs reach a certain threshold value, neurons fire off, just like learning paths are developed in the brain. In neural networks an activation function decides upon the state of activation. The output from the activation function is compared to the real value from the targeted response vector in a loss function, as shown at Frame 1 in Fig. 3.5. The output from the loss function is used to trace the global minima through stochastic gradient descent. The information is backpropagated through the network and used to adjust the weighted input of the network, and this process is found when training and validation losses converge and a stop criterion terminates the process.

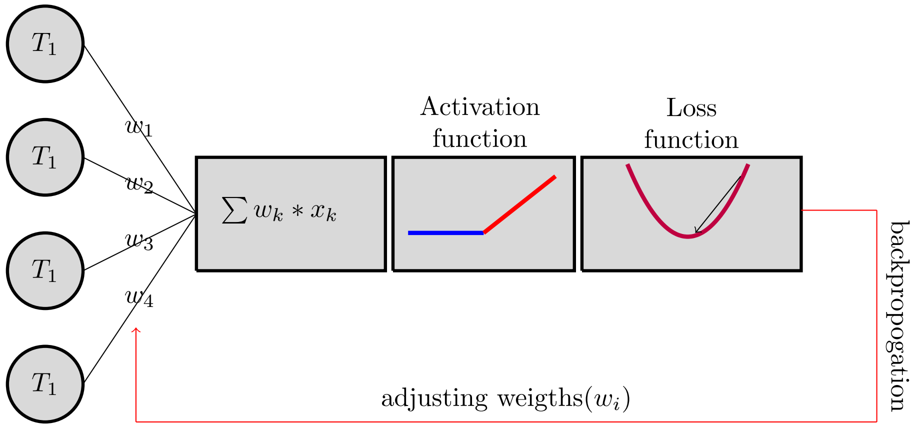


Figure 3.4: Schematic of Artificial Neural Networks

In ANNs weights are given to inputs. When the sum of these weights and inputs overcome a certain threshold value they give direction to the stochastic gradient descent of the loss function, as illustrated in Fig. 3.4. This information is then backpropagated and used to adjust the weights according to the information given to seek the global minima of the loss function.

ANNs are flexible and have been applied successfully to a great extent of various forecasting problems, yet their performance on real time series problems are not satisfactory, and it is important to create new models of neural networks [79] [80].

3.9 Recurrent Neural Networks

A Recurrent Neural Network (RNN) depend on the prior elements within the sequence, to perform its decision making. RNNs was first developed in natural language processing and the Vanilla RNN is a fully-connected RNN where the output from previous time step is to be fed to the next time step by an additional set of units. The units have also proven to be successful in other time series application, and for all problems constituted by sequences, such as electrical load demand.

In Frame 2 of Fig. 3.5, illutrates a Feed Forward Neural Network (FFNN) as a black box representation, with input, black box and a learn output.

Frame 3 in Fig. 3.5, shows a FFNN transposed to its vertical axis, to show the key concept of units in RNN's.

Recurrence that provides the key concepts behind Recurrent Neural Networks, the key idea, is that the RNN's remain the internal state, h_t, that is updated for each timestep, and keep the sense of recurrence in the network. The update is defined mathematically as shown in Equation (3.25):

$$h_t = f_w(x_t, h_{t-1}) \tag{3.25}$$

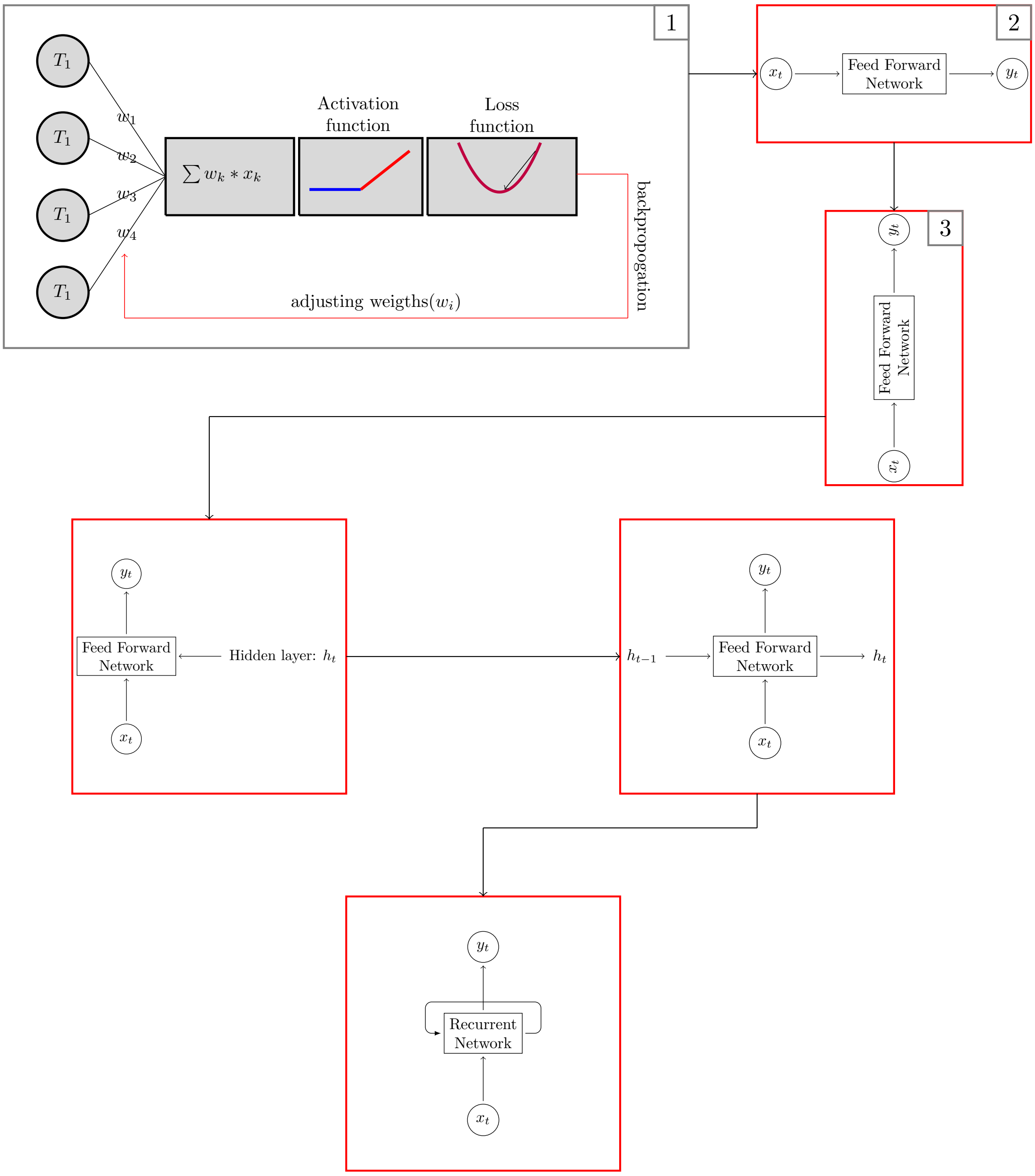


Figure 3.5: Reurrent Neural Network

This internal state, h_t, is a hidden layer used to define the state. When computed in the network is is used as shown in [81]:

$$h_{i,t} = \sigma_t(Uh_{i,t-1} + Wx_{i,t}) \tag{3.26}$$

In Equation (3.26), σ is activation function, U and H are learned weighted parameters for hidden states and input vectors.

The process then composes a set of learned weighted parameters in matrix V, which for a regression problems uses a linear activation function σ_y to give the result in the output layer:

$$y_{i,t} = \sigma_y(Vh_{i,t}) \tag{3.27}$$

3.10 Anomaly Detection using Machine Learning Algorithms

3.10.0.1 Isolation Forest

The Isolation Forest algorithm is composed of several isolation trees (iTres) Isolation forest takes advantage of the nature of anomalies which are less frequent than regular observations and different from those in terms of values to isolate those. Iforest can deal with large scale data quickly in a simplified way. It builds an ensemble of decision trees (iTrees) for a given data set. Clustering is done using binary tree clustering. Anomalies tend to be isolated closer to the root of the binary tree. Partitions are created using a split value between the minima and maxima of a randomly selected feature. The algorithm then tries to separate each point in the data. [82] [83] [84] [85].

3.10.0.2 Local Outlier Factor

Local Outlier Factor (LOF) is a density based anomaly detection algorithm introduced in 2000 [26]. LOF compares the local density of a point to the local density of k of its neighbors. By comparing the local density of a point to the local density of its neighbors one can identify point that have substantially lower density than its neighbors. These points are considered outliers. LOF uses the k-distance to a point as in kNN, to find the Local Reachability Density (LRD), where a point is most likely to be found. The sum of LRD is the used to find LOF for the point z, as in Equation (3.28):

$$\mathrm{LOF_k}(\mathbf{z}') = \sum_{\mathrm{z}\in \mathrm{N_k}(\mathbf{z}')} \frac{lrd_{\mathrm{k}}(\mathbf{z})}{lrd_k(\mathbf{z}')} / \left\| N_k(\mathbf{z}') \right\| \tag{3.28}$$

[86]

3.10.0.3 Normalising

The pre-processing of data is a transformed so that the machine learning algorithm can learn the patterns and generate a sound forecast. In a standard normalization process, input data are transformed with values from zero to one. This is done to make the predictive algorithm more robust [42].

$$\frac{\hat{X} - X_{min}}{X_{max} - X_{min}} \tag{3.29}$$

$$\frac{\hat{X}}{X_{sum}} \tag{3.30}$$

$$\frac{\hat{X}}{X_{max}} \tag{3.31}$$

$$\frac{\hat{X} - X_{avg}}{X_{max} - X_{avg}} \tag{3.32}$$

3.10.0.4 Performance metrics

To evaluate the performance of load forecasting, a performance metric is used, including mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), and symmetric mean absolute percentage error (SMAPE). They are defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}| \tag{3.33}$$

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}|}{(|y_i| + |\hat{y}|)/2}\right) * 100 \tag{3.34}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2 \tag{3.35}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{y}}{y_i}| * 100 \tag{3.36}$$

3.11 Detecting Patterns

The techniques provided in Machine Learning are aimed at enhancing the pattern recognition, and can be applied in a diversified range of fields (electrical demand forecasting, renewable energy production, condition monitoring, anomaly detection, energy market etc.). The specific applications of machine learning techniques require deep domain knowledge with improved novel methodologies that include subject specific knowledge.

Chapter 4

Methodology for Electrical Demand Prediction

The focus point of this works methodology is to understand how machine learning algorithms function to solve the task of electrical load demand forecasting. The two aspects is time organisation and how the algorithm perceive the more general information given. From the literature there are concerning aspects that when AI and ML becomes an integral part of decision making in society, there is a growing need to understand how to interpret and explain the algorithms. This paves the way for more fairness and awareness in artificial intelligence and information from experts in the domain where it is applied.

The complexity of models incorporates a trade-off between bias and variance when choosing the suitable model for a forecasting problem. The suitable model is found from this trade-off in complexity, as well as the level of exlainability and interpretability. There is a wide variety in models to choose from, and the algorithms shows different traits that will be beneficiary in choosing the suitable model.

In general for a electrical load demand forecasting problem, there is the incremental behaviour of each step in data governed by the time of use. This is incorporated by the autogressive modelling, yet for more complex models it is needed to deal with stationarity as well as seasonality. In time series modelling this is dealt with by introducing operators such as the nabla operator and adding layers that handles seasonality. In Neural Networks, the incremental steps of time series is dealt with in the recurrence and seasonality in long and short-term dependencies.

The methodology of this work is based in both a comprehensive feature engineering, based in in-depth correlation analysis and dimensionality reduction, also including principal component analysis. From the mentioned analysis, the most prominent features are selected and this has given the fundamentals to test the most suitable models. The novelty of this work is based in the way that the work also include a focus of domain knowledge, shown through the design of a vertical axis approach, see Fig. 4.1. The presented vertical time approach uses seasonal data for training and inference, as opposed to continuous time approach that utilizes all data in a continuum from the start of the dataset until the

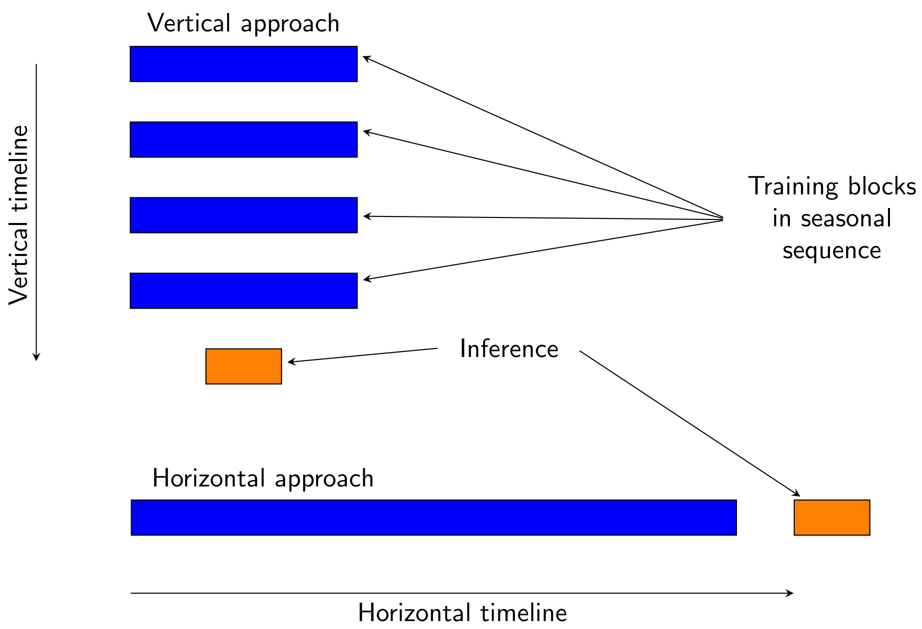


Figure 4.1: Illustration of vertical and horizontal approach.

time used for inference.

In Fig. 4.2 is illustrated how the regressor model is designed with inputs (Hour, Day, Previous load, Temperature), applied algorithm (kNN) and hyperparameters (Distance Function), to give the load consumption prediction result.

4.1 Main features influencing electrical load demand patterns

There is a consensus among the recent articles that the most important features influencing the load forecasting are[10][11][12][13][14][15][16][17][18][87][88][89][90][91][92]:

(i) Time
(ii) Weather
(iii) Random effects

The historical electrical load patterns has shown to be the most important factor to describe the changing load demand and identify electrical load demand patterns. This is also reflected in the series own lag, found through its autocorrelation, described in Chapter 3, section 3.1, using Equation (3.3). Electrical load demand forecasting is divided into categories by the time-term of the prediction window (short-/medium-/long-term). The literature is known to deviate in how to categorize the time-terms. The overall tendency is that short-term is everything below 24 Hours and medium-term is two or more days to months, and long-term is year. Short term load forecasting is the most popular research area, and occasionally short-term is divided into very-short-term, short term and

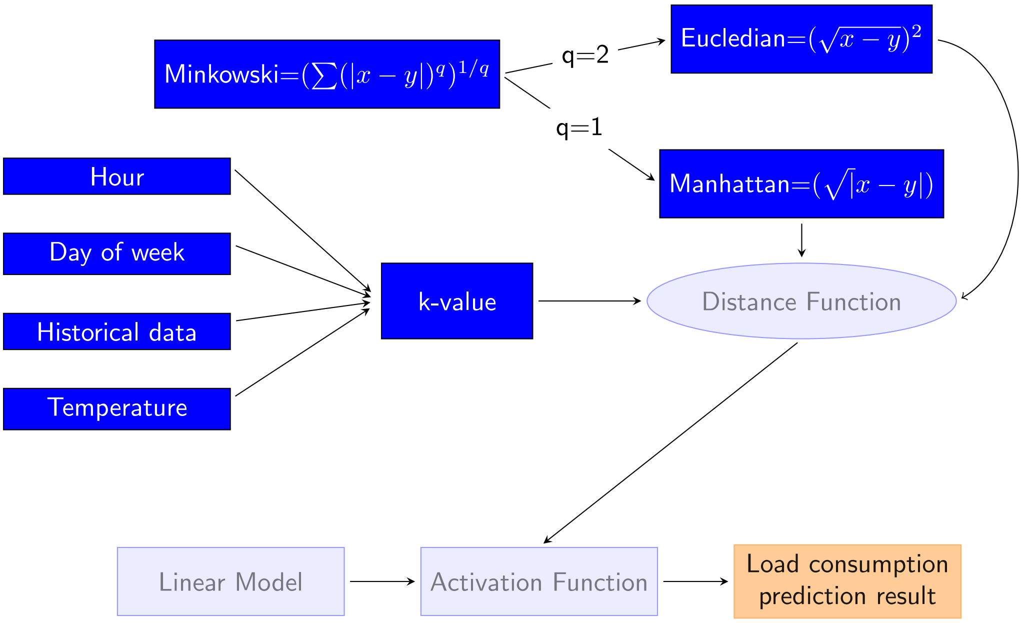


Figure 4.2: The regressor model for electric load demand forecasting

long-term, where very-short-term is minute to 5 minutes forecasting, short-term is half-hour to hourly and long-term is 24 hour forecast. In the following we will use the last definition, namely that short-term is half-hour to hourly and long-term is 24 hour forecast.

In the short-term (30 minute) forecast the time is the most important factor. This is shown through the mentioned autocorrelation, and hence forecast-methods such as autoregression, kNN, and Random Forest captures the trends of the short-term variations. To capture the seasonal variations in the datasets, techniques such as ARIMA has been devoloped into Seasonal Autoregressive Mean Average (SARIMA) [93]. A general term notation for ARIMA, includes parameters, p, d and, q: ARIMA(p,d,q). Where p is the autoregressive term, d is the number of differences (Δ - operator) to deal with stationarity, and q refers to the moving average term. For SARIMA the parameters is expanded to include for seasonal variations with (P,D,Q)s, where s is the number of periods per season and P,D and Q is the autoregressive, Δ and mean average parameter for the seasonal term.and the general equation:

$$\theta(B)\Theta(B^m)(1B)^d(1B^m)^D y_t = \alpha + \Theta(B)\Theta(B^m)\epsilon_t \tag{4.1}$$

where y_t is the value electrical load at current time; $\theta(B)$ represents the non-seasonal AR coefficients of the order p; $\Theta(B^m)$ is the seasonal AR coefficient of order P with seasonal degree of order m; $(1B)^d$ is non-seasonal differencing of order d; $(1B^m)^D$ is the seasonal differencing of order D with seasonal degree of order m; α is constant; $\Theta(B)$ is non-seasonal moving average; $\Theta(B^m)$ is the seasonal moving average and ϵ_t is the current time error [94].

A one-term SARIMA model is written $SARIMA(1,1,1)(1,1,1)_{12}$ [95]. The equation includes 12 number of periods per season, and has shown good result for monthly electrical load demand forecasting [95].

Neural Networks developed to include short term dependencies in the Vanilla Recurrent Neural Networks, to include long-term dependencies of seasonal parameters in LSTM.

4.2 Correlation Analysis of Electrical Load with Meteorological Parameters

In order to analyse how the external parameters influence the load and understand reasoning behind the development of extending models to include for seasonal behaviour correlation analysis is necessary. Correlation is a measurement to how two ranges of data move together, and will give us an indication of how to assess feature engineering. The Pearson Correlation Coefficient (r) computes the linear relationship between two datasets, in a range from -1 to +1. [36]. If the relationship is in the proximity of 1, it means that when x increases so does y and at exact linearity, the opposite is true for -1, it means that when a dataset is increasing the other dataset is decreasing.

$$r = r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2}\ \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \tag{4.2}$$

One of the means to improve prediction accuracy in spite the seasonal differences, is to create a dummy variable that increases the precision of the algorithm while differentiating the seasonal changes.

4.3 Principal Component Analysis for Electrical Load Forecasting

Principal components analysis (PCA) is a multivariate technique that can be applied to many fields for feature reduction. To find the intrinsic nature of linguistic representation Principal component analysis (PCA) has been performed on the hidden unit activation patterns to reveal that the network solves the task by developing complex distributed representations which encode the relevant time relations and hierarchical constituent structure [96]. It is the number of samples in the features that are reduced, not the entirety of a feature in itself. PCA has been found useful in many areas such as daily urban demand forecasting [97]. PCA is extracting the important information for later to represent it in a new set of orthogonal vector input constituting the principal components. These principal components is linear transformation of the data so that the first coordinate explains the most of the variation, the second coordinate the second most, and so on. The components are found through the eigen-decomposition and Singular Value Decomposition [98] [99].

To perform PCA the the input matrix is transposed and crossed with its non-transposed version, stored in matrix L. By diagonalising L, find a matrix M and diagonal matrix W:

$$L = M^T W M \tag{4.3}$$

The feature space is reduced by restricting inputs based on the number of columns that sums up M to make a rotated matrix. The eigenvalues from W are related to the variance of the principal components. PCA reduces the input feature space, yet remains to capture and keep the variation for future inputs and is a important step in the feature engineering.

The proportion of variance needed for optimal feature space may vary. The reference [97] refers to a meta-heuristic practice of principal components explaining 85% of the variance, yet their optimal value was found at 92%.

Chapter 5

Electrical Energy Data Description

The Big Data sources are divided into Operational and Non-Operational Data, where the first can be extracted directly from the distribution grid through devices such as Advanced Metering Infrastructure (AMI), time-referenced measures of voltage and current phasors (PMU), and remote monitoring and control through Supervisory Control and Data Aquisation (SCADA), and the latter comprises of additional information that help utility companies gain knowledge about external parameters such as weather, electricity market data, social media, geographical information systems (GIS) and customer behaviour data (number of occupants, square meters etc.) [10].

Big Data presents opportunities and challenges. Through the proper analysis of data the detection of underlying patterns help improve the electrical energy network, yet it is also changing energy production and consumption. These challenges involves collecting, storing and manage energy data, analysing, extracting and discovering patterns in large datasets using methods at the intersection of machine learning, statistics and database systems (known as data mining), using data information effective decision making and securing data and privacy of all players in the energy network [100].

In the era of Big Data, data have become a commodity in itself, and comes with a price. Offers of weather forecasts often comes with a price, and also recorded past values has a price tag. Some limited data might be freely available for downloading, but for specified granularity most of the cases the data has a price [101]. Governmental agencies offers data freely on web-services, but with limited usability due to the granularity of data and researchers have adopted techniques to deal with limited data [102].

In GIS information are stored and updated in layers with a geographic location [103]. The overlay technique computerized in the early 1970s and first used for siting power lines and roads. There are open source project that collect information about power systems, and many governmental agencies offer data through GIS. Improvements in GISs enabled environmental assessment and analysis. GIS allows for incorporating database capabilities, data visualization, and analytical tools in a single integrated environment [104].

5.1 Case study I: Urban Area Load - New South Wales

New South Wales, Sydney region electrical load profile data set [105] includes meteorological parameters (e.g. DryBulb and WetBulb Temperature, Humidity, Electricity price and time of use) [106]. Data is gathered from 2006-2011. The overall energy mix in New South Wales consists mainly of Coal, Natural Gas, Hydro and other renewable energy sources as shown in Table 5.1. Fig. 5.1 illustrates the New South Wales distribution network.

Power plant	Number	Installed Power (MW)
Hydro	24	4794
Wind	14	1250
Solar	9	228
Coal	8	11730
Biogas	11	56
Natural Gas	19	3766

Table 5.1: Energy Mix in New South Wales, Australia

New South Wales is operated in the National Electricity Market (NEM) that also comprises of five other states that includes Queensland, New South Wales (including the Australian Capital Territory), South Australia, Victoria, and Tasmania. Making it one of the worlds largest interconnected power systems covering a distance of 5000 km.

Data from Australian Government Department of Industry, Science, Energy and Resources [107], shows an average growth in Australia's energy consumption of 0.7 per cent a year over the past ten years. Fig. 5.2 illustrates the states with the highest electricity consumption (New South Wales, Queensland, Victoria, South Australia) in NEM, with data ranging from 1999-2020 [107]. The load profiles for the mentioned regions have been stable for tewnty years (1999-2020), at stable average values, see Table 5.2.

5.2 Case study II: Rural Area Load - Bjønntjønn Hyttegrend

From rural cabin area in Bjønntjønn, Telemark, Norway, the electrical load demand consumption profile is collected from smart meters. Weather data is collected from surrounding weather information statins in the surrounding area. The land owner of the area wants to realize the project 'Bjønntjønn Grønn' (Bjønntjønn Green). The project seeks through different initiatives to make the cabin area 'green', with power from local hydro power stations, possibility of electric vehicle charging and operation of the load consumption

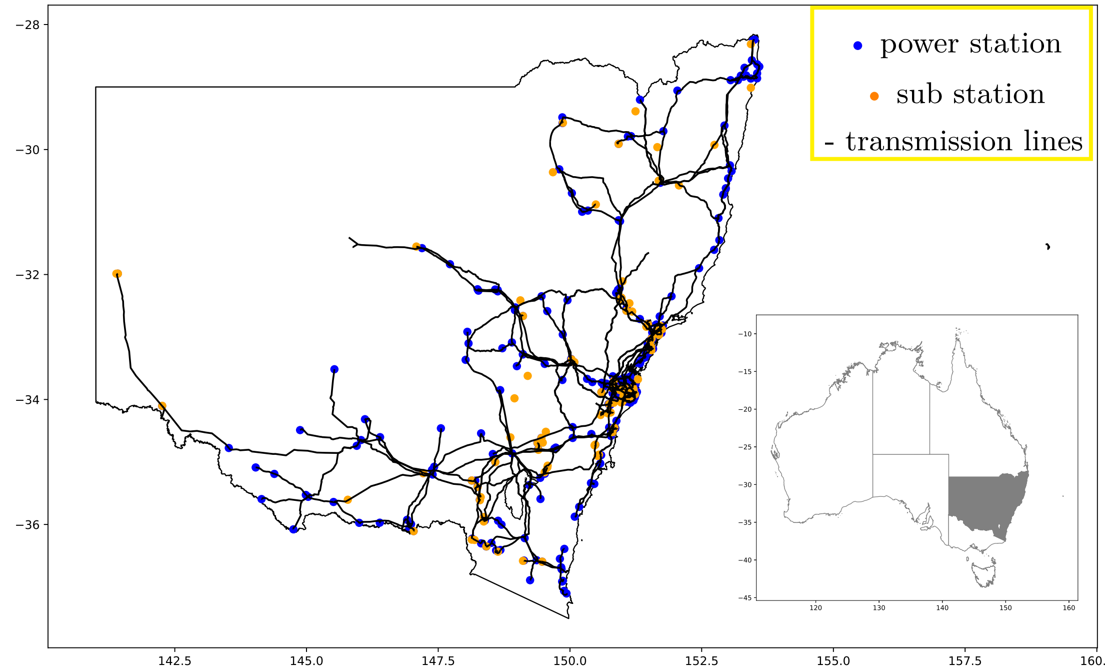


Figure 5.1: New South Wales Power system indicating location of transmission lines, power stations, and sub stations

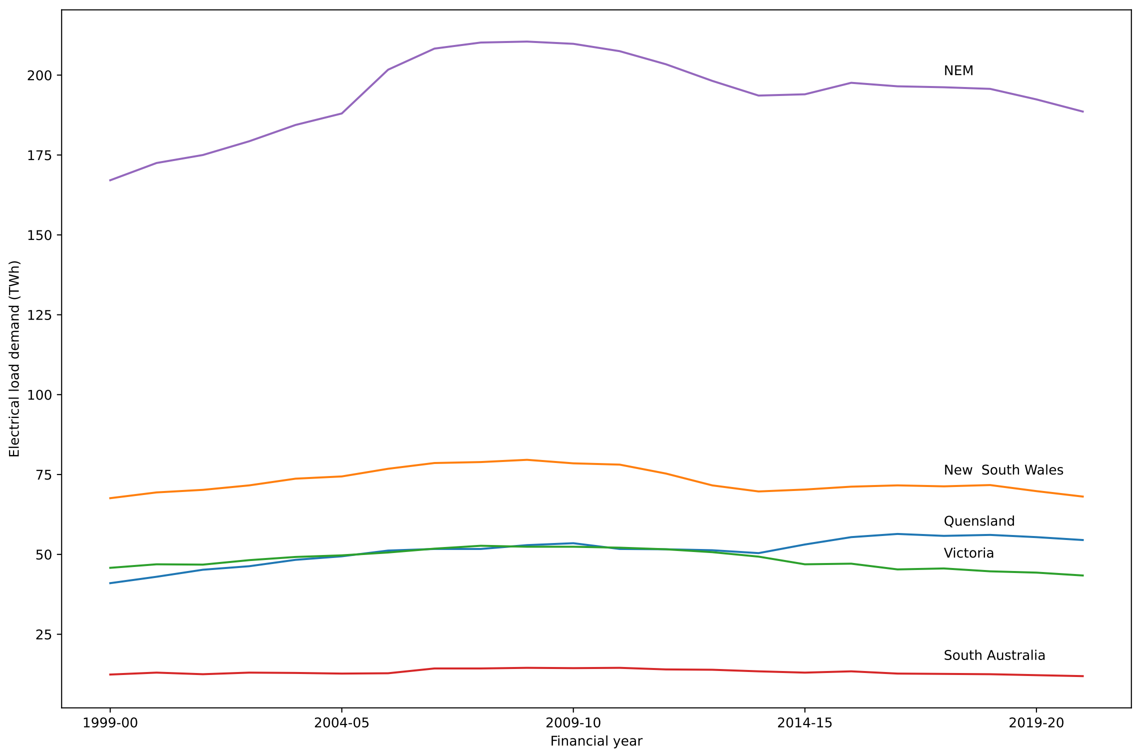


Figure 5.2: Load profile for the 4 regions in Australia, and the total load from NEM.

Region	Average Annual Electrical load demand (TWh)
New South Wales	73
Queensland	51
Victoria	49
South Australia	13
Tasmania	10
Snowy	0.6
NEM (Total)	194

Table 5.2: Regional Average Load demand 1999-2020, Australia

related to the power intensive usages. The land owner has currently an application to get license from The Norwegian Energy Regulatory Authority (NVE) to run hydro power stations in the area, with a total production of 10,08 GWh [108]. In the fall of 2021 NVE approved an appplication for a Tesla Supercharger from Tesla Norway, situated in the center of Treungen, an 8 km drive from the planned Bjønntjønn hydro power station [109] [110].

The rural area network of a typical Norwegian holiday resort cabin area, Bjønntjønn Cabin Area. It comprises 125 cottages with a peak demand of 478 kW. As for today, this cabin area is grid connected, but a microgrid solution involving photovoltaics and energy storage is also considered. In the summer of 2020 the land owner presented plan of building 445 new cabins in the area [111].

Rural electrification is very different from the urban area electrical consumption, due to diversified energy mix and overall conditions. A variety of case studies is necessary for a generic approach, although each system requires an independent approach. The Nordic market is much reliant on hydropower, as Norway's share of hydropower is 95.8 % [112]. Norway also has the highest integration of Electric Vehicles, and this faces challenges to the grid. This is especially a case in the rural area, where capacity is low, and the electrical vehicle charging poses a liability to the grid. In these cases, a micro-grid solution can aid the low-capacity network, with implementation of distributed generators, in combination with energy storage.

When examining the general load profile of all Norwegian Holiday Cabins, a clear trend is observed in the user behaviour. The load demand for Norwegian Cabins has increased their total consumption from 0.7 TWh in 1993 to 2.3 TWh in 2016. Although the consumption tripled and has been only 1.8 % of the total Norwegian load demand in 2016 [42]. Statistics Norway concludes in the 2018 report, that the increasing trend is due to the general development, and that more Norwegians have bought cottages in rural areas, such as mountains and seaside. Also, more cottages have been electrified in this period [112].

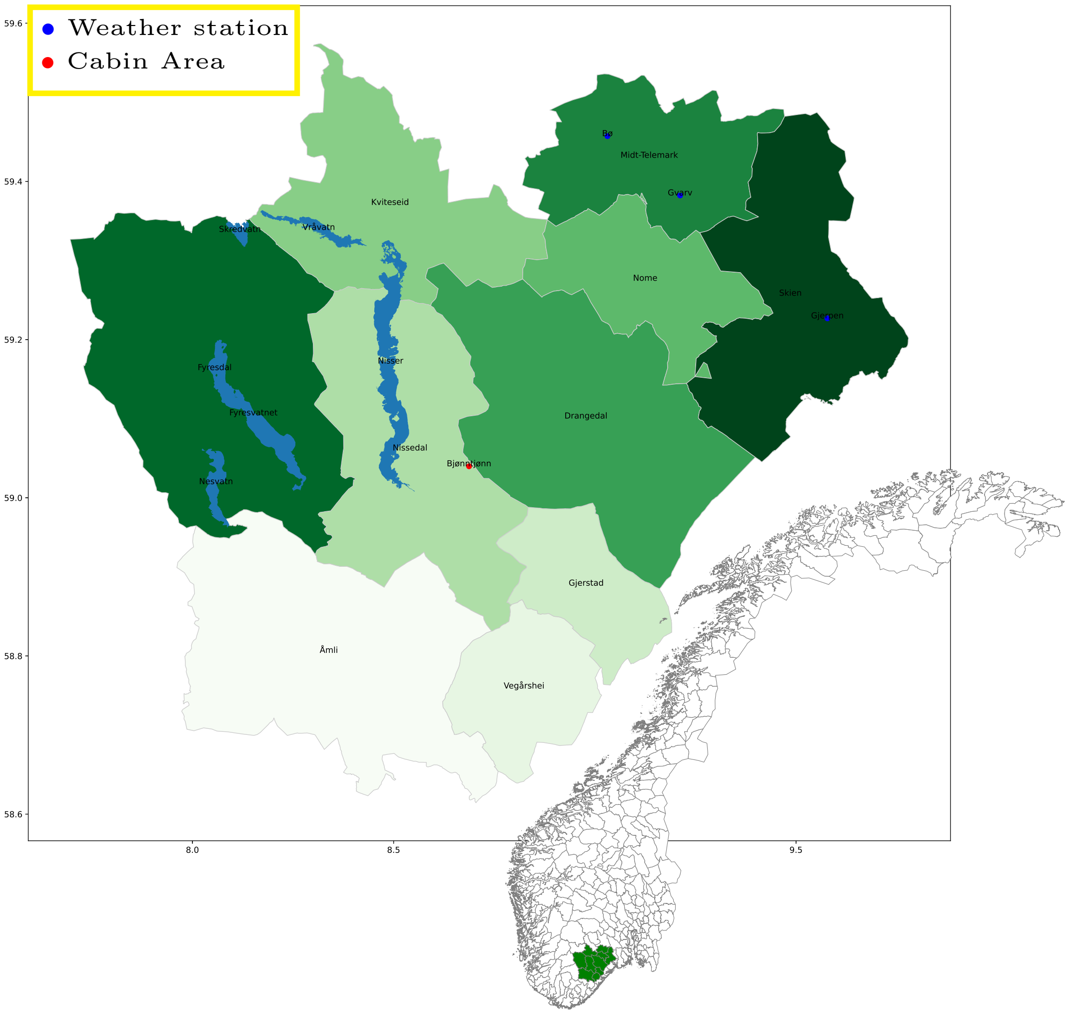


Figure 5.3: Rural Area in the South-East of Norway, with the situation of cabin area and nearby weather stations

A rural village in Siyambialanduwa (Sri Lanka) has a different composition of the basic elements and energy mix. Although it is comparable with the Nordic Holiday Resort in end-user size, with the villages 150 households, the daily energy load demand is 270 kWh [113]. The overall Sri Lankan energy mix is also like Norway heavily dependent on hydropower (40.5 %), besides a bigger part of thermal power (49 %). Due to economic conditions and geographical location (Sri Lanka is situated close to the equator) the most liable energy solution for Sri Lanka in the future will be renewable energy sources (RES). To electrify the rural area village population of Sri Lanka, through microgrid hybrid energy sources, may increase the electrical resiliency. A micro-grid organized network powered by RES hybrid systems can be considered for rural electrification and supported by the Sri Lankan government. Due to the intermittent nature of the RES, energy storage is an essential part of the system to maintain a continuity of supply and mitigate voltage fluctuations that might harm the electrical system. The optimal system for the village comprises photovoltaic (PV) system, wind turbines, diesel generation and a battery bank [113].

In the Bjønntjønn Cabin Area, to deal with the ever-increasing penetration of electric vehicles, photovoltaic system together with energy storage could be a scenario for the future rural electrification. For the Nordic rural area network, a microgrid solution can improve the electrical network capacity of the rural area, despite challenges from power demanding operations as electric vehicle charging. Since the electric vehicle will not be used mostly of the holiday resort area, the battery pack of the vehicle is be considered as the battery bank for the microgrid. When the state of charge (SOC) of the battery reaches a certain threshold level, it will be considered as a prosumer for the micro grid and be able to contribute to electrical supply and stability. In the further analysis of the rural electrification, it is necessary to have proper load analysis and forecasting.

5.3 Electricity Market Data - NordPool

The electricity 'spot market' is a day-ahead market, and not a continuous trade market. The process of deregulating the market and introducing a competitive trade market has reshaped what used to be a monopolistic government-controlled power sector. This process has been ongoing since the 1990's and electricity is now traded using spot and derivative contracts, hence the popular name 'spot-market'.

NordPool delivers power trading in the Nordic, Baltic and UK day-ahead markets, offering both day-ahead and intraday markets to customers and trade power in 13 markets as well as adding specific related services such as compliance and data. The Nordic trade market has since November 2011 been split into 12 bidding zones, 5 in Norway, 2 in Denmark, 4 in Sweden. Finland is operated as one bidding zone. Denmark has always been split in two bidding zones (Jutland has not been part of the Nordic synchronous area), until 2010 there was no interconnection between the two bidding zones and the connection is congested most of the time. Fluctuations in the hydrological balance make the Norwegian power system different from the rest of the region. Dynamic bidding zones create incentives for generators to optimize the use of storage capacity and ensures import when the hydrological situation is weak.

The electricity 'spot market' is typically a day-ahead market that does not allow for continuous trading. Day-ahead forecasting means that bids for the next 24 hours of day (d), are done before some given deadline the previous day (*d-1*). Then the market operator uses the bids to calculate a market clearing price for each of the 24 hours, once the deadline has passed. Then the bids are approved, and a contract between producer/consumer is established [114] [115].

5.3.0.1 Considered Regions - Denmark (DK1), Norway (NO2)

Although Norway and Denmark are neighbouring countries with sharing boarder, the topology and terrain of the countries are very indifferent. In Denmark the terrain is flat, contrasting the fjords and mountains of Norway. The differences of the two regions terrain make them rely on different power sources, as two different energy systems.

The Danish region DK1 is naturally defined as the peninsular Jutland, encapsulated by water and set apart by a the land border to Germany, see Fig 5.4. The Norwegian region NO2 is in the soutwestern part of Norway, limited by the 420 kV transmission line from Hasle to Rød, crossing the Oslo Fiord. Further, region NO2 is bordering NO5 by the 300 kV line Vemork-Flesaker.

The Danish grid is relying on wind power and it is the dominating source of energy, with 50.5% contribution to the Danish energy market. Norway's installed power is 3.4 GW, and 94.3% is hydropower. More than 1000 dams results in a storage capacity of 86.5 TWh, and this is considered as Europe's green battery in Norway.

The electric grid is joined by subsea DC-cable between Kristiansand in region 'NO2', of southern Norway, and Tjele on Jutland in Denmark in region 'DK1', with a total capacity of 1400 MW. It has been illustrated in Fig. 5.4.

The NordPool Market web service (www.nordpoolgroup.com) offers data of hourly values from 2013 to present date. The values are given in €/Mwh, which makes it effective compare the prices of the different regions of the NordPool market.

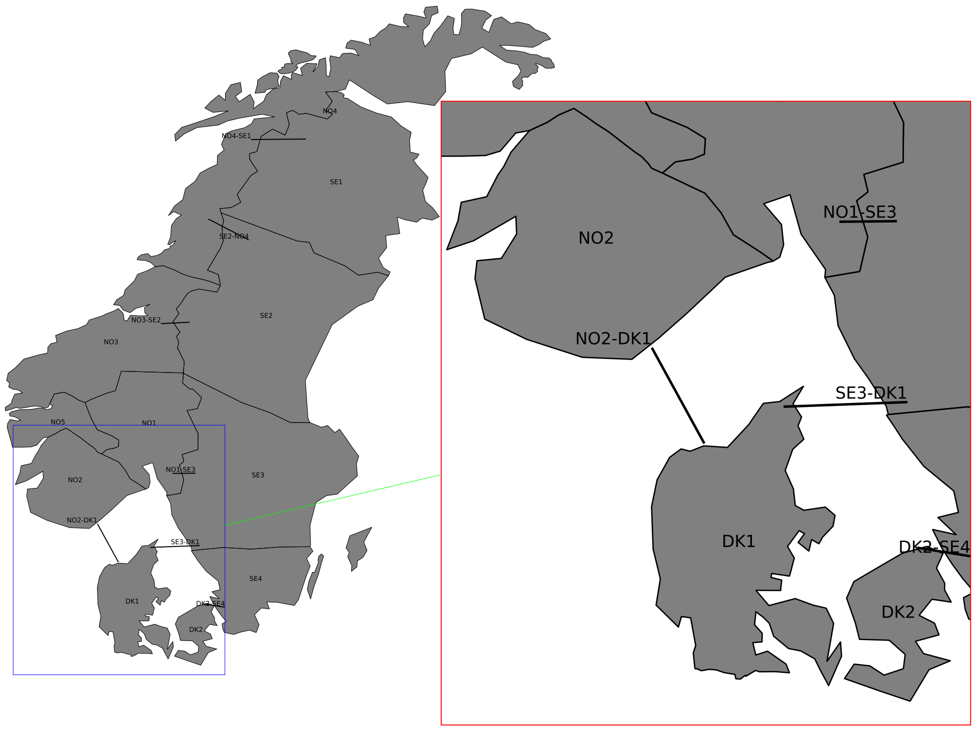


Figure 5.4: Overview of Nordpool Sandinavian Market regions, whith considered regions in frame.

Chapter 6

Summary of Results

In this chapter, the key findings are summarised from the research work which have been disseminated in the peer-reviewed published papers (Part II).

In **Paper A** the collected real-time data are analysed for predictions of energy consumption to manage and operate renewable energy based distribution network planning from a flexibility point of view. A novel vertical time approach are presented to help close the gap to the many methods on energy prediction mentioned in the literature, that utilise continuous time approach. **Paper A** investigates urban area electrical energy demand prediction with weather parameters and analyses their correlations. kNN, Random Forest and Linear Regression is analysed and evaluated both by using continuous and vertical time approach. It is observed that for 30 minutes predictions the Random Forest Regression has the best results, shown by a mean absolute percentage error (MAPE) in the range of 1-2 %. kNN show best results for the day-ahead forecasting with a best results with a MAPE of 2.61 %. The results, using vertical time approach, outperforms compared to continuous time approach. The results can compete with the more complex neural networks, with less amount of data. The regression tools for urban electric load forecasting have been presented in **Paper F**, and results from k-fold cross-validation are analysed, which have been further elaborated in the **Paper A**.

In **Paper B**, event-based demand prediction with impact of external parameters (e.g meteorological parameters, etc.) have been investigated using vertical time approach considering the seasonal impact of the external parameters. The **Paper B** investigates the potential of the pre-processing stage. To bridge this research gap, domain knowledge has been applied together with algorithmic development. The proposed solutions has been sought for in the pre-processing stage, developed in the novel vertical time approach, and by refining techniques from the domain of statistics and time series to be included in the model. The regression techniques are systematically investigated for smart load energy prediction analysis and correlating it to other impacting parameters. The techniques are developed with the knowledge of research mentioned in the literature, that utilise continuous time approach together with complex neural networks which requires huge amount of data. **Paper B** presents a systematic load synthesis and prediction analysis of rural areas seasonal occupancy for network expansion planning and integrated renewable energy sources. The work in **Paper B** is based on the findings in **Paper G**, where load analysis of a typical Nordic rural area with holiday cabins is done. **Paper I** is linked to **Paper B** and **Paper G**, through rural area energy consumption data. Three models are evaluated in **Paper G**, a base model, a heterogeneous model, and a homogeneous model. The base model uses k-nearest neighbor regression, random forest regression, and linear regression. In the heterogeneous model, XGBoost regression, LightGBM regression, and random forest regression are combined. The homogeneous model has three layers of linear regression. The meta estimator for all three models is linear regression. The performance is relative to the season, and it is consistent for both algorithms. Both algorithms perform best on the season with the highest granularity of data, meaning the season in which the electrical load consumption is the highest.

Paper C analyses dealing with the problems of irregularities and randomness in the time series considering urban and rural area case studies and it can be used for network topology optimisation. The work in **Paper C** uses both case studies presented in **Paper A** and **Paper B** into consideration, and presents the overall methodology and results from the two mentioned papers.

Paper D further investigates and analyses the pre-processing stage for Recurrent Neural Networks (RNN), and finds that reducing the dimensionality through principal components analysis (PCA) for improving the predictive performance. In the winter season, the correlations to weather parameters are higher than other seasons, as well as in general the winter season has a higher load demand. These are factors explaining the lower MAPE in winter season as opposed to other seasons. In the case of Gated Recurrent Unit (GRU) networks, the results for all the seasons are improved through PCA. Also for the Vanilla RNN, there is a benefit from reduced number of principal components in a lesser MAPE, and for the summer test on a week in July, it scores best of the compared RNNs. Yet for the Long Short-Term Memory (LSTM), it does not benefit from an improved MAPE from the PCA. The best results are measured in January, when also the electrical load demand is at the highest, and the impact of external weather parameters is influencing greatly on the load demand. The curvature of the load profile is dominated by a high peak at noon, and GRU captures it very good. The results from the week of April, has a lower load demand than January. In January, the load demand is highly correlated to the weather parameters readings in winter season. GRU with PCA achieves the best forecast MAPE result for the week in April, yet with a slightly higher MAPE than for January. This can be explained by the lower load demand in April, and that correlations to weather parameters are usually lower in spring and autumn. In the test week of October, which has the same range in load demand (6000 - 10000 MW), it is also GRU with PCA that scores best with a MAPE of 0.94. When comparing the results without and with the use of dimensionality redution through PCA, the MAPE is in the same range for Vanilla RNN (1.45 for April, and 1.38 for October), GRU (1.21 for April and 1.26 for October) and LSTM (1.25 for April, and 1.24 for October). The similarity in results from spring (observed from the test results for the week in April) and autumn (observed from the test results for the week in October) can be explained by similar load range and meteorological conditions. In the case of Vanilla RNN and GRU, the explanations of the compared results indicates the same, when investigating the results on the RNNs tested with PCA. The exception is the LSTM tested with PCA, that shows a higher MAPE. It is observed that LSTM is a more complex algorithm, than the Vanilla RNN and GRU, when it is trained with relatively lesser data. Although it is analysed using its principal components, it is not able to improve the predictions. It is observed that for the week in July with the lowest load demand, the simplest RNN (Vanilla RNN) with reduced principal components achieve the preferred MAPE, amongst all of the predictors. **Paper D** is the foundation of **Paper J** that covers Recurrent Neural Networks for electrical load forecasting to use in Demand Response.

Paper E considers abnormal operation due to various instances (e.g. random effect, intrusion, abnormal operation of smart devices, cyber-threats). In the results of kNN, iforest and LOF on urban area data and from rural region data, it is observed that the anomaly detection for the considered grid scenarios are different. For the rural region, most of the anomalies are observed in the latter timeline of the data concentrated in the last year of the collected data. For the urban area data, the anomalies are spread out over the entire timeline. The frequency of detected anomalies, where considerably higher for the rural area load demand than for the urban area load demand. Observing from case scenarios, the incidents of detected anomalies are more data driven, than exceptions in the algorithms. It is observed that there are some anomalies, where the recorded electrical load demand is zero, in the rural region dataset that the iforest and LOF did not detect. This was only detected by kNN. It is observed that from the domain knowledge of smart energy systems, the LOF is able to detect observations that could not have detected by visual inspection alone, in contrast to kNN and iforest. Whereas kNN and iforest excludes an upper and lower bound, the LOF is density based and separates out anomalies amidst in the data. The capability that LOF has to identify anomalies amidst the data, will together with the deep domain knowledge, be an advantage when detecting anomalies in energy system data. The results from **Paper E** is further analysed in **Paper K** on cyber security in power systems highlighting related topics, such as concept drift.

In **Paper H** analysis of Electricity Markets in Norway (NO2) and Denmark (DK1) are presented. From load and price analysis, it is shown that peaks normally happen once in the morning, and again in the evening with a valley during the night, and a shoulder leveling out the load consumption between the first peak in the morning and the second peak in the evening. The kNN-regressor surfaced with an autoregressor, is giving the finite gradient of the curvature, amd the MAPE improved. The Norwegian region is easier to predict, since it is a more stable system due to hydropower. However, the Danish system has a higher integration of peak and valleys in the dataset, hence more curvature, and therefore the algorithm improve its predictive accuracy for the Danish market, more than the Norwegian, when given the autoregressor.

Chapter 7

Concluding Remarks

The energy consumption data has been analysed using vertical axis approach, and it has been implemented and elaborated in this work. The vertical time axis approach has been proven to work effectively using instance based models, which also options for explainability and interpretability of the predictive algorithm. When regarding the computational time, which is less for instance based models that those of higher complexity (i. e. Neural Networks), added to the fact that vertical time axis approach uses less data, than the horizontal time organisation. It is observed that for both horizontal (i.e. continuous approach) and vertical approaches the correlation to weather parameters and multivariate input features helps in improving the predictive outcome. Especially introducing lower indicator variables (i.e. working/non-working day) is fruitful for improving the performance metrics. In vertical time axis approach when combined with lower indicator variables, the best performance is observed.

For rural area load demand forecasting the improvements by using vertical time axis approach is minimal, due to the low electrical load demand, hence it makes pattern recognition in general more difficult. It is observed that for areas with lower electrical load demand, aspects from statistics and time series analysis (such as autoregression and autocorrelation) improves the predictive outcome. In addition to improving the predictions it also gives a rapid response to the best input feature from the historical load values to use as input features.

This work has shown that the instance based models can compete with models of higher complexity, yet some methods in preprocessing (such as circular coding) does not function for an instance based learner such as kNearest Neighbor, and hence kNN can not option for this kind of complexity even in the feature engineering of the model.

It will be interesting for the future work of electrical load forecasting to develop solution that combines a high complexity in the feature engineering and have the explainability of instance based models. The short-term energy forecasting can be used for analysing possible abnormal operation of electrical energy network including cyber-intrusions. Day-ahead energy forecasting combining with energy market forecasting can be further investigated for finding more accurate demand considering user behaviour.

Bibliography

[1] Iain Soutar. Dancing with complexity: Making sense of decarbonisation, decentralisation, digitalisation and democratisation. *Energy Research Social Science*, 80:102230, 2021.

[2] Maria Luisa Di Silvestre, Salvatore Favuzza, Eleonora Riva Sanseverino, and Gaetano Zizzo. How decarbonization, digitalization and decentralization are changing key power infrastructures. *Renewable and Sustainable Energy Reviews*, 93:483–498, 2018.

[3] IEA. Key world energy statistics 2020. `https://www.iea.org/reports/key-world-energy-statistics-2020`, 2020. Accessed: 2021-12-14.

[4] IEA. Greenhouse gas emissions from energy – data explorer. `https://www.iea.org/articles/greenhouse-gas-emissions-from-energy-data-explorer`, 2021. Accessed: 2021-12-14.

[5] Hassan Haes Alhelou, Ehsan Heydarian-Forushani, and Pierluigi Siano. *Flexibility in Electric Power Distribution Networks*. CRC Press, 2021.

[6] Sinan Küfeoglu, Gaomin Liu, Karim Anaya, and Michael Pollitt. Digitalisation and new business models in energy sector. 2019.

[7] Vitor N. Coelho, Miri Weiss Cohen, Igor M. Coelho, Nian Liu, and Frederico Gadelha Guimarães. Multi-agent systems applied for energy systems integration: State-of-the-art applications and trends in microgrids. *Applied Energy*, 187:820–832, 2017.

[8] Qiuye Sun and Lingxiao Yang. From independence to interconnection — a review of ai technology applied in energy systems. *CSEE Journal of Power and Energy Systems*, 5(1):21–34, 2019.

[9] Peyman Razmi, Mahdi Ghaemi Asl, Giorgio Canarella, and Afsaneh Sadat Emami. Topology identification in distribution system via machine learning algorithms. *PLOS ONE*, 16(6):1–20, 06 2021.

[10] S. Barja-Martinez, M. Aragüés-Peñalba, Í. Munné-Collado, P. Lloret-Gallego, E. Bullich-Massagué, and R. Villafafila-Robles. Artificial intelligence techniques for enabling big data services in distribution networks: A review. *Renewable and Sustainable Energy Reviews*, 150, 2021. Cited By :2.

[11] F. M. Dahunsi, O. A. Somefun, A. A. Ponnle, and K. B. Adedeji. Compression techniques of electrical energy data for load monitoring: A review. *Nigerian Journal of Technological Development*, 18(3):194–208, 2021.

[12] P. Boza and T. Evgeniou. Artificial intelligence to support the integration of variable renewable energy sources to the power system. *Applied Energy*, 290, 2021. Cited By :7.

[13] O. A. Omitaomu and H. Niu. Artificial intelligence techniques in smart grid: A survey. *Smart Cities*, 4(2):548–568, 2021. Cited By :7.

[14] S. Aslam, H. Herodotou, S. M. Mohsin, N. Javaid, N. Ashraf, and S. Aslam. A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renewable and Sustainable Energy Reviews*, 144, 2021. Cited By :18.

[15] M. M. Rahman, M. Shakeri, S. K. Tiong, F. Khatun, N. Amin, J. Pasupuleti, and M. K. Hasan. Prospective methodologies in hybrid renewable energy systems for energy prediction using artificial neural networks. *Sustainability (Switzerland)*, 13(4):1–28, 2021. Cited By :8.

[16] D. Mahmood, N. Javaid, G. Ahmed, S. Khan, and V. Monteiro. A review on optimization strategies integrating renewable energy sources focusing uncertainty factor – paving path to eco-friendly smart cities. *Sustainable Computing: Informatics and Systems*, 30, 2021. Cited By :5.

[17] N. B. Vanting, Z. Ma, and B. N. Jørgensen. A scoping review of deep neural networks for electric load forecasting. *Energy Informatics*, 4, 2021.

[18] Y. L. Cheng, M. H. Lim, and K. H. Hui. Impact of internet of things paradigm towards energy consumption prediction: A systematic literature review. *Sustainable Cities and Society*, 78, 2022.

[19] Tao Hong, Pierre Pinson, Yi Wang, Rafał Weron, Dazhi Yang, and Hamidreza Zareipour. Energy forecasting: A review and outlook. *IEEE Open Access Journal of Power and Energy*, 7:376–388, 2020.

[20] George Gross and Francisco D Galiana. Short-term load forecasting. *Proceedings of the IEEE*, 75(12):1558–1573, 1987.

[21] Zhenyu Wang, Jun Li, Shilin Zhu, Jun Zhao, Shuai Deng, Shengyuan Zhong, Hongmei Yin, Hao Li, Yan Qi, and Zhiyong Gan. A review of load forecasting of the distributed energy system. *IOP Conference Series: Earth and Environmental Science*, 237:042019, mar 2019.

[22] Ali Lahouar and Jaleleddine Ben Hadj Slama. Random forests model for one day ahead load forecasting. In *IREC2015 The Sixth International Renewable Energy Congress*, pages 1–6, 2015.

[23] Jinkyu Hong and Won Sup Kim. Weather impacts on electric power load: partial phase synchronization analysis. *Meteorological Applications*, 22(4):811–816, 2015.

[24] Henrique S. Hippert and Carlos Eduardo Pedreira. Estimating temperature profiles for short-term load forecasting: neural networks compared to linear models. In *IEE Proceedings - Generation, Transmission and Distribution*, 2004.

[25] J. Valenzuela, M. Mazumdar, and A. Kapoor. Influence of temperature and load forecast uncertainty on estimates of power generation production costs. *IEEE Transactions on Power Systems*, 15(2):668–674, 2000.

[26] Fahad H. Al-Qahtani and Sven F. Crone. Multivariate k-nearest neighbour regression for time series data — a novel algorithm for forecasting uk electricity demand. *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2013.

[27] R. Afkhami and F.M. Yazdi. Application of neural networks for short-term load forecasting. In *2006 IEEE Power India Conference*, pages 5 pp.–, 2006.

[28] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[29] Vasudev Dehalwar, Akhtar Kalam, Mohan Lal Kolhe, and Aladin Zayegh. Electricity load forecasting for urban area using weather forecast information. *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*, pages 355–359, 2016.

[30] C. Chu and H. H. Iu. Complex networks theory for modern smart grid applications: A survey. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 7(2):177–191, June 2017.

[31] Sandro Nižetić, Nedjib Djilali, Agis Papadopoulos, and Joel J.P.C. Rodrigues. Smart technologies for promotion of energy efficiency, utilization of sustainable resources and waste management. *Journal of Cleaner Production*, 231:565 – 591, 2019.

[32] Adeshina Y. Alani and Isaac O. Osunmakinde. Short-term multiple forecasting of electric energy loads for sustainable demand planning in smart grids for smart homes. *Sustainability*, 9(11), 2017.

[33] Nian Liu, Qingfeng Tang, Jianhua Zhang, Wei Fan, and Jie Liu. A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. *Applied Energy*, 129:336 – 345, 2014.

[34] S. Yin and O. Kaynak. Big data for modern industry: Challenges and trends point of view. *Proceedings of the IEEE*, 103(2):143–146, Feb 2015.

[35] Yang Zhang, Tao Huang, and Ettore Francesco Bompard. Big data analytics in smart grids: a review. *Energy Informatics*, 1(11), 2018.

[36] A. Songpu. External parameters contribution in domestic load forecasting using neural network. *IET Conference Proceedings*, pages 6 .–6 .(1), January 2015.

[37] Ernesto Aguilar Madrid and Nuno Antonio. Short-term electricity load forecasting with machine learning. *Information*, 12(2), 2021.

[38] Abderrezak Laouafi, Mourad Mordjaoui, Salim Haddad, Taqiy Eddine Boukelia, and Abderahmane Ganouche. Online electricity demand forecasting based on an effective forecast combination methodology. *Electric Power Systems Research*, 148:35–47, 2017.

[39] S. Mirasgedis, Y. Sarafidis, E. Georgopoulou, D.P. Lalas, M. Moschovits, F. Karagiannis, and D. Papakonstantinou. Models for mid-term electricity demand forecasting incorporating weather influences. *Energy*, 31(2):208–227, 2006.

[40] Yongxiu He, Jie Jiao, Qian Chen, Sifan Ge, Yan Chang, and Yang Xu. Urban long term electricity demand forecast method based on system dynamics of the new economic normal: The case of tianjin. *Energy*, 133:9–22, 2017.

[41] Peter Lusis, Kaveh Rajab Khalilpour, Lachlan Andrew, and Ariel Liebman. Short-term residential load forecasting: Impact of calendar effects and forecast granularity. *Applied Energy*, 205:654–669, 2017.

[42] Nils Jakob Johannesen, Mohan Lal Kolhe, and Morten Goodwin. Load demand analysis of nordic rural area with holiday resorts for network capacity planning. In *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–7. IEEE, 2019.

[43] Nils Jakob Johannesen, Mohan Kolhe, and Morten Goodwin. Deregulated electric energy price forecasting in nordpool market using regression techniques. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 1932–1938. IEEE, 2019.

[44] Muhammad Amri Yahya, Sasongko Pramono Hadi, and Lesnanto Multa Putranto. Short-term electric load forecasting using recurrent neural network (study case of load forecasting in central java and special region of yogyakarta). In *2018 4th International Conference on Science and Technology (ICST)*, pages 1–6, 2018.

[45] Lei Zhang, Linghui Yang, Chengyu Gu, and Da Li. Lstm-based short-term electrical load forecasting and anomaly correction. In *E3S Web of Conferences*, volume 182, page 01004. EDP Sciences, 2020.

[46] Kunjin Chen, Kunlong Chen, Qin Wang, Ziyu He, Jun Hu, and Jinliang He. Short-term load forecasting with deep residual networks. *IEEE Transactions on Smart Grid*, 10(4):3943–3952, 2019.

[47] Weixian Li, Thillainathan Logenthiran, and Wai Lok Woo. Multi-gru prediction system for electricity generation's planning and operation. *IET Generation, Transmission & Distribution*, 13(9):1630–1637, 2019.

[48] Hermine N. Akouemo and Richard J. Povinelli. Probabilistic anomaly detection in natural gas time series data. *International Journal of Forecasting*, 32(3):948–956, 2016.

[49] Hermine N. Akouemo and Richard J. Povinelli. Time series outlier detection and imputation. *2014 IEEE PES General Meeting — Conference & Exposition*, pages 1–5, 2014.

[50] Xue Liu, Yong Ding, Hao Tang, and Feng Xiao. A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data. *Energy and Buildings*, 231:110601, 2021.

[51] Yassine Himeur, Abdullah Alsalemi, Faycal Bensaali, and Abbes Amira. Smart power consumption abnormality detection in buildings using micromoments and improved k-nearest neighbors. *International Journal of Intelligent Systems*, 36(6):2865–2894, 2021.

[52] Hadis Karimipour, Sandra Geris, Ali Dehghantanha, and Henry Leung. Intelligent anomaly detection for large-scale smart grids. In *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, pages 1–4, 2019.

[53] Youbiao He, Gihan J. Mendis, and Jin Wei. Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism. *IEEE Transactions on Smart Grid*, 8(5):2505–2516, 2017.

[54] Manikant Panthi. Anomaly detection in smart grids using machine learning techniques. In *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*, pages 220–222, 2020.

[55] João Gama, Indrundefined Žliobaitundefined, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM Comput. Surv.*, 46(4), mar 2014.

[56] Giuseppe Fenza, Mariacristina Gallo, and Vincenzo Loia. Drift-aware methodology for anomaly detection in smart grid. *IEEE Access*, 7:9645–9657, 2019.

[57] Alexandra L'heureux, Katarina Grolinger, Hany F Elyamany, and Miriam AM Capretz. Machine learning with big data: Challenges and approaches. *Ieee Access*, 5:7776–7797, 2017.

[58] Song-Qinbao Du, Lei and Xiaolin. Jia. 'detecting concept drift: An information entropy based method using an adaptive sliding window'. *Intelligent Data Analysis*, 18(3):pp. 337–364, june 2014.

[59] Husheng Guo, Hai Li, Qiaoyan Ren, and Wenjian Wang. Concept drift type identification based on multi-sliding windows. *Information Sciences*, 585:1–23, 2022.

[60] Rongbin Xu, Yongliang Cheng, Zhiqiang Liu, Ying Xie, and Yun Yang. Improved long short-term memory based anomaly detection with concept drift adaptive method for supporting iot services. *Future Generation Computer Systems*, 112:228–242, 2020.

[61] Nils Jakob Johannesen, Mohan Lal Kolhe, and Morten Goodwin. Comparing recurrent neural networks using principal component analysis for electrical load predictions. In *2021 6th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–6. IEEE, 2021.

[62] Mostafa Mohammadpourfard, Yang Weng, Mykola Pechenizkiy, Mohsen Tajdinian, and Behnam Mohammadi-Ivatloo. Ensuring cybersecurity of smart grid against data integrity attacks under concept drift. *International Journal of Electrical Power Energy Systems*, 119:105947, 2020.

[63] Michael L. Bender, Bruno Malaize, Joseph Orchardo, Todd Sowers, and Jean Jouzel. *High Precision Correlations of Greenland and Antarctic Ice Core Records Over the Last 100 Kyr*, pages 149–164. American Geophysical Union (AGU), 1999.

[64] Federal Energy Regulatory Commission. Reports on demand response and advanced metering. `https://www.ferc.gov/industries-data/electric/power-sales-and-markets/demand-response/reports-demand-response-and`, 2021. Accessed: 2021-12-16.

[65] Aileen Nielsen. *Practical Time Series Analysis : Prediction with Statistics and Machine Learning*. O'Reilly Media, Inc, USA, 2019.

[66] George Udny Yule. Vii. on a method of investigating periodicities disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 226(636-646):267–298, 1927.

[67] Angeliki Menegaki. Chapter 2 - stationarity and an alphabetical directory of unit roots often used in the energy-growth nexus. In Angeliki Menegaki, editor, *A Guide to Econometrics Methods for the Energy-Growth Nexus*, pages 31–61. Academic Press, 2021.

[68] J. Franklin. *The Science of Conjecture: Evidence and Probability Before Pascal*. A Johns Hopkins paperback. Johns Hopkins University Press, 2015.

[69] Parsimony (in as few words as possible). `https://philosophynow.org/issues/81/Parsimony_In_as_few_words_as_possible`. Accessed: 2021-12-25.

[70] D. Patidar, Bhavin C. Shah, and M. Mishra. Performance analysis of k nearest neighbors image classifier with different wavelet features. *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, pages 1–6, 2014.

[71] Sholom M. Weiss and Nitin Indurkhya. Rule-based machine learning methods for functional prediction. *J. Artif. Int. Res.*, 3(1):383–403, December 1995.

[72] T Mitchell. Machine learning. new york: Mcgraw-hill, 1997.

[73] Leo Breiman. Random forests. *Springer, Machine Learning*, 2001.

[74] Stasys Jukna. *The Entropy Function*, pages 313–326. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[75] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189 – 1232, 2001.

[76] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[77] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, page 214–226, New York, NY, USA, 2012. Association for Computing Machinery.

[78] Sorelle A. Friedler, Carlos Scheidegger, Suresh Venkatasubramanian, Sonam Choudhary, Evan P. Hamilton, and Derek Roth. A comparative study of fairness-enhancing interventions in machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 329–338, New York, NY, USA, 2019. Association for Computing Machinery.

[79] Mehdi Khashei and Mehdi Bijari. An artificial neural network (p,d,q) model for timeseries forecasting. *Expert Systems with Applications*, 37(1):479–489, 2010.

[80] Ahmed Tealab. Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal*, 3(2):334–340, 2018.

[81] Zhe Wu, David Rincon, Quanquan Gu, and Panagiotis D. Christofides. Statistical machine learning in model predictive control of nonlinear processes. *Mathematics*, 9(16), 2021.

[82] Ehdieh Khaledian, Shikhar Pandey, Pratim Kundu, and Anurag K. Srivastava. Real-time synchrophasor data anomaly detection and classification using ¡italic¿isolation forest, kmeans, and loop¡/italic¿. *IEEE Transactions on Smart Grid*, 12(3):2378–2388, 2021.

[83] Wei Mao, Xiu Cao, Qinhua zhou, Tong Yan, and Yongkang Zhang. Anomaly detection for power consumption data based on isolated forest. In *2018 International Conference on Power System Technology (POWERCON)*, pages 4169–4174, 2018.

[84] Armin Aligholian, Mohammad Farajollahi, and Hamed Mohsenian-Rad. Unsupervised learning for online abnormality detection in smart meter data. In *2019 IEEE Power Energy Society General Meeting (PESGM)*, pages 1–5, 2019.

[85] Saeed Ahmed, Youngdoo Lee, Seung-Ho Hyun, and Insoo Koo. Unsupervised machine learning-based detection of covert data integrity assault in smart grid networks utilizing isolation forest. *IEEE Transactions on Information Forensics and Security*, 14(10):2765–2777, 2019.

[86] Yifan Ou, Bin Deng, Xuan Liu, and Ke Zhou. Local outlier factor based false data detection in power systems. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 2003–2007, 2019.

[87] A. Azeem, I. Ismail, S. M. Jameel, and V. R. Harindran. Electrical load forecasting models for different generation modalities: A review. *IEEE Access*, 9:142239–142263, 2021.

[88] A. Román-Portabales, M. López-Nores, and J. J. Pazos-Arias. Systematic review of electricity demand forecast using ann-based machine learning algorithms. *Sensors*, 21(13), 2021. Cited By :1.

[89] R. Dashti, M. Daisy, H. Mirshekali, H. R. Shaker, and M. Hosseini Aliabadi. A survey of fault prediction and location methods in electrical energy distribution networks. *Measurement: Journal of the International Measurement Confederation*, 184, 2021. Cited By :4.

[90] A. Moradzadeh, A. Mansour-Saatloo, M. Nazari-Heris, B. Mohammadi-Ivatloo, and S. Asadi. *Introduction and Literature Review of the Application of Machine Learning/Deep Learning to Load Forecasting in Power System.* Power Systems. 2021.

[91] B. Völker, A. Reinhardt, A. Faustine, and L. Pereira. Watt's up at home? smart meter data analytics from a consumer-centric perspective. *Energies*, 14(3), 2021. Cited By :7.

[92] J. Dhanalakshmi and N. Ayyanathan. A systematic review of big data in energy analytics using energy computing techniques. *Concurrency and Computation: Practice and Experience*, 34(4), 2022.

[93] Alan Pankratz. *Forecasting with univariate Box-Jenkins models: Concepts and cases.* John Wiley & Sons, 2009.

[94] Tasarruf Bashir, Chen Haoyong, Muhammad Faizan Tahir, and Zhu Liqiang. Short term electricity load forecasting using hybrid prophet-lstm model optimized by bpnn. *Energy Reports*, 8:1678–1686, 2022.

[95] Y. H. Kareem and A. R. Majeed. Monthly peak-load demand forecasting for sulaimany governorate using sarima. In *2006 IEEE/PES Transmission Distribution Conference and Exposition: Latin America*, pages 1–5, 2006.

[96] Jeffrey L Elman. Distributed representations, simple recurrent networks, and grammatical structure. *Machine learning*, 7(2):195–225, 1991.

[97] Baigang Du, Qiliang Zhou, Jun Guo, Shunsheng Guo, and Lei Wang. Deep learning with long short-term memory neural networks combining wavelet transform and principal component analysis for daily urban water demand forecasting. *Expert Systems with Applications*, 171:114571, 2021.

[98] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1):37–52, 1987. Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.

[99] Hervé Abdi and Lynne J. Williams. Principal component analysis. *WIREs Computational Statistics*, 2(4):433–459, 2010.

[100] Kaile Zhou, Chao Fu, and Shanlin Yang. Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56:215–225, 2016.

[101] Openweathermap.org. Detailed pricing - current weather and forecasts collection. https://openweathermap.org/full-price, 2022.

[102] Rui Zhang, Yan Xu, Zhao Yang Dong, Weicong Kong, and Kit Po Wong. A composite k-nearest neighbor model for day-ahead load forecasting with limited temperature forecasts. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pages 1–5, 2016.

[103] Stan Aronoff. Geographic information systems: A management perspective. *Geocarto International*, 4(4):58–58, 1989.

[104] Mordechay Haklay, Eran Feitelson, and Yerahmiel Doytsher. The potential of a gis-based scoping system: An israeli proposal and case study. *Environmental Impact Assessment Review*, 18(5):439–459, 1998.

[105] AEMO. National electricity market data - nem, 2021.

[106] Vasudev Dehalwar, Akhtar Kalam, Mohan Lal Kolhe, and Aladin Zayegh. Electricity load forecasting for urban area using weather forecast information. In *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*, pages 355–359, 2016.

[107] Australian Energy Regulator. Annual electricity consumption - nem. https://www.aer.gov.au/wholesale-markets/wholesale-statistics/annual-electricity-consumption-nem, 2021.

[108] NVE. Bjønntjønn kraftverk og pumpestasjon. https://www.nve.no/konsesjon/konsesjonssaker/konsesjonssak?id=8386type=V-1, 2021.

[109] NVE. Tesla ladestasjon i treungen. https://www.nve.no/konsesjon/konsesjonssaker/konsesjonssak/?id=7606type=A-1, 2021.

[110] Jarle Pedersen. Tesla bygger ny supercharger i telemark. https://www.tvedestrandsposten.no/tesla-bygger-ny-supercharger-i-telemark/s/5-50-1252227, 2022.

[111] Camilla Moen. Planlegger gigantutbygging med 445 nye hytter på gautefallheia: - noen mener det er kontroversielt. https://www.kv.no/planlegger-gigantutbygging-med-445-nye-hytter-pa-gautefallheia-noen-mener-det-er-kontroversielt/s/5-63-271204, 2021.

[112] M. Holstad T. Aanensen. Supply and consumption of electricity in the period 1993-2016. https://www.ssb.no/en/energi-og-industri/artikler-og-publikasjoner/supply-and-consumption-of-electricity-in-the-period-1993-2016, 2017.

[113] Mohan L. Kolhe, K.M. Iromi Udumbara Ranaweera, and A.G.B. Sisara Gunawardana. Techno-economic sizing of off-grid hybrid renewable energy system for rural electrification in sri lanka. *Sustainable Energy Technologies and Assessments*, 11:53–64, 2015.

[114] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030 – 1081, 2014.

[115] Jesus Lago, Fjo [De Ridder], and Bart [De Schutter]. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, 221:386 – 405, 2018.

Part II
Appended Papers

Appendix A

Paper A - Relative Evaluation of Regression Tools for Urban Area Electrical Energy Demand Forecasting

Nils Jakob Johannesen and Mohan Kolhe and Morten Goodwin

Faculty of Engineering and Science, University of Agder, PO Box 422, NO 4604 Kristiansand, Norway.

Abstract **- Load forecasting is the most fundamental application in Smart-Grid, which provides essential input to Demand Response, Topology Optimization and Abnormally Detection, facilitating the integration of intermittent clean energy sources. In this work, several regression tools are analyzed using larger datasets for urban area electrical load forecasting. The regression tools which are used are Random Forest Regressor, k-Nearest Neighbor Regressor and Linear Regressor. This work explores the use of regression tool for regional electric load forecasting by correlating lower distinctive categorical level (season, day of the week) and weather parameters. The regression analysis has been done on continuous time basis as well as vertical time axis approach. The vertical time approach is considering a sample time period (e.g seasonally and weekly) of data for four years and has been tested for the same time period for the consecutive year. This work has uniqueness in electrical demand forecasting using regression tools through vertical approach and it also considers the impact of meteorological parameters. This vertical approach uses less amount of data compare to continuous time-series as well as neural network techniques. A correlation study, where both the Pearson method and visual inspection, of the vertical approach depicts meaningful relation between pre-processing of data, test methods and results, for the regressors examined through Mean Absolute Percentage Error (MAPE). By examining the structure of various regressors they are compared for the**

lowest MAPE. Random Forest Regressor provides better short-term load prediction (30 min) and kNN offers relatively better long-term load prediction (24 hours).

Keywords - *Electrical Energy Demand Forecasting, Impact of meteorological parameters on demand forecasting, Smart-Grid management, Machine Learning, Regression Tools, Random Forest Regressor, k-Nearest Neighbor Regressor, Linear Regressor.*

A.1 Introduction

Urban area electrical energy demand forecasting is necessary for optimizing the electrical power generation scheduling in coordination with distributed generators including intermittent renewable energy sources. It will also be beneficial for demand side management considering grid constraints. In the literature, most of the electrical energy prediction studies are using shallow neural networks and support vector [2, 3]. Popular stochastic models, such as hidden Markov models, are also used for energy prediction [4, 5]. In the EU FP7 SEMIAH (Scalable Energy Management Infrastructure for Aggregation of Households) project [5], the domestic demand has been predicted using a two-stage linear stochastic optimization for managing operation of non-critical power intensive loads (for example, thermal load).

Recent research from 2018 on Computational Intelligence Approaches for Energy Load Forecasting [6], that reviewed more then 50 research papers related to the subject outlines the complexity of demand patterns as potentially influenced by factors like climate, time periods, holiday or working days and other factors such as social activities, economic factors including power market policies. Electrical energy demand is influenced by meteorological weather conditions, therefore it is necessary to include the impact of meteorological weather parameters in electrical energy demand forecasting also renewable electrical energy production is nature dependent. The future electrified grid will increasingly depend on renewable intermittent energy sources (solar, wind), and the individual load profiles of such a system will change radically as home appliances includes new energy demanding appliances (e.g. heat pump, electric vehicles and induction stove) [7]. The new electrified grid is Smart Grid System, as it is a complex whole of two-way communication aided by intelligent agents. The information will be used to provide demand side management such as peak shaving, where non-critical load demands are shifted to other periods where the stress on the grid is less intense. Electric load forecasting by machine learning will be useful in the operation of load shifting, with an accurate prediction of the load demand. Machine learning falls into two categories of Supervised, where the data points have a known value, and Unsupervised, where data points have unknown outcome. The types of supervised learning is divided into regression and classification. The first where the outcome is continuous (numerical), the latter categorical. Regression models considers the relationship to independent variables, predictors, and a dependent variable, known as target.

The regression models k-Nearest Neighbor (kNN), Linear Regression (LR), and Random Forests (RF) are supervised machine learning algorithms with a numerical outcome. The model is trained to find rules for pattern recognition in the input to output relation. The input to the model are known as features. Neural Networks is the preferred machine learning tool and are known as both feedforward and backpropagating networks, where a number of inputs are weighted in order to provide a predictive outcome. Neural networks are good for detecting non-linearities and therefore preferred as a predictive tool in electrical load forecasting, yet also often criticized for low transparency and lack of interpretability because of the black box approach, and using large amount of data. Overfitting is still a challenging issue when applying Neural Networks to electrical demand prediction [8]. The literature distinguishes between short term prediction and long term prediction time. In this article short term is defined as the 30 minute prediction time interval, and long term prediction is defined as 24 hour time prediction interval.

Urban area load is influenced by meteorological conditions therefore it is important to include impact of weather parameters on load prediction, yet this impact is governed by the prediction time, greater for long term and decreases as the prediction time is narrowed. The electrical energy demand is influenced by the user behavior as well as weather conditions. Individual human behavior and weather are so random that a complex neural network would not predict the outcome better then a coin toss. Hence, if one has to analyze the load demand of larger area like the urban area, systematic load behavior with correlation to weather parameters and continuous load profile, should be investigated.

This work has uniqueness in electrical demand forecasting using regression tools through vertical approach and it also considers the impact of meteorological parameters. This vertical approach uses less amount of data compare to continuous time-series as well as neural network techniques.

The objectives of this work are to explore the use of regression tools for regional electrical load forecasting by correlating lower distinctive categorical levels (season, day of the week) and weather parameters. The vertical time approach is considering a sample time period (e.g seasonally and weekly) of data for four years and has been tested for the same time period for the consecutive year. A vertical axis approach, showed to be competitive to Artificial Neural Networks (ANN), with a low amount of data.

The paper is organized as follows: Review of electrical load forecasting is presented in Section 2. In Section 3, various parameters (e.g. weather parameters, seasonal impact and time as well as random effects) are discussed for urban area electrical energy demand forecasting. Section 4 shows analysis both by Pearson correlation method and visual inspection to find correlation of meteorological parameters and previous load patterns on Urban Area Load Forecasting and shows the Regressor Model and gives regression model analysis. Results and Discussions are provided in Section 5. Finally, in Section 6, the conclusions are presented.

A.2 Review of electrical load forecasting

In most of the work, hourly electrical energy predictions are considered. It is important to have precise prediction for short-term (e.g. 30 min) using less amount of data as well as for long-term (e.g. 24 hours) for urban area electrical demand for electrical power generation coordination. The small area of Tunis (with only installed capacity of 4425 MW) is considered for analysis of load prediction with seasonal variations [22]. A variation in load due to season is only once a year during heat wave in the summer. For training set they have used horizontal time-series approach, where almost 10 months (more than 14400 datapoints) of training was used for testing on one week. According to Lahouar and Slama (2015) [22], who used random forest for day-ahead load forecast for the Tunisian market with historical data from 2009-2014, they obtained an average MAPE of 2.24% when crediting for the next 24 hours. Presented method of [22] does not improve, when predicting for the heat wave season, as the average MAPE for heat wave period (7-13 July) has increased to 2.6899%. During the Arabic spring in Tunis 2011, Tunis experienced a Random effect caused by a much lower power demand during the Tunisian Revolution, the MAPE for some 24 hour intervals of prediction as high as 19.61%. It was even worse during the Blackout of August 2014 where the MAPE rose to 398.09%. This show the machine learning algorithms inabilities in forecasting rare events. [22] also makes a comparison with ANN, and for the testing period of 7-13 of July it scores 2.9140 MAPE. They state that the main advantage with Random Forest over other methods is that there are few hyper-parameters to set and generalize by saying default settings is normally enough to compete with ANN and Support Vector Machine (SVM)/Support Vector Regressor (SVR), which accuracy depends on the tuning of their hyper-parameters. In our work we have used the experiences from Tunis to understand the random effects and their input on electrical energy demand forecasting as well as the understanding of hyper-parameters.

Jinkyu and Sup (2015) [23] recognizes artificial intelligence techniques like ANN or Kalman filter, to show promising results in the load forecasting predictions, although the hidden structures in AI might limit the understanding of the complex spatiotemporal developments in correlation between meteorological conditions and electricity demand. Electrical load demand and the temperature effects have been studied and short term load forecast needs to take temperature effects into consideration for day-ahead predictions. In the very short time load prediction the time scale is to short for the temperature to have any effect, and in the long run the effect tend to even out [24, 25]. On the load forecasting for the UK electricity demand Al-Qahtani and Crone (2013), proposes a multivariate k-NN approach that, opposed to the univariate model that does not take into account the underlying sub-categories of the calender, create a binary dummy variable where dt = 1 for all nonworking days and dt = 0 for working days. The load forecast MAPE of both univariate and multivariate show improved results by the use of dummy variables. A MAPE of 2.3284 was found using the univariate model, and a 1.8133 was found using the multivariate model. The dataset contained data for more then 7 years (2001-2008). The complete year of 2004 was used for training and 2005 used for validation [26]. Based on their research we developed the relevance of doing multiple correlation analysis with different time factors, where we can observe that meteorological parameters increase their importance on the prediction output as time window increases. In this context we regarded the work of Afkhami and Yazdi who proposes a way to quantisize the day into 3 periods of 8 hours for neural networks to enhance their performance [27].

Local Interpretable Model-Agnostic Explanations (LIME) aims to reflect the behavior in proximity to the predicted outcome, and does so by offering a interpretation that can explain doubts about the model. By explaining here means to provide some mean of qualitative understanding in the relation between a decision making and the predictive outcome. In medical diagnosis LIME highlights what features in the dataset that led to the prediction, and what was evidential against it [28].

ANN studies have shown an MAPE of 1.9, resulting in a Mean Absolute Error (MAE) of 167.91 MW, based on training data for a whole year. The research includes studies of temperature effects and introduces two threshold values where the load and temperature exhibits close correlation, at below 10 degrees Celsius due to heating, and above 23 degrees because of cooling [29].

The focus of this work is to verify the regression tools for electrical energy demand forecasting and we have not considered the prediction from the supply side. We considered regional area electrical energy demand forecasting with impact of weather parameters. And we have the availability of the required data for mentioned period.

A.3 Urban Area Electrical Demand Forecasting

The purpose has been to test the regression tools on the available real data. Urban area electrical energy demand forecasting is very important for generation scheduling, as well as effectively taking contribution from renewable energy sources and demand side management. Urban area electrical energy demand predictions for short term (30 min) and long term (24 hours) are necessary for scheduling power generation units as well as participating them in short term and day ahead energy market. When predicting the electrical load demand for a particular time window, in this case the next 30 minutes or 24 hours, the machine learning algorithms search for patterns and rules for the predictive outcome in the Supervised category with a continuous numerical output.

The following three parameters are important for system electrical energy demand:
(i) Time
(ii) Weather
(iii) Random effects
The seasonal patterns are repeating with the same upper and lower limits (e.g repeating on annual basis) and therefore considered as no economic effects are influencing the load behavior in the urban area of Sydney during the years 2006-2010. When investigating the Sydney dataset, see Figure A.1, we find that the load curves, yet containing cyclic

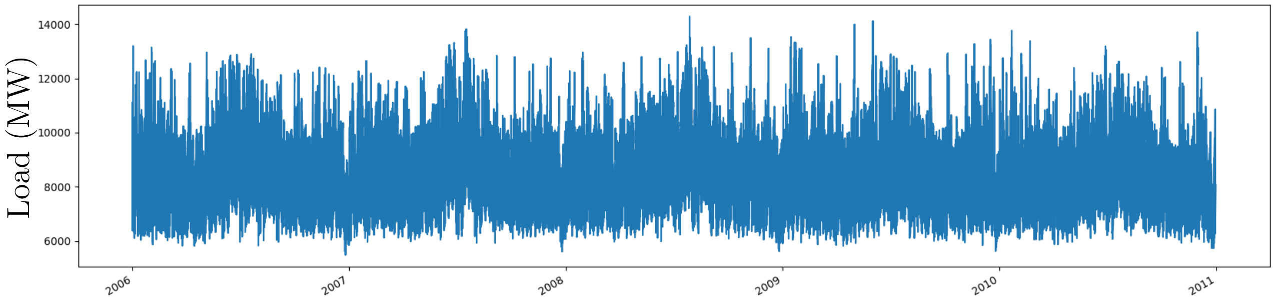


Figure A.1: Load curve of Sydney dataset containing five years of half hour values.

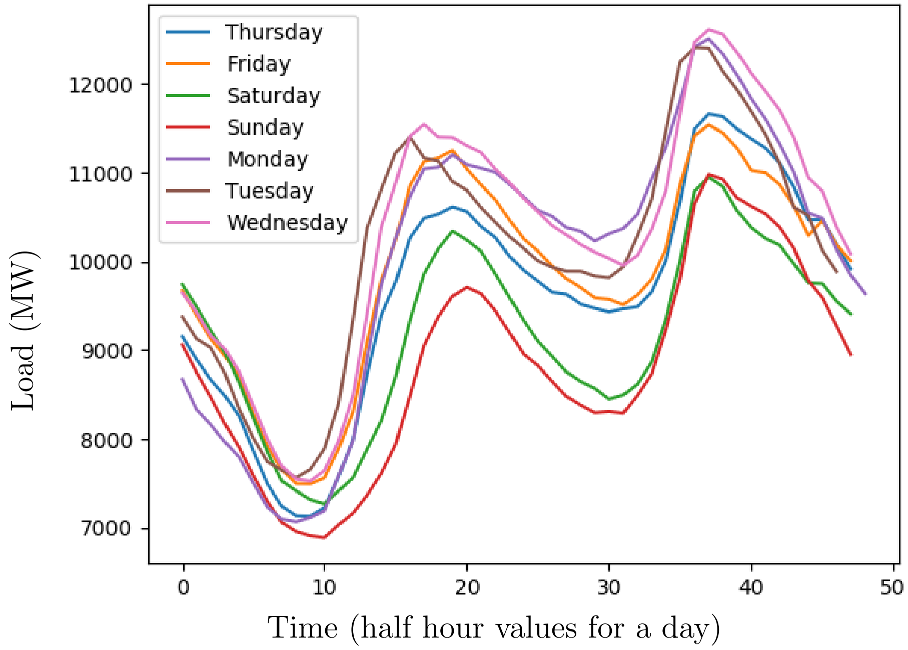
and seasonal differences, do not contain significant changes on the system load due to changing economic trends [57].

A.3.1 Time

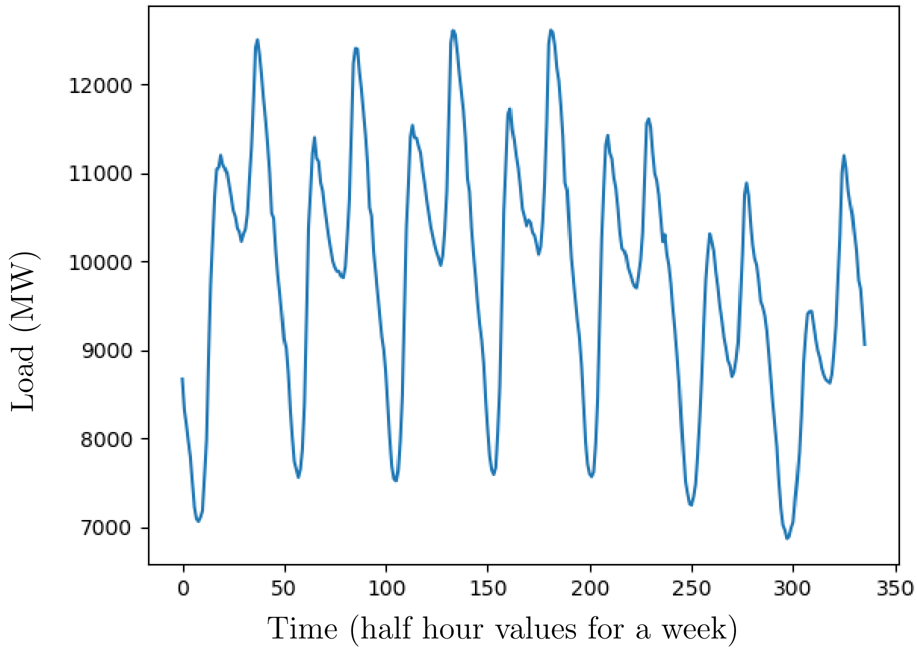
Apart from the seasonal effects, shown in Figure A.1, underlying patterns emerges in the system load demand. There are different peaks throughout the seasons, whether it is a winter peak or a summer peak. Emerging under this seasonal patterns are daily- and a weekly-cycles. The daily routines of human behavior are manifested in systematic load patterns on a daily basis. Day of the week is also significant. Public buildings and offices demands large amount of electrical load and whether it is a working day or not, influences load patterns.

When inspecting the daily- and weekly-cycle in Figure A.2, we can clearly see a load pattern emerging from a very low activity during the early hours of the day, into one peak at morning (between 8-10 hours), and another peak in the evening (between 19-21 hours) in Subfigure A.2a. The same daily repeating pattern, with a low activity followed by two peaks, are also evident in the weekly cycle, seen in Subfigure A.2b, except for that the two last days of the week (Saturday and Sunday) the peaks and general load is lower. It can be seen that urban area load predominantly reflects the domestic load and it can be correlated to human behavior. The periodicity in the load patterns reveal a load demand that reflects consumer-lifestyle.

The periodicity reflected in the daily load curve is significant in weekly cycles as well as monthly, seasonal and yearly load curves, as seen in Figure A.1 and A.2. Sub-categorical levels like working/non-working days are referred to in the literature as an indicator variable. In this work the time has been used as a variable which can be categorized as day of the week or working/non-working days or time of the day. To give this properties to our algorithms are very important as it makes prediction of forecast load more efficient [48]. The use of such type of variables has been successfully employed in electric market forecasting in the Tunis as well as the UK [22, 26].



(a) Daily cycle



(b) Weekly cycle

Figure A.2: Load patterns in daily- and weekly- cycles

A.3.2 Weather

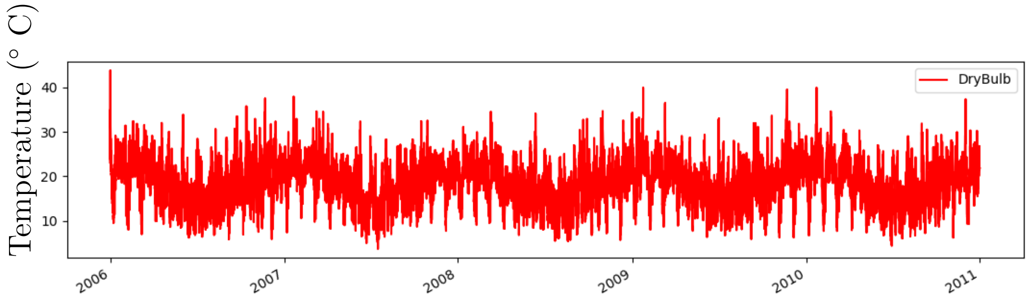


Figure A.3: DryBulb temperature curve of Sydney dataset containing five years of half hour values.

The features enlisted in the Sydney dataset, has two time indicators 'Date' and 'Hour', four weather parameters, information about the electricity price, 'ElecPrice' and information about the electricity load consumption, 'SYSLoad', these features have been developed in the pre-processing to match the requirements of the prediction tool, see Figure A.4.

The four weather parameters enlisted are DryBulb, DewPnt, WetBulb and Humidity. Dry Bulb Temperature (DBT) is temperature measured from the air, yet not exposed to solar radiation or moisture. Wet Bulb Temperature (WBT) is measured from a thermometer where the bulb of the measurement device is soaked by a wet cloth. As long as the air is not saturated, evaporation from the moist cloth keeps the WBT lower then the DBT. From the DBT and WBT one can then derive the relative humidity of the air and the dew point from a Mollier Chart by psychometrics.

Many electrical utilities are weather-sensitive, such as heating and air conditioning. Temperature, as well as past temperature effect on the load are important effect on the electrical demand, the temperature on a hot summer day may reach its peak after sunset due to heat buildup in the construction materials of buildings. In addition to the daily heat buildup, will a sequence of days with high temperature create new system peak.

The complexity in the control system engineering of maintaining thermal comfort as well as optimizing for energy is important to know. At the same time it is important to acknowledge that most houses are designed to resist the worst meteorological conditions[50]. There are also limitations in the heating system itself that might cause load peaks, like the inertia in the floor heating system, known as thermal lag [51].

In humid and hot places it is likely that humidity will effect the load pattern in similar ways as temperature. Humidity explains the complex relation between temperature and load, and therefore mathematical models is not enough in a thorough analysis. Humidity is the amount of water vapor in the air and might increase the gap between actual and apparent or felt temperature. When regulating temperature the body utilizes evaporate cooling, and the rate of evaporation through the skin is correlated to humidity, and because of the conductive properties of water, we feel warmer at high humid conditions.

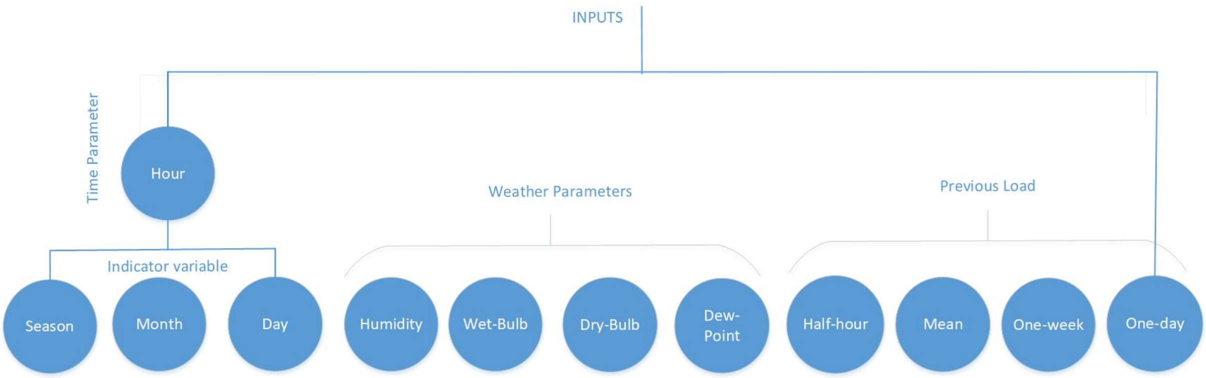


Figure A.4: Model input parameters: indicator variables, weather parameters and previous load consumption.

Also, due to the seasonal changes of weather data, the correlation to the electrical load will vary during the year.

A.3.3 Random effects

Infrastructural changes in the urban area and maintenance work are random effects that will not be detected by pattern recognition. When examining the Sydney dataset load curve as shown in Fig. A.1, there is consisting seasonal patterns. Load pattern are consistent from year to year, and show reoccurring seasonal pattern. When the yearly load curves do not vary from year to year show that there are no economic trends.

A.3.4 Relevance

It is important to investigate the main effects on the system load pattern as these are the main predictors in load forecasting.

To look for causalities in load and effect has been the topic of previous studies in load forecasting. Knowledge about the cause and effect about external parameters and system load is needed for accurate prediction. In the literature concerns have been voiced about more complicated forecast scenarios based on deregulated markets [22] and demand side management.

When energy consumers are free to choose suppliers the varying energy prices are incentives to attempt to shift non-critical load demands to periods where the stress on the grid is less intense, otherwise known as peak shaving. The other aspect is the integration of the district level environment friendly power plant, relying on intermittent renewable energy sources. In Figure 2.4, the load and temperature are plotted in the same plot. The plot will help searching for linearity among the features. The upper side of the plot forms a v-shape, separating the plot into two linear relationships at around 21°C. While the lower end has a more round u-shape.

A.4 Correlation Analysis of Electrical Load with Meteorological Parameters

Correlation is a measurement to how two ranges of data move together, and will give us an indication of how to assess feature engineering. Other means to measure the relevance between variables is Shannons concept of Mutual Information (MI), a method based in the entropy function that gives the certainty of a variable [21]. Correlation is widely used in contemporary research, where regression tools and other machine learning methods are applied to various engineering features (e.g. power transformers health index [22], emission prediction of Combined Cycle Gas Turbine [23], wind power prediction based on weather data and local terrain [24]). The Pearson Correlation Coefficient (r) computes the linear relationship between two datasets, in a range from -1 to +1. [36]. If the relationship is in the proximity of 1, it means that when x increases so does y and at exact linearity, the opposite is true for -1, it means that when a dataset is increasing the other dataset is decreasing.

$$r = r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\ \sqrt{n\sum y_i^2 - (\sum y_i)^2}} \tag{A.1}$$

One of the means to improve prediction accuracy in spite the seasonal differences, is to create a dummy variable that increases the precision of the algorithm while differentiating the seasonal changes.

A dummy variable or Indicator variable is an variable created to represent more distinct categorical level. In this paper one was made to categorize on day of week:

$$df['season'] = (df['month']\%12 + 3)// \tag{A.2}$$

The use of dummy variables has been successfully employed by forecasting on the UK electricity market, to categorize days into working and non-working days [26].

Other papers conclude what this research also experienced that the most accurate prediction comes from either from predicting on the same hour or for 24 hours, probably due to the habitual individual behaviour like showering and putting on the coffee at the same time every day [25].

A.4.1 Regressor Model

The input for the model are based on tree parameters, time, weather and previous load consumption, see Figure A.4. The time parameters are divided into sub-categories in lower categorical level as the indicating variables day of the week, working-/non-working days and season. Included are also the previous load consumption are organized by the lag method and weather parameters.

The preprocessed inputs are then computed using regression tools, in Figure A.5, represented by the k-Nearest Neighbour regressor. Figure A.5, is showing the k-Nearest

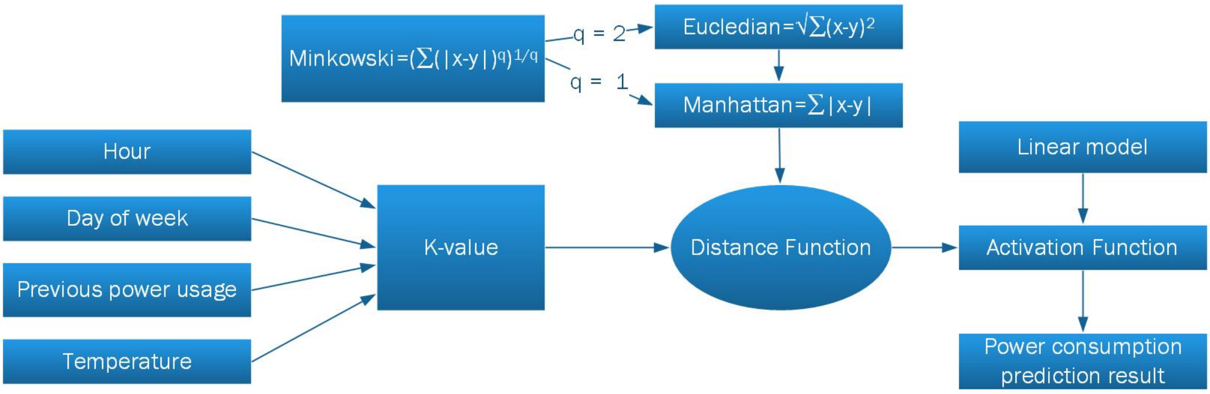


Figure A.5: The regressor model

Neighbour algorithm, where the model shows the algorithm consider a k-value and distance function based on the inputs. The regressors are taken from the scikit learn library [29], and further the hyper-parameters are tuned for optimal performance.

A.4.2 Regression Models Analysis

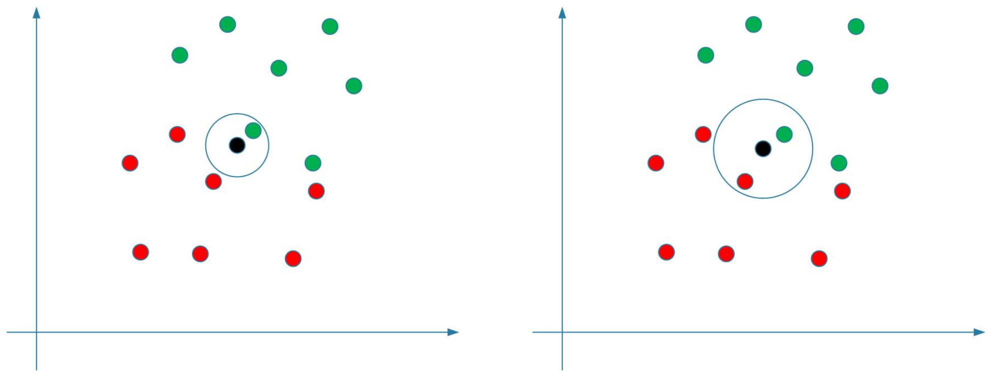
The regression-tools considered in this article are kNearestNeighborRegressor, LinearRegressor and RandomForestRegressor. To elaborate further on the model used in this research in the following the kNearestNeighborRegressor is explained: The k-Nearest Neighbour is an algorithm that computes the numerical value of the distance between given features or data points and a query point in an multi-dimensional array, and then find the point in vicinity to the query point [26].

In Figure A.5, the model takes a set of inputs, based on time, date, previous power consumption and weather parameters, based on the features of the Sydney dataset and further created.

A.4.2.1 kNN Regression Tool

The kNN-classifier is illustrated in Figure A.6, where Subfigure A.6a, depicts a nearest neighbour of k=1, where simply the nearest neighbour decides the class of prediction, whilst in Subfigure A.6b, the number of k is increased to more then one [70].

Using k=1 can lead to false prediction, and a set of k-Nearest Neighbours are often used. When classifying the dependent variable is categorical can easily been made numerical by regression. The k-NN regressor makes a regression based on the number of k-Nearest Neighbours to minimize the false predictions. The model considers a range of different k-values to find the optimal value.



(a) k=1 (b) k=2

Figure A.6: kNN-classifier

A.4.2.2 Distance

A variety of distances is used in the algorithm. As seen in Equation C.10, C.11, C.12, and C.13, they are most used since it is easy to intersect by changing the variable q. The variable q is also considered to find the optimal value.

A.4.2.3 Manhattan/City Block Distance

$$d(x, y) = \sum_{i=1}^{k} |x_i - y_i| \tag{A.3}$$

A.4.2.4 Euclidean distance

$$d(x, y) = \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2} \tag{A.4}$$

A.4.2.5 Minkowski Distance

$$d(x, y) = \left(\sum_{i=1}^{k} (|x_i - y_i|)^q \right)^{\frac{1}{q}} \tag{A.5}$$

A.4.2.6 Chebychev Distance

$$d(x, y) = \lim_{q \to \infty} \left(\sum_{i=1}^{k} (|x_i - y_i|)^q \right)^{\frac{1}{q}} \tag{A.6}$$

Similarly the all the regression-tools have parameters viable for optimisation. This research employs a systematic grid-search on selected parameters.

A.5 Results and Discussions

In k-fold cross validation the dataset D is divided into an equally adjusted amount of k's. For the Sydney dataset the subsets are D_{2006}, D_{2007}, D_{2008}, D_{2009} and D_{2010}. One subset when is taken apart for testing D_i, and the remaining four is used for training. The method is repeated until all the subsets are tested on an equal shifted amount of training data [28]. In the case of the Sydney dataset containing 87648 datapoints, each k-subset will contain approximately 17530 datapoints depending if there is a leap year or not.

The cross validation was done on various regressors from the Python library scikit-learn [29]. All regressors were set to the default values. In Table A.1, the validation is done for short-term (30 min) time prediction window, denoted *t-1*, and long-term (24 hour) time prediction window, denoted *t-48*. The MAPE of the cross validation show little variation between the subsets. In this work the weather parameters and the load data from the urban area of Sydney city is used. The results are analyzed for correlation among the dataset variables, graphical inspection for understanding some patterns between load and temperature, impact analyses of q-values on load prediction, and analyses of results for load and indicator variables.

A.5.1 Correlation Analysis

Correlation analysis between the variables enlisted in the Sydney dataset (Date and Hour, four weather parameters; DryBulb, DewPnt, WetBulb and Humidity, information about the electricity load consumption, 'SYSLoad') are presented in Table A.2.

A.5.2 Graphical Inspection between Load and Temperature

In section 8.1 it is observed that there is significant impact of temperature on the load. Therefore it is also investigated through graphical depiction the complex relation between DryBulb Temperature and the load patterns emerging from human lifestyle behavior, influenced by the weather conditions. The correlation of System Load to Last half hour value correlates highest at 0.98, and is also the most effective variable for short-term load forecast. Preceding the last half hour value is the variable Hour at 0.48, giving high impact on the periodicity. It has been observed that among the weather parameters, DryBulb has a better correlation with the load. The correlation for DryBulb to the load improves further when it is correlated to the previous 24 hour load data. This might explain why the 24 hour prediction results improves when impact of the weather parameters are included. When investigating the correlation between load and temperature from the graphical depiction, as seen in Figure A.7, where seasonal effects influences the load patterns we find complex patterns, but also periodicity. From these observations it can be seen that the vertical approach (considering a sample time period - e.g seasonally and weekly - of data for four years, and tested for the same time period of the consecutive year) enables the algorithm to reveal the complexity of load and temperature for better prediction results [30].

Table A.1: k-fold validation result in MAPE

Regr.	*2006 t-1*	*2006 t-48*	*2007 t-1*	*2007 t-48*	*2008 t-1*	*2008 t-48*	*2009 t-1*	*2009 t-48*	*2010 t-1*	*2010 t-48*
Rand.	1.02	4.47	1.00	4.07	0.98	3.96	1.02	4.35	1.05	4.35
k-NN	1.83	4.93	1.73	4.65	1.64	4.43	1.79	4.92	1.82	4.88
Linear	2.22	5.49	2.12	5.07	2.13	4.95	2.17	5.24	2.11	5.11
Bayes	2.22	5.49	2.12	5.07	2.13	4.95	2.17	5.24	2.11	5.11

Table A.2: Correlation of Dataset

	Hour	DryBulb	WetBulb	Humidity	SYSLoad	weekday	LastHalfHr
Hour	1.00	0.20	0.11	-0.23	0.48	0.00	0.51
DryBulb	0.20	1.00	0.90	-0.22	0.09	0.01	0.09
WetBulb	0.11	0.90	1.00	0.20	-0.02	0.00	-0.02
Humidity	-0.23	-0.21	0.20	1.00	-0.30	-0.02	-0.30
SYSLoad	0.48	0.09	-0.02	-0.30	1.00	-0.14	0.98
weekday	0.00	0.01	0.00	-0.02	-0.14	1.00	-0.14
LastHalfHr	0.51	0.10	-0.02	-0.30	0.98	-0.14	1.00

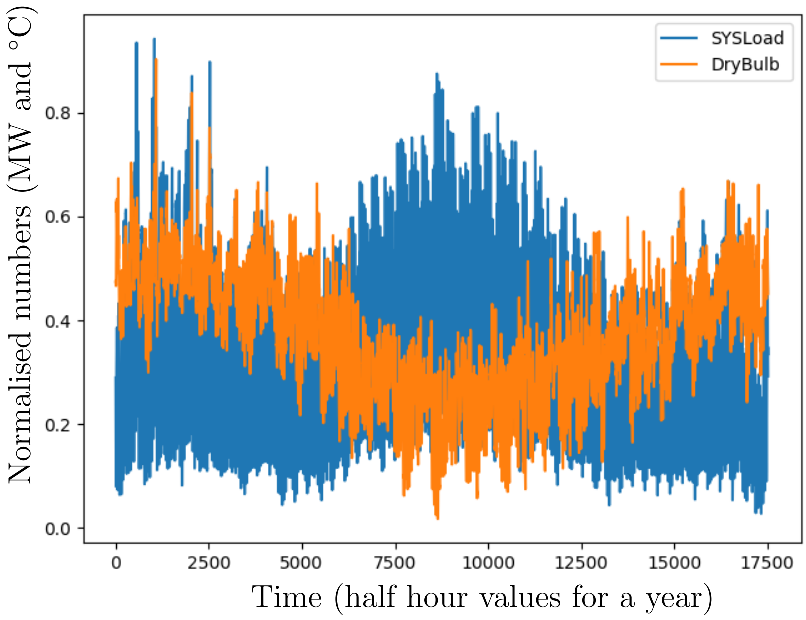


Figure A.7: Correlation of DryBulb Temperature and Electric Load consumption.

Table A.3: Seasons

Season	Months		
Season 1	December	January	February
Season 2	March	April	May
Season 3	June	July	August
Season 4	September	October	November

A.5.3 Impact Analysis of q-value on MAPE

In this work the annual load profile has been divided in four seasons and time frames are given in Table A.3. Observing the results of the impact of q-values on prediction, the preferred value is 1, which is the absolute value. Only occasionly are other q-values the preferred output, meaning the one with the lowest MAPE. On these occasions the highest q-value was 4. Load prediction has been analyzed for all seasons for different regressors and the MAPE for short-term (30 minutes) and long term (24 hours) are presented in Table A.4. In this analysis only the previous load pattern were taken into account. MAPE analysis has been carried out for horizontal (continous time series) as well as vertical approach. It has been observed that Random Forest Regressor provides better results for 30 minutes prediction in horizontal as well as vertical approach for all seasons. For 24 hours prediction it has been observed that in most of the season k-Nearest Neighbour Regressor performs well compared to other regression tools. But in season one for vertical approach Linear Regression has given better result. In season three k-Nearest Neighbour regressor performs well especially considering the vertical approach.

The load prediction using Random Forest Regressor, k-Neareast Neighbor Regressor and Linear Regression has been presented in Figure A.8. These regression results for 24 hour load prediction in season three using vertical approach. Tests conducted by including previous load consumption, weather parameters and indicator variables.

A.5.4 Lowest MAPE for short term and long term prediction

The relative comparison of the MAPE for different regression tools for 30 minutes and 24 hours have been done using both horizontal and vertical approach for all seasons, as shown in Table C.4. It has been found that the the lowest MAPE was achieved with the use of previous load patterns together with indicator variables, and noticeably disregarding weather variables. This goes well with the previous analysis of correlation, which confirms that previous load patterns and indicator variables have higher correlation to the actual load, then the weather parameters.

It has been observed from the test results the lowest MAPE is found through Random Forest Regressor for 30 minutes prediction using the vertical approach. For the 24 hour time period k-Nearest Neighbor is providing lowest MAPE, again through the vertical approach. The lowest MAPE for 30 minutes prediction in season three using vertical ap-

Table A.4: q-Value Results

Time	Random Forest	k-Nearest Neighbour	Linear Regression
	Regressor		
Season One Horizontal Approach			
30 minutes	1.12(16*)	1.29(5**,1***)	2.02
24 hours	5.21(13*)	4.30(16**,4***)	5.55
Season One Vertical Approach			
30 minutes	1.01(16*)	1.44(11**,1***)	1.78
24 hours	6.75(13*)	6.63(19**,1***)	6.29
Season Three Horizontal Approach			
30 minutes	1.13(15*)	1.43(7**,1***)	2.29
24 hours	4.00(15*)	3.60(19**,3***)	5.03
Season Three Vertical Approach			
30 minutes	0.93(18*)	1.17(7**,1***)	2.22
24 hours	3.73(13*)	*3.58(7**,1***)*	5.09

* n-estimator
** k-value
***q-value

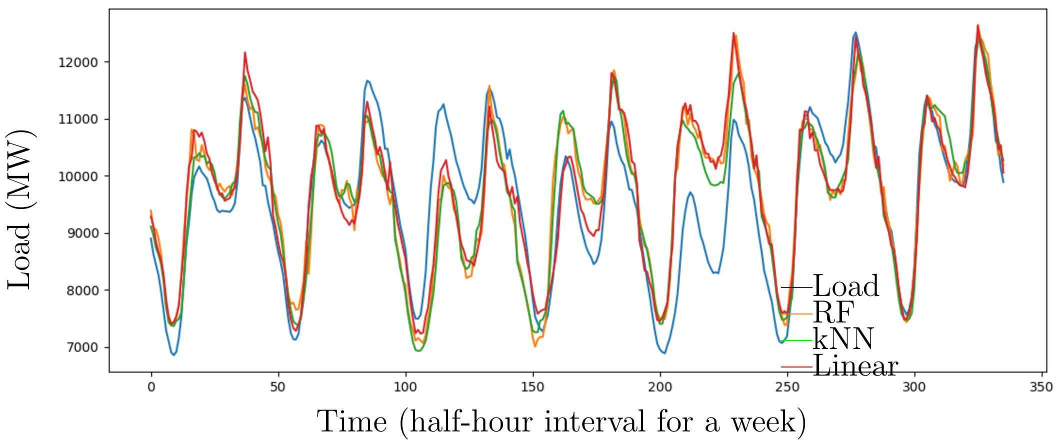


Figure A.8: Regression results on 24 hour prediction for season tree using vertical approach. Tests conducted by including previous load consumption, weather parameters and indicator variables.

proach is shown in Figure A.9, and similarly for 24 hours in Figure A.10. The MAPE for 30 min prediction results using 'random forest regressor' is varying between 1-2%, as seen in Figure A.9, and providing very good results compare to other regressions techniques, which have been used in this work. The 24 hours predictions results using 'k-Nearest Neighbor Regressor' technique has MAP of 2.61%, as seen in Figure A.10, which is much

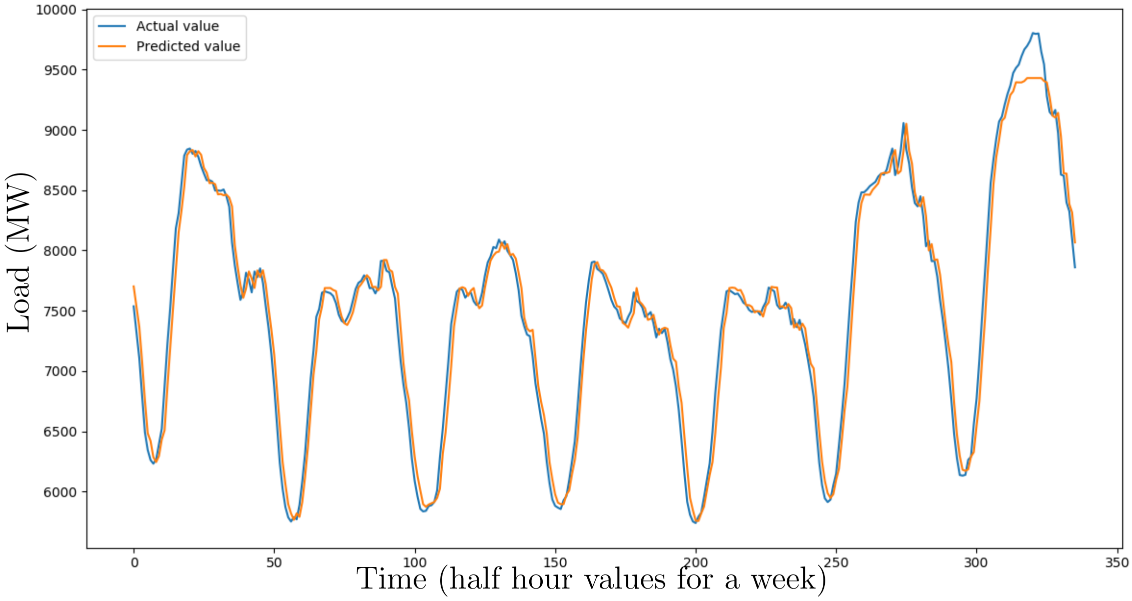


Figure A.9: Best performance for 30 min prediction by Random Forest Regressor

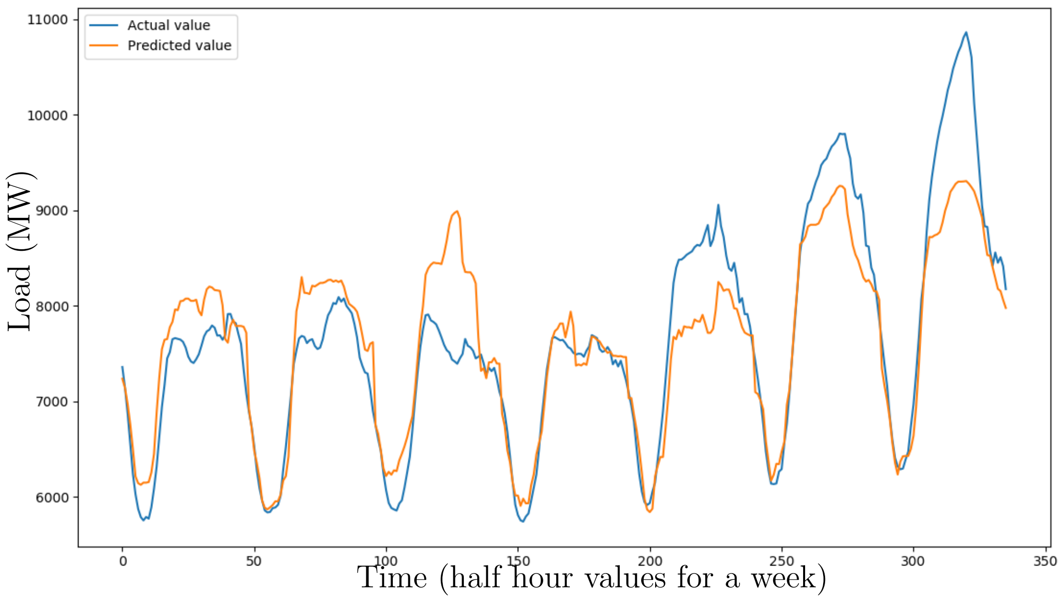


Figure A.10: Best performance for 24 hour prediction by kNN regressor

better compare to other regressors, which have been studied in his work. From the results, it has been observed that for short-term predictions (30 min) the 'random forest regressor' should be used; and for long-term predictions (24 hours) the 'k-Nearest Neighbor Regressor' should be considered.

Table A.5: BEST RESULTS (MAPE Load and Indicator aggregated version test results)

	Regressor		
Time	Random Forest	k-Nearest Neighbour	Linear Regression
Season One Horizontal Approach			
30 minutes	1.11(9*)	1.98(7**,1***)	2.04
24 hours	5.32(13*)	6.53(4**,1***)	5.15
Season One Vertical Approach			
30 minutes	0.94(16*)	1.85(8**,1***)	1.76
24 hours	5.88(13*)	5.49(5**,2***)	5.83
Season Three Horizontal Approach			
30 minutes	1.12(17*)	2.36(5**,1***)	2.29
24 hours	4.76(9*)	5.41(19**,1***)	5.27
Season Three Vertical Approach			
30 minutes	*0.86(17*)*	1.19(6**,1***)	2.15
24 hours	2.71(17*)	*2.61(17**,1***)*	4.26

* n-estimator
** k-value
***q-value

A.6 Conclusion

In this paper the regression tools, Random Forest Regressor, k-Nearest Neighbor Regressor and Linear Regression are used for analyzing the urban area electrical energy demand forecasting. Using larger dataset of Sydney region. This work has explored the use of regression tools for electrical energy load forecasting through correlating weather parameters as well as the time period. Load prediction analysis using regression tools have been done continuous time basis (horizontal) as well as vertical time approach.

A correlation study, where both the Pearson method and visual inspection, of the vertical approach depicts meaningful relation between pre-processing of data, test methods and results, for the regressors examined. Data correlation over seasonal changes have been argued by means of improving Mean Absolute Percentage Error (MAPE). By examining the structure of various regressors they are compared for the lowest MAPE. The regressors show good MAPE for short term (30 min) prediction and Random Forest Regressor scores best in the range of 1-2 % MAPE. kNN show best results for 24 hour prediction, with a MAPE of 2.61%.

Results of this work is going to be useful for predicting the short term 30 minutes electrical energy using vertical approach and considering Random Forest Regression Tool. For long term prediction of 24 hours kNN Regression Tool can provide better results using vertical approach. It is also important to consider further investigations of the impact of various weather parameters on load prediction.

The presented regression techniques can forecast electrical energy demand for short-term (30 min) and long-term (24 hours) using limited datasets. Vertical axis approach has shown competitiveness to ANN due to use of low amount of data and considering the impact of meteorological parameters. Load forecasting is the most fundamental application in Smart-Grid, which provides essential input to other applications such as Demand Response, Topology Optimization and Abnormally Detection, facilitating the integration of intermittent clean energy sources. Presented regression techniques can also be used for predicting energy output (short- and long-term) from the intermittent renewable energy sources.

Acknowledgement:

Authors are very grateful to the scientific committee of IEEE SpliTech 2018 for recommending this article for publication in a Special Issue of Journal of Cleaner Production. Authors appreciatively acknowledge the anonymous reviewers for providing suggestions for improving the manuscript.

Bibliography

[1] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

[2] Nikolaos G. Paterakis, Elena Mocanu, Madeleine Gibescu, Bart Stappers, and Walter van Alst. Deep learning versus traditional machine learning methods for aggregated energy demand prediction. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6, 2017.

[3] S.L. Wong, Kevin K.W. Wan, and Tony N.T. Lam. Artificial neural networks for energy analysis of office buildings with daylighting. *Applied Energy*, 87(2):551–557, 2010.

[4] Tehseen Zia, Dietmar Bruckner, and Adeel Zaidi. A hidden markov model based procedure for identifying household electric loads. In *IECON 2011-37th Annual Conference of the IEEE Industrial Electronics Society*, pages 3218–3223. IEEE, 2011.

[5] AG Azar, E Olivero, J Hiller, K Lesch, L Jiao, M Kolhe, N Asanalieva, P Ferrez, Q Zhang, R Jacobsen, et al. Algorithms for demand response and load control. *SEMIAH, SEMIAH-WP5-D5*, pages 1–v0, 2015.

[6] Seyedeh Narjes Fallah, Ravinesh C. Deo, Mohammad Shojafar, Mauro Conti, and Shahaboddin Shamshirband. Computational intelligence approaches for energy load forecasting in smart energy management grids: state of the art, future challenges, and research directionsand research directions. *Energies*, 11, 2018.

[7] Dario Pevec, Jurica Babic, Martin A Kayser, Arthur Carvalho, Yashar Ghiassi-Farrokhfal, and Vedran Podobnik. A data-driven statistical approach for extending electric vehicle charging infrastructure. *International journal of energy research*, 42(9):3102–3120, 2018.

[8] Tao Hong, Min Gui, Mesut E Baran, and H Lee Willis. Modeling and forecasting hourly electric load by multiple linear regression with interactions. In *Ieee pes general meeting*, pages 1–8. IEEE, 2010.

[9] Ali Lahouar and Jaleleddine Ben Hadj Slama. Random forests model for one day ahead load forecasting. In *IREC2015 The Sixth International Renewable Energy Congress*, pages 1–6, 2015.

[10] Jinkyu Hong and Won Sup Kim. Weather impacts on electric power load: partial phase synchronization analysis. *Meteorological Applications*, 22(4):811–816, 2015.

[11] Henrique S. Hippert and Carlos Eduardo Pedreira. Estimating temperature profiles for short-term load forecasting: neural networks compared to linear models. In *IEE Proceedings - Generation, Transmission and Distribution*, 2004.

[12] J. Valenzuela, M. Mazumdar, and A. Kapoor. Influence of temperature and load forecast uncertainty on estimates of power generation production costs. *IEEE Transactions on Power Systems*, 15(2):668–674, 2000.

[13] Fahad H. Al-Qahtani and Sven F. Crone. Multivariate k-nearest neighbour regression for time series data — a novel algorithm for forecasting uk electricity demand. *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2013.

[14] R. Afkhami and F.M. Yazdi. Application of neural networks for short-term load forecasting. In *2006 IEEE Power India Conference*, pages 5 pp.–, 2006.

[15] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[16] Vasudev Dehalwar, Akhtar Kalam, Mohan Lal Kolhe, and Aladin Zayegh. Electricity load forecasting for urban area using weather forecast information. *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*, pages 355–359, 2016.

[17] Tao Hong, Min Gui, Mesut E. Baran, and H. Lee Willis. Modeling and forecasting hourly electric load by multiple linear regression with interactions. *IEEE PES General Meeting*, pages 1–8, 2010.

[18] Muhammad Usman Fahad and Naeem Arbab. "factor affecting short term load forecasting,". *Journal of Clean Energy Technologies vol. 2*, pages 305–309, 2014.

[19] M. Mastouri and N. Bouguila. A methodology for thermal modelling and predictive control for building heating systems. *2017 18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, pages 568–573, 2017.

[20] T Salque, D Marchio, and P Riederer. Neural predictive control for single-speed ground source heat pumps connected to a floor heating system for typical french dwelling. *Building Services Engineering Research and Technology*, 35(2):182–197, 2014.

[21] C. Hansen D. Becker Peter Hirsch D.Paraven, A. Debs and R Golob. "influence of temperature on short-term load forecasting using epri-annstlf". 2002.

[22] P. Sarajcev, D. Jakus, J. Vasilj, and M. Nikolic. Analysis of transformer health index using bayesian statistical models. In *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–7, 2018.

[23] Chai Phing Chen, Sieh Kiong Tiong, Siaw Paw Koh, Fong Yu Chooi Albert, and Fauzan K. Mohd Yapandi. Gas emission prediction for environmental sustainability via heterogeneous data sources correlation with support vector regression. In *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–7, 2018.

[24] Oleg A. Yakimenko and William W. Anderson. Challenges in modeling wind power generation based on available weather data. In *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–9, 2018.

[25] Morten Goodwin and Anis Yazidi. A pattern recognition approach for peak prediction of electrical consumption. In Lazaros Iliadis, Ilias Maglogiannis, and Harris Papadopoulos, editors, *Artificial Intelligence Applications and Innovations*, pages 265–275, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[26] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. When is "nearest neighbor" meaningful? In Catriel Beeri and Peter Buneman, editors, *Database Theory — ICDT'99*, pages 217–235, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.

[27] Dharmendra Kumar Patidar, Bhavin C. Shah, and Manoj R. Mishra. Performance analysis of k nearest neighbors image classifier with different wavelet features. *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, pages 1–6, 2014.

[28] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI*, 1995.

[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[30] Nils Jakob Johannesen, Mohan Kolhe, and Morten Goodwin. Comparison of regression tools for regional electric load forecasting. In *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–6, 2018.

Appendix B

Paper B - Smart Load Prediction Analysis for Distributed Power Network of Holiday Cabins in Norwegian Rural Area

Nils Jakob Johannesen and Mohan Kolhe and Morten Goodwin

Faculty of Engineering and Science, University of Agder, PO Box 422, NO 4604 Kristiansand, Norway.

***Abstract* - The Norwegian rural distributed network is designed for Holiday Cabins with limited loading capacity. Load prediction analysis, of such type of network, is necessary for effective operation and to manage increasing demand of new appliances (e. g. electric vehicles and heat pumps). In this paper, load prediction of distributed network (a typical Norwegian rural area power network with 125 cottages with 478 kW peak demand) is carried out using regression analysis for making autocorrelations and correlations among weather parameters and occurrence time in the period of 2014 to 2018. In this study, the regression analysis for load prediction is done considering vertical and continuous time approach for day-ahead prediction. The vertical time approach uses seasonal data for training and inference, compared to continuous time approach that utilizes all data in a continuum from the start of the dataset until the time period used for inference. The vertical approach does this with even fewer data than continuous approach. The regression tools can perform using the low amount of data, and the prediction accuracy matches with other techniques. It is observed through load predictive analysis that the autocorrelation by vertical approach with kNN-regressor gives a low Symmetric Mean Absolute Percentage Error. The kNN-regressor is compared with Random Forest Regressor and, also it uses autoregression. Autoregression is the simplest and the most straightforward predictive model based on the targeted vector itself. The autoregression indicates the decline and incline of the**

time-series, and thus gives a finite gradient for the curvature of load profile. It is observed that joint learning of regression tools with autoregression can predict time-series components of different load profile characteristics. The presented load prediction analysis is going to be useful for distributed network operation, demand-side management, integration of renewable energy sources and distributed generator.

Keywords - *Load Predictive Analysis, Distributed Network Operation, Machine Learning, Regression Analysis*

B.1 Introduction

The Norwegian rural distributed network is designed for Holiday Cabins with limited loading capacity. Load prediction analysis, of such type of network, is necessary for effective operation and to manage increasing demand of new appliances (e. g. electric vehicles and heat pumps). Change in user behavior due to installed heat pumps and electric vehicle charging stations are expected to increase the electric load demand. Such type of rural distributed network can be operated as micro-grid with integration of renewable energy sources and distributed generators. The rural distributed network may face voltage instability due to increasing demand of power intensive loads, therefore appropriate operation and management of rural distributed network are required. The rural area distribution network performance can be improved by operating it as a micro-grid with integration of energy storage, renewable energy sources and distributed generators. The smart micro-grid (i.e. smart distributed network) is a complex system encompassing of various sub-systems at various stages of aggregation. Smart micro-grid is going to accommodate multi-directional power flow to go together with multi-directional information flows between all the vectors (e.g. power generations, transmission and distribution system operators, distributed intermittent renewable energy sources, demand response aggregations, end-users, etc.). Over the past decade the power system is changing from centralized grid to more decentralized and its operational management is going to be real-time monitored smart and micro-grids [30]. Reference [31] has reviewed energy technologies for application in smart distributed network using IOT technologies, various different types of solar technologies has been reviewed in the same paper and discusses control strategies PV's and hybrid energy systems. For effective operation of micro-grid and demand side management, the load prediction analysis with impact of external parameters is required.

Machine learning algorithms can be electively used for electrical energy demand as well as predicting the output from the renewable energy sources. It is important to do the prediction of future load consumption to balance the electrical energy supply and demand [32]. Existing research into micro-grid electric energy load demand forecasting is scarce. The majority of the existing research selected micro-grids of large power scale with electric energy load demand ranging from 10 MW scale, to larger ones at 1000 MW. The GW-scale which is the size of a medium city and forecasting results from such a large scale micro-grid is comparable to urban area load forecasting. Hence the smaller scale micro-grid is more difficult to predict due to higher load fluctuations and randomness. At a smaller scale the load fluctuations within the same time period may be higher than for bigger more stable load. A comprehensive study compares small and large scale micro-grids in China. The chineese case study uses five different scale of micro-grid where the two smallest micro-grids have subsequently maximum load of 273 and 463.8 kW. To efficiently predict the electric energy load demand for these micro-grids they propose to use different hybrid forecasting models based on Empirical Mode Decomposition (EMD), Extended Kalman Filter (EKF), Extreme Learning Machine with Kernel (KELM) and Particle Swarm Optimization (PSO). For the small scale micro-grid the hybrid models achieves acceptable MAPE of 7 to 10 % [33]. Existing research on network capacity planning deal with much larger data samples. The term Big Data is a relative concept and not an absolute definition, at best it is ambiguous and to quantify dataset is a difficult task as the capacity and computational power is continuously increasing. Typical Big Data is regarded as that quantification of collected data in different sampling rates is in the Terabyte (TB) area [35] [34].

The main objectives for this research work is to investigate the vertical axis approach, described in our paper [7] by studying user behavior and applying vertical time approach that uses seasonal data for training and inference. Potential research will be analyzing micro-grid architecture (adaptive) based on local renewable energy prediction as well as demand forecasting. This architecture will consider techno-economic operational characteristics of dispatchable distributed generators, and focus on analyzing predictive techniques and performance metrics for maintaining the system reliability and stability in practical operation and management.

In a review article [8], the performance metrics mean squared error (MSE), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), are evaluated. The last three decades the popular performance metrics has changed from MSE to MAPE, bringing MAPE to be the preferred metric in recent years. MAPE works well for load forecasting, as long as the real value is unlike zero, that is causing a computational error as described in [43]. The review on electric price forecasting (EPF) [10] points out there is no standardized method for evaluating prediction performance. Absolute errors, although widely used, make it hard to compare among different dataset, and measures, based on absolute percentage errors, are used. With point forecast for low values the MAPE values become very large, even though the absolute value is not. MAPE comparisons must be done with caution. In the case of low values, a symmetric mean absolute percentage (SMAPE) can be used. The Makridakis or M-Competitions conducted by the International Institute of Forecasters (IIF) for evaluating the participating methods by focus of empirical validation, [11] recognizes that the metric SMAPE penalizes large positive errors.

In our previous study [7], we have used regression techniques for urban area load forecasting and it has been validated by correlation analysis to external parameters with the vertical approach. The regression techniques are used in this work for the rural area load prediction with autocorrelation analysis. From previous study [7], it has concluded that the vertical approach predicts well with fewer data. In the rural area, where data is limited, hence the vertical approach is a preferred method for the rural area electric load demand forecast.

In this paper, load prediction of a distributed network (a typical Norwegian rural power network of 125 cottages with 478 kW peak demand) is carried out using regression analysis for making autocorrelation and correlations among weather parameters and time of usage in the time period of 2014 to 2018. In this study the regression analysis for load prediction is done using vertical and continuous time approach for day-ahead planning with 24 hour prediction. The load prediction analysis is going to be useful for distributed network operation, demand-side management, integration of renewable energy sources and distributed generator.

Selection and description of load profile of the data are presented in Section B.2. The quick and easy application of optimized autocorrelation based feature selection is presented in Section B.3. The regression techniques are explained analytically in Section B.4. The obtained results of the considered rural area are analyzed in Section E.5. The usefulness of the presented load prediction techniques is summarized in Section B.6.

B.2 Load Profile of Selected Rural Area Network

The electric energy load demand for holiday resorts have increased radically the last two decades. Since 1996 the load demand in Norwegian Cabin Areas has been growing into tree times its original size. Most of this is due to a change in standard, from bio-fueled ovens to electric heating, therefore load analysis and forecasting is important due to the enlarged power dependent installations like heat pumps and chargers for electric vehicles. This is an important field of research and has been neglected since the holiday resort electric energy load consumption is only 1.8 % of the 2016 Norwegian electric energy load demand [12]. The weekly electric load cycles of Bjønntjønn Cabin Area is direct opposite to that of larger urban areas, where the electric energy load demand is considerable lower during weekdays, where businesses are not demanding energy. In Fig. B.1, where the total kilowatt consumption is aggregated and showing high load demand on typical (holiday) weekends, from Friday to Sunday.

The selected rural area network is used for Holiday Cabins and there is a potential for integrating solar photovoltaic system with energy storage. In Norway the penetration of electric vehicles is increasing more then in any other countries, and is a potential challenge for the operation and management of the entire grid, therefore the load prediction analysis of such type of rural network is necessary. Bjønntjønn Cabin Area is a typical rural area low capacity network in the south-east part of Norway, see Fig B.2. The load demand of Bjønntjønn Cabin Area from 2014 to 2018, as illustrated in Fig. B.3 shows a peak load demand in typically holiday winter seasons, and low load during summer time, where temperature is higher, and evenings are brighter and thus less time for indoor activities. To study correlation between load and external parameters data from Norwegian Institute of Bioeconomy Research (NIBIO) with weather information from 3 closest meteorological stations, to Bjønntjønn Cabin Area (Bø, Gvarv and Gjerpen) are picked for correlation analysis. Through correlation analysis the highest correlating weather station, is found. Most of the pattern that constitutes the electric load profile is dependent on individual user behavior. The individual human activities is not enough to make substantial patterns on its own accord, yet together with the influence of the changing weather the impact is growing, and an important component of feature engineering in load forecasting.

The Dry-bulb temperature is the most fundamental external parameter debated in the load forecasting literature [13]. Comprehensive correlation analysis of load demand to weather has historically proven to be important [14]. Previous developed research makes inquires into seasonal load demand variation for the amount used on space heating and reveals that the amount is substantial, and hence contributes to the correlation to electric energy load demand. The technique proposed by [15] indicates that individual activities (Television/Radio, heating water, lights) are negatively correlated temperature. The electric energy load demand reaches a peak demand in the end of typical holiday season, and this period is not particularly colder then out of holiday season period, as seen in Fig. B.4, that illustrates the complex relation of temperature and load demand. Time occurrence dependence relationship is a fundamental asset for optimal feature extraction based on correlations between independent features and are described in Section B.3.

For load analysis of electric energy demand it is important to look into the characteristics of the data; trends, seasonality and cycles [16]. Trend is when the load consumption in the total time-series from start to finish shows an inclination to increase or decrease with a longer-term change of the mean value. On a lower level there might be reoccurring phenomenon due to seasonality, whether it is a higher load demand during winter due to increased heating and indoor activities as opposed to summer. Seasonality can also take shape from a lower indicative level such as month, and can be the change in monthly arrival of residencers at a cabin area. Cycles can be patterns that are observed for more than a year for various reasons (e. g. droughts, famine or financial crisis). Cycles can also be observed at lower time levels as daily and weekly cycles [7] [17] .

B.3 Feature Engineering

The efficient and transparent predictive model is extracting a focus set of informative features from a bigger dataset. The process of removing redundant and irrelevant features has many names; feature extraction, feature selection or feature engineering. Leaving the decision making to a small feature space reduce data dimensionality to evoke faster

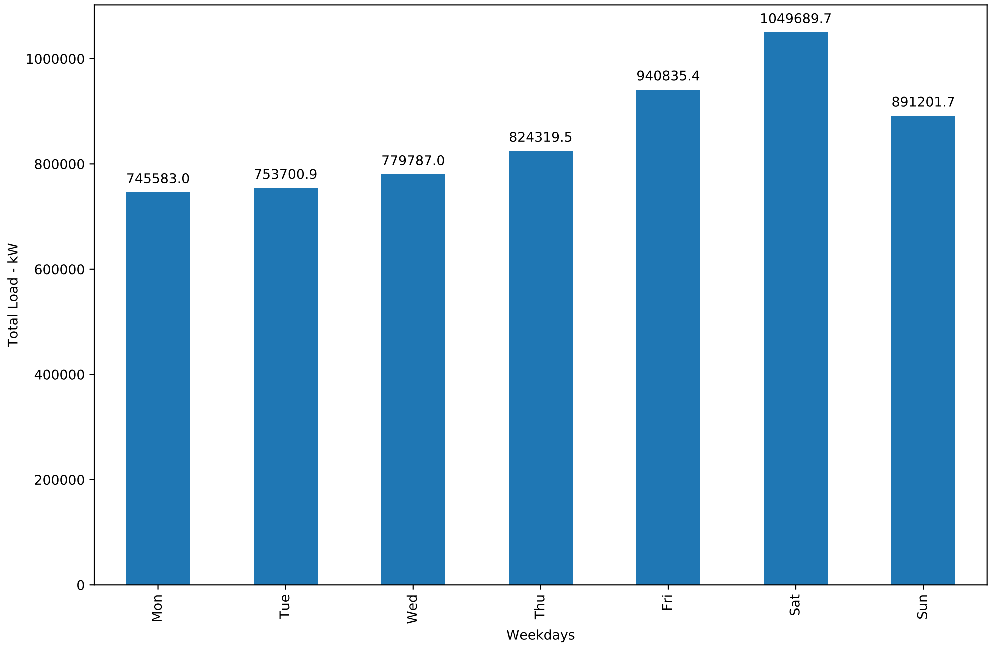


Figure B.1: The total sum of load in kWh for the days of the week in Bjønntjønn Cabin Area 2014-2018

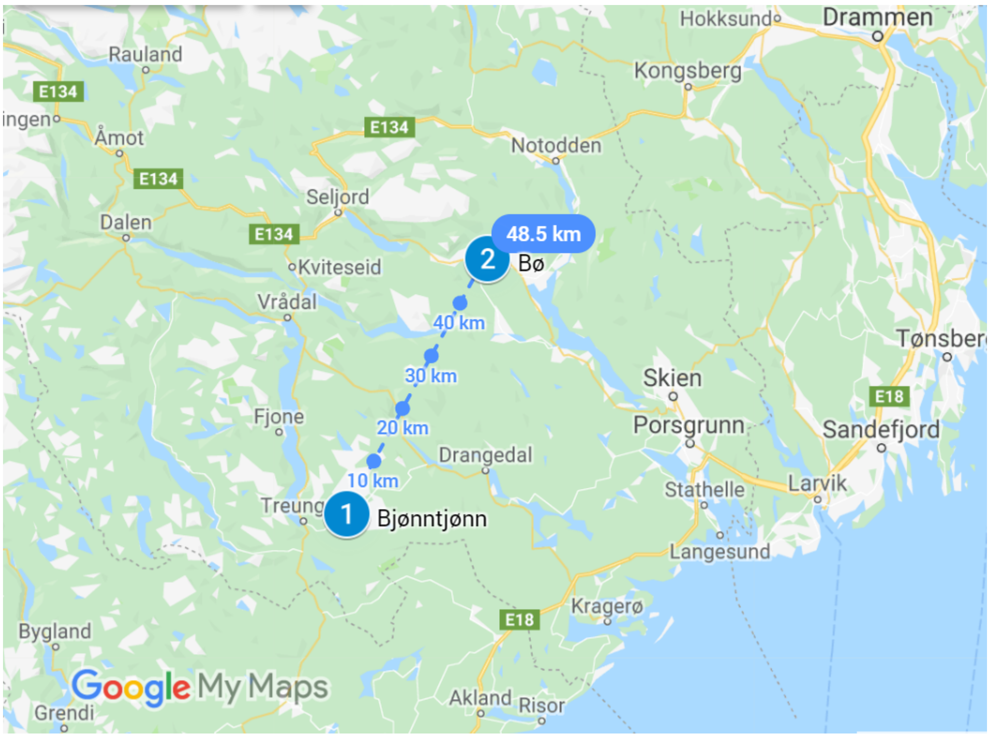


Figure B.2: The Bjønntjønn Cabin Area and weather station Bø. Map data © 2019 Google.

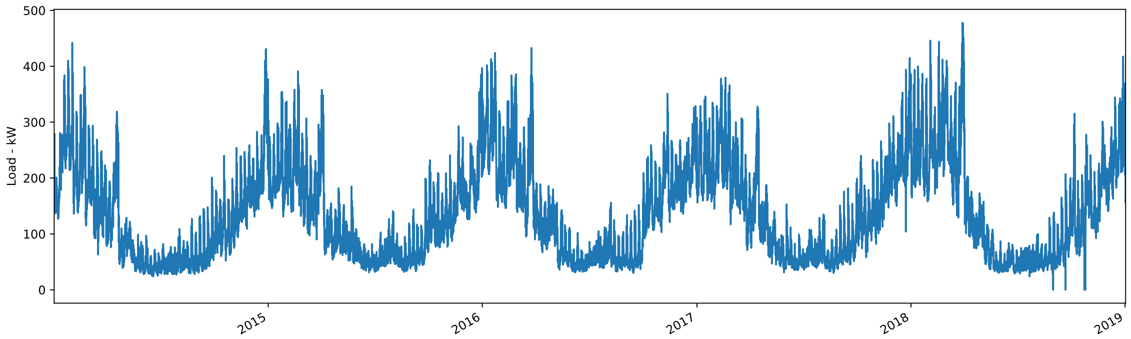


Figure B.3: Electric Energy Load Demand at Bjønntjønn Cabin Area in south of Norway from 2014 to 2018.

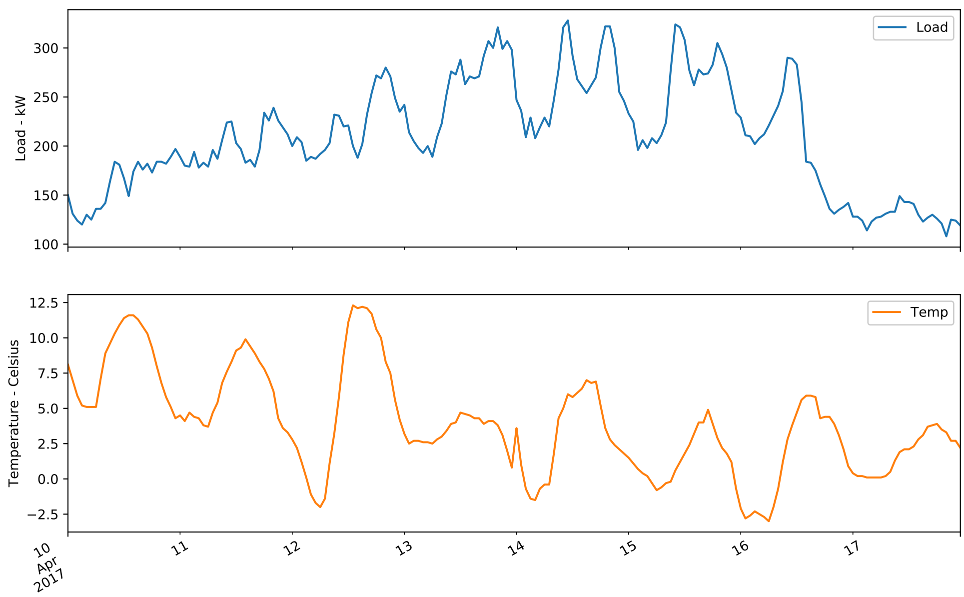


Figure B.4: Load consumption and temperature readings during Easter holiday 2017 computation time, avoid overfitting and induce model transparency.

B.3.1 Autocorrelation

Autocorrelation is a type of serial dependence, and it shows how a time-series is related to its own lagged version. By plotting the autocorrelation, information on the temporal component of the data is given and unfolds the fundamental construction of time-series; unraveling trends, seasonality and their inherent structure [18]. Features of previous load information is selected through analysis of autocorrelation of the previous 200 hourly timelags, see Fig. B.5 and Fig. C.13, by equation B.1.

Norwegian meteorological web service, yr.no, offers first hand downloadable data through their service. The data is limited, informing the daily minimum, maximum and mean values. The sparse information have no practical use in hourly prediction. This is a known problem, other national meteorological forecasters like the Bureau of Meteorology of Australian Government (BMAG) only release the minimum and maximum value have limited information available. The authors of [19] offer a way to mitigate this problem, through k-Nearest Neighbor algorithm and searching for nearest neighbors among the external parameters, by taking the square root and adding the difference of two squared sums of daily minimum and maximum temperatures.

$$r_k = \frac{\frac{\sum_{t=1}^{N-k}(x_t-\bar{x})(x_{t+k}-\bar{x})}{N-k-1}}{\frac{\sum_{t=1}^{N}(x_t-\bar{x})^2}{N-1}} \tag{B.1}$$

B.3.2 External Parameters

B.3.2.1 Weather Parameters

Based on correlation analysis the weather station with the strongest correlation of temperature to the load data from Bjønntjønn Cabin Area is identified, and used for the further research. Previous research found Bø weather station with the highest negative correlation to the electric energy load demand at Bjønntjønn Cabin Area [20]. The heuristics of good correlation-based feature selection is based on the level of intercorrelation within the class and subset features. A good feature set contains independent variables that have high positive or negative correlation to the dependent variable, and no correlation amongst the other dependent variables [55]. The correlation of the variables in the Bjønntjonn Cabin Area dataset, see Table D.1, shows a high negative correlation of load to temperature, positive correlation of load and holiday and no correlation between the dependent variables holiday and temperature. In Fig. C.11 the variation of temperature and load are illustrated for the seasonal information of Bjønntjønn Cabin Area.

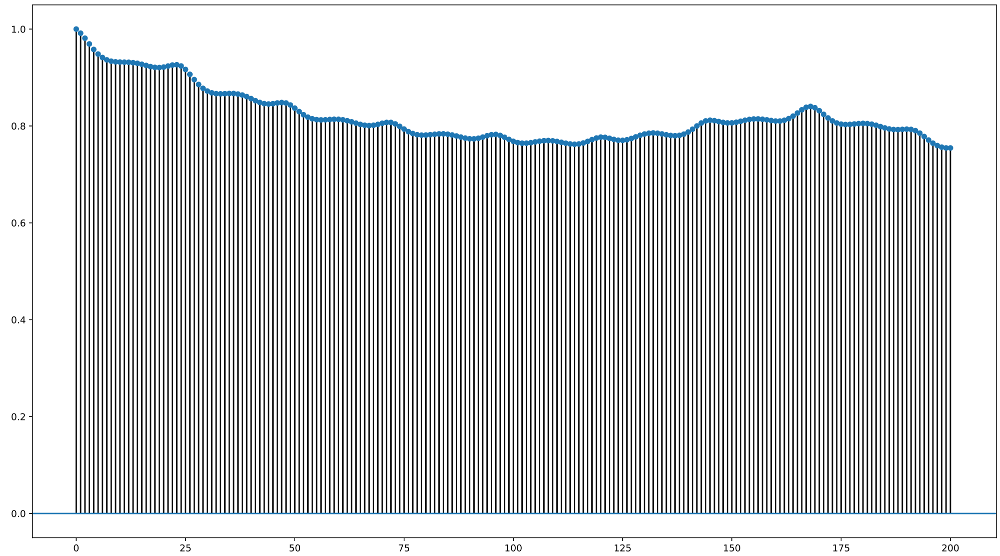


Figure B.5: Autocorrelation of load consumption by 200 lags for Bjønntjønn Cabin Area 2014-2018

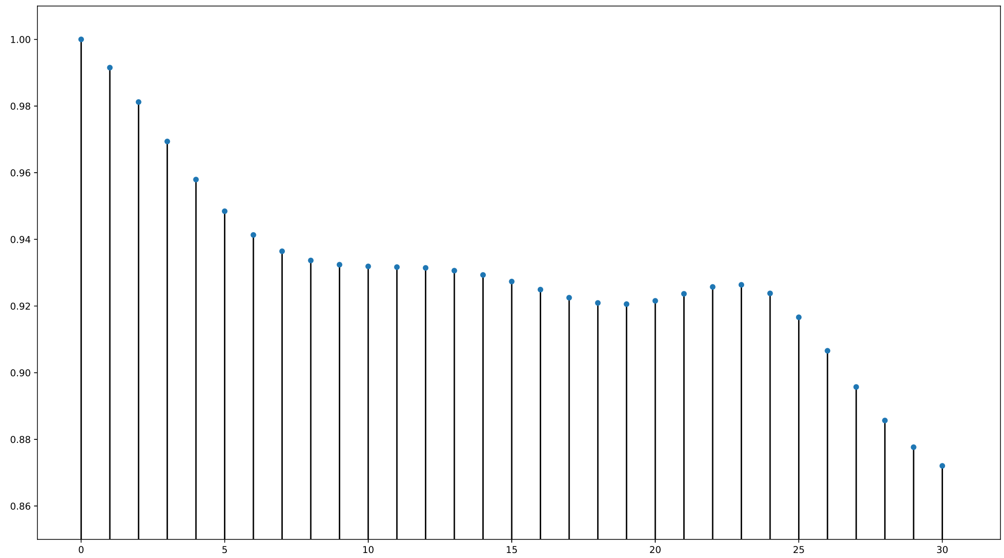


Figure B.6: Autocorrelation of load consumption of the first 30 lags for Bjønntjønn Cabin Area 2014-2018

Table B.1: Correlation of features

	Load	***Temperature***	***Holiday***
Load	1	-0.82	0.18
Temperature	-0.82	1	0
Holiday	0.18	0	1

B.3.2.2 Working-/Non-Working Days

To search for patterns among the days of the week, all kilowatthours-usage based on the respective day of the week are summed together, and illustrated in a bargraph in Fig. B.1. From Friday to Sunday the sum of kilowatthours for the total years of 2014-2018 is above 890 MWh, with a top consumption on Saturdays with surpassing 1 GWh. The rest of the week, from Monday to Wednesday is stable in the 700 MWh region. The weekly pattern follows a very neat curve of increasing electric energy load demand from Monday to Saturday, before there is a slight decline on Sunday. This coincides with the holiday patterns of holiday resorts users, in Norway people travel to their cabin after lunch on Friday and return home Sunday evening.

B.3.2.3 Public Holidays

In the comprehensive study of German market the authors [22] found improvement of forecasting accuracy by 80 % by including holiday effects. This underpins the usefulness of including the effects of public holidays as they are usually known in advance, by law, and one can therefore anticipate the affect of human activity. National or state authorities agree upon holidays and state them as law. We identified all Norwegian holidays; Easter, labor day, national day, ascension day, Pentecost and X-mas. Identification of holidays as well at studying holiday behavior given by Statistics Norway, we categorize holidays as one. The days in the holiday periods also included working-/non-working as defined in the Section B.3.2.2, regardless of this definition all the days of holiday period is coded with the value 1, meaning a non-working day.

B.3.3 Validation

Cross-validation (CV) is a simple and universal tool for estimating expected accuracy of the predictive algorithm by taking the mean value of all errors of the independent samples of the dataset. For data with temporal dependencies, the validation and training samples are no longer independent. Leave one out or hold-out k-fold validation, uses one fold for testing and the remaining folds for training, where for the NordPool dataset k equals five (for data from 2014-2018), see Fig. B.8. Leave-one-out validation is also called jackknife due to the jackknifes ability to be used as a 'quick and dirty' replacement tool for more sophisticated tools. Leave-one out method, is compared to crogging, a method aimed at preserving the temporal dependencies of a time series. Crogging combines cross-

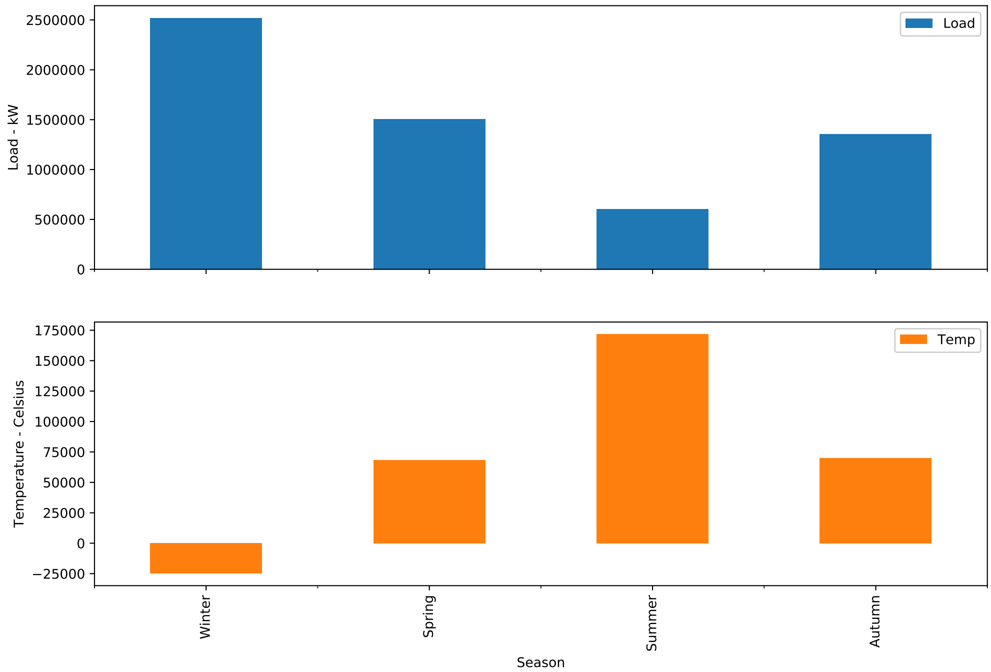


Figure B.7: Sum of load consumption and temperature on a seasonal basis

validation and forecast aggregation, where each fold aggregates training data whilst all the time validating against new test data, see Fig. B.9.

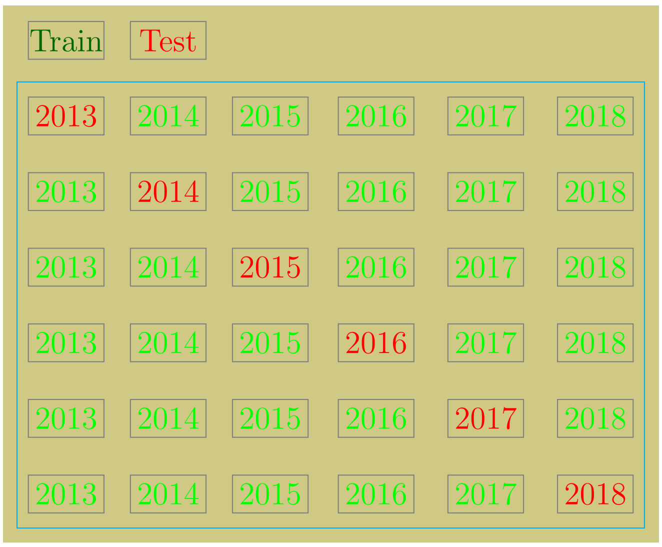


Figure B.8: Leave-one-out, or jackknife, leaves the test sample out of the training and trains the algorithm on the remaining

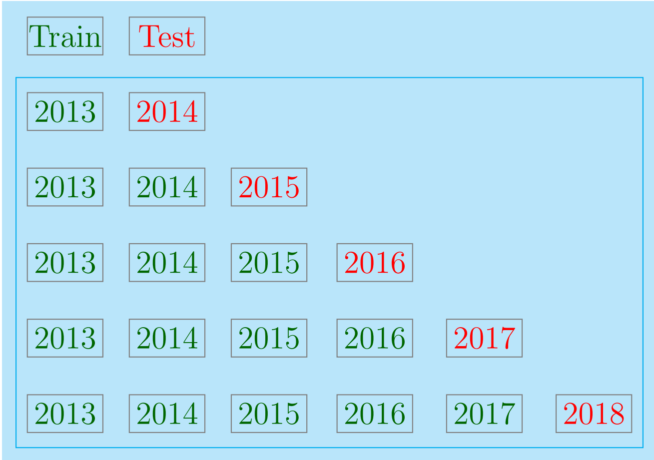


Figure B.9: Crogging combines cross validation and forecast aggregation to capture the temporal dependency of time-series

B.4 Methodology

The vertical time approach uses seasonal data for training and inference, as opposed to continuous time approach that utilizes all data in a continuum from the start of the dataset until the time period used for inference. In this work the regression analysis is done on continuous time basis as well as using vertical time axis approach. The kNN-regressor is compared to Random Forest Regressor and also used autoregression. Autoregression is the simplest and most straightforward predictive model, based on the targeted vector itself and a certain time-window. It indicates the decline and incline of the time-window, and thus gives a finite gradient for the curvature of load profiles. The joint learning of regression tools with autoregression predicts time-series components of the different characteristics.

B.4.1 Performance Metrics

To evaluate the rural area electric energy load forecasting, several performance metric can be used where the real value y is compared over equations C.20, C.23 and C.21 by the predicted value $\hat{y}$.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{B.2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}| * 100 \tag{B.3}$$

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2}\right) * 100 \tag{B.4}$$

Data correlation over seasonal changes will be argued by means of improving MAPE, SMAPE and MAE.

B.4.2 Regression Tools

The methodology of this work is based on consideration of limited dataset, therefore the vertical approach is appropriate. The research work is using on k-Nearest Neighbor Regressor (kNN) and the Random Forest Regressor (RF). Prior research finds kNN and RF can perform best in short time load forecasting in a comparison of different regressors [23].

B.4.2.1 k-Nearest Neighbor

The kNN computes the difference of the sum of the inputs, and finds the number of nearest neighbors from the designated k-value. And it provides the numerical continuous output based on regression considering nearest neighbors.

B.4.2.2 Random Forest Regressor

RF is a magnitude of different decision trees that uses a majority vote to rule the best class. For the RF, the trees are grown dependent on a random vector, and the outputs are numerical scalars. One sole decision tree encompasses attributes and classes in the datasets and uses an entropy function to find the best classifier as well as gain function to build the best structured tree.

B.4.2.3 Autoregression

The autoregressor finds the curvature and gives a finite gradient based on the latest updates from the targeted vector. In this case it is the load, based on equation C.24.

$$c = (L_{t-1} - L_{t-2})^{\frac{1}{p}} \tag{B.5}$$

The methodology used in this research work is developed to deal with the problems of irregularities and randomness in the time series. RF-regressor yield good result on hourly time prediction in load forecasting. The kNN-regressor has shown precise prediction in time-series, due to its capability to capture the nearest step in a time series based on the nearest neighbor principle. The two regressors need to be investigated independently, to search for their independent qualities, and finally as a hybrid model to fully utilize their joint potential. Previous study shows that the combination of qualities in hybrid models are able to capture the stationary linearity of the time series and capture the peaks of the time series to enhance the forecasting precision [10].

B.4.3 Test/Inference

The testing and inference to finalize the chosen parameters are done by cross validation methods of Leave-One-Out and Crogging, as explained in Section B.3.3. Meaning that based on these results we find the final model used for further testing and inference. The last fold of both of the mentioned cross-validation methods, is the continuous approach. Since the folds are divided into separate years, test periods is extracted based on seasons, to effectively compare to the vertical approach. The seasonal performance is then verified by weekly MAE, MAPE and SMAPE, as explained by [10]. The weeks are chosen by the mid-week of each season, so that for the winter season (December, January, February) the week for verification is considered mid-January, and so on for all the seasons. It is important that the algorithm has never seen the inference data, e.g. that this data has not been used for training. For continuous approach, we are training the algorithm with all the data from 2014 up before the week in mid-January 2018. By this way, we ensure that training- and test- data are carefully separated. We are using the same manner of verification for the continuous approach on all four seasons.

In the vertical approach, we aggregate the data by concatenating each season as a training set. The vertical approach is taking winter season from 2014 to 2017, and then test for the mid-week of January 2018, we are following the same pattern for all four seasons.

The continuous approach have the advantage to be trained by more data in sequence, then the vertical approach.

B.4.4 Test set-up regime

We are testing for two algorithms, kNN and RF Regressor for day-ahead forecasting (24 hour). They are tested both for the vertical approach as well as continuous approach (as described in section B.4.3). Hyperparameter tuning based on cross-validation is tested for a range of nearest neighbors (2-12) and n-estimators (2-12), we the best option based on performance are selected to be neighbors 12 and n-estimator of 10.

Since a time-series is related to the same lagged version of itself, we select it as a feature always to be tested since the values of autocorrelation are showing high significance. We analyze the autocorrelating behavior of the time-series of electric energy load demand for cabin-users at Bjønntjønn, and find that the preceding-day (24 hours), preceding-two day (48 hours) and preceding-week (168 hours) are the prominent previous load features of the data. They are always embedded as features for the test set-up. When presented in tables this feature is notated as AC for autocorrelation.

We want to analyze how the kNN and RF Regressors behave when given the information of the autoregression. We test for this feature together with the features given from the autocorrelation (AC). This feature is notated as AR for autoregression.

A matter of interest is how well the external parameters, weather and time of occurence contribute to the predictive outcome, and we have tested them. This features is notated as T for temperature and H for holidays.

Table B.2: Forecasting Results (24 hours prediction) by seasons trained with time feature lags of 24-, 48- and 168-hours

Features	Vertical						Continous					
	summer			winter			summer			winter		
	SMAPE	**MAPE**	**MAE**	**SMAPE**	**MAPE**	**MAE**	**SMAPE**	**MAPE**	**MAE**	**SMAPE**	**MAPE**	**MAE**
kNN AC	12.74	12.74	**6.87**	9.88	10.06	26.07	13.17	13.35	7.17	9.72	9.74	25.60
RF AC	14.70	14.78	8.07	10.43	10.67	27.85	15.27	15.47	8.49	9.56	9.49	25.24
kNN AC AR	13.17	13.24	7.11	10.05	10.20	26.39	13.28	13.43	7.23	9.25	9.24	24.42
RF AC AR	14.16	14.14	7.70	10.87	11.03	28.67	13.89	14.07	7.54	10.34	10.34	26.91
kNN AC T H	14.79	14.46	7.94	9.48	9.66	25.09	15.07	14.75	8.08	9.05	9.09	23.89
RF AC T H	16.53	16.10	8.80	11.39	11.53	29.86	17.05	16.48	9.14	11.50	11.53	29.81
kNN AC AR T H	14.27	14.07	7.68	9.75	9.92	25.65	14.41	14.14	7.71	**8.88**	**8.86**	23.45
RF AC AR T H	16.98	16.66	9.02	12.03	12.18	31.56	17.21	16.91	9.19	10.88	10.96	28.06

B.5 Results and Discussion

The load profile of the considered holiday resort is categorized season wise. In this work Regression Tools are used for load predictive analysis. In the load predictive analysis the vertical time approach is used for a particular holiday time period. Vertical approach can perform with minimum amount of data compared to continuous approach. Also, the vertical time approach predictive results are compared with the prediction based on continuous time-series data. The presented methodology can also deal with the problems of irregularities and randomness in the dataset.

The kNN with autocorrelation (kNN AC), the SMAPE for summer season using vertical approach is 12.74 % and in winter season 9.88 %, but for the continuous data SMAPE is 13.17 % in summer season and 9.72 % in winter season. Although both SMAPE and MAPE values are relatively high. The kNN with autocorrelation performs by far the best in terms of MAE, as illustrated in Fig. B.10. The kNN with autocorrelation, for vertical approach for summer season is giving the lowest amount of information as well as a low amount of data, meaning there is a minimum ability to recognize a pattern. Except from a low dip at the very end of the week (as seen in Fig. B.10) the load is fluctuating in the same low load interval. For generality the results show a low MAE for all the different versions of regressors and hybrid models with various features when trained with low amount of data. With the low load consumption, due to summer season, MAE scores comparatively good for all instances. The best is the simplest version of kNN only, as the time features of previous load are 24, 48 and 168 time lags. The 24, 48 and 168 time lags is found to autocorrelate higher than any other time lag. Similarly RF with autocorrelation (RF AC), the SMAPE for summer season with vertical approach is 14.70 % and in winter season 10.43 %, but for continuous data SMAPE is 15.27 % in summer season and 9.56 % in winter season. Results from an altered time dependent feature (containing time lags at 24 and 168) are different from the findings in the autocorrelation analysis, and they have impacted the predictive outcome negatively. With these different time-features, the vertical approach for the winter season results in a SMAPE of 12.22% (kNN AC) and 13.43% (RF AC), a more than 2% difference from the results presented in the Table B.2 using time dependent features from the autocorrelation analysis.

Through the kNN with autocorrelation and autoregression (kNN AC AR), the SMAPE for summer season using vertical approach is 13.17 %, and in winter season 10.05 %. For continuous data SMAPE is 13.28 % in summer season and 9.25 % in winter season. Similarly RF with autocorrelation and autoregression (RF AC AR), the SMAPE for summer season given vertical approach is 14.16% and in winter season 10.87 %, but for continuous data SMAPE is 13.89 % in summer season and 10.34 % in winter season. The comparative analysis of various regression techniques on load prediction for summer and winter seasons load is presented in Table B.2. The forecasting results of electric loads are compared for vertical and continuous approach for both seasonal loads.

When training kNN regressor hybrid model with autoregressor, weather parameter and

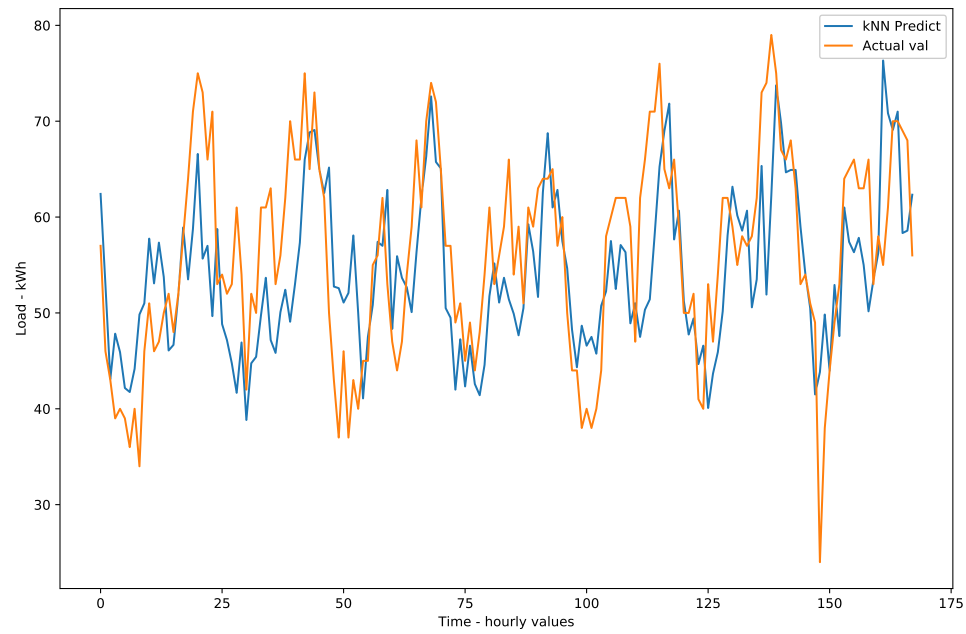


Figure B.10: Prediction for week of July 2018 scoring MAE 6.87 using vertical approach kNN-regressor only trained with time features.

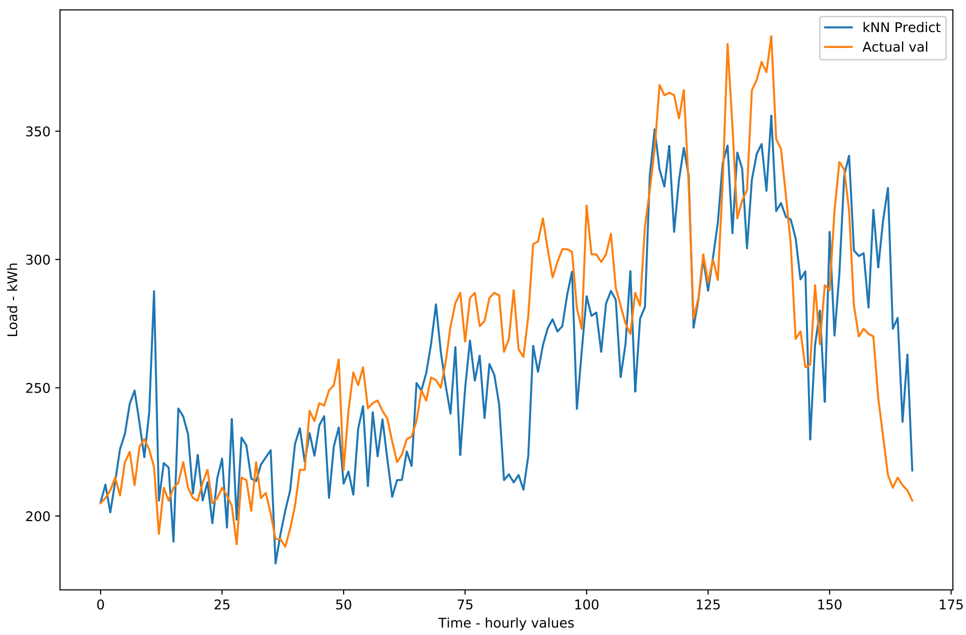


Figure B.11: kNN regressor with autoregressor by continuous approach scoring 8.86 SMAPE and 8.88 MAPE for a week in January 2018. Trained with time feature, weather parameters and holiday information.

holiday information, it is observed that the prediction can follow the load into the longer term holiday period, where the load is peaking (see Fig. B.11). It is observed for all the regression techniques during summer season, the vertical approach has better prediction compared to continuous approach, as measured by all the performance metrics including SMAPE, MAPE as well as MAE.

B.6 Conclusion

The regressors, kNN and RF, are used with autoregression as well as autocorrelation and correlation among parameters for the relative comparison for prediction accuracy. Autocorrelation is a neat and practical approach to feature engineering that saves time for the appropriate actions to be made for feature extraction. The regression tools can handle the low amount of data for day-ahead forecasting and the prediction measurements through MAPE is relatively much better compared to other techniques.

In this study, the regression analysis for load prediction of rural area Norwegian network is done using vertical and continuous time approach for day-ahead planning with 24 hour prediction. The vertical time approach uses seasonal data for training and inference, as opposed to continuous time approach that utilizes all data in a continuum from the start of the dataset until the time period used for inference. The regression tools can handle the low amount of data, and the prediction accuracy through MAPE matches other techniques. The vertical approach does this with even fewer data than continuous approach. It is observed that through load predictive analysis the autocorrelation by vertical approach with kNN-regressor gives a low SMAPE. The methodology used in this research work is developed to deal with the problems of irregularities and randomness in the time series, RF-regressor yield good result on day-ahead (24 hours) time prediction in load forecasting.

The presented load prediction analysis is going to be useful for distributed network operation, demand-side management, integration of renewable energy sources and distributed generator. To establish more accuracy for this work, the research is continued into the Deep Learning, exploring neural networks with capability of long short term memory.

B.7 Acknowledgement

The authors are grateful for the kindness, help from Dr. Sverre Tveit and Nissedal Cabins, and permitting us to use the data for this case study.

Bibliography

[1] C. Chu and H. H. Iu. Complex networks theory for modern smart grid applications: A survey. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 7(2):177–191, June 2017.

[2] Sandro Nižetić, Nedjib Djilali, Agis Papadopoulos, and Joel J.P.C. Rodrigues. Smart technologies for promotion of energy efficiency, utilization of sustainable resources and waste management. *Journal of Cleaner Production*, 231:565 – 591, 2019.

[3] Adeshina Y. Alani and Isaac O. Osunmakinde. Short-term multiple forecasting of electric energy loads for sustainable demand planning in smart grids for smart homes. *Sustainability*, 9(11), 2017.

[4] Nian Liu, Qingfeng Tang, Jianhua Zhang, Wei Fan, and Jie Liu. A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. *Applied Energy*, 129:336 – 345, 2014.

[5] Yang Zhang, Tao Huang, and Ettore Francesco Bompard. Big data analytics in smart grids: a review. *Energy Informatics*, 1(11), 2018.

[6] S. Yin and O. Kaynak. Big data for modern industry: Challenges and trends point of view. *Proceedings of the IEEE*, 103(2):143–146, Feb 2015.

[7] Nils Jakob Johannesen, Mohan Kolhe, and Morten Goodwin. Relative evaluation of regression tools for urban area electrical energy demand forecasting. *Journal of Cleaner Production*, 218:555 – 564, 2019.

[8] Alexei Botchkarev. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. *ArXiv*, abs/1809.03006, 2018.

[9] Nils Jakob Johannesen, Mohan Kolhe, and Morten Goodwin. Deregulated electric energy price forecasting in nordpool market using regression techniques. *IEEE*, unpublished, 2019.

[10] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030 – 1081, 2014.

[11] Spyros Makridakis, Evangelos Spiliotis, and Vassilios Assimakopoulos. Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13, 03 2018.

[12] T Aanensen and M. Holstad. *Statistics Norway*, 2018.

[13] J. Xie, Y. Chen, T. Hong, and T. D. Laing. Relative humidity for load forecasting models. *IEEE Transactions on Smart Grid*, 9(1):191–198, Jan 2018.

[14] Tao Hong. Energy forecasting: Past, present and future. 2013.

[15] J. Zimmermann. 'end-use metering campaign in 400 households in sweden'. *SEA*, 2009.

[16] Z. Ivanovski, A. Milenkovski, and Z. Narasanov. Time series forecasting using a moving average model for extrapolation of number of tourist. *UTMS Journal of Economics*, 9(2):121–132, 2018.

[17] Irena Koprinska, Mashud Rana, and Vassilios G. Agelidis. Correlation and instance based feature selection for electricity load forecasting. *Knowledge-Based Systems*, 82:29 – 40, 2015.

[18] Alexander Tureczek, Per Sieverts Nielsen, and Henrik Madsen. Electricity consumption clustering using smart meter data. *Energies*, 11(4), 2018.

[19] Rui Zhang, Yan Xu, Zhao Yang Dong, Weicong Kong, and Kit Po Wong. A composite k-nearest neighbor model for day-ahead load forecasting with limited temperature forecasts. *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pages 1–5, 2016.

[20] N. J. Johannesen, M. Lal Kolhe, and M. Goodwin. Load demand analysis of nordic rural area with holiday resorts for network capacity planning. In *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–7, June 2019.

[21] Mark A. Hall. Correlation-based feature selection for discrete and numeric class machine learning. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, pages 359–366, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[22] Florian Ziel. "modeling public holidays in load forecasting: a german case study". *"Journal of Modern Power Systems and Clean Energy"*, 6(2):191–207, 2018.

[23] N. J. Johannesen, M. Kolhe, and M. Goodwin. Comparison of regression tools for regional electric load forecasting. In *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–6, June 2018.

Appendix C

Paper C - Application of Regression Tools for Load Prediction in Distributed Network for Flexible Analysis

Nils Jakob Johannesen and Mohan Kolhe

Faculty of Engineering and Science, University of Agder, PO Box 422, NO 4604 Kristiansand, Norway.

Abstract **- The electrical load prediction is necessary for distributed network energy management and finding opportunity for flexibility in shifting the operation of non-critical power intensive loads. The application of regression tools has showed to be promising for predicting electric load within distributed network as well as for flexibility analysis. The distributed electrical energy network is low-capacity networks with low amount of data that need flexible operation and analysis. Random forest regressor, k-nearest neighbor (kNN) regressor, and linear regression are considered for analyzing electrical energy demand forecasting. The methodology, used in this chapter, dealing with the problems of irregularities and randomness in the time series considering urban and rural area case studies. Random forest-regressor yields good results on hourly time prediction in load forecasting. The kNN shows precise prediction due to its capability to capture the nearest step in a time series based on the nearest neighbor principle. The presented vertical time approach uses seasonal data for training and inference, as opposed to continuous time approach that utilizes all data in a continuum from the start of the dataset until the time used for inference. The regression tools can handle the low amount of data, and the prediction accuracy matches with other techniques.**

Keywords *Power system flexibility, load prediction, distributed network, regression tools.*

C.1 Load Flexibility and Management

Load flexibility relates to the ability of power system to shift the operation. The flexibility has to respond to the variability and uncertainty of the net load. The increasing penetration of variable renewable generation increases the need for flexibility in the load demand. A flexible power system can adapt to rapid change in supply and demand. The flexibility of resources is defined by their dynamic capabilities such as ramp time, start-up/shut-down time, operating range (minimum and maximum operating level) as well as minimum up and down times of the energy generation system.

The regression tools can be used to understand the variation and uncertainty in load and supply, as well as to analyze and forecast the expected output. Regression techniques can be used to model the past behavior, to understand and help to predict the future scenarios both on demand and generation.

Flexible electric power system operation is going to help in integrating a mix of energy sources that can respond to the varying demand for electricity. This demand is met with three types of plants typically referred to as baseload (meeting the constant demand), intermediate load (meeting the diurnal changes), and peaking (meeting the peak demand). At very high penetration of RG, a key element of system flexibility is the ability of baseload generators, as well as generators providing operating reserves, to reduce output to very low levels while maintaining system reliability. Although baseload generators are a capital incentive, but inexpensive small-unit generators are favored [1, 2, 3, 4, 5].

Demand side management is an umbrella term that describes the utility company efforts to improve energy consumption at customer site, the demand side of the meter [6]. Demand response (DR) is the customers' adaptation to alter their normal electricity usage in response to the adjusted electricity prices with grid constraints or other incentives created to decrease energy consumption at times of shortage or when system reliability is at risk [64]. The introduction of advanced metering system in the form of smart energy meters (SEM) allows for an unprecedented granularity in data gathering, and hence unlocking the potential of DR. The SEM implements an advanced measurement infrastructure, a two-way communication between the end-user and the distribution management system. SEM monitors, measures, and reports electric energy load demand in near real-time [8]. Traditionally, utilities have used three types of generating facilities to serve the diurnal and seasonal changes in load demand: Baseload, intermediate load, and peak load plants [9]. A load demand curve for a sample European country shown in Fig C.2 illustrates typical load demand patterns, where the segments indicate natural threshold level typical for baseload, intermediate load, and peak load. Yearly seasonal load demand of a selected European country is given in Fig. C.3.

The diurnal changes start with a surge demand in the morning when industrial companies commence activity and domestic end-users start their home appliances; it is the first peak in the load demand curve. Following the early morning activities, load demand

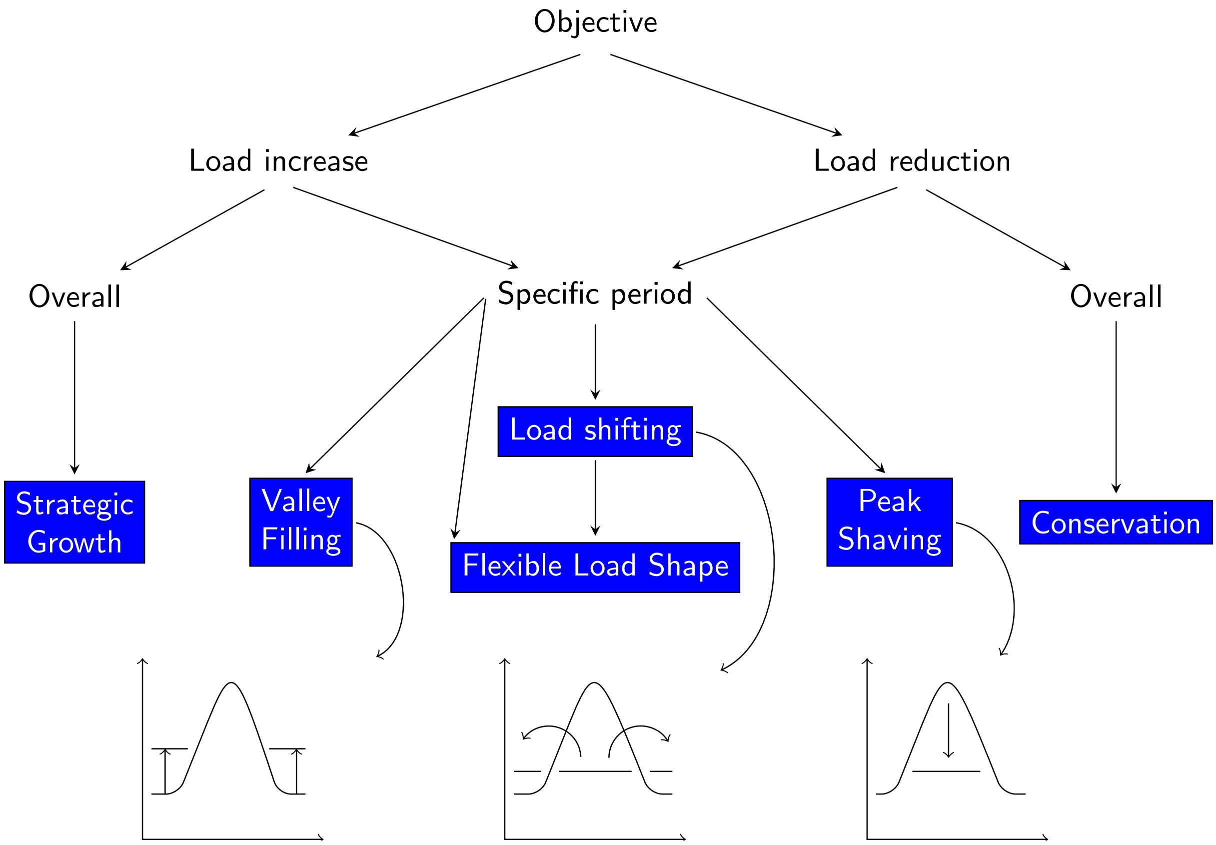


Figure C.1: Objective of load shapes

stabilizes; there is a dip in the load demand creating a valley in the load curve. When the working day is over, another surge load follows when people return to home and start cooking. The last diurnal valley in the load demand curve commences in the night time when people go to bed.

Depending on the operative flexibility of generators, they serve different load demand [10]. Efforts have been done to advance more flexible operation for managing the range between peak power and minimum load. Load cycling has a degenerating effect on units, impairs power production and leads to frequent breakdowns and unplanned maintenance [11, 12]. Different techniques are used to create a better match between load and supply. Peak clipping or peak load shaving is to reduce the peak demand. Another incentive is to fill up the valleys where demand is low Load shifting as seen in Fig C.1, combines the two previous techniques by shaving of the peak demand and filling the low-demand valleys [1].

Load shifting regime is crucial to development of microgrids within the distributed network. Microgrids are designed without peaking generator, thus reserve their capacity and up to 10% of load is not utilised [13]. These tasks can be solved by robust electric energy load demand forecasting. Demand forecasting is done by understanding how the past influences the future by learning from the past in order to prophesy the future.

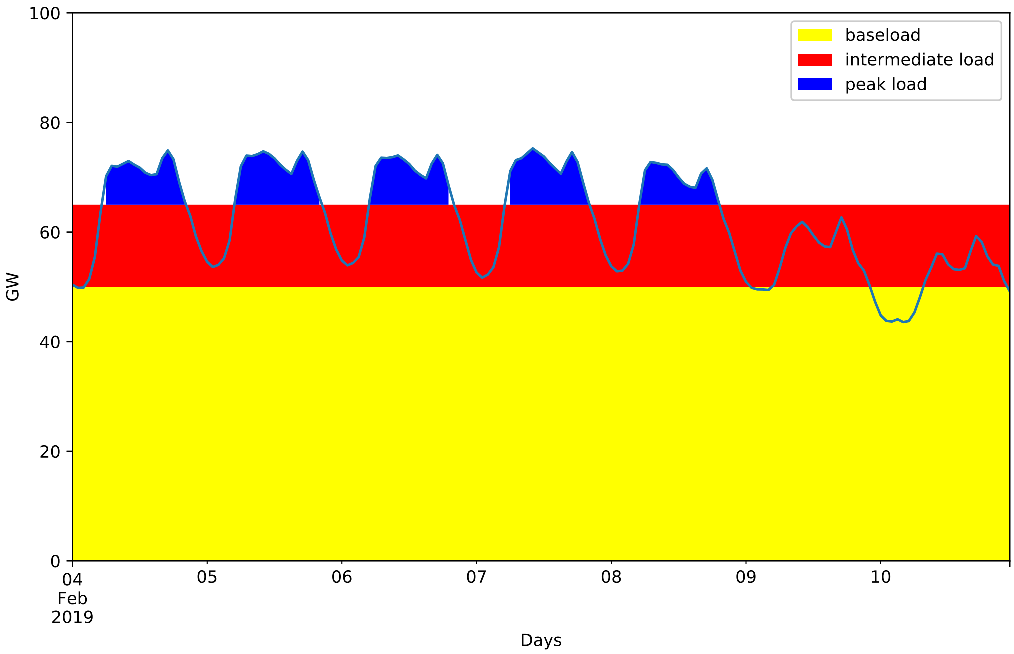


Figure C.2: The electric load demand curve of a sample European country for one week, indicating level of load curves. Source: ENTSOE-E

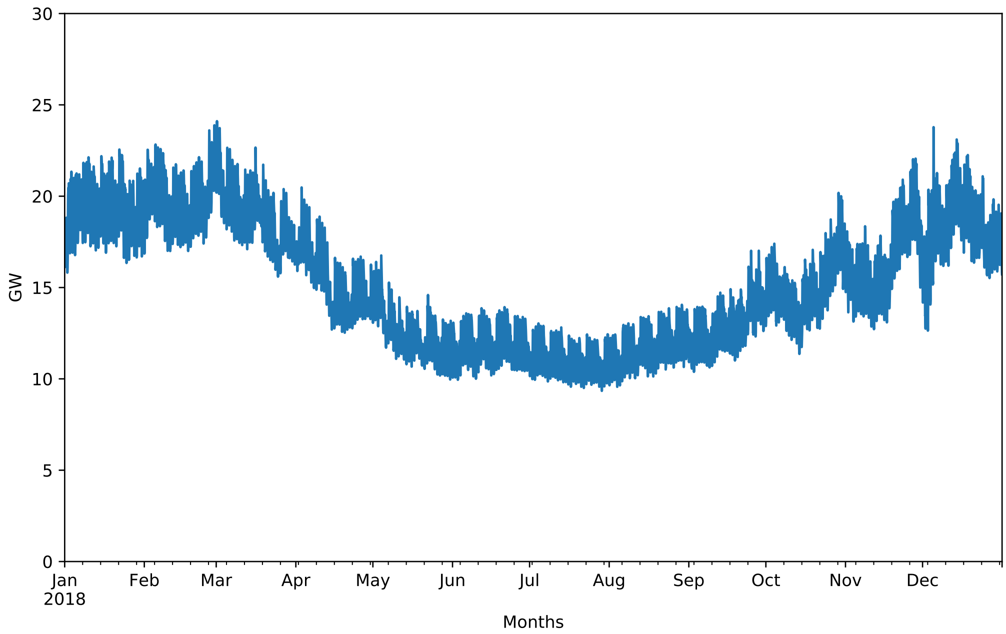


Figure C.3: The yearly electric load demand curve of a sample European country, depicting seasonal changes. Source: ENTSOE-E

C.2 Conventional Electric Load Forecasting Techniques

The electrical load forecasting has been carried out using conventional mathematical techniques. The traditional forecasting techniques are based on linear regression series. Most of them use statistical techniques. A time series is a collected sequence of events, based on the assumption of an inherent structure. The inherent structure is analytically observed trhough means such as autocorrelation, trend, and seasonal behaviour. There are many different scenarios of how these sequences of events are collected and described. The most often used time series techniques are in particular autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), autoregressive integrated moving average with exogenous variables (ARIMAX). For stationary processes ,ARMA is usually used, and it has been extended to ARIMA for non-stationary processes. ARIMAX is the most natural tool since electrical load generally depends on exogenous variables such as weather and historical time series data. Time series forecasting, its data and analysis will in the future be increasingly important as the availability and scaling of such data is growing through Internet of things (IOTs), the rise of smart cities, and due to the advanced infrastructure metering. The continuous monitoring and data mining will pave the way for adequate time series analysis, both statistical and machine learning techniques, as well as hybrid models will increase.

Time series analysis has traditionally been performed in meteorology, energy, and economics. The era of modern time series analysis started and the Box-Jenkins model was introduced [4]. The Box-Jenkins method has been further developed by the research community to a robust parsimonious ARMA for multivariate forecasting, requiring less human intervention [5]. Additional improvement has been reached with a combined Box-Jenkins econometric approach to forecast monthly peak system load. By observing changes in economic and weather-related variables in a Box-Jenkins time series model, refined forecasts are obtained [6]. It is common for these approaches that they use multiplex mathematical computations and possess a heavy computational burden [7].

Machine learning models seriously contested the classical statistics with the artificial neural networks (ANN) [18]. The neural networks can aid dispatchers deal with uncertain loads [19]. ANN is used with updating network parameters, generating plant control and economic power dispatch problem [20, 21, 22, 23]. A typical neural network model with back propagating adjusted weights is presented in Fig. C.4. In the following years during the 1990s, the research on ANN in electric load forecasting was mainly concerned with regional loads in the MW-scale, resembling the load consumption of a medium size European country and including multivariate time series analysis [24, 25, 26].

Focus has also been attuned towards case and system dependency of ANN [27], the explainable and interpretative ANN, and the “black box nature” of neural networks. This has paved the way for ensembles of trees, linear fits, Support Vector Machines (SVM), and other machine learning models. Some of these models find their origin in the statistics and overlapping with machine learning (see discussion [28]) [65, 30]. Deep learning techniques based on long- or short-term memory and recurrent neural networks have shown promising results for optimal scheduling of microgrids [31]. Also, the convolutional neural networks (CNN) show good results, but need big load schemes in GW-scale to perform well [32].

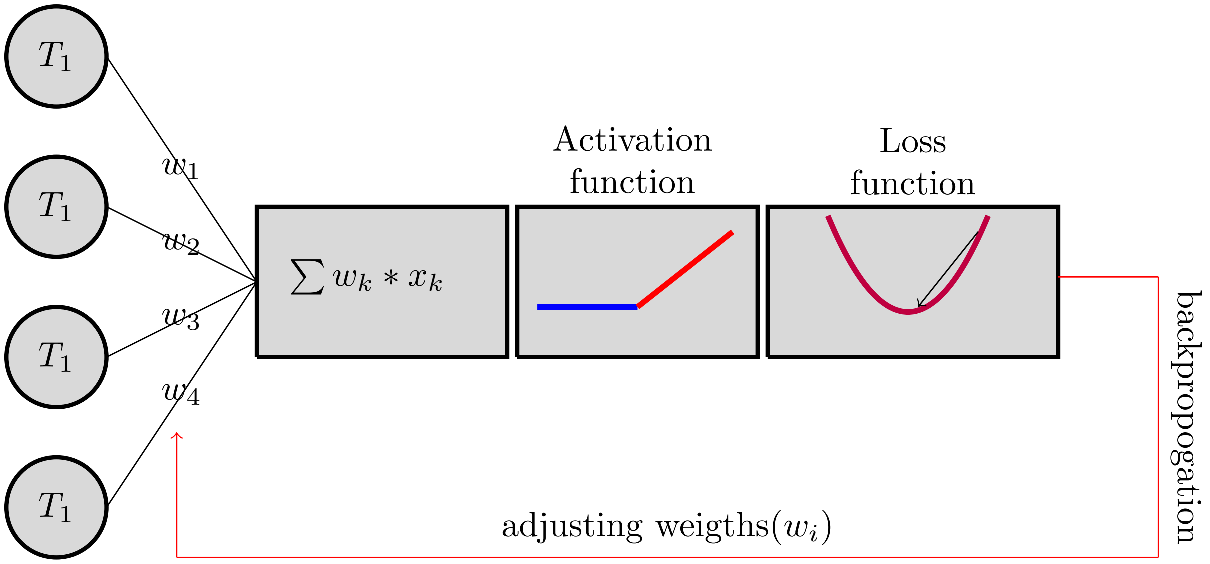


Figure C.4: Neural network model with the propagating adjusted weights

C.3 Learning Systems

Machine learning provides a framework for estimating from the observed data to form an appropriate model in the time dependencies. Machine learning is a subcategory of artificial intelligence and usually divided into two main types, supervised and unsupervised learning. Unsupervised learning is learning without any prior knowledge of the aim of learning, and is also named as knowledge discovery. Hence, the unsupervised learning can be state dependent or clustering. For the supervised learning, the aim or independent variable is known. In supervised learning, data is orchestrated in such a way that it fits the aim.

In supervised learning, x and y are preserved in a train and test set. Here, D is called the training set and N is the number of training examples. Test-set, is preserved for inference purposes. When the inference is performed, the algorithm is normally verified according to a performance metrics. In the predictive or supervised learning approach, the goal is to learn a mapping from inputs x to outputs y, given a labeled set of input-output pairs $D = (x_i, y_i)_{i=1}^{N}$. Given the inputs, $D = x_{i}{}_{i=1}^{N}$, the aim is recognizing patterns in the data. The problem at hand is undefined, and we don't know what to look for, and no use of performance metric as we do not have a given x to the observed value; the response vector y [33].

C.4 Regression tools

Regression is distinguishable from classification by the response vector (y), which is a continuous output of time, whilst in classification, the y vector is categorical. In this sense, the classification is a subdivision of regression [71]. For this reason, regression has been known by machine learning practitioners "learning how to classify among continuous classes” [35].

Regression methods vary from purely statistical methods, machine learning techniques to hybrid models that combine two methods. The regression tools can be parametric, where a particular distribution constitutes the method, either by direct measures or when posing a relationship to external parameters. The non-parametric regression methods do not prescribe any certain distribution, hence regress on pure mathematical foundations. The semi-parametric regression models combine an underlying distribution with a pure mathematical relation. A feature used in many of the regression tools is correlation techniques, either to research the data for their general function, or in multivariate time series that correlates to external parameters. Correlation is a measurement to how two ranges of data move together. The Pearson Correlation Coefficient (r) computes the linear relationship between two variables, in a range from – 1 to + 1 [36]. If the relationship is in the proximity of 1, it means that when x increases so does y, and at exact linearity, the opposite is true for – 1, which means that when one variable increases, the other decreases.

$$a = 1 \tag{C.1}$$

Autocorrelation function (ACF) shows how a time series is correlated to its own lagged version at each lag_k [37]:

$$\rho_k(t) = \frac{\sum_{i=1}^{n-k}(x_t - \hat{x})\sum_{i=1}^{n-k}(y_{t+k} - \hat{y})}{\sqrt{\sum_{i=1}^{n-k}(x_i - \hat{x})^2}\sqrt{\sum_{i=1}^{n-k}(y_{t+k} - \hat{y})^2}} \tag{C.2}$$

Cross-correlation can be found when one of the variables is shifted in time (t), and can be used to alter the time lags between the variables for a reshaped perspective of the relationship between them. As the times series are cross-correlated, an evaluation of temporal similarity is made [38]:

$$\rho_{xy}(t) = \frac{\sum_{i=1}^{n}(x_i - \hat{x})\sum_{i=1}^{n}(y_{i-t} - \hat{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \hat{x})^2}\sqrt{\sum_{i=1}^{n}(y_{i-t} - \hat{y})^2}} \tag{C.3}$$

Autoregression (AR) is a simple and straightforward regression technique, where past values of the univariate time series are dependent on their own lagged version defined by a parameter weighting of each input, ϕ, and therefore a parametric model. The current value of y(t) is expressed by previous values of time $y_{t-1}, y_{t-2}, ..., y_{t-p}$. The order of an AR process is defined by the number of past values of y(t) it is regressed on. AR(p) is defined by the last y_{t-p}, considered in the process, denoted as:

$$y(t) = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t \tag{C.4}$$

Where the error term ϵ_t, is white noise defined by a constant mean and some unknown fixed variance $\sigma_\epsilon^2(t)$, a stationary process. The ACF of a white noise process is zero at all lags other than lag zero where it is unity, to indicate that the nature of its process is completely uncorrelated. By using backshift operator (B), the previous value of the time series is related to the current value $y_{t-1} = By_t$, and thus; $y_{t-m} = B^m y(t)$, and the error term is explained as:

$$\phi(B)y_t = \epsilon_t \tag{C.5}$$

An AR process p-value is defined by the autocorrelation of residuals of the AR process. If the residuals autocorrelation falls within a confidence interval, normally considered as 95%, the autocorrelation function of the residuals are considered to be white noise. If not, the AR process will still continue to find another parameter, until its residuals satisfy the criteria of white noise. If the current and previous values of a white noise series $\epsilon_t, \epsilon_{t_1}$ are expressed linearly, it is known as moving average process (MA), and an equivalent implementation of backshift operator (B) would be:

$$y(t) = \theta(B)\epsilon(t) \tag{C.6}$$

A combination of the two processes is the ARMA. If the mean or covariance of the time series observations change with time, the series is defined as non-stationary, and a differencing process makes it stationary by introducing the ∇ operator, and the AR, MA and ARMA processes are transformed into ARI, IMA or ARIMA process.

C.4.1 Linear Regression

Another parametric model is multiple linear regression (MLP) that assumes a linear relationship in the training data and to explanatory variables to explain relationship to the response-vector (y):

$$y(t) = a_0 + \beta_1 x_1(t) + ... + \beta_n x_n(t) + \epsilon(t) \tag{C.7}$$

where $x_1(t), ..., x_n(t)$ are independent explanatory variables correlated with the dependent load variable y(t). The independent variables are found through correlation analysis, and coefficient estimation normally found through least square estimation, or iteratively reweighted least squares (IRWLS). All parameters start at 0 and is step-wise improved using backpropagation through a loss function to find appropriate weights, or through finding the intercept a_0. Each explanatory variable finding its coefficient based on the covariance and standard deviation of dependent and independent variables is defined as:

$$\beta_x = \frac{\sigma_{xy}}{\sqrt{\sigma_x}} \tag{C.8}$$

C.4.2 k-Nearest Neighbor Regression

Opposite to the linear regression (LR) is the k-nearest neighbor (kNN) regressor, which is non-paramteric, relying on its own table look-up and mathematical foundation, and highly non-linear.

$$y_{knn}(x) = \frac{1}{K}\sum_{k=1}^{K} y_k \text{ for K nearest neighbours of x} \tag{C.9}$$

The kNN-classifier is illustrated in Fig. E.1, where the left diagram with a small encirclement options for k = 1, where simply the nearest neighbor decides the class of prediction, whilst in the right diagram in Fig. E.1, the number of k is increased to more then one [70].

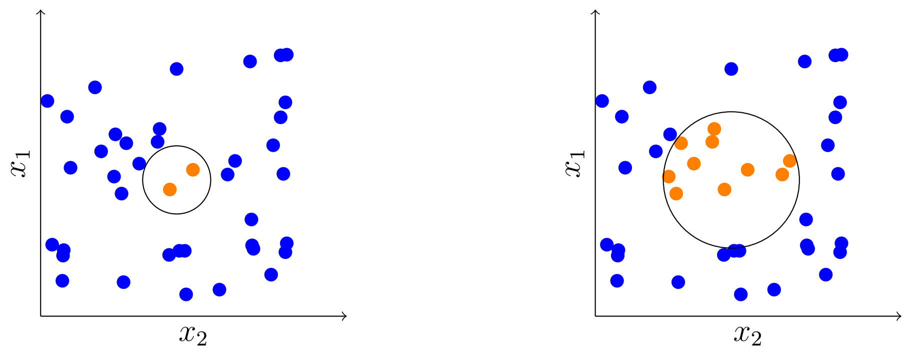


Figure C.5: k-Nearest Neighbour classifying based on the k'th observation.

Using k = 1 can lead to false prediction, and a set of kNNs is often used. When classifying the dependent variable is categorical, it can easily be made numerical by regression. The kNN regressor makes a regression based on the number of kNNs to minimize false predictions. The model considers a range of different kvalues to find the optimal value. The kNN regressor needs thorough pre-processing and feature engineering to limit the effect of noise caused by irrelevant features, and is, therefore, dependent on finding the appropriate distance model [71]:

C.4.3 Distance

A variety of distances is used in the algorithm. As seen in Equations C.10, C.11, C.12, and C.13, they are mostly used, since it is easy to intersect by changing the variable q. The variable q is also considered to find the optimal value.

C.4.3.1 Manhattan/City Block Distance

$$d(x, y) = \sum_{i=1}^{k} |x_i - y_i| \tag{C.10}$$

C.4.3.2 Euclidean distance

$$d(x, y) = \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2} \tag{C.11}$$

C.4.3.3 Minkowski Distance

$$d(x, y) = \left(\sum_{i=1}^{k} (|x_i - y_i|)^q \right)^{\frac{1}{q}} \tag{C.12}$$

C.4.3.4 Chebychev Distance

$$d(x,y) = \lim_{q\to\infty}\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{\frac{1}{q}} \tag{C.13}$$

C.4.4 Random Forest Regression

Random forest (RF) regression is a combination of decision trees, found through recursive partitioning to build a piece-wise linear model. From these tree models, it uses a majority vote for the most popular class. The trees grow dependant on a random vector, and the outputs are numerical scalars [73]. Each leaf on the tree is a linear model constructed for the cases at each node by regression techniques. One sole decision tree encompasses attributes and classes in the data and uses an entropy function gain function to distinguish its structure. Entropy is known from thermodynamics as a measure of disorder, and later adopted by the information theory. In information theory, entropy is a measure of uncertainty of a variable, and defines a pure classifier [74]. In equation (5) p is positive and n is negative:

$$Entropy(S) = -p * log2(p) - n * log2(n) \tag{C.14}$$

The entropy function is then used to evaluate the information gathered (gain) of an attribute, and thus to know how to choose the highest gaining attribute as the next branch in the decision tree. The equation yields the expected reduction in entropy, by imposing another branch in the decision tree.

$$Gain(S,A) = Entropy(S) - \sum_{\nu\in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{C.15}$$

In equation (C.15), A are attributes used for splitting the data into subsets (S). S is the sum of subsets, and Sv is the value of subsets. Using prior known input/output relationships, the algorithm searches for a model for the best prediction in the training set. The mathematical equations are structured in the algorithm, see Fig. C.6, based on the past knowledge.

C.4.4.1 Normalising

The pre-processing of data is a transformed so that the machine learning algorithm can learn the patterns and generate a sound forecast. In a standard normalization process, input data are transformed with values from zero to one. This is done to make the predictive algorithm more robust [42].

$$\frac{\hat{X} - X_{min}}{X_{max} - X_{min}} \tag{C.16}$$

$$\frac{\hat{X}}{X_{sum}} \tag{C.17}$$

$$\frac{\hat{X}}{X_{max}} \tag{C.18}$$

$$\frac{\hat{X} - X_{avg}}{X_{max} - X_{avg}} \tag{C.19}$$

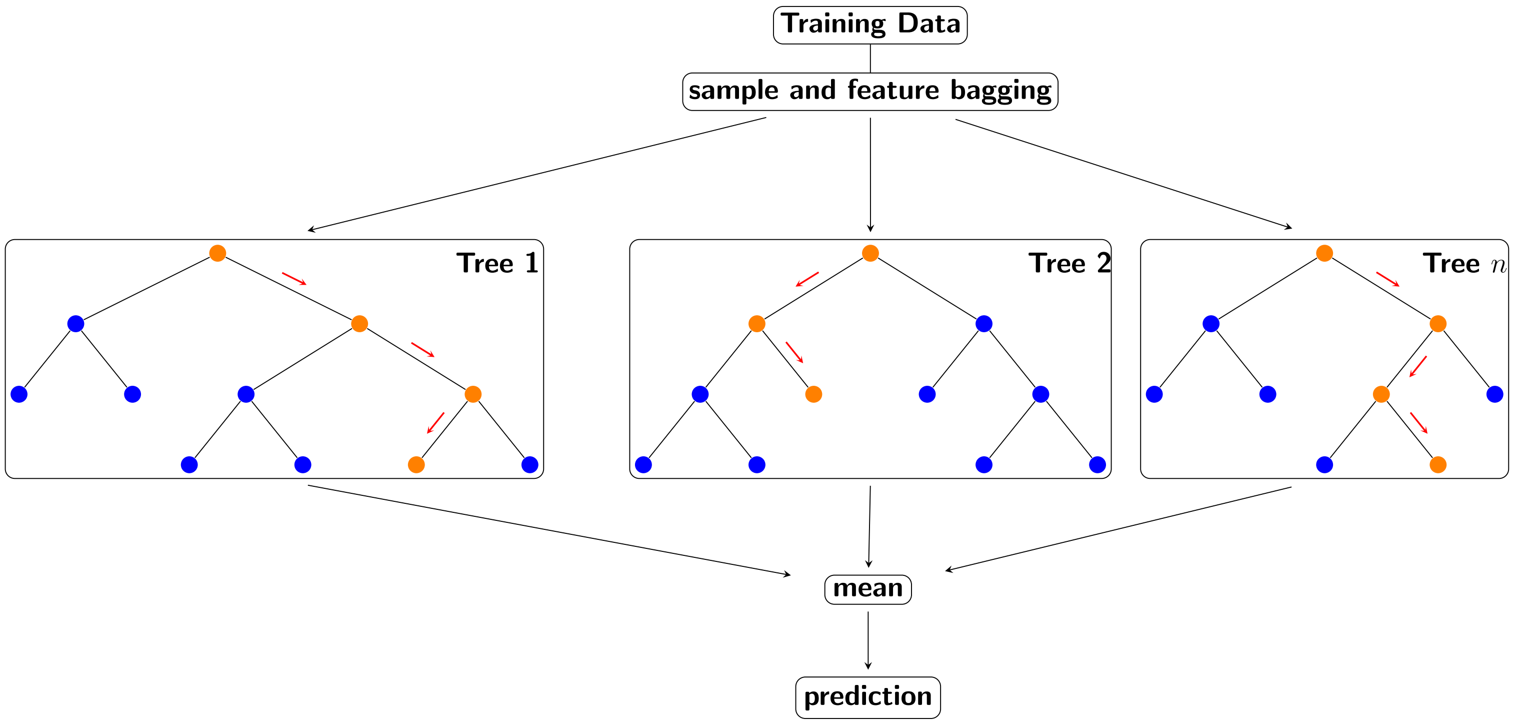


Figure C.6: Random Forest Regression diagram sampling and voting from n trees

C.4.4.2 Performance metrics

To evaluate the performance of load forecasting, a performance metric is used, including mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), and symmetric mean absolute percentage error (SMAPE) [43]. They are defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}| \tag{C.20}$$

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n} \left(\frac{|y_i - \hat{y}|}{(|y_i| + |\hat{y}|)/2} \right) * 100 \tag{C.21}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y})^2 \tag{C.22}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} |\frac{y_i - \hat{y}}{y_i}| * 100 \tag{C.23}$$

C.4.5 Visual Inspection

The first thing is to plot the time series of the data shown in Fig. C.7 and C.8. In these plots, the time series are plotted as univariate time series with y-axis representing the univariate or dependent variable, and x-axis being the time axis. By visual inspection, these plots are giving the main features of the time series. Important information such as time span, trends, and cycles are emerging in the figures. When applying intuition to visually inspect these time series, they certainly display some repetitive patterns, as in Fig. C.7, where load pattern seems to be taken a U-wave form that repeats itself over time. Fig. C.8 is much more dense then Fig. C.7, and looks to contain more information.

In some instances, a univariate time series can be explained by itself as is the case for univariate analysis; even then a univariate series can and most likely will be affected by other influences, but remains self-explanatory for this purpose. For the multivariate case where explanatory independent features are added, they are not directly connected to the dependent/response variable such as weather parameters, yet correlation exists to aid the time series analysis.

C.5 Applications of Regression Techniques for Electric Load Forecasting

Recent research from 2018 on computational intelligence approaches for energy load forecasting that reviewed more than 50 research papers related to the subject outlines the complexity of demand patterns as potentially influenced by factors such as climate, time periods, holiday or working days and other factors such as social activities, economic

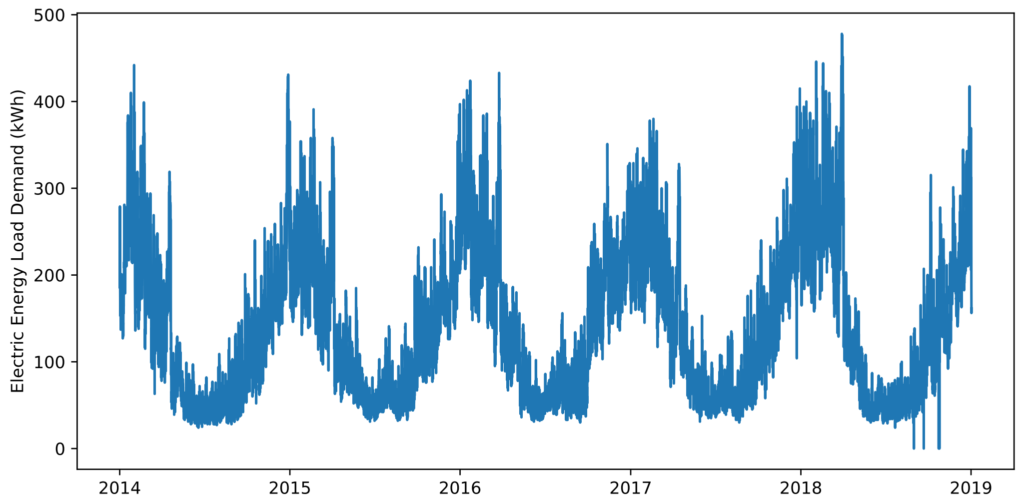


Figure C.7: Rural Load

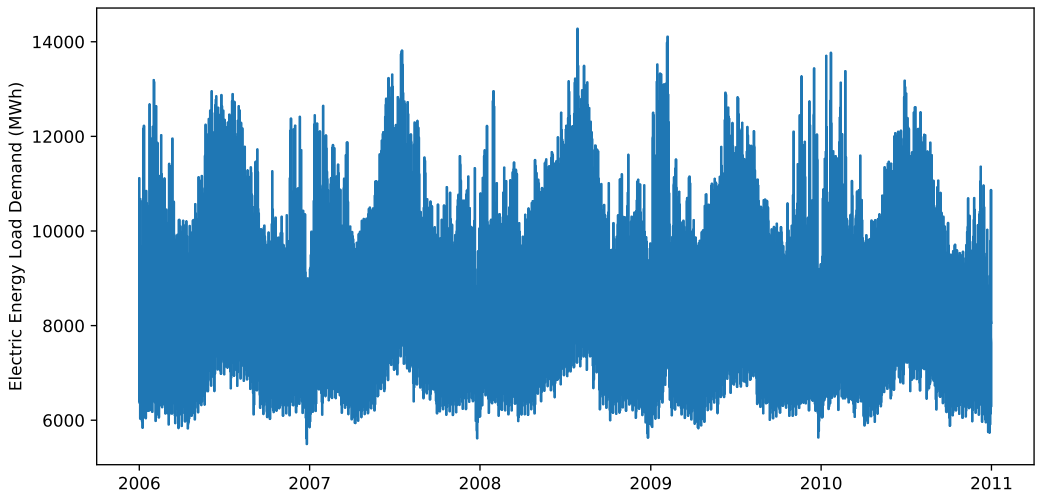


Figure C.8: Urban Load

factors, including power market policies. Electrical energy demand is influenced by meteorological weather conditions; therefore, it is necessary to include the impact of meteorological weather parameters on electrical energy demand forecasting; also renewable electrical energy production is nature-dependent. The future electrified grid will increasingly depend on renewable intermittent energy sources (solar, wind), and the individual load profiles of such a system will change radically as home appliances include new energy demanding appliances (e.g., heat pump, electric vehicles, and induction stove) [44].

The regression models kNN, LR, and RF are supervised machine learning algorithms with a numerical outcome. The model is trained to find rules for pattern recognition in the input to output relation. The inputs to the model are known as features. Neural networks are the preferred machine learning tool and are known as both feedforward and back propagating networks, where a number of inputs are weighted in order to provide a predictive outcome. Neural networks are good for detecting non-linearities, and therefore preferred as a predictive tool in electrical load forecasting, yet also often criticized for low transparency and lack of interpretability because of the black box approach and using a large amount of data. Overfitting is still a challenging issue when applying neural networks to electrical demand prediction. It is known as the bias-variance trade-off. When a model is of very low complexity and yet scores well, it is highly biased, which signifies that the data fits the model accurately (the training set), and it will often perform poorly on new data (from the test set). The model should contain a complexity that is in coherence with the level of information embedded in the data. Somewhere in between is the optimal model, also referred to as the suitable model [45].

Urban area load is influenced by meteorological conditions; therefore, it is important to include impact of weather parameters on load prediction, yet this impact is governed by the prediction time, greater for long term, and decreases as the prediction time is narrowed. The electrical energy demand is influenced by the user behavior as well as weather conditions. Individual human behavior and weather are so random that a complex neural network would not predict the outcome better than a coin toss. Hence, if one has to analyze the load demand of larger area such as the urban area, systematic load behavior with correlation to weather parameters and continuous load profile should be investigated. This work has uniqueness in electrical demand forecasting using regression tools through vertical approach, and it also considers the impact of meteorological parameters. This vertical approach uses less amount of data compared to continuous time series as well as neural network techniques. The objectives of this work are to explore the use of regression tools for regional electrical load forecasting by correlating lower distinctive categorical levels (season, day of the week) and weather parameters, see Fig.C.9. The vertical time approach is to consider a sample time period (e.g., seasonally and weekly) of data for four years, which will be tested for the same time period for the consecutive year. A vertical axis approach is shown to be competitive to ANN.

C.5.1 Feature engineering for electric load demand forecasting

The following three parameters are important for system electrical energy demand:
(i) Time
(ii) Weather
(iii) Random effects

C.5.1.1 Time

Apart from the seasonal effects, underlying patterns emerge in the system load demand. There are different peaks throughout the seasons, whether it is a winter peak or a summer peak. Emerging under this seasonal patterns are daily- and weekly-cycles. The daily routines of human behavior are manifested in systematic load patterns on a daily basis.

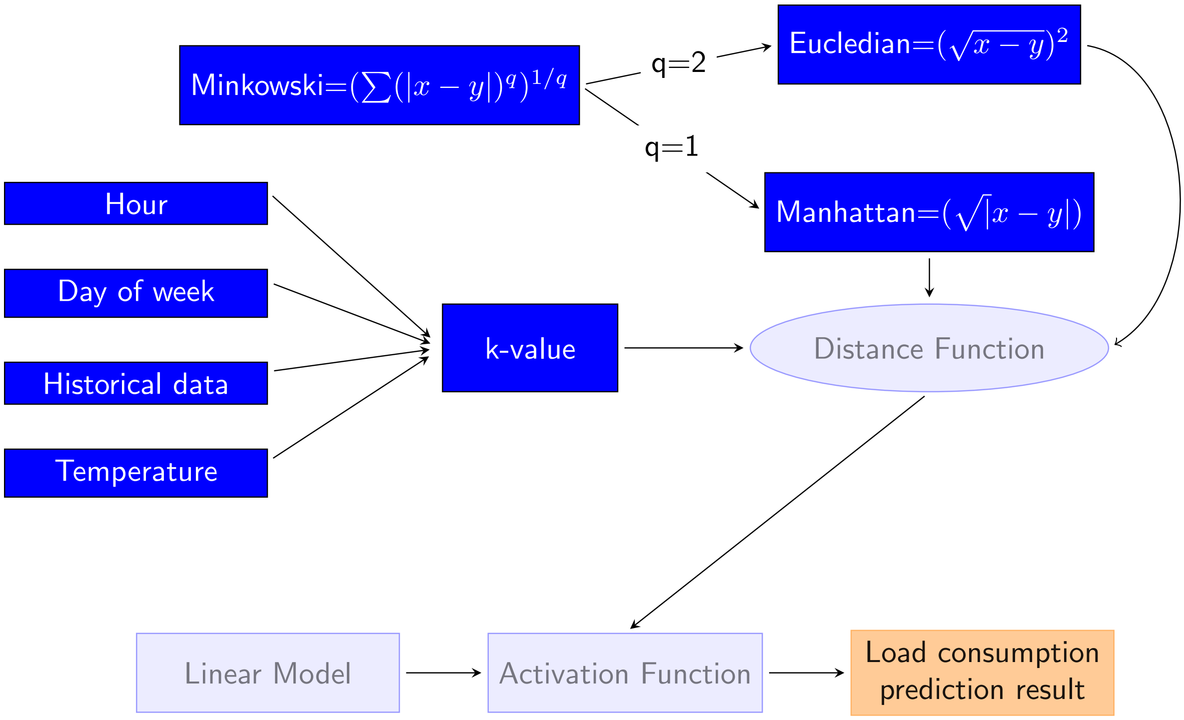


Figure C.9: The regressor model for electric load demand forecasting

Day of the week is also significant. Working day or off day or non-working day (weekend or other calendar event) changes human activities, and whether it is a working day or not, influences load patterns. People might also during weekends shift their sleeping habits, as to wake up later, and thus change the diurnal load demand to delay the morning peak load demand. Sub-categorical levels such as working/non-working days are referred to in the literature as an indicator variable. Such an indicator variable composes a lower indicator level, with a binary switch of working days and non-working days/holidays (0 and 1). To give this property to our algorithms is very important as it makes prediction of forecast load more efficient. The use of such type of variables has been successfully employed in the forecasting of electric market [42, 22, 47, 48, 49].

C.5.1.2 Weather

Weather variables play an important rule in changing load patterns. The effect of ambient temperature as well as past temperature on the load is necessary for prediction analysis; the indoor temperature on a hot summer day may reach its peak after sunset due to heat buildup in the construction materials of buildings. In addition to the daily heat buildup, a sequence of days with high temperature creates a new system peak. The time delay from shift in temperature until the change in electric is observed and should be evaluated through the temporal similarity of cross-correlation between the load and different weather parameters: DryBulb, DewPnt, WetBulb, and Humidity. Dry bulb temperature (DBT) is the temperature measured from air, yet not exposed to solar radiation or moisture. Wet bulb temperature (WBT) is measured from a thermometer where the bulb of the measurement device is soaked by a wet cloth. As long as the air is not saturated, evaporation from the moist cloth keeps the WBT lower than the DBT. From the DBT and WBT, one can then derive the relative humidity of the air and the dew point from a Mollier Chart by psychometric. In humid and hot conditions, it is likely that humidity will effect the load pattern in similar ways as temperature. Humidity explains the complex relation between temperature and load, and therefore mathematical models are not enough in a thorough analysis. Humidity is the amount of water vapor in the air, and might increase the gap between the actual and the apparent or felt temperature. When regulating temperature, the body utilizes evaporative cooling, and the rate of evaporation through the skin is correlated to humidity, and because of the conductive properties of water, we feel warmer at high humid conditions. Also, due to the seasonal changes of weather data, the correlation to the electrical load will vary during the year. Many electrical utilities are weather-sensitive such as heating and air conditioning. Electric loads are often classified as weather-sensitive load and non-weather-sensitive load. Temperature data is obviously a very important factor affecting the load. However, its value is often limited to the confidence level on weather forecasting. Therefore, unless the weather forecasting is very accurate, an underlying deterministic model is its premise. The complexity in the control system engineering of maintaining thermal comfort as well as optimizing for energy is important to know. At the same time, it is important to acknowledge that most houses are designed to resist the worst meteorological conditions. There are also limitations in the heating system itself that might cause load peaks, such as the inertia in the floor heating system, known as thermal lag. Therefore machine learning can help to use the weather parameters for load predictions in the built-environment [50, 51, 52, 53].

C.5.1.3 Random effects

Random disturbances lead to increase the number of electricity consumers due to many factors. Infrastructural changes in the urban area and maintenance work are random effects that are not detected by pattern recognition. Load patterns are consistent from year to year, and show reoccurring seasonal pattern. When the yearly load curves do not vary from year to year, it means that there are no economic trends. Load prediction analysis using machine learning can take care of random effects.

The effect of external parameters on load predictions can be considered through the machine learning approaches for different type of loads (e.g. rural area and urban area loads).

C.6 Case study 1: Rural Area Electric Energy Load

In this study, the dataset for rural area electric energy load is the data collected by a smart meter at a electric substation providing Nissedal Cabin Area in Bjønntjønn with power. It is a typical Norwegian rural power network with 125 cottages, and 478 kW peak demand. The dataset is hereby referred to as the Bjønntjønn dataset. The rural area load profile is illustrated in Figure 4.7. The smart meter collects data at every hour, as a point value, making it a dataset of hourly values. The weather information by Norwegian Institute of Bioeconomy Research (NIBIO) runs 52 weather stations with detailed information down to hourly resolution and freely downloadable on their web service (lmt.nibio.no). Among the 52 weather stations, three weather stations closest to Bjønntjønn Cabin Area are Bø, Gvarv, and Gjerpen. Based on the correlation analysis, the weather station with the strongest correlation of temperature to the load data from Bjønntjønn Cabin Area is identified, and used for the analysis.

C.7 Case study 2: Urban Area Electric Load

The dataset for urban area electric load contains 87648 collected datapoints from the urban area of Sydney in the region of New South Wales in Australia. It is called the Sydney dataset. These datapoints are collected at every 30 minutes, spanning from five years. Since it is the granularity of collected data observations that decides the lower limit of forecast window, this dataset gives the oppurtunity of 30 minutes predictions. The historical data is gathered by Australian Energy Market Operator (AEMO) and Bureau of Meteorology (BOM) from years 2006 to 2010, and hereafter referred to as the Sydney dataset. During the years 2006–2010, the maximum load was 14274.2 MW. In this study the purpose is to test the regression tools on the available real data of urban area.

C.8 Results and Discussion

In this work, several regression tools have been analyzed and compared for different datasets. Based on the analysis of the data and regressors, a new vertical approach has been further developed and inferred to deal with the relatively low amount of data and load pattern; it has been in particular validated for the case studies (i) in the rural area and (ii) in the urban area.

The vertical time approach also uses seasonal data for training and inference. The horizontal approach uses continuous datasets, i.e., it utilizes all data in a continuum from the start of the dataset until the time period used for inference. The illustration of horizontal and vertical approaches is presented in Fig. C.10.

Vertical approach can be performed with minimum amount of data compared to continuous approach. Also, the vertical time approach predictive results are compared with prediction based on continuous time series data. In vertical approach, the training set, $D = \{x_i\}_{i=1}^{N}$, is partitioned into subsets by each season of the year, and then are merged together only containing seasonally information about the load pattern. In a dataset containing time observation for five years (e.g., 2016–2020), time is separately selected season-wise, and then merged to contain only the specific season for training, $D = \{x_{spring_i}\}_{i=2016}^{2019}$. In this study, the inferred test-set is for a week in the middle of the selected trained season for the following year $D = \{x_{week}\}_{i=monday}^{sunday}$. Seasons are divided by months, as seen in Table C.1, where Season 1 is Winter, and Season 4 is Autumn.

C.8.1 Case Study 1: Rural Area

In the case study of rural area load prediction, the regression analysis has been done on continuous time basis as well as using vertical time axis approach. The correlation analysis of load and weather parameters has been analyzed to study the relation between meteorological parameters and electricity consumption. The hourly electrical loads of each season have been juxtaposed to the seasonal temperature, and negative correlation has been observed (Fig C.11).

From this observation, it can be seen that vertical approach enables the algorithm to reveal complexity of load and temperature for better prediction results [54]. The relation between working days and non-working days affects the cycles of load consumption, and is noticeable in the latter part of of the holiday where load demand increases even more (Fig. C.12).

The load pattern shows autocorrelation (AC) to previous lags, as seen in Fig. C.13. The AC aids the feature extraction procedure in engineering for the optimal previous k-lag values to be selected for the predictive algorithm. The observed results from the the autocorrelation function (ACF) plot (Fig C.13), shows a steep linear decline in lags 0–5; after that the slope is almost horizontal (lags 6–15) before it makes a small bump at lag 17–20, for then again to increase its value for the 23rd lag (which is the 24th hour since unity lag is zero), and then a deep decrease. The ACF plot also shows strong dependencies on historical data values, which indicate that the time series is autoregressive. The further correlation analysis of the rural electrical load demand patterns reveals also a strong dependency on the day of the week. For the considered Norwegian rural load of holiday cabins, the Norwegian holidays are identified as Easter, labor day, national day, ascension day, Pentecost, and X-mas. The observed correlations between the load and temperature, load and working days/non-working days, and the intercorrelation of temperature and working days/non-working days for the rural area have been well within the good heuristic model for correlation-based feature selection. The heuristics of good correlation-based feature selection is based on the level of intercorrelation within the class and subset features. In the rural area, there is no correlation between the working days and temperature. A good feature set contains independent variables that have high positive or negative correlation to the dependent variable, and no correlation amongst the other dependent variables [55].

Season		Months	
Season 1	December	January	February
Season 2	March	April	May
Season 3	June	July	August
Season 4	September	October	November

Table C.1: Seasons

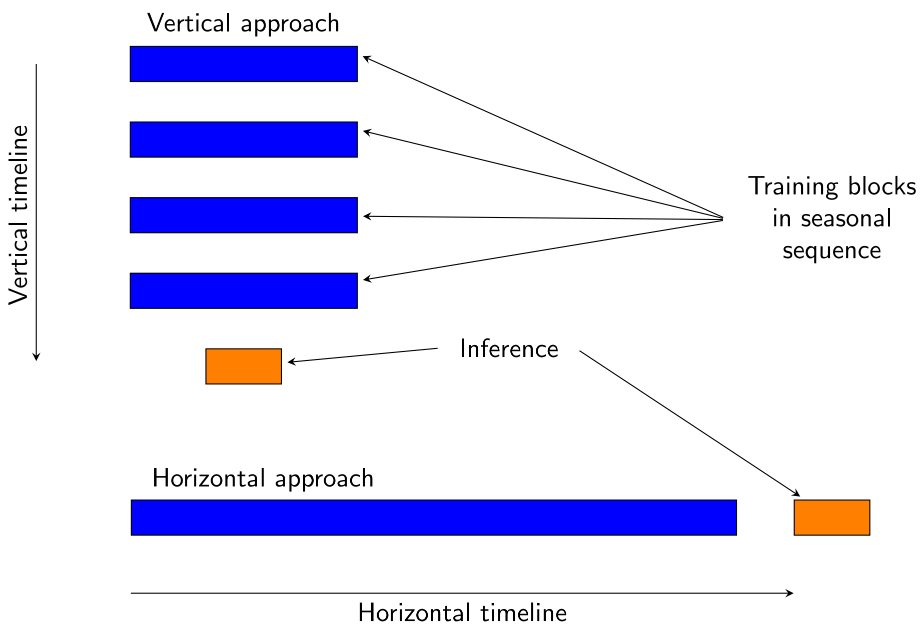


Figure C.10: Illustration of vertical and horizontal approach.

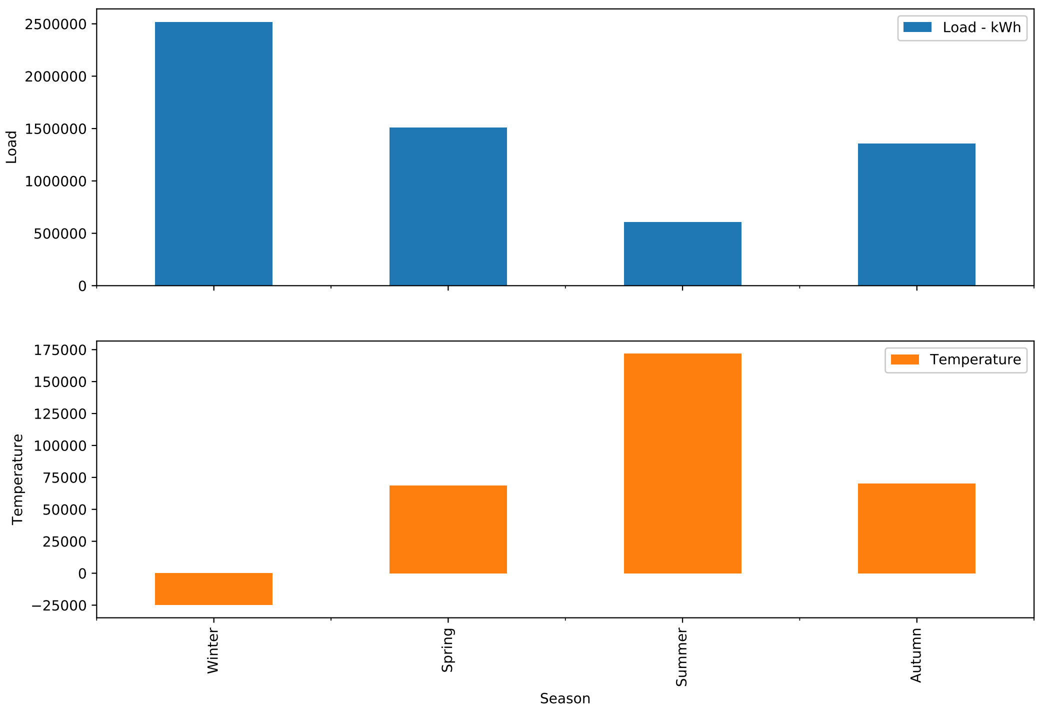


Figure C.11: Load consumption and temperature profiles on seasonal basis

In the further evaluation of the regressors performance metrics are used (Table C.2 and C.3)

In this work, different features in the regression tools (kNN and RF) have been studied to

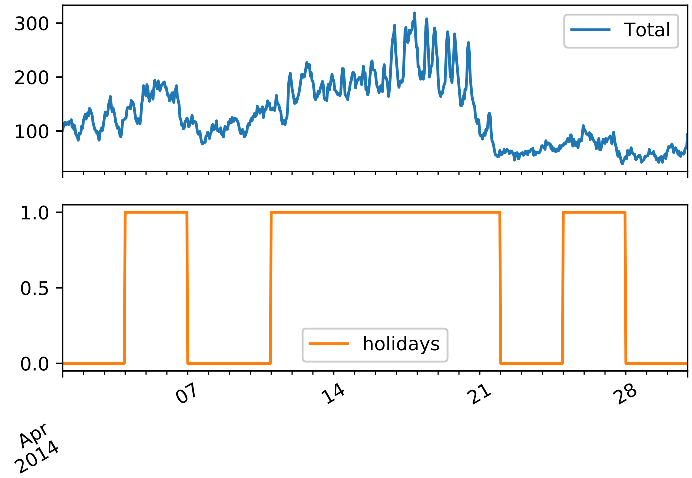


Figure C.12: Load consumption related to working and non-working days

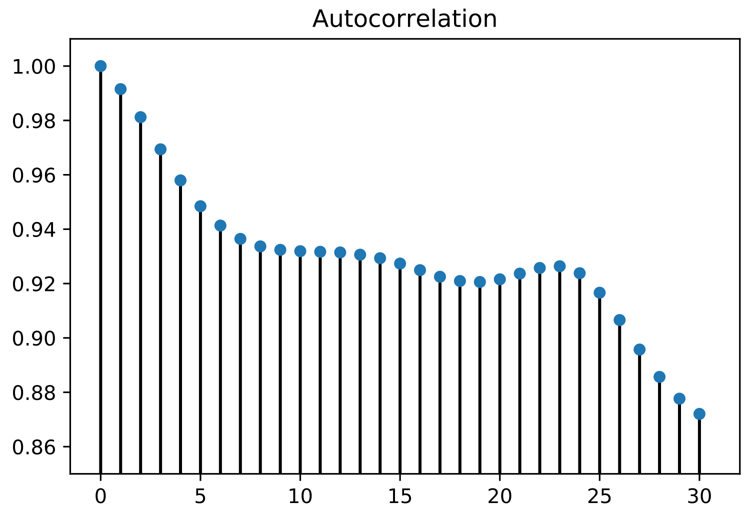


Figure C.13: Autocorrelation of load consumption of the first 30 lags for Bjønntjønn Cabin Area 2014-2018

analyze how they perform. In Tables (C.2, C.3), the autocorrelation (AC), autoregression (AR), temperature (T) and holiday effects (H) have been studied separately and together (AC, AR, T, H) combined with the regressors. The performance metrics SMAPE, MAPE, and MAE have been chosen to make appropriate analysis of their performance (see paragraph C.4.4.2). MAE is the most straightforward error estimation, but is poor in order to understand the context it is given; therefore MAPE is more used, since it is normalized to the true value of time series. Typially for the rural area, the load demand is low, opposite to the urban area, and occasionally the rural area load reaches zero. At zero values the MAPE is obsolete and the performance is also measured by SMAPE.

The Table C.2, compares the vertical and continuous approach for the winter season, whilst Table C.3, compares the vertical and continuous approach for summer season.

Table C.2: Forecasting Results (24 hours prediction) for season 1 (winter) trained with time feature lags of 24-, 48- and 168-hours

Features	Vertical winter			Continous winter		
	SMAPE	**MAPE**	**MAE**	**SMAPE**	**MAPE**	**MAE**
kNN AC	9.88	10.06	26.07	9.72	9.74	25.60
RF AC	10.43	10.67	27.85	9.56	9.49	25.24
kNN AC AR	10.05	10.20	26.39	9.25	9.24	24.42
RF AC AR	10.87	11.03	28.67	10.34	10.34	26.91
kNN AC T H	9.48	9.66	25.09	9.05	9.09	23.89
RF AC T H	11.39	11.53	29.86	11.50	11.53	29.81
kNN AC AR T H	9.75	9.92	25.65	**8.88**	**8.86**	23.45
RF AC AR T H	12.03	12.18	31.56	10.88	10.96	28.06

Table C.3: Forecasting Results (24 hours prediction) for season 3 (summer) trained with time feature lags of 24-, 48- and 168-hours

Features	Vertical summer			Continous summer		
	SMAPE	**MAPE**	**MAE**	**SMAPE**	**MAPE**	**MAE**
kNN AC	12.74	12.74	**6.87**	13.17	13.35	7.17
RF AC	14.70	14.78	8.07	15.27	15.47	8.49
kNN AC AR	13.17	13.24	7.11	13.28	13.43	7.23
RF AC AR	14.16	14.14	7.70	13.89	14.07	7.54
kNN AC T H	14.79	14.46	7.94	15.07	14.75	8.08
RF AC T H	16.53	16.10	8.80	17.05	16.48	9.14
kNN AC AR T H	14.27	14.07	7.68	14.41	14.14	7.71
RF AC AR T H	16.98	16.66	9.02	17.21	16.91	9.19

Note the big difference in MAE between the seasons; however, MAPE and SMAPE have more or less the same values. This is due to relatively higher load consumption in winter time that leads to a higher absolute error, but when compared in absolute percentage error, the error is not noticeable.

The kNN regressor is compared to RF regressor, and it also uses autoregression. In the analysis, a visual inspection might aid to understand the predictive outcome. Prediction results are compared with and without error estimation (see Fig. C.14 and C.15). The kNN and RF alone has no information about the finite gradient of the curvature. In Fig. C.14, the two graphs mostly appear to merely be shifted in time. To overcome this, the real value was compared to the error estimation (see Fig. C.15)), and increasingly peaking errors were shown. A simple form of autoregression is tried in order to mitigate the

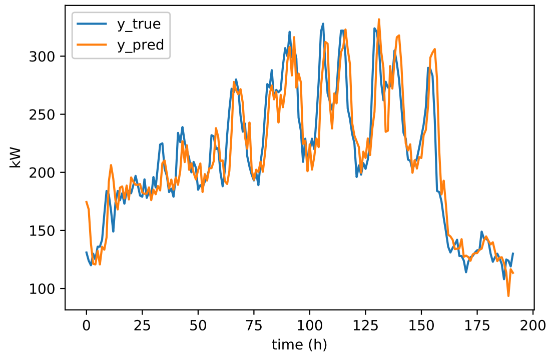


Figure C.14: Prediction result without error estimation

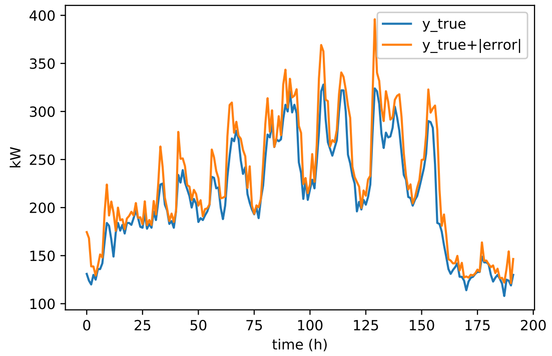


Figure C.15: Prediction result with error estimation

problem of peaking errors. It is a possible remedy, since the correlation analysis showed a strong autocorrelation to the first historical instances of the time series. Instead of a cumbersome ordinary least square-search (OLS) for the backshift operator parameter, only a backshift value is found based on Equation C.24. The autoregressor is used to find the curvature and gives a finite gradient based on the latest update from the targeted vector (in this case, the load). The autoregressor is used to find the curvature and give a finite gradient based on the latest update from the targeted vector, in this case the load.

$$c = (L_{t-1} - L_{t-2})^{\frac{1}{p}} \tag{C.24}$$

Autoregression is the simplest and most straightforward predictive model, based on the targeted vector itself, and at certain time window, it indicates the decline and incline of the time window, and gives a finite gradient for the curvature of load profiles. The joint learning of regression tools with autoregression predicts time series components of different characteristics. Other hybrid combinations can be done with MA, ARMA, and ARIMA models, to aid the regressor model in the predictions.

The load profile of the considered holiday resort (rural area) is categorized seasonally. In this work, regression tools are used for load predictive analysis. In the load predictive analysis, vertical time approach is used for a particular holiday time period. Vertical approach can be performed with minimum amount of data compared to continuous approach. Also, in vertical time approach, predictive results are compared with the prediction based on continuous time series (i.e., horizontal approach). The presented vertical approach methodology can also deal with the problems of irregularities and randomness in the dataset [56].

C.8.2 Case Study 2: Urban Area

The dataset for urban area electric load contains 87648 collected datapoints from the urban area of Sydney in the region of New South Wales in Australia. The relative comparison of load prediction with MAPE for considered regression tools for 30 minutes and

24 hours is done using both horizontal and vertical approaches for all seasons. The results are shown in Table C.4. It is found that the lowest MAPE is achieved with the use of previous load patterns together with indicator variables, and noticeably disregarding weather variables. This goes well with the previous analysis of correlation, which confirms that previous load patterns and indicator variables have higher correlation to the actual load than the weather parameters.

It has been observed from the test results, the lowest MAPE is found through RF regressor for 30 minutes prediction using vertical approach. For the 24-hour time period, kNN provides the lowest MAPE through vertical approach.

MAPE for 30 minutes prediction results using RF regressor varies between 1% and 2%, and provides very good results compared to other regressions techniques, which have been used in this work. The 24 hours prediction results using kNN regressor technique have MAPE of 2.61%, which is much better compared to other regressors. From the results, it has been observed that for short-term predictions (30 minutes), RF regressor should be used; and for long-term predictions (24 hours), kNN regressor should be considered [53].

Urban area electrical energy demand forecasting is very important for generation scheduling and flexibility with consideration of renewable energy sources and possible demand side management. Urban area electrical energy demand predictions for short term (30 minutes) and long term (24 hours) are necessary for scheduling power generation units as well as for participating them in short term and day ahead energy market.

The seasonal patterns are repeating with the same upper and lower limits (e.g., repeating on annual basis), and can be further investigated for economic effects on the load behavior in the urban area of Sydney during the years 2006–2010. When investigating the Sydney dataset, we find that the load curves, yet containing cyclic and seasonal differences, do not contain significant changes on the system load due to changing economic trends [57]. When inspecting the daily and weekly load cycle, we can clearly see a load pattern emerging from a very low activity during the early hours of the day, into one peak at morning (between 8 and 10 hours), and another peak in the evening (between 19 and 21 hours). The same daily repeating patterns, with a low activity followed by two peaks, are also evident in the weekly cycle, except for the last two days of the week (Saturday and Sunday) when the peaks and general load are lower. It can be seen that urban area load predominantly reflects the domestic load, and it can be correlated to human behavior. The periodicity in the load patterns reveals a load demand that reflects a consumer lifestyle. When examining the features enlisted in the Sydney dataset, it has indicators “Date” and “Hour”, four weather parameters, information about the electricity price, “ElecPrice” and information about the electricity load consumption, “SYSLoad”. These features have been developed in the pre-processing to match the requirements of the prediction tool.

RF regressor, kNN regressor, and LR are used for analyzing the urban area electrical

Table C.4: MAPE for Urban Area Load and Indicator aggregated version test results)

Time	Random Forest	k-Nearest Neighbour	Linear Regression
		Regressor	
	Season One Horizontal Approach		
30 minutes	1.11(9*)	1.98(7**,1***)	2.04
24 hours	5.32(13*)	6.53(4**,1***)	5.15
	Season One Vertical Approach		
30 minutes	0.94(16*)	1.85(8**,1***)	1.76
24 hours	5.88(13*)	5.49(5**,2***)	5.83
	Season Three Horizontal Approach		
30 minutes	1.12(17*)	2.36(5**,1***)	2.29
24 hours	4.76(9*)	5.41(19**,1***)	5.27
	Season Three Vertical Approach		
30 minutes	*0.86(17*)*	1.19(6**,1***)	2.15
24 hours	2.71(17*)	*2.61(17**,1***)*	4.26

* n-estimator
** k-value
***q-value

energy demand forecasting, using larger dataset of Sydney region. Data correlation over seasonal changes have been argued by means of improving MAPE. By examining the structure of various regressors, they are compared for the lowest MAPE. The regressors show good MAPE for short term (30 minutes) prediction, and RF regressor scores best in the range of 1–2% MAPE. kNN shows the best results for 24 hours prediction, with a MAPE of 2.61%. The prediction of the short-term 30 minutes electrical energy using vertical approach is relatively better through RF regression tool. For long-term prediction of 24 h, kNN regression tool can provide better results using vertical approach.

C.9 Conclusions

This work has explored the use of regression tools for electrical energy load forecasting through correlating weather parameters as well as the time period. Load prediction analysis using regression tools has been done on continuous time basis (horizontal) as well as using vertical time approach. The Pearson method and visual inspection of the vertical approach depict meaningful relation among pre-processing of data, test methods, and results for the examined regressors.

The application of regression tools has shown to be promising for predicting electric load within distributed network as well as for flexibility analysis. The distributed network are low-capacity networks with low amount of data that need flexible operation and analysis. RF regressor, kNN regressor, and are considered for analyzing the rural area and urban area electrical energy demand forecasting. In addition, LR is used for urban area due to the continous load patterns.

The methodology presented is developed to deal with the problems of irregularities and randomness in the time series. RF regressor yields good result on hourly time prediction in load forecasting. The kNN regressor has shown precise prediction in time series due to its capability to capture the nearest step in a time series based on the nearest neighbor principle.

Autocorrelation is a neat and practical approach to feature engineering that saves time for the appropriate actions to be made for feature extraction. The regression tools can handle the low amount of data, typical for the rural area, for day-ahead forecasting. In this work, the regression analysis for load prediction of rural area is done using vertical and continuous time approaches for day-ahead planning with 24 hours prediction. The vertical time approach uses seasonal data for training and inference, as opposed to continuous time approach that utilizes all data in a continuum from the start of the dataset until the time period used for inference. The regression tools can handle the low amount of data, and the prediction accuracy (through MAPE) matches with other techniques. It is observed that through load predictive analysis, the autocorrelation by vertical approach with kNN-regressor gives a low SMAPE. The kNN captures the lower boundaries of the load demand quite well. When analyzing the error, we find that the algorithms struggle for identifying and predicting the high peaks of the load demand. When the autoregression is given, it helps the algorithm to find the curvature of high peaks; even without capturing the overall trend of the load peak demand, MAPE can be improved by autoregression.

RF regressor, kNN regressor, and LR are used for analyzing the urban area electrical energy demand forecasting. The presented regression techniques can forecast electrical demand for short term (30 minutes) and long term (24 hours) using limited datasets. Vertical axis approach can have more competitiveness to ANN due to the use of low amount of data and considering the impact of meteorological parameters.

Load forecasting is the most fundamental application of smart grid, which provides essential input for flexibility such as demand response, topology optimization, and abnormally detection, facilitating the integration of intermittent clean energy sources.

Bibliography

[1] K. Kostková, Ľ. Omelina, P. Kyčina, and P. Jamrich. An introduction to load management. *Electric Power Systems Research*, 95:184–191, 2013.

[2] Paul Denholm and Maureen Hand. Grid flexibility and storage required to achieve very high penetration of variable renewable electricity. *Energy Policy*, 39(3):1817 – 1830, 2011.

[3] Mark G. Lauby and Rhaiza Villafranca. Flexibility requirements and metrics for variable generation: Implications for system planning studies. Technical report, North American Electric Reliability Corporation, 116-390 Village Blvd., Princeton, NJ 08540, August 2010.

[4] C. Gorria, J. Jimeno, I. Laresgoiti, M. Lezaun, and N. Ruiz. Forecasting flexibility in electricity demand with price/consumption volume signals. *Electric Power Systems Research*, 95:200 – 205, 2013.

[5] Eric Hirst. Flexibility benefits of demand-side programsin electric utility planning. *The Energy Journal*, 11(1), 1990.

[6] Craig B. Smith and Kelly E. Parmenter. Chapter 1 - introduction. In Craig B. Smith and Kelly E. Parmenter, editors, *Energy Management Principles (Second Edition)*, pages 1 – 12. Elsevier, Oxford, second edition edition, 2016.

[7] Federal Energy Regulatory Commision. Federal energy regulatory commissin reports on demand response and advanced metering, [online], 2017.

[8] Luis I. Minchala-Avila, Jairo Armijos, Daniel Pesántez, and Youmin Zhang. Design and implementation of a smart meter with demand response capabilities. *Energy Procedia*, 103:195 – 200, 2016. Renewable Energy Integration with Mini/Microgrid – Proceedings of REM2016.

[9] NATO Advanced Study Institute on Demand-Side Management and Electricity End-Use Efficiency, and Almeida, Anibal T. de. and Rosenfeld, Arthur H., and North Atlantic Treaty Organization. . *Demand-side management and electricity end-use efficiency / edited by Anibal T. De Almeida and Arthur H. Rosenfeld.* Kluwer Academic Publishers Dordrecht ; Boston, 1988.

[10] Abhishek Shivakumar, Constantinos Taliotis, Paul Deane, Johann Gottschling, Rajesh Pattupara, Ramachandran Kannan, Dražen Jakšić, Karmen Stupin, Rudolf V.

Hemert, Bo Normark, and Aurélie Faure-Schuyer. Chapter 21 - need for flexibility and potential solutions. In Manuel Welsch, Steve Pye, Dogan Keles, Aurélie Faure-Schuyer, Audrey Dobbins, Abhishek Shivakumar, Paul Deane, and Mark Howells, editors, *Europe's Energy Transition*, pages 149 – 172. Academic Press, 2017.

[11] Alberto Benato, Anna Stoppato, and Stefano Bracco. Combined cycle power plants: A comparison between two different dynamic models to evaluate transient behaviour and residual life. *Energy Conversion and Management*, 87:1269 – 1280, 2014.

[12] Kenneth Van den Bergh and Erik Delarue. Cycling of conventional power plants: Technical limits and actual costs. *Energy Conversion and Management*, 97:70 – 77, 2015.

[13] Moslem Uddin, M.F. Romlie, M.F. Abdullah, ChiaKwang Tan, GM Shafiullah, and A.H.A. Bakar. A novel peak shaving algorithm for islanded microgrid using battery energy storage system. *Energy*, 196:117084, 2020.

[14] George.E.P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

[15] S. Vemuri, W. L. Huang, and D. J. Nelson. On-line algorithms for forecasting hourly loads of an electric utility. *IEEE Transactions on Power Apparatus and Systems*, PAS-100(8):3775–3784, 1981.

[16] Noel D. Uri. Forecasting peak system load using a combined time series and econometric model. *Applied Energy*, 4(3):219 – 227, 1978.

[17] S. Rahman and R. Bhatnagar. An expert system based algorithm for short term load forecast. *IEEE Transactions on Power Systems*, 3(2):392–399, 1988.

[18] M. J. Damborg, M. A. El-Sharkawi, M. E. Aggoune, and R. J. Marks. Potential of artificial neural networks in power system operation. In *IEEE International Symposium on Circuits and Systems*, pages 2933–2937 vol.4, 1990.

[19] D. J. Sobajic and Y. . Pao. Artificial neural-net based dynamic security assessment for electric power systems. *IEEE Transactions on Power Systems*, 4(1):220–228, 1989.

[20] R. Fischl, M. Kam, J. . Chow, and S. Ricciardi. Screening power system contingencies using a back-propagation trained multiperceptron. In *1989 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 486–489 vol.1, 1989.

[21] H Mori and S Tsuzuki. Power system topological observability analysis using a neural network model. In *Proc. of Second Symposium on Expert Systems Application to Power Systems*, pages 385–391, 1989.

[22] A Marthur, T Samad, and K Anderson. Neural networks and how the utility industry can benefit from them. In *EPRI Conference on Expert Systems Applications for the Electric Power Industry, Scottsdale, Arizona*, 1989.

[23] Satoshi Matsuda and Yoshiakira Akimoto. The representation of large numbers in neural networks and its application to economical load dispatching of electric power. In *International 1989 Joint Conference on Neural Networks*, pages 587–592. IEEE, 1989.

[24] D. C. Park, M. A. El-Sharkawi, R. J. Marks, L. E. Atlas, and M. J. Damborg. Electric load forecasting using an artificial neural network. *IEEE Transactions on Power Systems*, 6(2):442–449, 1991.

[25] A. G. Bakirtzis, V. Petridis, S. J. Kiartzis, M. C. Alexiadis, and A. H. Maissis. A neural network short term load forecasting model for the greek power system. *IEEE Transactions on Power Systems*, 11(2):858–863, 1996.

[26] Che-Chiang Hsu and Chia-Yon Chen. Regional load forecasting in taiwan—applications of artificial neural networks. *Energy Conversion and Management*, 44(12):1941 – 1949, 2003.

[27] C. . Lu, H. . Wu, and S. Vemuri. Neural network based short term load forecasting. *IEEE Transactions on Power Systems*, 8(1):336–342, 1993.

[28] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.

[29] Aileen Nielsen. *Practical Time Series Analysis : Prediction with Statistics and Machine Learning*. O'Reilly Media, Inc, USA, 2019.

[30] Nesreen K. Ahmed, Amir F. Atiya, Neamat El Gayar, and Hisham El-Shishiny. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6):594–621, 2010.

[31] Lulu Wen, Kaile Zhou, Shanlin Yang, and Xinhui Lu. Optimal load dispatch of community microgrid with deep learning based solar power and load forecasting. *Energy*, 171:1053 – 1065, 2019.

[32] Hossein Javedani Sadaei, Petrônio Cândido de Lima e Silva, Frederico Gadelha Guimarães, and Muhammad Hisyam Lee. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. *Energy*, 175:365 – 377, 2019.

[33] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[34] Sholom M. Weiss and Nitin Indurkhya. Rule-based machine learning methods for functional prediction. *J. Artif. Int. Res.*, 3(1):383–403, December 1995.

[35] J Ross Quinlan. Combining instance-based and model-based learning. In *Proceedings of the tenth international conference on machine learning*, pages 236–243, 1993.

[36] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[37] Mohamed N. Nounou and Bhavik R. Bakshi. Chapter 5 - multiscale methods for denoising and compression. In Beata Walczak, editor, *Wavelets in Chemistry*, volume 22 of *Data Handling in Science and Technology*, pages 119 – 150. Elsevier, 2000.

[38] Timothy R. Derrick and Joshua M. Thomas. *Time Series Analysis: The Cross-Correlation Function*. Kinesiology Publications, 2004.

[39] D. Patidar, Bhavin C. Shah, and M. Mishra. Performance analysis of k nearest neighbors image classifier with different wavelet features. *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, pages 1–6, 2014.

[40] Leo Breiman. Random forests. *Springer, Machine Learning*, 2001.

[41] Stasys Jukna. *The Entropy Function*, pages 313–326. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[42] Muhammad Qamar Raza and Abbas Khosravi. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renewable and Sustainable Energy Reviews*, 50:1352 – 1372, 2015.

[43] N. J. Johannesen, M. Kolhe, and M. Goodwin. Deregulated electric energy price forecasting in nordpool market using regression techniques. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 1932–1938, 2019.

[44] Seyedeh Fallah, Ravinesh Deo, Mohammad Shojafar, Mauro Conti, and Shahaboddin Shamshirband. Computational intelligence approaches for energy load forecasting in smart energy management grids: State of the art, future challenges, and research directions. *Energies*, 11(3):596, Mar 2018.

[45] Peder Bacher and Henrik Madsen. Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings*, 43(7):1511 – 1522, 2011.

[46] A. Lahouar and J. Ben Hadj Slama. Random forests model for one day ahead load forecasting. *IREC2015 The Sixth International Renewable Energy Congress*, pages 1–6, 2015.

[47] Fahad H. Al-Qahtani and Sven F. Crone. Multivariate k-nearest neighbour regression for time series data@ a novel algorithm for forecasting uk electricity demand. *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2013.

[48] Muhammad Usman and Naeen Arbab. Factor affecting short term load forecasting. *Journal of Clean Energy Technologies*, 2, 2014.

[49] N. J. Johannesen, M. Lal Kolhe, and M. Goodwin. Load demand analysis of nordic rural area with holiday resorts for network capacity planning. In *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–7, 2019.

[50] M. Mastouri and N. Bouguila. A methodology for thermal modelling and predictive control for building heating systems. *2017 18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, pages 568–573, 2017.

[51] T Salque, D Marchio, and P Riederer. Neural predictive control for single-speed ground source heat pumps connected to a floor heating system for typical french dwelling. *Building Services Engineering Research and Technology*, 35(2):182–197, 2014.

[52] K. Y. Lee, T. I. Choi, C. C. Ku, and J. H. Park. Short-term load forecasting using diagonal recurrent neural network. In *[1993] Proceedings of the Second International Forum on Applications of Neural Networks to Power Systems*, pages 227–232, 1993.

[53] Nils Jakob Johannesen, Mohan Kolhe, and Morten Goodwin. Relative evaluation of regression tools for urban area electrical energy demand forecasting. *Journal of Cleaner Production*, 218:555 – 564, 2019.

[54] N. J. Johannesen, M. Kolhe, and M. Goodwin. Comparison of regression tools for regional electric load forecasting. In *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–6, 2018.

[55] Mark A. Hall. Correlation-based feature selection for discrete and numeric class machine learning. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, pages 359–366, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[56] Nils Jakob Johannesen, Mohan Lal Kolhe, and Morten Goodwin. Smart load prediction analysis for distributed power network of holiday cabins in norwegian rural area. *Journal of Cleaner Production*, 266:121423, 2020.

[57] T. Hong, M. Gui, M. E. Baran, and H. L. Willis. Modeling and forecasting hourly electric load by multiple linear regression with interactions. In *IEEE PES General Meeting*, pages 1–8, 2010.

Appendix D

Paper D - Comparing Recurrent Neural Networks using Principal Component Analysis for Electrical Load Predictions

Nils Jakob Johannesen and Mohan Kolhe and Morten Goodwin

Faculty of Engineering and Science, University of Agder, PO Box 422, NO 4604 Kristiansand, Norway.

***Abstract* - Electrical demand forecasting is essential for power generation capacity planning and integrating environment-friendly energy sources. In addition, load predictions will help in developing demand-side management in coordination with renewable power generation. Meteorological conditions influence urban area load pattern; therefore, it is vital to include weather parameters for load predictions. Machine Learning algorithms can effectively be used for electrical load predictions considering impact of external parameters. This paper explores and compares the basic Recurrent Neural Networks (RNN); Simple Recurrent Neural Networks (Vanilla RNN), Gated Recurrent Units (GRU), and Long Short-Term Memory networks (LSTM). Vanilla RNNs are fully connected neural networks where the output from the previous time step is being fed to the next time step. GRUs are networks with a gating mechanism: a forget gate. LSTM networks also, in addition to a forget gate, include an output gate. Even though the recurrent structure in itself is robust for efficient forecasting, pre-processing of data (including load, weather) is important to enhance the performance. Principal Component Analysis (PCA) reduces and extracts the main components of available data. This work shows that PCA improves the performance of RNNs with use of weather parameters. The historical electrical load dataset from Sydney region is used to test the load forecasting using these techniques considering meteorological parameters. Through load forecasting, it is observed that for the 30 minutes predictions,**

GRU trained with a reduced number of principal components performs best for a typical period with a mean absolute percentage error (MAPE) of 0.74%.

Keywords - *load forecasting, principal component analysis, smart grid, load time series, recurrent neural networks*

D.1 Introduction

The smart electrical energy network grid requires more accurate demand and prediction for control and managing the demand in coordination with intermittent renewable energy sources [1]. The smart grid will require advanced control and management, including reliable forecasting to anticipate the events involved in dispatching, control and management of the operating grid. The accurate load prediction can help in managing peak demand and to reduce overall capital cost investment [2]. Demand prediction is important for short term load forecasting. The aim of demand response in the long term is to reduce overall plant and capital cost investments and to postpone the need for network upgrades. For effectively implement demand response programs, short-term load forecasting will provide useful information [3].

Time series analysis has traditionally been performed in meteorology, energy and economics [4]. The Box Jenkins method for time series analysis has been further developed by the research community to a robust parsimonious Autoregressive Moving Average (ARMA) for multivariate forecasting, requiring less human intervention [5]. By observing changes in economic and weather related variables in a Box-Jenkins time series model, refined forecasts are obtained [6]. The AutoRegressive Integrated Mean Average (ARIMA) model was introduced to deal with trends in the dataset. For the multivariate case the exogenous variable is introduced in AutoRegressive Integrated Moving Average with Exogenous variables (ARIMAX). This is further developed into Seasonal AutoRegressive Moving Average with Exogenous variables (SARIMAX), that also accounts for seasonal behaviour [7]. These methods are useful for the modeling of time series and aids the electrical load analysis. Cycles, trends and periodicity can be found through tests provided by time series analysis [11].

Stack Generalization functions on the principle that two minds work better than one. When Geoffrey Hinton first introduced 'Deep Learning' in 2006 composing artificial neurons in stacked layers [9]. The stacking layers of neurons showed that Deep Learning is possible, with the aid of computer power and big amounts of data [10].

State of the art research in electrical load demand forecasting focuses on three main aspects in order to make sound predictions. These inputs are from weather parameters, holidays and time of day. The mentioned relations has been found equally important both for simpler instance based machine learning models to the more complex black box neural networks [11] [36]. And the results of this are provided in the research for short term [10] [37] [38] , mid-term [39], as well as long-term forecasting [40]. The impact of external weather parameters has proven also to be important for forecasting on limited data, such as for households and buildings [41], as well as cabin areas [35]. Hybrid forecast combining neural networks with autoregression has proven to aid in tracing the curvature of the peak in the volatile electricity markets [5].

In short-term electric load demand forecasting, Recurrent Neural Networks (RNN) by Levenberg-Marquardt and Bayesian regularization on 30 minutes predictions had achieved a mean absolute percentage error (MAPE) of an average in one week 1.4792 [44]. One hour ahead prediction, has been performed on hourly power consumption in Toronto Canada using Long Short-Term Memory (LSTM), achieving a MAPE of 2.639, which was an improvement of the Vanilla RNN of 3.712 MAPE [45]. The Resnetplus model for the ISO-NE dataset proposed a day-ahead load forecasting model based on deep residual networks. A basic structure of several fully connected layers to produce preliminary forecasts of 24 hours. A forecast is then made on the residuals of the preliminary forecast provided with a formulation of Monte Carlo dropout for probabilistic forecasting, achieving an average MAPE of 1.447 [46]. Gated Recurrent Unit (GRU) was used to predict the electricity market in Singapore. Multi-features input models of different time structural architecture named Multi-GRU has been used to give 30 minutes predictions [47].

This article is organised in the following sections: Section D.2 the principal components analysis is explained, the Section E.4 outlines the methodology, Section D.4 includes the data pre-processing, results are discussed in Section E.5, and finally the Conclusion is provided in Section E.6.

D.1.1 Scaling data, normalising

Data is scaled. The general method of calculation is to determine the distribution mean and standard deviation for each feature. Next we subtract the mean from each feature. Then we divide the values (mean is already subtracted) of each feature by its standard deviation.

$$x'_{ij} = \frac{x_i - \hat{x}_j}{\sigma_j} \tag{D.1}$$

x'_{ij} is the value of the input variable of row i and column j, $\hat{x}_j$ is the mean of the values in column j, and finally σ_j is the standard deviation of the values in column j [24].

D.2 Principal Component Analysis for Electrical Load Forecasting

Principal components analysis (PCA) is a multivariate technique that can be applied to many fields for feature reduction. It is the number of samples in the features that are reduced, not the entirety of a feature in itself.

PCA has been found useful in many areas such as daily urban demand forecasting [97].

PCA is extracting the important information for later to represent it in a new set of orthogonal vector input constituting the principal components. These principal components is linear transformation of the data so that the first coordinate explains the most of the variation, the second coordinate the second most, and so on. The components are found through the eigen-decomposition and Singular Value Decomposition [98] [99].
In this work, Sydney region load profile data set is used, which includes meteorological parameters (e.g. DryBulb and WetBulb Temperature, Humidity, weekday and time of use) [106]. In the further feature engineering, a lower indicator variable is designed to differentiate over working-days / non-working days with a binary switch [29]. The RNNs purposefully search in a higher category space to find meaningful relations between the vectors, and therefore the time input is coded using circular coding. The circular coding identifies the time of day according to the unit circle, giving both a sine and cosine coordination as its parameters. They are used as training inputs for the target vector, the electric load demand. The data pre-prossesing in this case leaves the entire feature space with 9 principal components.

Fig. D.1, depicts the proportion of variance that are captured by each number of principal components after feature engineering for the Sydney Data. The red dashed line signifies that when we include the 6 principal components the PCA-process capture 95 % of the variance.

To perform PCA the the input matrix is transposed and crossed with its non-transposed version, stored in matrix L. By diagonalising L, find a matrix M and diagonal matrix W:

$$L = M^T W M \tag{D.2}$$

The feature space is reduced by restricting inputs based on the number of columns that sums up M to make a rotated matrix. The eigenvalues from W are related to the variance of the principal components. PCA reduces the input feature space, yet remains to capture and keep the variation for future inputs and is a important step in the feature engineering.

The proportion of variance needed for optimal feature space may vary. The reference [97] refers to a meta-heuristic practice of principal components explaining 85% of the variance, yet their optimal value was found at 92%.

D.3 Method

The traditional deep neural networks learn patterns on the assumption that inputs and outputs are independent of each other. A RNN depend on the prior elements within the sequence, to perform its decision making. The RNNs used in this work are all based on Keras [30]. RNNs was first developed in natural language processing and the Vanilla RNN is a fully-connected RNN where the output from previous time step is to be fed to next time step by an additional set of units. These units provide for limited recurrence, hence the name 'simple'. The units have also proven to be successful in other time

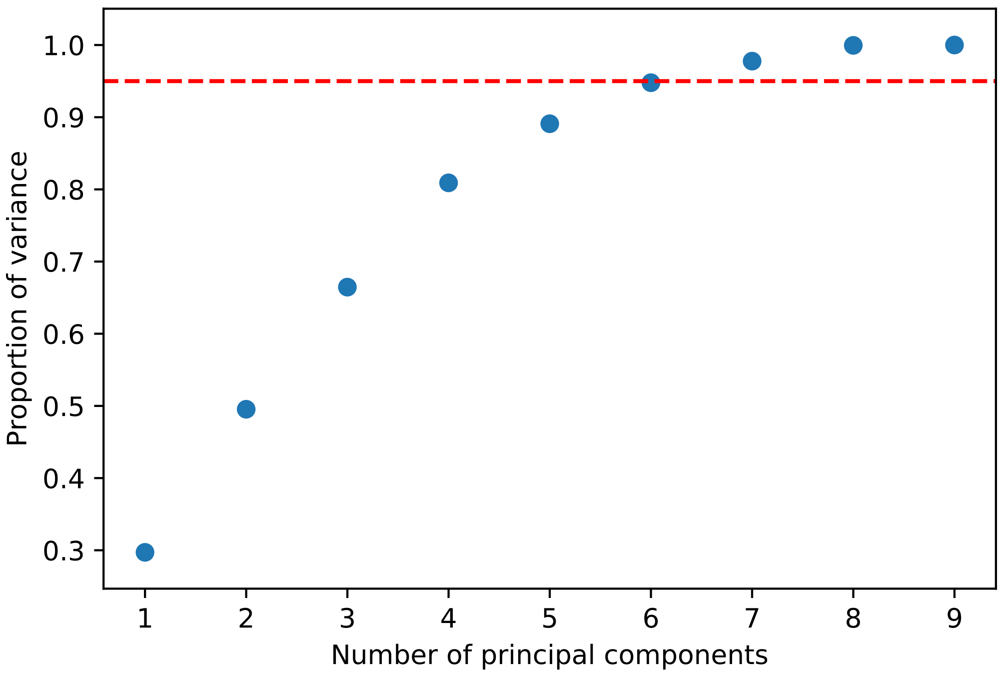


Figure D.1: The cumulative variance per introduced principal components, with red dashed line indicating 95% variance.

series application, and for all problems constituted by sequences, such as electrical load demand. To find the intrinsic nature of linguistic representation Principal component analysis (PCA) has been performed on the hidden unit activation patterns to reveal that the network solves the task by developing complex distributed representations which encode the relevant time relations and hierarchical constituent structure [96]. Vanilla RNN's are fully-connected neural network where the output from previous time step is being fed to the next time step. GRU's are networks with a gating mechanism, a forget gate. Long short-term memory networks also, in addition to a forget gate includes an output gate.

In a recurrent network, in addition the weight layer is combined with the previous state, called the recurrent weight layer U [32]:

$$net_j(t) = \sum_i^n x_i(t) + w_{ji} + \sum_h^m x_h(t-1)u_{jh} + \theta_j \tag{D.3}$$

The set of weights in net_j^t is a candidate value, and through learning finds a candidate solution, $\hat{h}_t^j$, that combines the present state with the previous state. The Vanilla RNN remembers the near future quite well due to the introduction of the hidden state, h, in practice they seem to forget quickly. In LSTM network a memory state is introduced alongside the hidden states, to evaluate long term state dependencies. As illustrated in Fig. D.2, at the bottom the input comes in together with the hidden state (as explained by Vanilla RNN), at the bottom left forget gate f_n:

$$f_n = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{D.4}$$

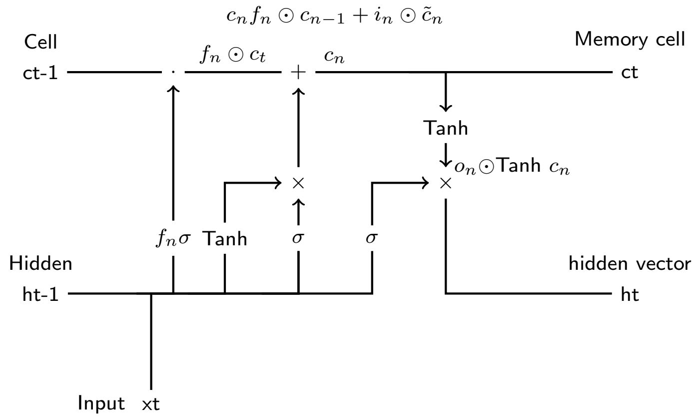


Figure D.2: LSTM network with hidden state and memory cell

and input gate:

$$i_n = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{D.5}$$

In the top memory timeline of Fig. D.2 is a memory cell or cell state, c, the new memory cell is concatenated with previous cell state, added to the input concatenated with cell $\tilde{c}_n$:

$$c_n = f_n \odot c_{n-1} + i_n \odot \tilde{c}_n \tag{D.6}$$

The c_n is updated by forgetting memory as well as adding new memory content $\tilde{c}_n$. For each LSTM unit there exists a memory attached to it c_n at time t. The activation, of the LSTM unit is:

$$h_t = o_n * \tanh\left(c_n\right) \tag{D.7}$$

Where the output gate o_t is computed as:

$$o_t = \sigma\left(W_o \cdot [U_o h_{t-1}, x_t] + V_o c_t\right) \tag{D.8}$$

GRU has only two gates, reset gate r, and update gate z. The first determines the relation of new input to previous memory, and the latter defines to what degree of previous memory is kept. The reset gate is directly applied to the hidden state:

$$r = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right) \tag{D.9}$$

$$z = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right) \tag{D.10}$$

When r=1 and z=0, it equals the Vanilla RNN [33].

D.4 Load data pre-processing with Time organisation and training, validation and testing

Time dependent structures are composed as vectors and fed as inputs to the RNNs. To avoid biases and overfitting the data is to be divided amongst training, validation and testing. In particular the algorithm must capture trends and seasonal variations. If the time series can claim to be stationary, no means needs to be taken. To prove stationarity a search for no trend, constant variance and constant autocorrelation is conducted. Testing for stationarity is done by introducing the null hypothesis H_o: Time series is non-stationary due to trend. By the Augmented Dickey-Fuller (ADF) test, if certain criteria are met the null hypothesis is rejected and the time series is assumed to be stationary. The ADF basically searches for trends in the dataset by evaluating mean and variance over time. Based on this assumption that the time series is stationary, a division into training, validation and test set are done (Fig. D.3).

The training set ranges from the beginning of the recorded data on 01.01.2006 until 31.12.2008. The entire 2009 is used for validation and finally 2010 is for testing. The RNN is learned through a time-lag vector, also known as lookback, that for the multivariate case is a 3D-vector, containing he amount of data (samples), lags and number of inputs (features). Equally on the output, it aims for the target vector. In the training phase this is the next step ahead relative to the input vector.

The proposed model in this work finds suitable training, validation and test-sets by searching for stationarity through Augmented Dickey Fuller Test. The original training set is then reduced feature space and variation representation by performing its principal components analysis, reducing the principal components from an offset features of 9 to be represented by 8 principal components according for 99% of the variance. The training set has then been scaled, and trained on three different RNNs, Vanilla RNN, GRU, and LSTM. These different models have been tested for different seasons to analyse how they assimilate for seasonal variations. Finally the models using PCA, are compared to a version that does not reduce its feature space through PCA.

It is observed from training the RNNs with PCA that during 50 epochs of training and validation, the training loss and validation loss decreases to a point of stability with a minimal gap between the two final loss values, in the Fig. D.5 illustrated with the GRU with PCA, for the Vanilla RNN and LSTM the loss curves show the same convergence.

The RNNs have been tested for a week in January, April, July and October, respectively, and MAPE has been averaged. The results show that all of the RNNs are capturing the inherent structure of the electric load demand quite well, resulting in an acceptable MAPE around 1-2% through all seasons, see Table D.1.

D.5 Results and Discussion

In the winter season the correlations to weather parameters are higher than other seasons, as well as in general the winter season has a higher load demand. These are factors explaining the lower MAPE in winter season as opposed to other seasons.

In the case of GRU networks, the results for all the seasons are improved through PCA

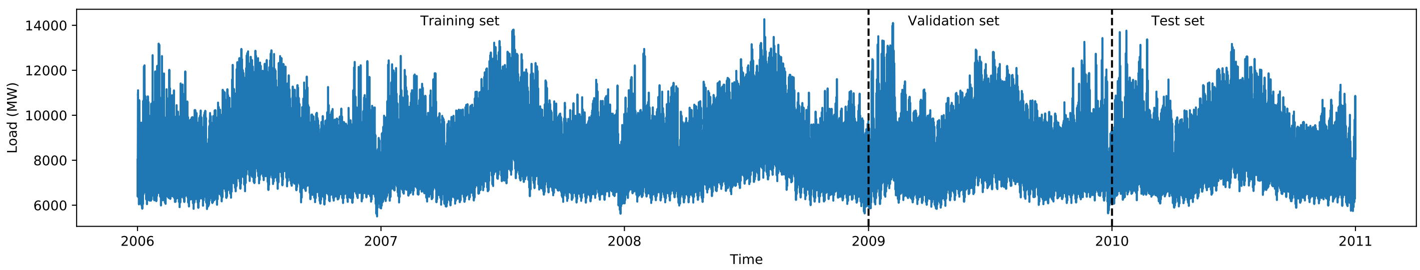


Figure D.3: The Sydney Region data with load measurements for every 30 minutes from 2006 to 2010. Dashed black lines indicates the separation into train-, validation- and test-set.

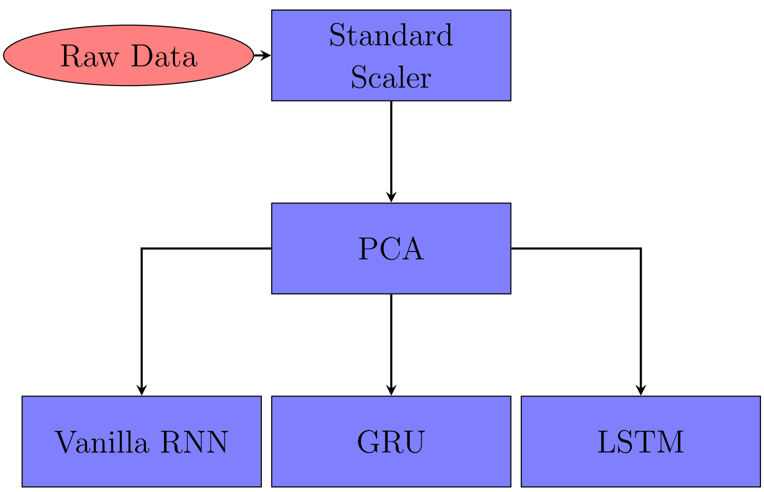


Figure D.4: The Model applied scales the raw Sydney Data, and through PCA predicted by different RNNs

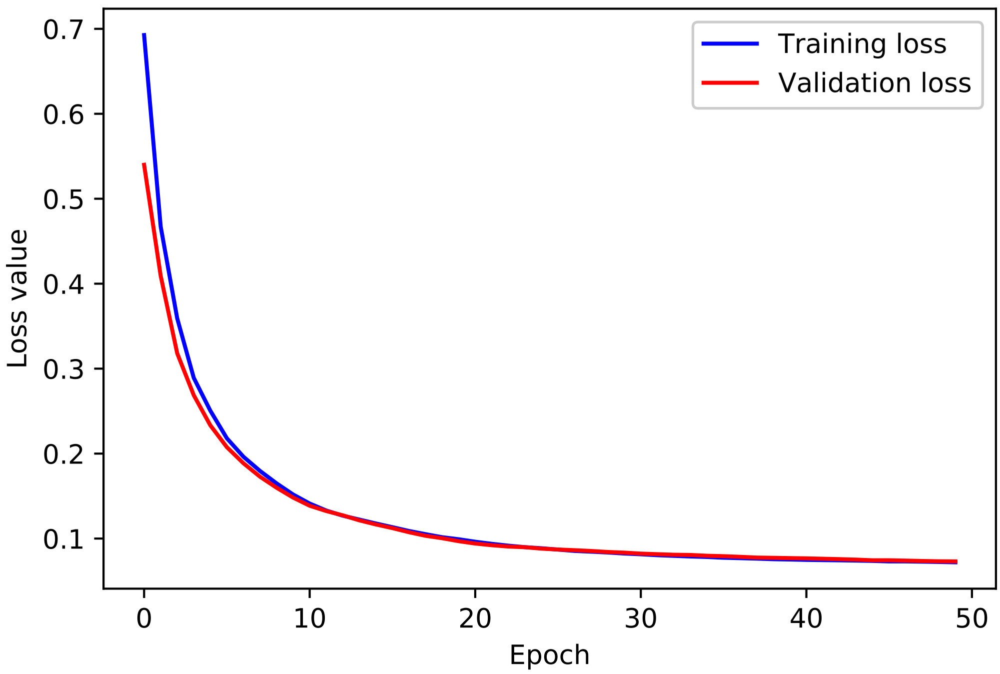


Figure D.5: GRU networks training and validation loss decreases to a point of stability

concluding with 99% of the variation captured by the 8 principal components, see Table D.2. Also for the Vanilla RNN there is a benefit from reduced number of principal components in a lesser MAPE, and for the summer test on a week in July (Fig. D.8), it scores best of all RNNs. Yet for the LSTM it does not benefit from an improved MAPE from the PCA. The best results are measured in January when also the electrical load demand is at the highest (Fig. D.6), and the impact of external weather parameters is influencing greatly on the load demand. The curvature of the load profile is dominated by a high peak at noon, and GRU captures this very good.

The results from the week of April (Fig. D.7), has a lower load demand than January. In January the load demand is highly correlated to the weather parameters readings in winter season. In April, as in January, GRU with PCA achieves the best forecast MAPE result for the week in April, yet with a slightly higher MAPE than for January. This can be explained by the lower load demand in April, and that correlations to weather parameters are usually lower in spring and autumn. In the test week of October (Fig. D.9), which has the same range in load demand (6000 - 10000 MW), it is also GRU with PCA that scores best with a MAPE of 0.94, see Table D.2.

When comparing the results in Tables D.1 with D.2, the MAPE is in the same range for Vanilla RNN (1.45 for April, and 1.38 for October), GRU (1.21 for April and 1.26 for October) and LSTM (1.25 for April, and 1.24 for October). The similarity in results from spring (observed from the test results for the week in April) and autumn (observed from the test results for the week in October) can be explained by similar load range and

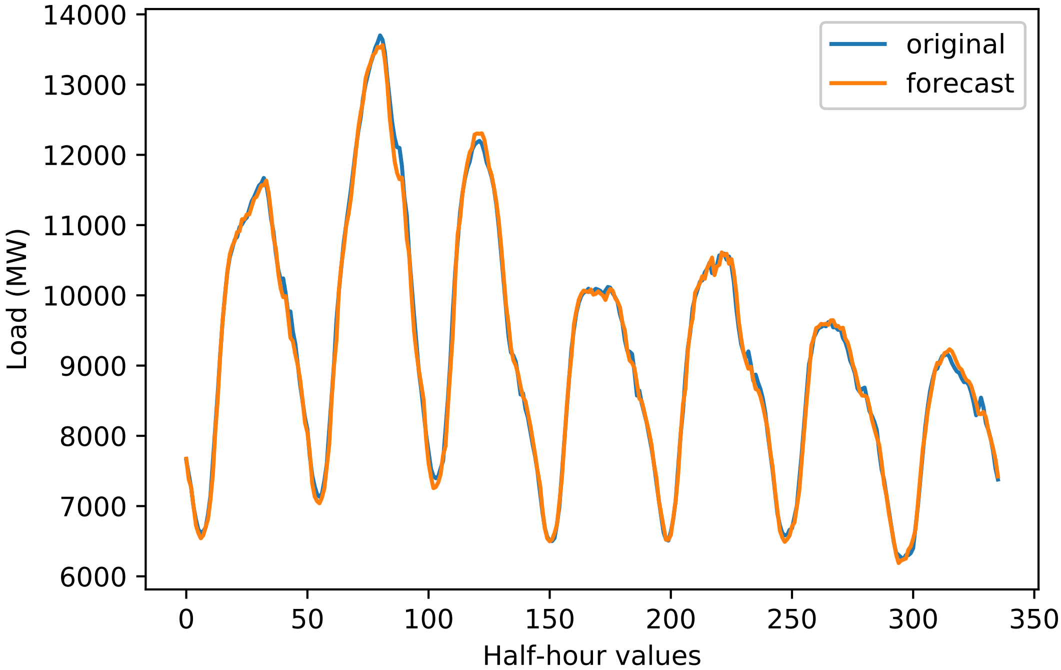


Figure D.6: GRU with PCA tested on a week in January, with a MAPE of 0.74

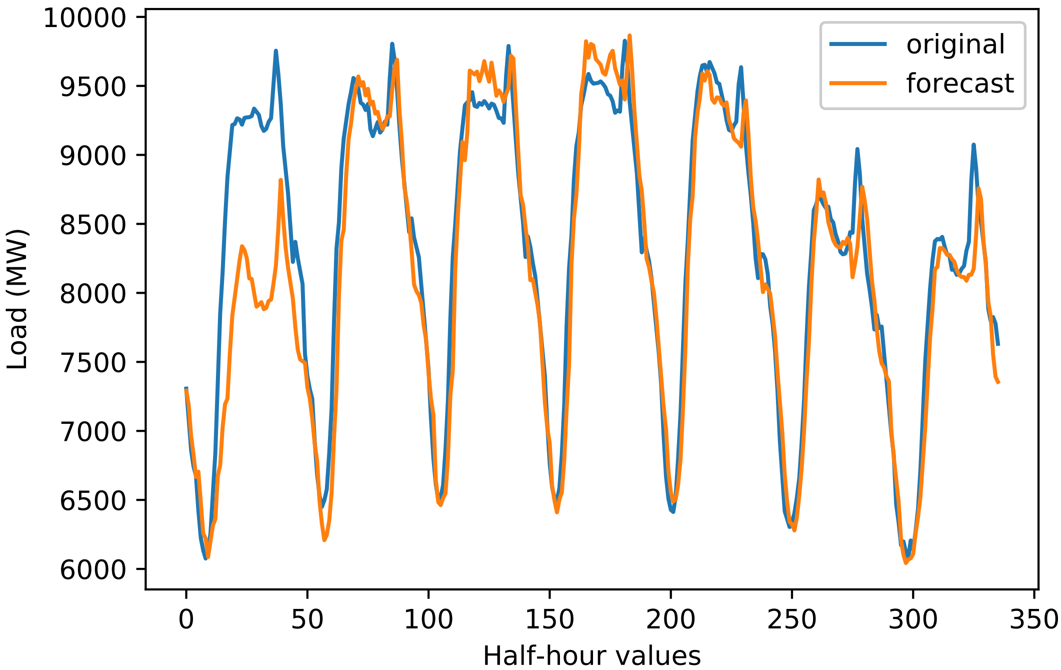


Figure D.7: GRU with PCA performing best of the RNNs for the test week in April

meteorological conditions. In the case of Vanilla RNN and GRU, the explanations of the compared results indicates the same when investigating the results on the RNNs tested with PCA. The exception is the LSTM tested with PCA, that shows a higher MAPE. It is

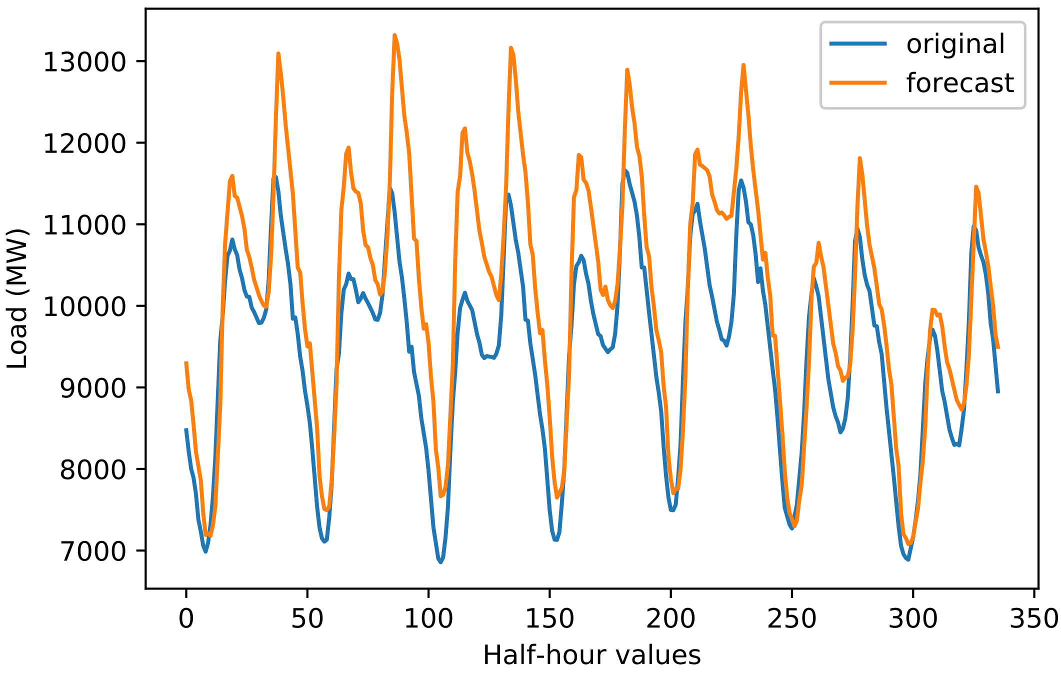


Figure D.8: Vanilla RNN is performing best of all the RNNs on the test week with the lowest load demand, July.

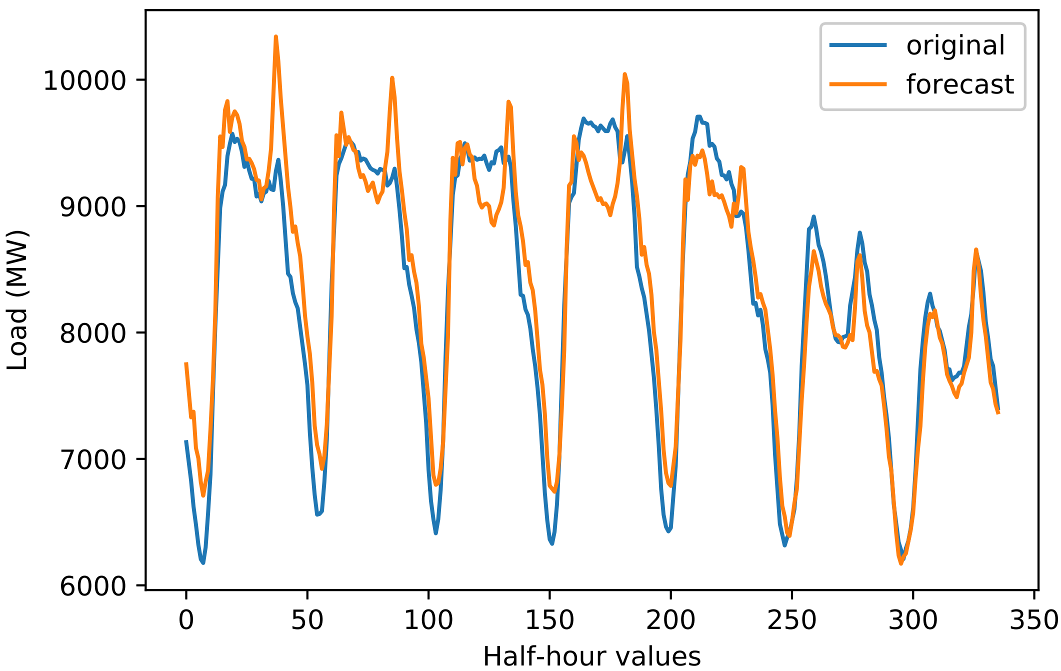


Figure D.9: GRU with PCA performing best of the RNNs for the test week in October

observed that LSTM is a more complex algorithm, than the Vanilla RNN and GRU, and when it is trained with relatively lesser data, although it is analysed using its principal components, it is not able to improve the predictions. It is observed that for the for the

Table D.1: Performance (MAPE)

MAPE	Recurrent Neural Networks		
–	***Vanilla RNN***	***GRU***	***LSTM***
Jan	0.95	0.87	0.90
April	1.45	1.21	1.25
July	1.84	1.64	1.30
October	1.38	1.26	1.24

Table D.2: Performance using PCA (MAPE)

MAPE	Recurrent Neural Networks		
PCA	***Vanilla RNN***	***GRU***	***LSTM***
Jan	0.87	0.74	0.89
April	1.11	1.16	1.60
July	1.39	1.53	1.75
October	1.06	0.94	1.27

week in July with the lowest load demand the simplest RNN (Vanilla RNN) with reduced principal components achieves the preferred MAPE, amongst all of the predictors.

D.6 Conclusion

This paper explores and compares the load prediction analysis through basic RNN; Vanilla RNN, GRU, and LSTM, using PCA. The winter season load behaviour is more influenced by weather parameters, which explains why in the winter season the RNNs scores relatively higher than in other seasons. It is found that PCA can be used to reduce the number of principal components for Vanilla RNN, GRU and LSTM networks. Not only is the reduced feature input space the preferred option in terms of dimensionality reduction, yet also the predictive output is improved. For the electric load demand forecasting the preferred RNN is GRU trained with a principal component of 8, and it is shown through MAPE. After comparing with the version without PCA, the results show that MAPE is reduced when using PCA. For the 30 minutes forecasting GRU with PCA performs best MAPE of 0.74%. This work will benefit the reliable forecasting to anticipate the events involved in dispatching, control and management of the operating grid.

Bibliography

[1] Pierluigi Siano. Demand response and smart grids—a survey. *Renewable and Sustainable Energy Reviews*, 30:461–478, 2014.

[2] Stephen Haben, Siddharth Arora, Georgios Giasemidis, Marcus Voss, and Danica Vukadinovic Greetham. Review of low-voltage load forecasting: Methods, applications, and recommendations. *arXiv preprint arXiv:2106.00006*, 2021.

[3] Xavier Serrano-Guerrero, Marco Briceño-León, Jean-Michel Clairand, and Guillermo Escrivá-Escrivá. A new interval prediction methodology for short-term electric load forecasting based on pattern recognition. *Applied Energy*, 297:117173, 2021.

[4] George.E.P. Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

[5] S. Vemuri, W. L. Huang, and D. J. Nelson. On-line algorithms for forecasting hourly loads of an electric utility. *IEEE Transactions on Power Apparatus and Systems*, PAS-100(8):3775–3784, 1981.

[6] Noel D. Uri. Forecasting peak system load using a combined time series and econometric model. *Applied Energy*, 4(3):219 – 227, 1978.

[7] S. Rahman and R. Bhatnagar. An expert system based algorithm for short term load forecast. *IEEE Transactions on Power Systems*, 3(2):392–399, 1988.

[8] Nils Jakob Johannesen and Mohan Lal Kolhe. Application of regression tools for load prediction in distributed network for flexible analysis. In *Flexibility in Electric Power Distribution Networks*. CRC Press, 2021.

[9] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[10] Aurlien Gron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., 1st edition, 2017.

[11] Nils Jakob Johannesen, Mohan Kolhe, and Morten Goodwin. Relative evaluation of regression tools for urban area electrical energy demand forecasting. *Journal of Cleaner Production*, 218:555–564, 2019.

[12] A. Songpu. External parameters contribution in domestic load forecasting using neural network. *IET Conference Proceedings*, pages 6 .–6 .(1), January 2015.

[13] Ernesto Aguilar Madrid and Nuno Antonio. Short-term electricity load forecasting with machine learning. *Information*, 12(2), 2021.

[14] Abderrezak Laouafi, Mourad Mordjaoui, Salim Haddad, Taqiy Eddine Boukelia, and Abderahmane Ganouche. Online electricity demand forecasting based on an effective forecast combination methodology. *Electric Power Systems Research*, 148:35–47, 2017.

[15] S. Mirasgedis, Y. Sarafidis, E. Georgopoulou, D.P. Lalas, M. Moschovits, F. Karagiannis, and D. Papakonstantinou. Models for mid-term electricity demand forecasting incorporating weather influences. *Energy*, 31(2):208–227, 2006.

[16] Yongxiu He, Jie Jiao, Qian Chen, Sifan Ge, Yan Chang, and Yang Xu. Urban long term electricity demand forecast method based on system dynamics of the new economic normal: The case of tianjin. *Energy*, 133:9–22, 2017.

[17] Peter Lusis, Kaveh Rajab Khalilpour, Lachlan Andrew, and Ariel Liebman. Short-term residential load forecasting: Impact of calendar effects and forecast granularity. *Applied Energy*, 205:654–669, 2017.

[18] Nils Jakob Johannesen, Mohan Lal Kolhe, and Morten Goodwin. Load demand analysis of nordic rural area with holiday resorts for network capacity planning. In *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–7, 2019.

[19] Nils Jakob Johannesen, Mohan Kolhe, and Morten Goodwin. Deregulated electric energy price forecasting in nordpool market using regression techniques. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 1932–1938, 2019.

[20] Muhammad Amri Yahya, Sasongko Pramono Hadi, and Lesnanto Multa Putranto. Short-term electric load forecasting using recurrent neural network (study case of load forecasting in central java and special region of yogyakarta). In *2018 4th International Conference on Science and Technology (ICST)*, pages 1–6, 2018.

[21] Lei Zhang, Linghui Yang, Chengyu Gu, and Da Li. Lstm-based short-term electrical load forecasting and anomaly correction. In *E3S Web of Conferences*, volume 182, page 01004. EDP Sciences, 2020.

[22] Kunjin Chen, Kunlong Chen, Qin Wang, Ziyu He, Jun Hu, and Jinliang He. Short-term load forecasting with deep residual networks. *IEEE Transactions on Smart Grid*, 10(4):3943–3952, 2019.

[23] Weixian Li, Thillainathan Logenthiran, and Wai Lok Woo. Multi-gru prediction system for electricity generation's planning and operation. *IET Generation, Transmission & Distribution*, 13(9):1630–1637, 2019.

[24] Xavier Serrano-Guerrero, Guillermo Escrivá-Escrivá, and Carlos Roldán-Blay. Statistical methodology to assess changes in the electrical consumption profile of buildings. *Energy and Buildings*, 164:99–108, 2018.

[25] Baigang Du, Qiliang Zhou, Jun Guo, Shunsheng Guo, and Lei Wang. Deep learning with long short-term memory neural networks combining wavelet transform and principal component analysis for daily urban water demand forecasting. *Expert Systems with Applications*, 171:114571, 2021.

[26] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1):37–52, 1987. Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.

[27] Hervé Abdi and Lynne J. Williams. Principal component analysis. *WIREs Computational Statistics*, 2(4):433–459, 2010.

[28] Vasudev Dehalwar, Akhtar Kalam, Mohan Lal Kolhe, and Aladin Zayegh. Electricity load forecasting for urban area using weather forecast information. In *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*, pages 355–359, 2016.

[29] Nils Jakob Johannesen, Mohan Lal Kolhe, and Morten Goodwin. Smart load prediction analysis for distributed power network of holiday cabins in norwegian rural area. *Journal of Cleaner Production*, 266:121423, 2020.

[30] Francois Chollet et al. Keras, 2015.

[31] Jeffrey L Elman. Distributed representations, simple recurrent networks, and grammatical structure. *Machine learning*, 7(2):195–225, 1991.

[32] Mikael Boden. A guide to recurrent neural networks and backpropagation. *the Dallas project*, 2002.

[33] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014.

Appendix E

Paper E - Evaluating Anomaly Detection Algorithms through different Grid scenarios using k-Nearest Neighbor, iforest and Local Outlier Factor

Nils Jakob Johannesen and Mohan Kolhe and Morten Goodwin

Faculty of Engineering and Science, University of Agder, PO Box 422, NO 4604 Kristiansand, Norway.

Abstract **- Detection of anomalies based on smart meter data is crucial to identify potential risks and unusual events at an early stage. The available advanced information and communicating platform and computational capability renders smart grid prone to attacks with extreme social, financial and physical effects. The smart network enables energy management of smart appliances contributing support for ancillary services. Cyber threats could affect operation of smart appliances and hence the ancillary services, which might lead to stability and security issues. In this work, an overview is presented of different methods used in anomaly detection, performance evaluation of 3 models, the k-Nearest Neighbor, local outlier factor and isolated forest on recorded smart meter data from urban area and rural region.**

Keywords cybersecurity, anomaly detection, smart grid, local outlier factor, isolated forest

E.1 Introduction

The smart electrical energy network grid requires more accurate demand and prediction for control and managing the demand in coordination with intermittent renewable energy sources [1]. The smart grid will require advanced control and management, including reliable forecasting to anticipate the events involved in dispatching, control and management of the operating grid. The accurate load prediction can help in managing peak demand and to reduce overall capital cost investment [2].

From the field of Artificial Intelligence (AI) a tool to process meaningful relation of complex big data by uncovering structures and patterns is learned through training with Machine Learning (ML). When presented with new data the machine can learn to perform a task without the need of re-programming [9]. ML can provide electrical load demand forecasting, giving information about future loads, which provides essential input to other applications such as Demand Response, Topology Optimization and Anomaly Detection, facilitating the integration of intermittent clean energy sources. Anomaly detection can been used as a first step in data cleaning process and has been known to enhance any forecasting algorithm [4][5].

The data used is of such an amount, that it is not possible to do so manually or by visual inspection, and there is a need for an efficient, automated and accurate anomaly detection methods [48]

An anomaly is defined as a deviation from an established normal pattern. Spotting an anomaly depends on the ability to defy what is normal. Anomaly detection systems aim at finding these anomalies. Anomaly detection systems are in high demand, despite the fact that there is no clear validation approach. These systems rely on deep domain expertise. Cyber threats could affect the ancillary services that are being delivered from the aggregators, which might lead to stability and security issues resulting in brownout or massive blackouts [7]. Large scale monitoring using the supervisory control and data acquisition (SCADA) makes it vulnerable to cyber attacks. Anomaly detection can be used for preventing possible cyber-attacks.

The buses in a power system is in normal operation in the same state, it is reasonable that an anomaly exists if one bus deviates from the others [8] The implementation of two way communication by the use of sensors and intelligent agents such as advanced metering infrastructure as well as load aggregation, make these attractive objects for cyber attacks. Sensors can be penetrated using a Trojan Horse, to manipulate the adversary inside the control platform, and change reference inputs in controllers of components. The attacker can here change acquisition gains, that create bias in the measurements report.

In the distributed power network the attack can disrupt the frequency regulation, voltage stability and the power flow management [9].

It is necessary to investigate different computing methods, and their applications in anomaly detection. In this work the performance evaluation of 3 models is analysed on recorded smart meter data from urban area and rural region.

This article is organised in sections: Section E.2 the literature review. Theory in Section E.3, user scenarios in Section E.4, results in Section E.5, and conclusion in Section E.6.

E.2 Review on Anomaly Detection

Anomaly detection is done on any time series data. Various anomalies can be detected in historic time series data, due to human error, false meter measurement, inaccuracies in data processing and failure of delivery due to extreme weather or other failures. A two-stage method is proposed in reference [48] combining two probabilistic anomaly detection approaches for identifying anomalies in time series data of natural gas. Exogenous variables are known to influent the electrical load consumption [10], and loads are identified accordingly as baseload, intermediate load and peak load [11]

An autoregressive integrated moving average with exogenous inputs (ARIMAX) model is used to extract weather dependency to find the residuals, then through hypothesis testing the extremities, maximum and minimums are found [49]. This procedure was reproduced, with linear regression finding the residuals and a Bayesian maximum likelihood classifier to identify anomalies [48].

A data-mining based framework using DBSCAN was used to detect anomalies in office buildings. The framework is aimed to identify typical electricity load patterns and gain knowledge hidden in the patterns and to potentially be used in an early fault detection of anomalous electricity load profiles[50]. Also to detect anomalies of electricity consumption in office buildings an improved kNN is proposed, ikNN, to automatically classify consumption footprints as normal or abnormal [51].

Dynamic Bayesian Networks and Restricted Boltzman Machine has been proposed for anomaly detection in large-scale smart grids. Simulated on the IEEE 39, 118, and 2848 bus systems the results were verified [52]. Real-Time Mechanism for detecting FDIA analyzed the change of correlation between two phasor measurement units parameters using Pearson correlation coefficient on IEEE 118 and 300-bus sytems [53]. Machine learning techniques have been highlighted for theire ability to differentiate between cyberattacks and natural disturbances. By a simulating a variety of scenarios the ability for One R, Random Forest, Naive Bayes and J-Ripper to recoginze attacks was investigated: Short Circuit faults; location is represented by the percentage range, Line maintenance; identified through remote relay trip command, Remote tripping command injection; the attacker operates the relay remotely that causes a breaker to open, Relay setting change; the attacker misconfigures the relay settings to cause maloperation of relays, FDIA; attacker manipulates measurements sensors. The simulated scenarios was grouped into classes; natural events, attack events, and no events [54].

In concept drift, models are inaccurate due to change in the underlying data [56]. Thus the observation can be a result of an improved energy system, and not anomaly[57].

E.3 Anomaly Detection using Machine Learning Algorithms

3 different models is compared for anomaly detection in the different grid scenarios:

E.3.0.1 k-Nearest Neighbor

The k-nearest neighbor (kNN) regressor, which is non-paramteric, relying on its own table look-up and mathematical foundation, and highly non-linear.

$$y_{knn}(x) = \frac{1}{K}\sum_{k=1}^{K} y_k \text{ for K nearest neighbours of x} \tag{E.1}$$

The kNN-classifier is illustrated in Fig. E.1, where the left diagram with a small encirclement options for k = 1, where simply the nearest neighbor decides the class of prediction, whilst in the right diagram in Fig. E.1, the number of k is increased to more then one [70].

Using k = 1 can lead to false prediction, and a set of kNNs is often used. When classifying the dependent variable is categorical, it can easily be made numerical by regression. The kNN regressor makes a regression based on the number of kNNs to minimize false predictions. The model considers a range of different kvalues to find the optimal value. The kNN regressor needs thorough pre-processing and feature engineering to limit the effect of noise caused by irrelevant features, and is, therefore, dependent on finding the appropriate distance model [71].

E.3.0.2 Isolation Forest

The Isolation Forest algorithm is composed of several isolation trees (iTres) Isolation forest takes advantage of the nature of anomalies which are less frequent than regular observations and different from those in terms of values to isolate those. Iforest can deal with large scale data quickly in a simplified way. It builds an ensemble of decision trees (iTrees) for a given data set. Clustering is done using binary tree clustering. Anomalies tend to be isolated closer to the root of the binary tree. Partitions are created using a split value between the minima and maxima of a randomly selected feature. The algorithm then tries to separate each point in the data [82] [83] [84] [85].

E.3.0.3 Local Outlier Factor

Local Outlier Factor (LOF) is a density based anomaly detection algorithm introduced in 2000 [26]. LOF compares the local density of a point to the local density of k of its neighbors. By comparing the local density of a point to the local density of its neighbors one can identify point that have substantially lower density than its neighbors. These

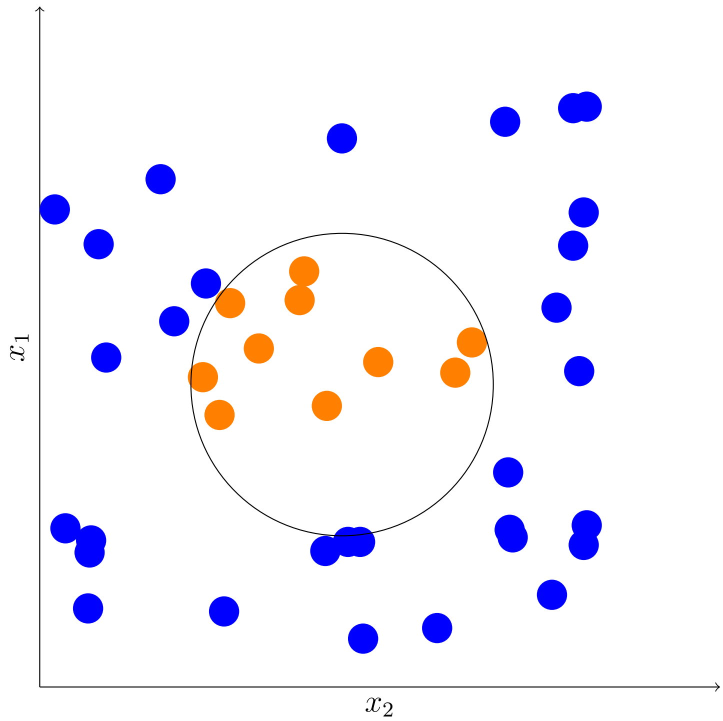


Figure E.1: k-Nearest Neighbour classifying based on the k'th observation.

points are considered outliers. LOF uses the k-distance to a point as in kNN, to find the Local Reachability Density (LRD), where a point is most likely to be found. The sum of LRD is then used to find LOF for the point z, as in Equation (E.2):

$$\mathrm{LOF_k}(\mathbf{z}') = \sum_{\mathrm{z} \in \mathrm{N_k}(\mathbf{z}')} \frac{lrd_{\mathrm{k}}(\mathbf{z})}{lrd_k(\mathbf{z}')} / \|N_k(\mathbf{z}')\| \tag{E.2}$$

[86]

E.4 User case scenarios

In this work 3 different models is used to detect anomalies in two different grid scenarios:

E.4.1 Scenario 1

New South Wales, Sydney region electrical load profile data set [105] includes meteorological parameters (e.g. DryBulb and WetBulb Temperature, Humidity, Electricity price and time of use) [106]. Data is gathered from 2006-2011. The overall energy mix in New South Wales consists mainly of Coal, Natural Gas, Hydro and other renewable energy sources. Fig. E.2 illustrates the New South Wales distribution network.

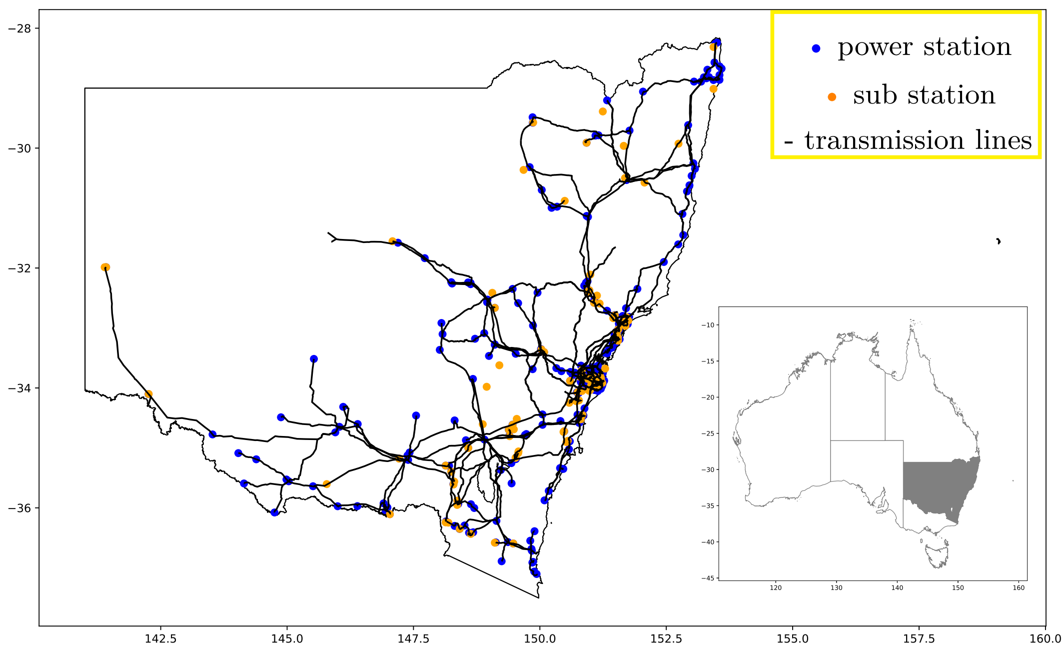


Figure E.2: New South Wales Power system, indicating transmission lines, power stations, and substations

E.4.2 Scenario 2

From rural cabin area in Bjønntjønn, Telemark, Norway, the electrical load demand consumption profile is collected from smart meters. Weather data is collected from surrounding weather information statins in the surrounding area. The land owner of the area wants to realize the project 'Bjønntjønn Grønn' (Bjønntjønn Green). The project seeks through different initiatives to make the cabin area 'green', with power from local hydro power stations, possibility of electric vehicle charging and operation of the load consumption related to the power intensive usages. The land owner has currently an application to get license from The Norwegian Energy Regulatory Authority (NVE) to run hydro power stations in the area, with a total production of 10,08 GWh [108]. In the fall of 2021 NVE approved an appplication for a Tesla Supercharger from Tesla Norway, situated in the center of Treungen, an 8 km drive from the planned Bjønntjønn hydro power station [109] [110].

The rural area network of a typical Norwegian holiday resort cabin area, Bjønntjønn Cabin Area. It comprises 125 cottages with a peak demand of 478 kW. As for today, this cabin area is grid connected, but a microgrid solution involving photovoltaics and energy storage is also considered. In the summer of 2020 the land owner presented plan of building 445 new cabins in the area [111].

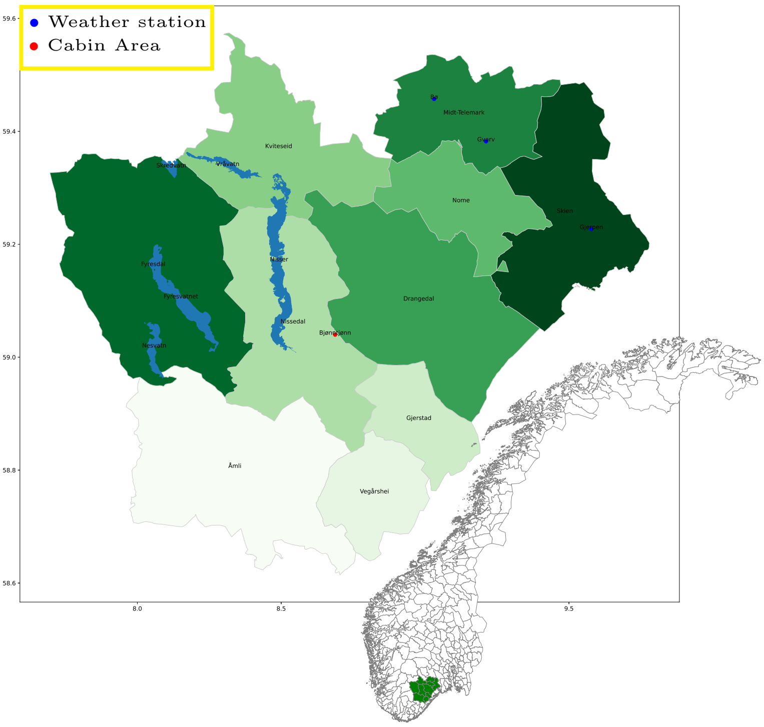


Figure E.3: Rural Area in the South-East of Norway, with the situation of cabin area and nearby weather stations

Rural electrification is very different from the urban area electrical consumption, due to diversified energy mix and overall conditions. A variety of case studies is necessary for a generic approach, although each system requires an independent approach. The Nordic market is much reliant on hydropower, as Norway's share of hydropower is 95.8 % [112]. Norway also has the highest integration of Electric Vehicles, and this faces challenges to the grid. This is especially a case in the rural area, where capacity is low, and the electrical vehicle charging poses a liability to the grid. In these cases, a micro-grid solution can aid the low-capacity network, with implementation of distributed generators, in combination with energy storage.

When examining the general load profile of all Norwegian Holiday Cabins, a clear trend is observed in the user behaviour. The load demand for Norwegian Cabins has increased their total consumption from 0.7 TWh in 1993 to 2.3 TWh in 2016. Although the consumption tripled and has been only 1.8 % of the total Norwegian load demand in 2016 [35]. Statistics Norway concludes in the 2018 report, that the increasing trend is due to the general development, and that more Norwegians have bought cottages in rural areas, such as mountains and seaside. Also, more cottages have been electrified in this period [112].

In the Bjønntjønn Cabin Area, to deal with the ever-increasing penetration of electric vehicles, photovoltaic system together with energy storage could be a scenario for the future rural electrification. For the Nordic rural area network, a microgrid solution can improve the electrical network capacity of the rural area, despite challenges from power demanding operations as electric vehicle charging. Since the electric vehicle will not be used mostly of the holiday resort area, the battery pack of the vehicle is be considered as the battery bank for the microgrid. When the state of charge (SOC) of the battery reaches a certain threshold level, it will be considered as a prosumer for the micro grid and be able to contribute to electrical supply and stability.

E.5 Results and Discussion

The results of kNN, iforest and LOF on urban area data, are shown in Fig. E.4, E.5 and E.6, and from rural region data in Fig. E.7, E.8 and E.9. The results are depicted with a 0.0005 amount of contamination of the data set, this is the proportion of outliers in the data set. Used when fitting to define the threshold on the decision function [36].

It is observed that the anomaly detection for the two grid scenarios are different, for the rural region most of the anomalies where observed in the latter timeline of the data concentrated in the last year of the collected data. For the urban area data the anomalies are spread out over the entire timeline. In Table E.1, it is shown that the frequency of detected anomalies where considerably higher for the rural area load demand than for the urban area load demand. When observing the anomalies detected based on the algorithm the results in Table E.1 are consistent.

algorithm	urban	rural
kNN	44	10
iforest	35	25*
lof	44	21

Table E.1: Results using fraction 0.0005, except * = 0.0006

Observing from these case scenarios the incidents of detected anomalies are more data driven, then exceptions in the algorithms. It is observed that there are 3 anomalies, where the recorded electrial load demand is zero, in the rural region dataset that the iforest and LOF did not detect. This was only detected by kNN, see Fig. E.7.

When comparing the 3 algorithms tested on the urban area data it is observed that kNN and isolated forest finds a threshold value, based in the mentioned fraction of contamination, and separates a lower and upper bound, whilst the density based LOF finds anomalies at several ranges of the dataset, see Fig. E.4, E.5 and E.6.

When visually inspecting results in Fig. E.4, E.5, E.6, E.7, E.8 and E.9, it is observed that from the domain knowledge of smart energy systems the LOF is able to detect observations that could not have detected by visual inspection alone, in contrast to kNN and iforest. Whereas kNN and iforest excludes an upper and lower bound, the LOF is density based and separates out anomalies amidst in the data. The capability that LOF has to identify anomalies amidst the data will together with the deep domain knowledge is an advantage when detecting anomalies in smart meter data.

E.6 Conclusion

Detection of anomalies based on smart meter data is crucial to identify potential risks and unusual events at an early stage. An anomaly is defined as a deviation from an established normal pattern. Spotting an anomaly depends on the ability to defy what is normal.Cyber threats could affect operation of smart appliances and hence the ancillary services, which might lead to stability and security issues. In this work is evaluated the performance of 3 models, the k-Nearest Neighbor, local outlier factor and isolated forest on recorded smart meter data from urban area and rural region. Observed that from the domain knowledge of smart energy systems the LOF is able to detect observations that could not have detected by visual inspection alone, in contrast to kNN and iforest. Whereas kNN and iforest excludes an upper and lower bound, the LOF is density based and separates out anomalies amidst in the data. The capability that LOF has to identify anomalies amidst the data will together with the deep domain knowledge is an advantage when detecting anomalies in smart meter data. The anomaly detection based on machine learning algorithms gives a fast response to potential anomalies.

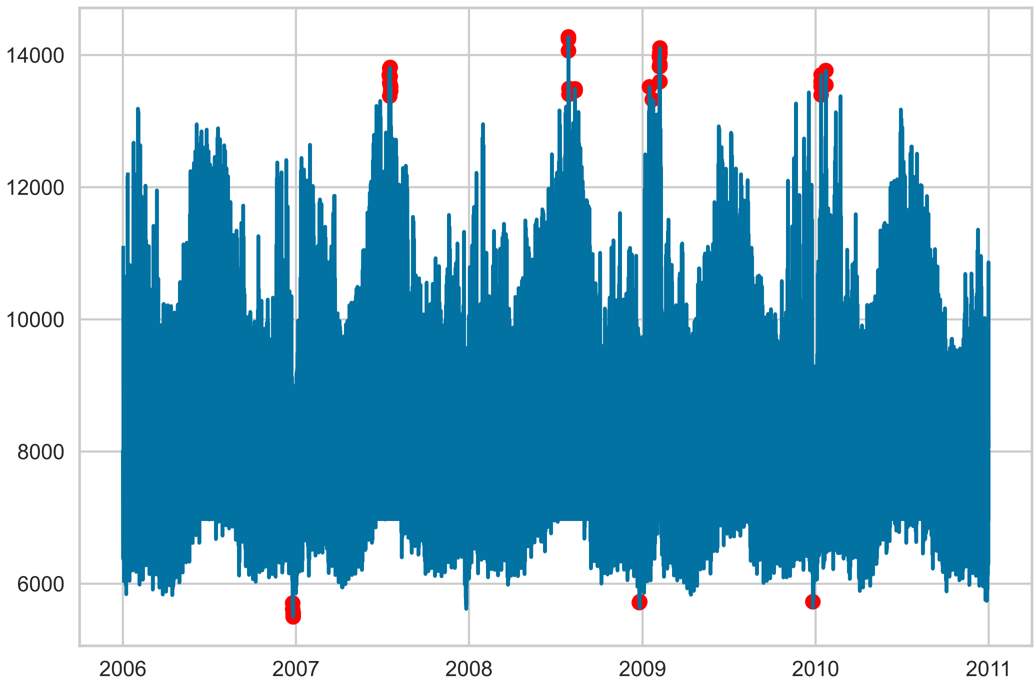


Figure E.4: Anomaly detected outliers marked in red using kNN, fraction = 0.0005

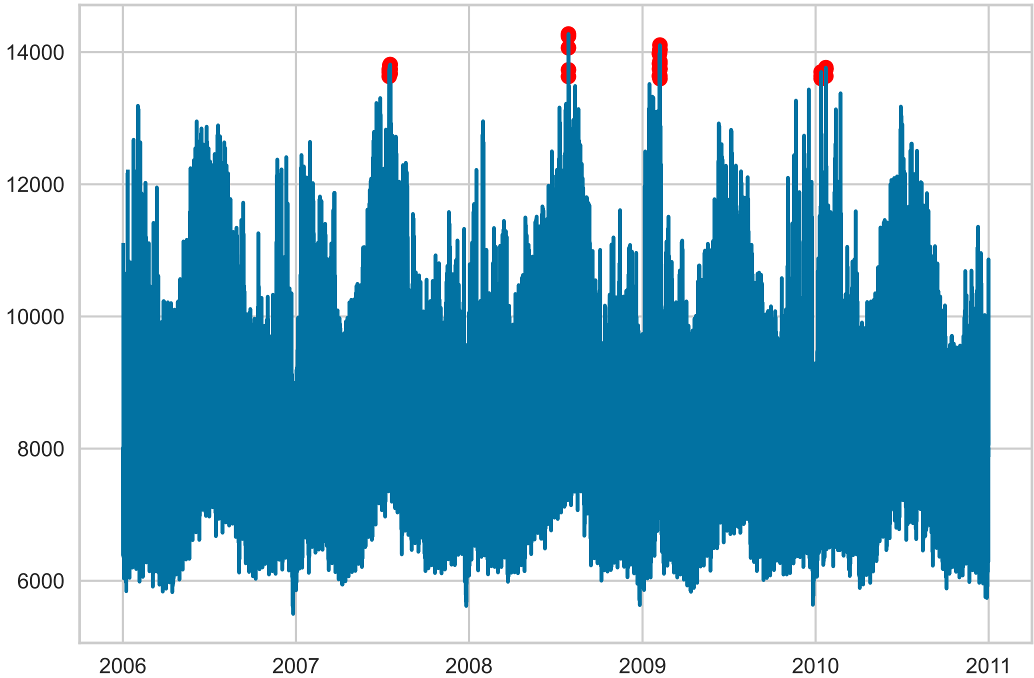


Figure E.5: Anomaly detected outliers marked in red using iforest, f = 0.0005

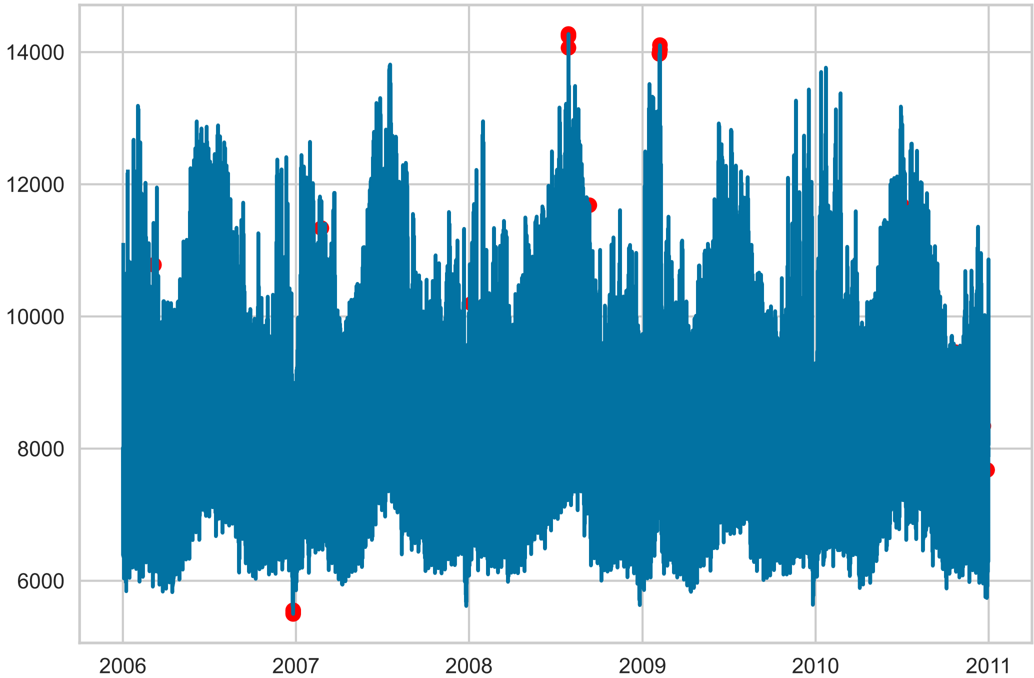


Figure E.6: Anomaly detected outliers marked in red using LOF, f= 0.0005

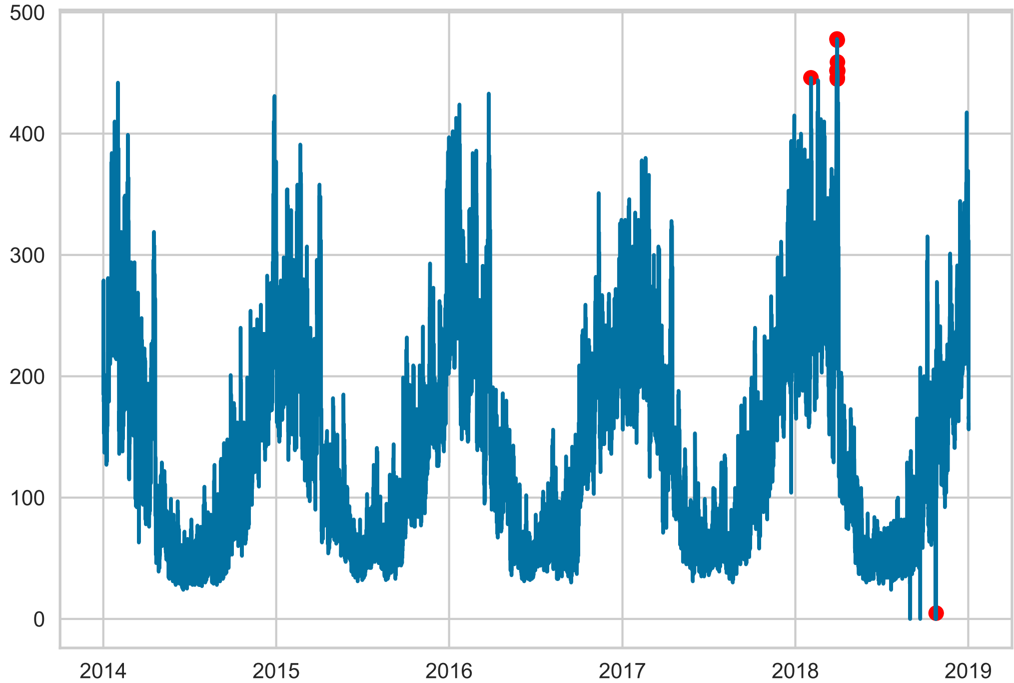


Figure E.7: Anomaly detected outliers marked in red using kNN, fraction = 0.0005

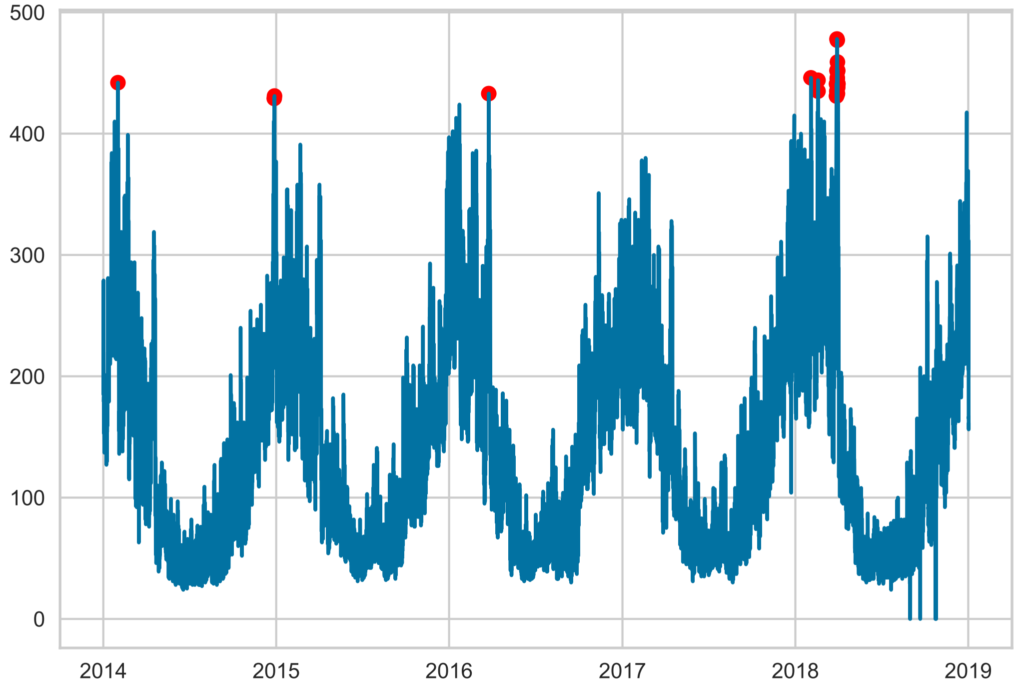


Figure E.8: Anomaly detected outliers marked in red using iforest, f = 0.0006

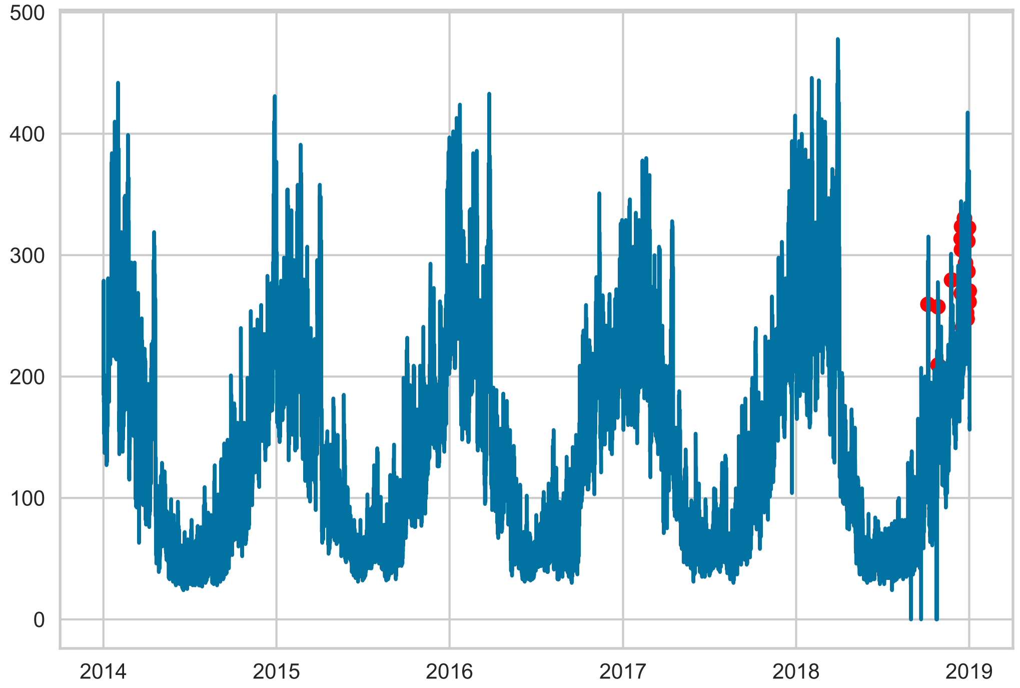


Figure E.9: Anomaly detected outliers marked in red using LOF, f= 0.0005

Bibliography

[1] Pierluigi Siano. Demand response and smart grids—a survey. *Renewable and Sustainable Energy Reviews*, 30:461–478, 2014.

[2] Stephen Haben, Siddharth Arora, Georgios Giasemidis, Marcus Voss, and Danica Vukadinovic Greetham. Review of low-voltage load forecasting: Methods, applications, and recommendations. *arXiv preprint arXiv:2106.00006*, 2021.

[3] Peyman Razmi, Mahdi Ghaemi Asl, Giorgio Canarella, and Afsaneh Sadat Emami. Topology identification in distribution system via machine learning algorithms. *PLOS ONE*, 16(6):1–20, 06 2021.

[4] Ih Chang, George C. Tiao, and Chung Chen. Estimation of time series parameters in the presence of outliers. *Technometrics*, 30(2):193–204, 1988.

[5] Nils Jakob Johannesen, Mohan Kolhe, and Morten Goodwin. Deregulated electric energy price forecasting in nordpool market using regression techniques. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 1932–1938, 2019.

[6] Hermine N. Akouemo and Richard J. Povinelli. Probabilistic anomaly detection in natural gas time series data. *International Journal of Forecasting*, 32(3):948–956, 2016.

[7] Kaixing Huang, Chunjie Zhou, Yu-Chu Tian, Shuanghua Yang, and Yuanqing Qin. Assessing the physical impact of cyberattacks on industrial cyber-physical systems. *IEEE Transactions on Industrial Electronics*, 65(10):8153–8162, 2018.

[8] Shengyuan Liu, Yuxuan Zhao, Zhenzhi Lin, Yilu Liu, Yi Ding, Li Yang, and Shimin Yi. Data-driven event detection of power systems based on unequal-interval reduction of pmu data and local outlier factor. *IEEE Transactions on Smart Grid*, 11(2):1630–1643, 2020.

[9] S. Sahoo, T. Dragičević, and F. Blaabjerg. Cyber security in control of grid-tied power electronic converters–challenges and vulnerabilities. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, pages 1–1, 2019.

[10] Nils Jakob Johannesen, Mohan Kolhe, and Morten Goodwin. Relative evaluation of regression tools for urban area electrical energy demand forecasting. *Journal of Cleaner Production*, 218:555–564, 2019.

[11] Nils Jakob Johannesen and Mohan Lal Kolhe. Application of regression tools for load prediction in distributed network for flexible analysis. In *Flexibility in Electric Power Distribution Networks*. CRC Press, 2021.

[12] Hermine N. Akouemo and Richard J. Povinelli. Time series outlier detection and imputation. *2014 IEEE PES General Meeting — Conference & Exposition*, pages 1–5, 2014.

[13] Xue Liu, Yong Ding, Hao Tang, and Feng Xiao. A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data. *Energy and Buildings*, 231:110601, 2021.

[14] Yassine Himeur, Abdullah Alsalemi, Faycal Bensaali, and Abbes Amira. Smart power consumption abnormality detection in buildings using micromoments and improved k-nearest neighbors. *International Journal of Intelligent Systems*, 36(6):2865–2894, 2021.

[15] Hadis Karimipour, Sandra Geris, Ali Dehghantanha, and Henry Leung. Intelligent anomaly detection for large-scale smart grids. In *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, pages 1–4, 2019.

[16] Youbiao He, Gihan J. Mendis, and Jin Wei. Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism. *IEEE Transactions on Smart Grid*, 8(5):2505–2516, 2017.

[17] Manikant Panthi. Anomaly detection in smart grids using machine learning techniques. In *2020 First International Conference on Power, Control and Computing Technologies (ICPC2T)*, pages 220–222, 2020.

[18] Giuseppe Fenza, Mariacristina Gallo, and Vincenzo Loia. Drift-aware methodology for anomaly detection in smart grid. *IEEE Access*, 7:9645–9657, 2019.

[19] Alexandra L'heureux, Katarina Grolinger, Hany F Elyamany, and Miriam AM Capretz. Machine learning with big data: Challenges and approaches. *Ieee Access*, 5:7776–7797, 2017.

[20] D. Patidar, Bhavin C. Shah, and M. Mishra. Performance analysis of k nearest neighbors image classifier with different wavelet features. *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)*, pages 1–6, 2014.

[21] Sholom M. Weiss and Nitin Indurkhya. Rule-based machine learning methods for functional prediction. *J. Artif. Int. Res.*, 3(1):383–403, December 1995.

[22] Ehdieh Khaledian, Shikhar Pandey, Pratim Kundu, and Anurag K. Srivastava. Real-time synchrophasor data anomaly detection and classification using ¡italic¿isolation forest, kmeans, and loop¡/italic¿. *IEEE Transactions on Smart Grid*, 12(3):2378–2388, 2021.

[23] Wei Mao, Xiu Cao, Qinhua zhou, Tong Yan, and Yongkang Zhang. Anomaly detection for power consumption data based on isolated forest. In *2018 International Conference on Power System Technology (POWERCON)*, pages 4169–4174, 2018.

[24] Armin Aligholian, Mohammad Farajollahi, and Hamed Mohsenian-Rad. Unsupervised learning for online abnormality detection in smart meter data. In *2019 IEEE Power Energy Society General Meeting (PESGM)*, pages 1–5, 2019.

[25] Saeed Ahmed, Youngdoo Lee, Seung-Ho Hyun, and Insoo Koo. Unsupervised machine learning-based detection of covert data integrity assault in smart grid networks utilizing isolation forest. *IEEE Transactions on Information Forensics and Security*, 14(10):2765–2777, 2019.

[26] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, page 93–104, New York, NY, USA, 2000. Association for Computing Machinery.

[27] Yifan Ou, Bin Deng, Xuan Liu, and Ke Zhou. Local outlier factor based false data detection in power systems. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 2003–2007, 2019.

[28] AEMO. National electricity market data - nem, 2021.

[29] Vasudev Dehalwar, Akhtar Kalam, Mohan Lal Kolhe, and Aladin Zayegh. Electricity load forecasting for urban area using weather forecast information. In *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*, pages 355–359, 2016.

[30] NVE. Bjønntjønn kraftverk og pumpestasjon. https://www.nve.no/konsesjon/konsesjonssaker /konsesjonssak?id=8386type=V-1, 2021.

[31] NVE. Tesla ladestasjon i treungen. https://www.nve.no/konsesjon/konsesjonssaker/konsesjonssak/ 1, 2021.

[32] Jarle Pedersen. Tesla bygger ny supercharger i telemark. https://www.tvedestrandsposten.no/tesla-bygger-ny-supercharger-i-telemark/s/5-50-1252227, 2022.

[33] Camilla Moen. Planlegger gigantutbygging med 445 nye hytter på gautefallheia: - noen mener det er kontroversielt. https://www.kv.no/planlegger-gigantutbygging-med-445-nye-hytter-pa-gautefallheia-noen-mener-det-er-kontroversielt/s/5-63-271204, 2021.

[34] M. Holstad T. Aanensen. Supply and consumption of electricity in the period 1993-2016. https://www.ssb.no/en/energi-og-industri/artikler-og-publikasjoner/supply-and-consumption-of-electricity-in-the-period-1993-2016, 2017.

[35] Nils Jakob Johannesen, Mohan Lal Kolhe, and Morten Goodwin. Load demand analysis of nordic rural area with holiday resorts for network capacity planning. In *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–7, 2019.

[36] Moez Ali. *PyCaret: An open source, low-code machine learning library in Python*, April 2020. PyCaret version 1.0.0.