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

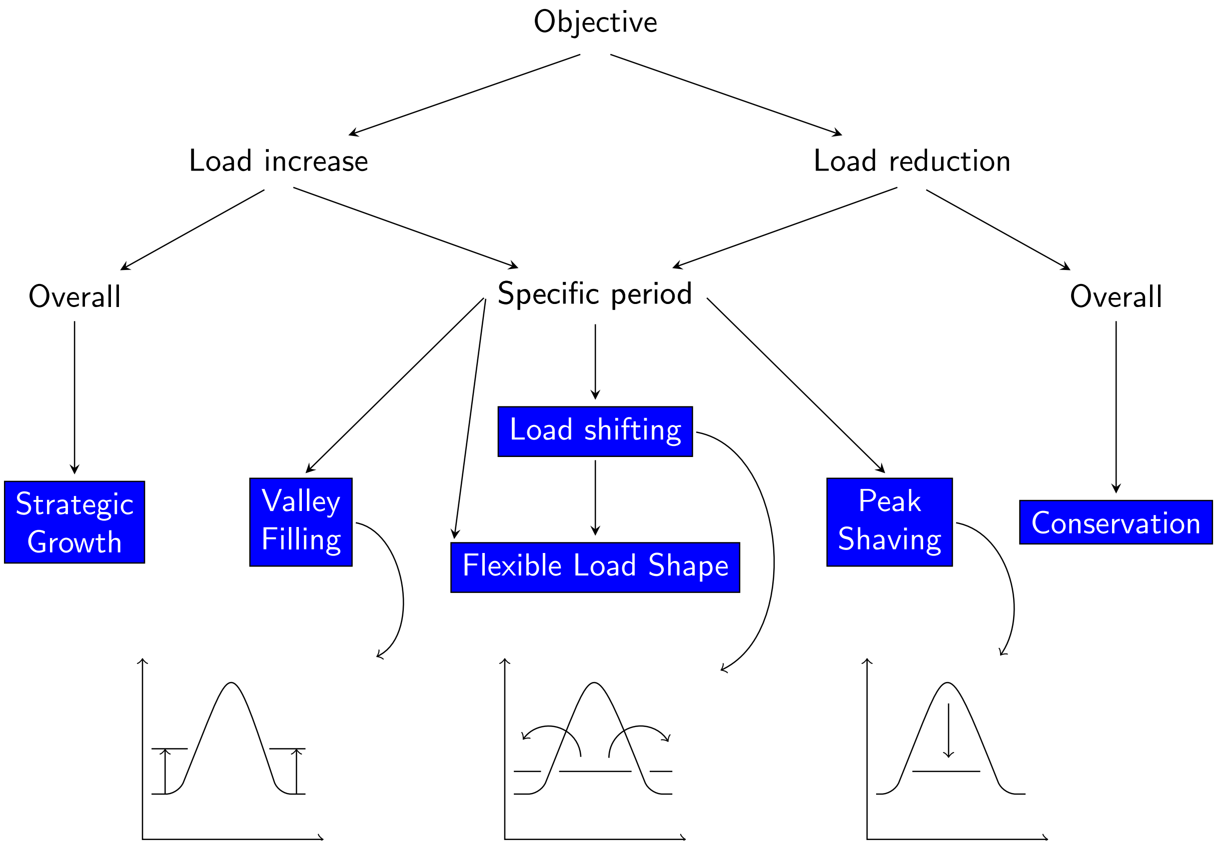

Figure C.1: Objective of load shapes

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.

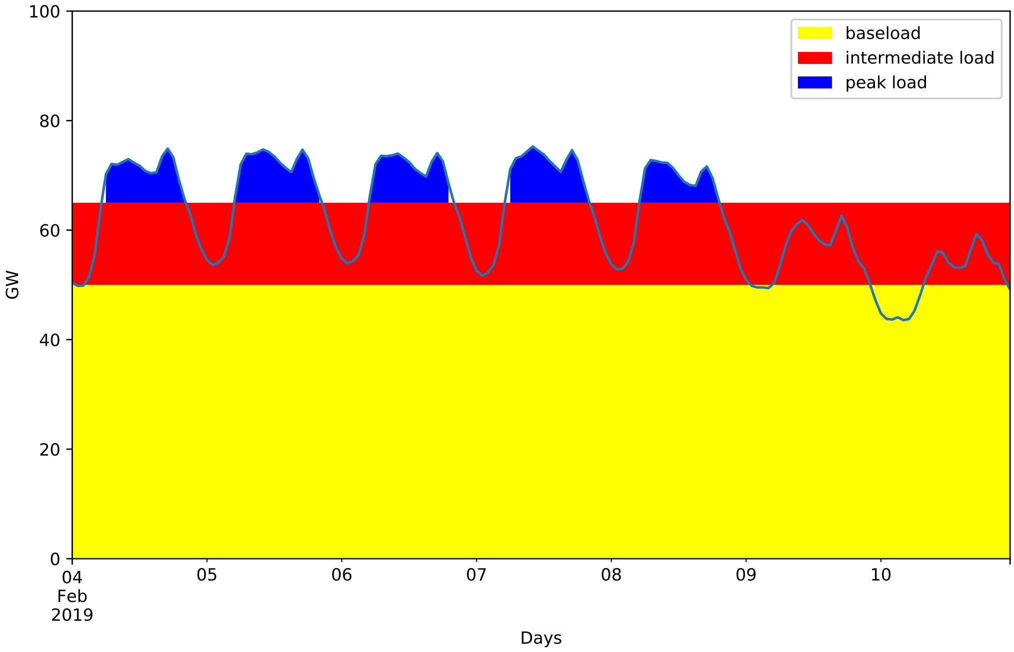

Figure C.2: The electric load demand curve of a sample European country for one week, indicating level of load curves. Source: ENTSOE-E

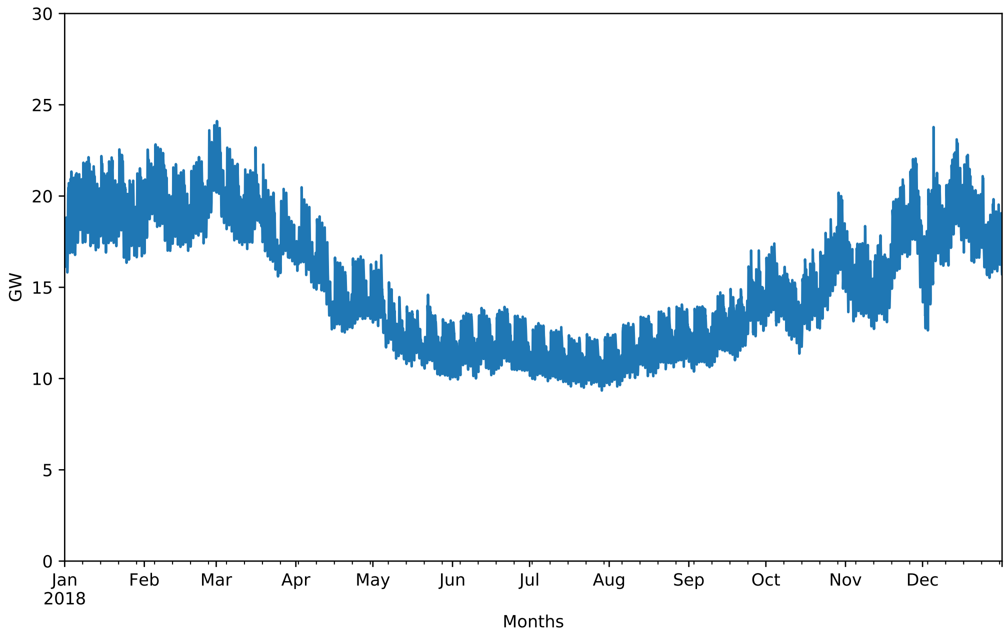

Figure C.3: The yearly electric load demand curve of a sample European country, depicting seasonal changes. Source: ENTSOE-E

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].

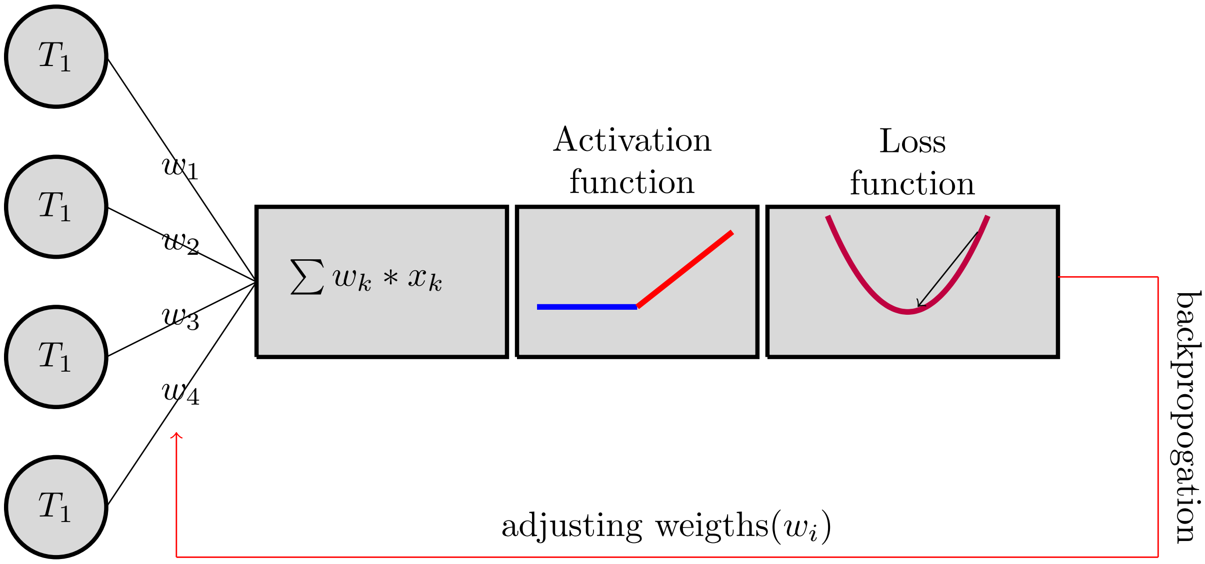

Figure C.4: Neural network model with the propagating adjusted weights

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 $\sigma^2_\epsilon(t)$, 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].

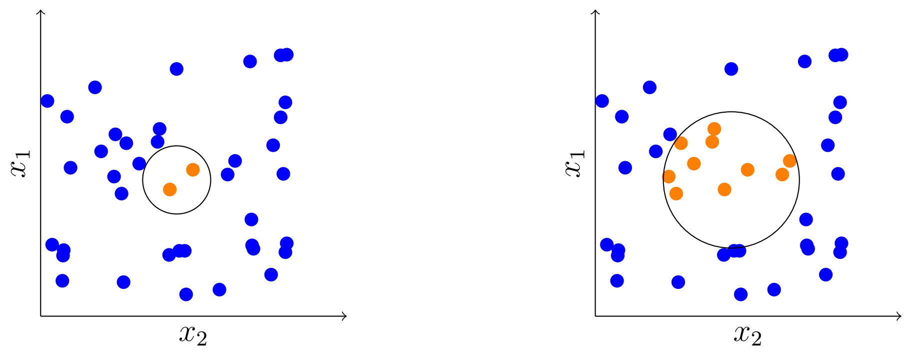

Figure C.5: k-Nearest Neighbour classifying based on the k'th observation.

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].

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

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

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

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

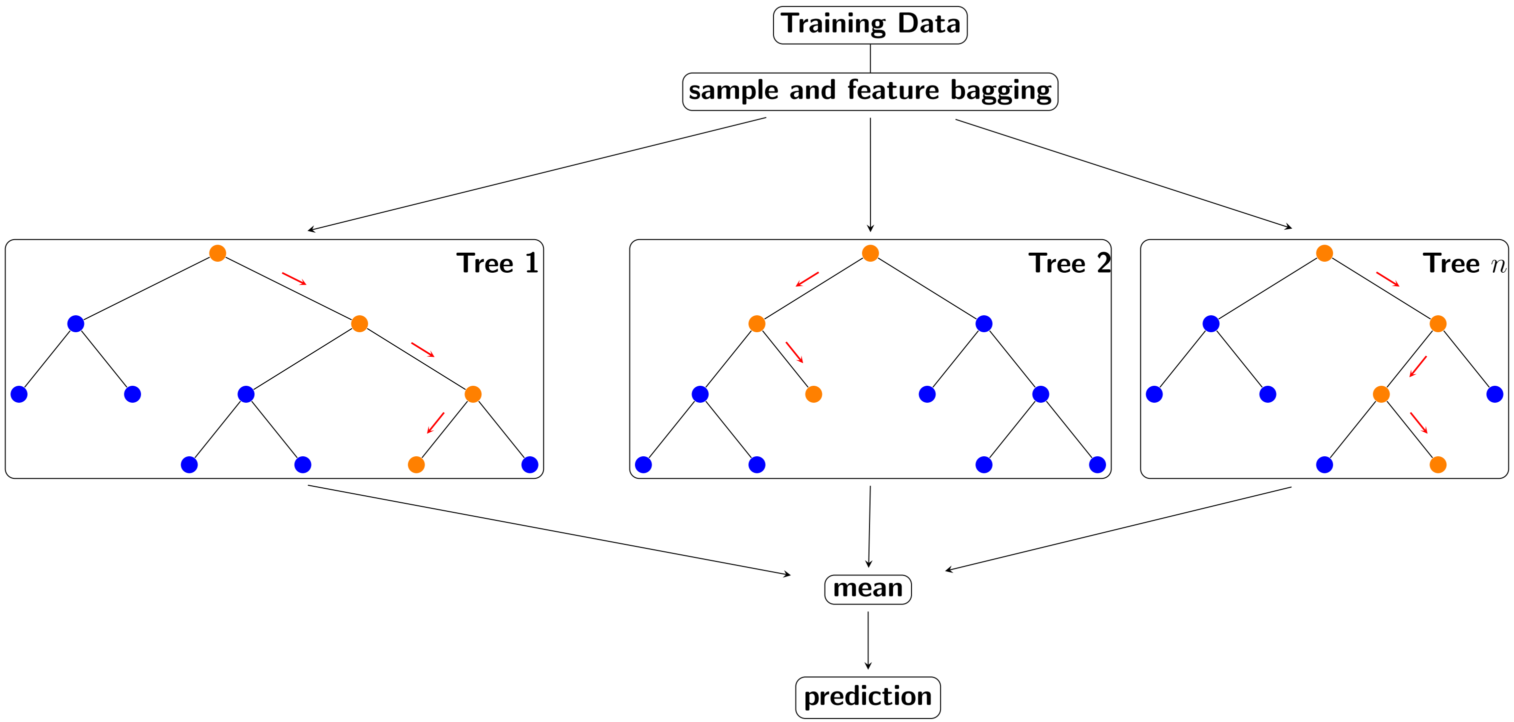

Figure C.6: Random Forest Regression diagram sampling and voting from n trees

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

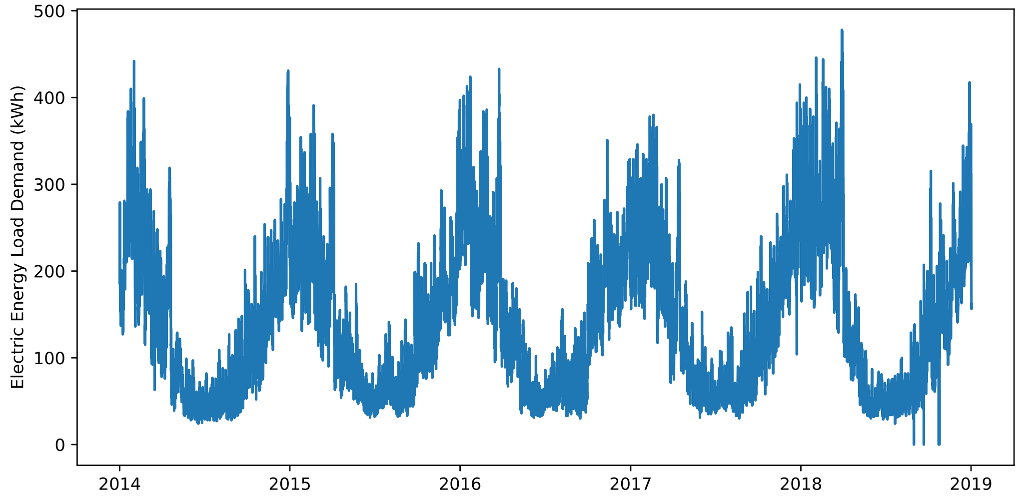

Figure C.7: Rural Load

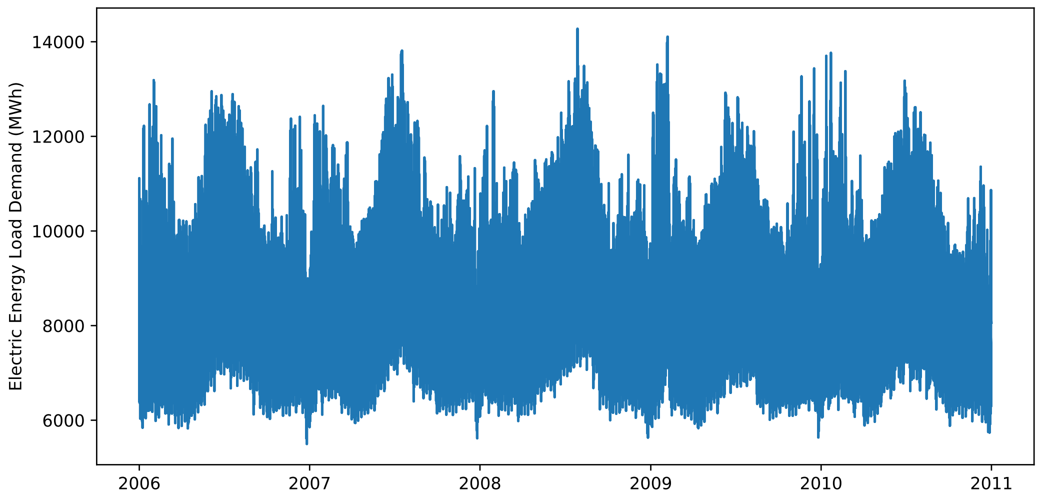

Figure C.8: Urban Load

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.

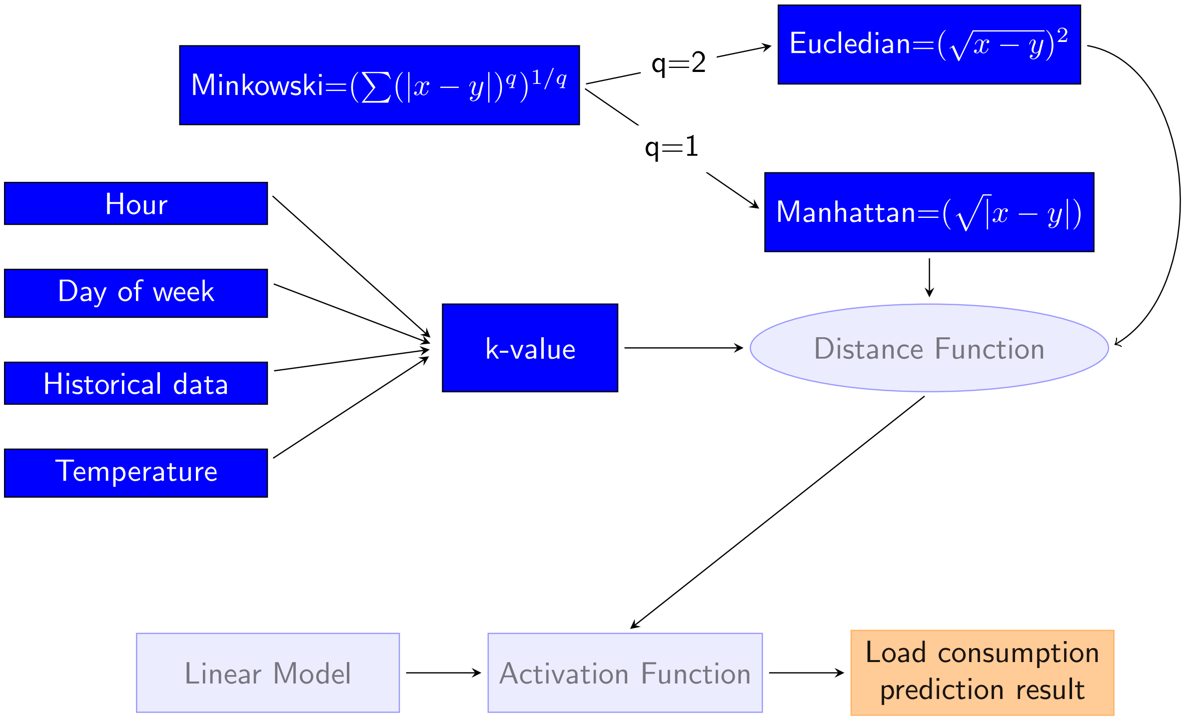

Figure C.9: The regressor model for electric load demand forecasting

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 seasonwise, 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

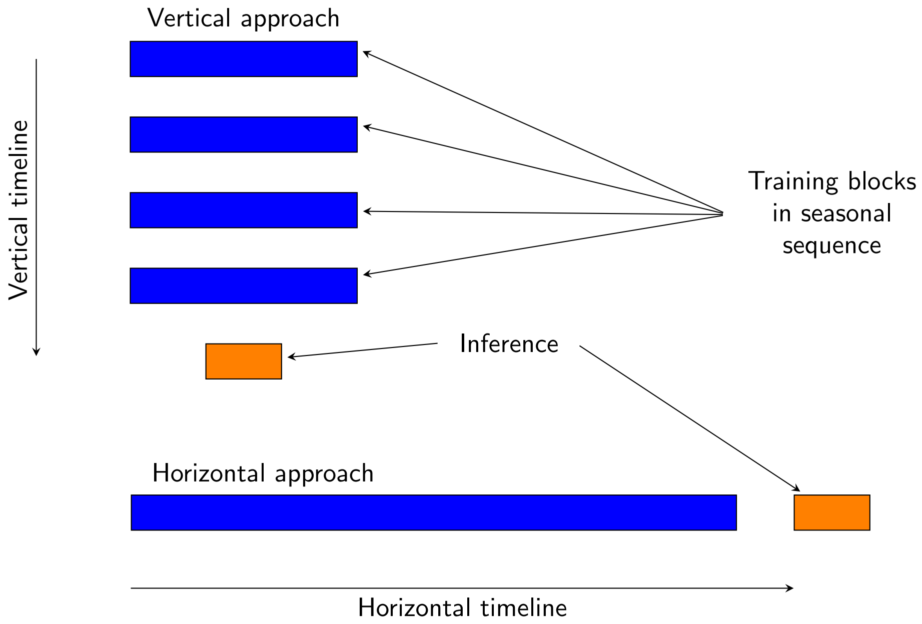

Figure C.10: Illustration of vertical and horizontal approach.

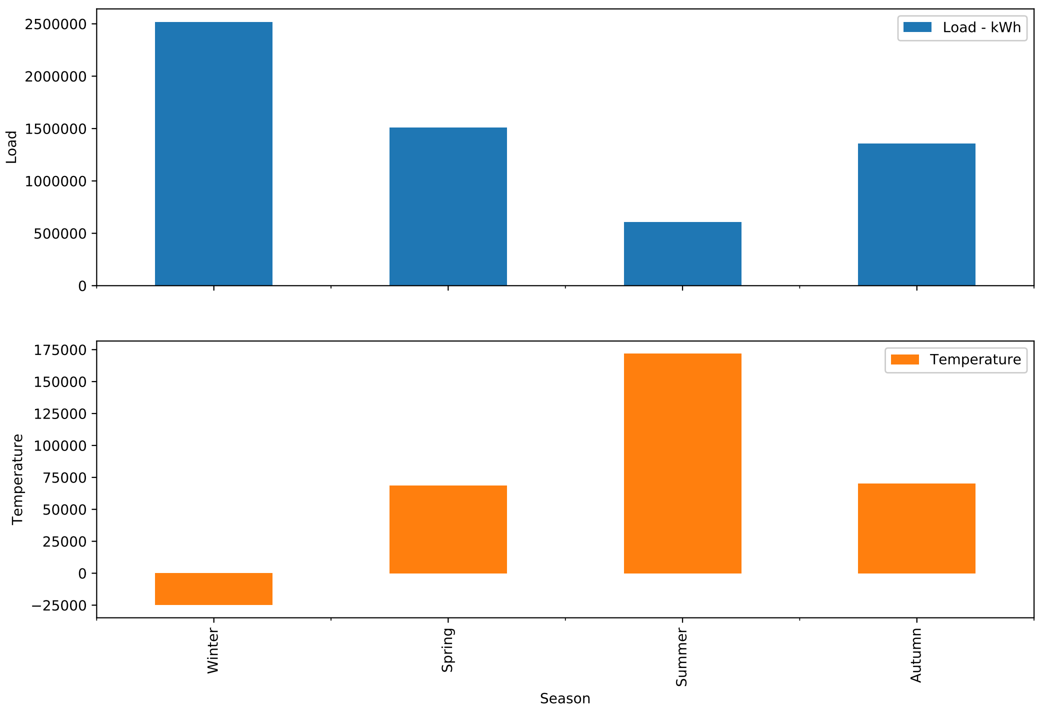

Figure C.11: Load consumption and temperature profiles on seasonal basis

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

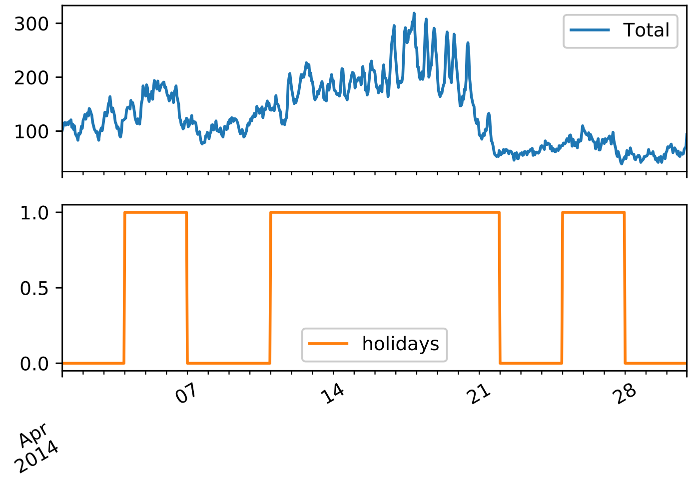

Figure C.12: Load consumption related to working and non-working days

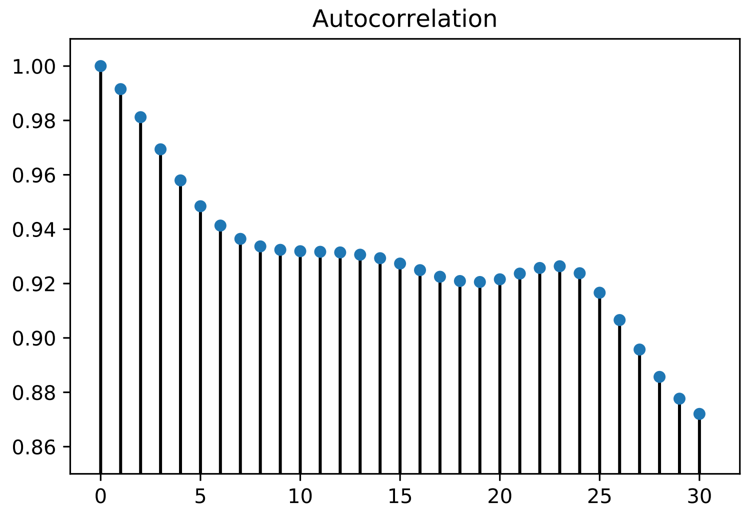

Figure C.13: Autocorrelation of load consumption of the first 30 lags for Bjønntjønn Cabin Area 2014-2018

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

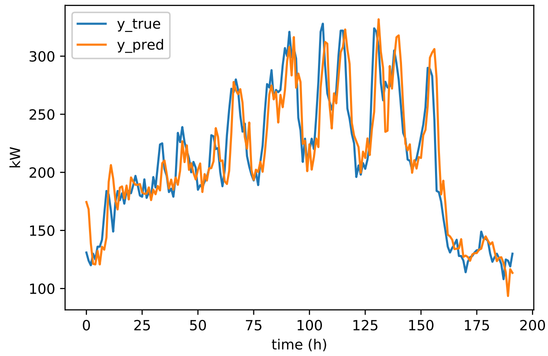

Figure C.14: Prediction result without error estimation

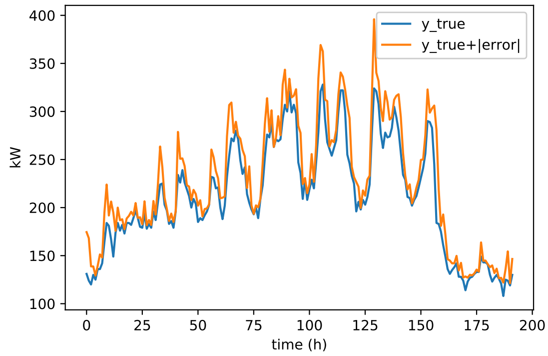

Figure C.15: Prediction result with error estimation

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)

	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

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.