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

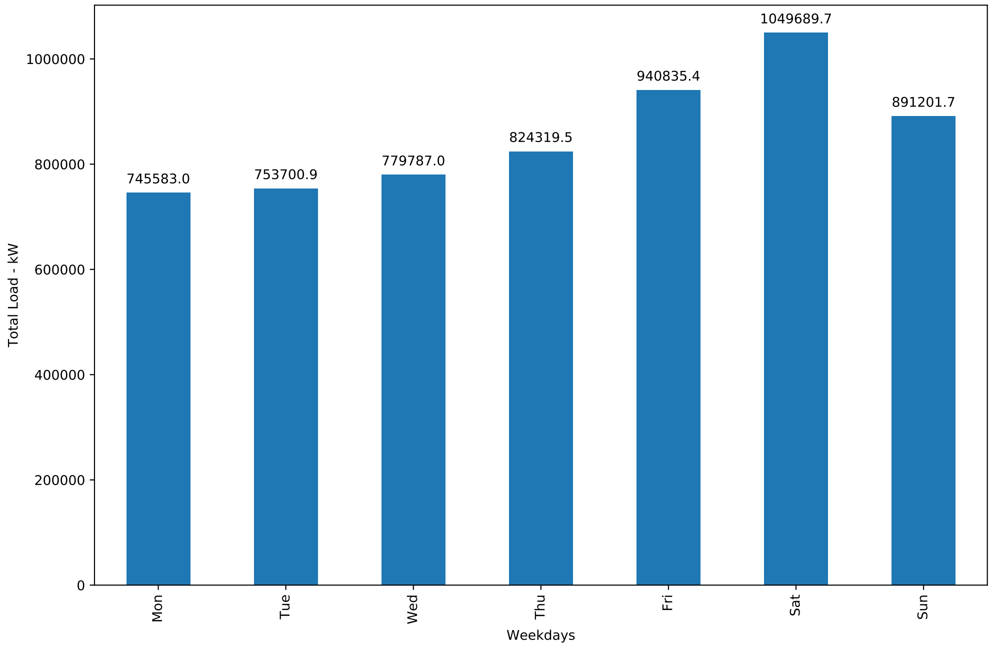

Figure B.1: The total sum of load in kWh for the days of the week in Bjønntjønn Cabin Area 2014-2018

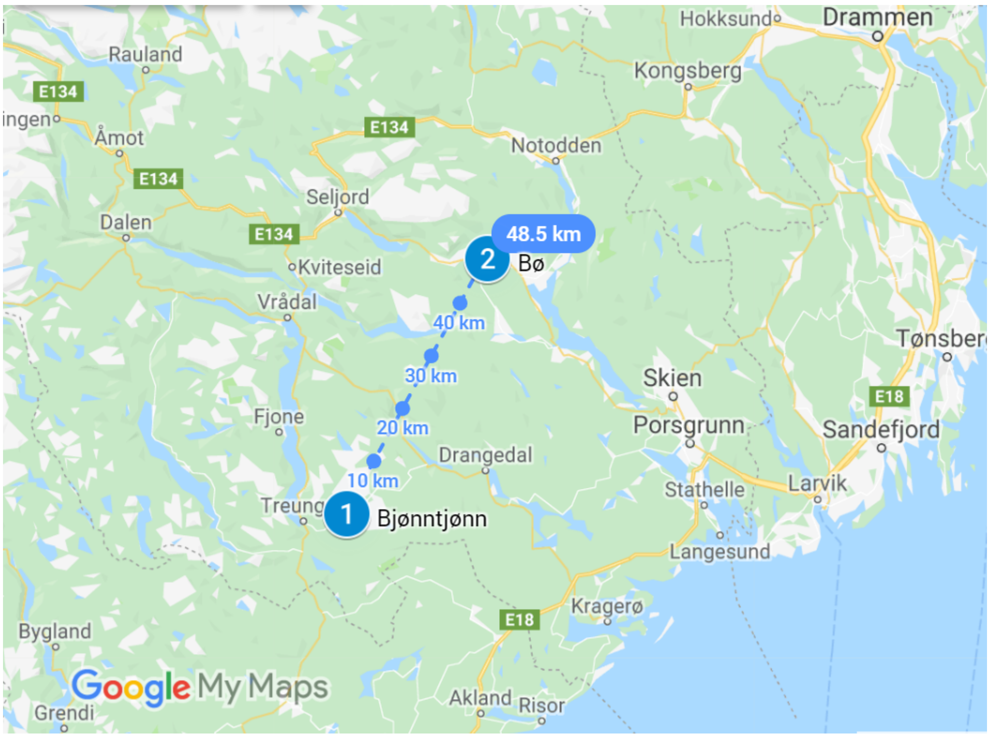

Figure B.2: The Bjønntjønn Cabin Area and weather station Bø. Map data © 2019 Google.

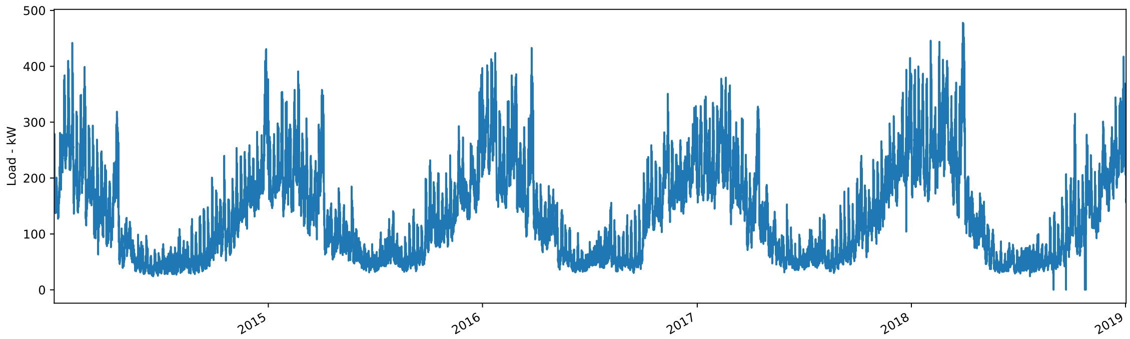

Figure B.3: Electric Energy Load Demand at Bjønntjønn Cabin Area in south of Norway from 2014 to 2018.

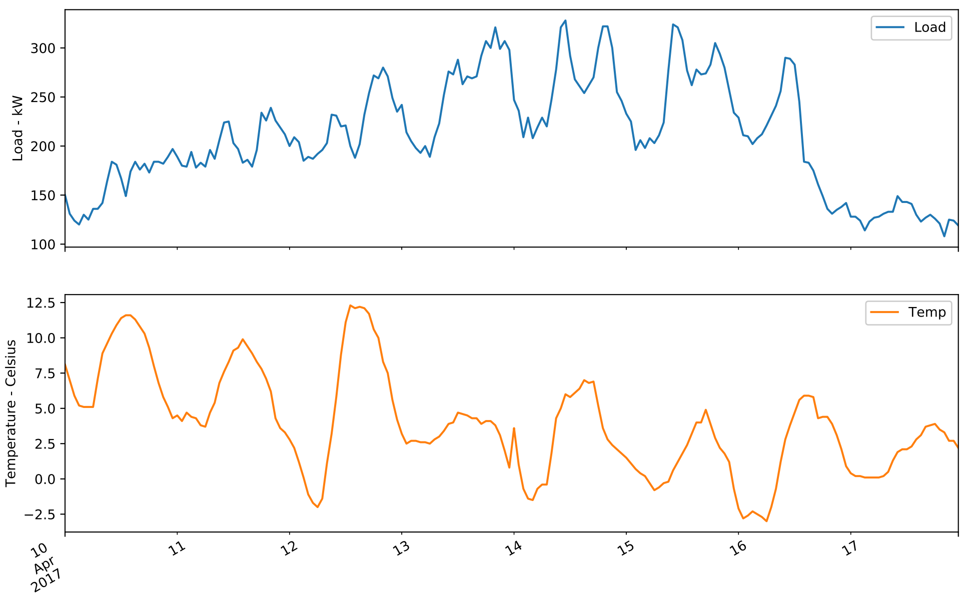

Figure B.4: Load consumption and temperature readings during Easter holiday 2017 computation time, avoid overfitting and induce model transparency.

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.

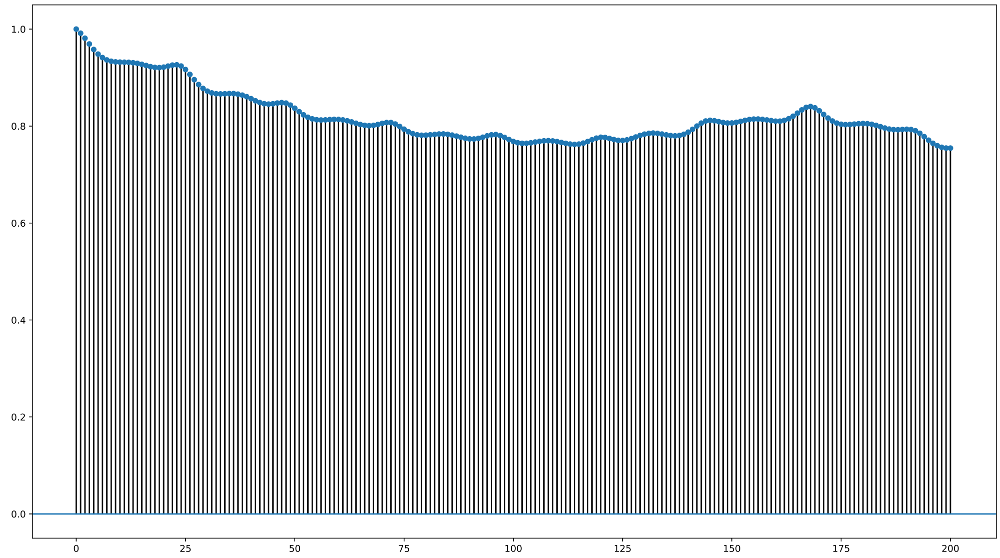

Figure B.5: Autocorrelation of load consumption by 200 lags for Bjønntjønn Cabin Area 2014-2018

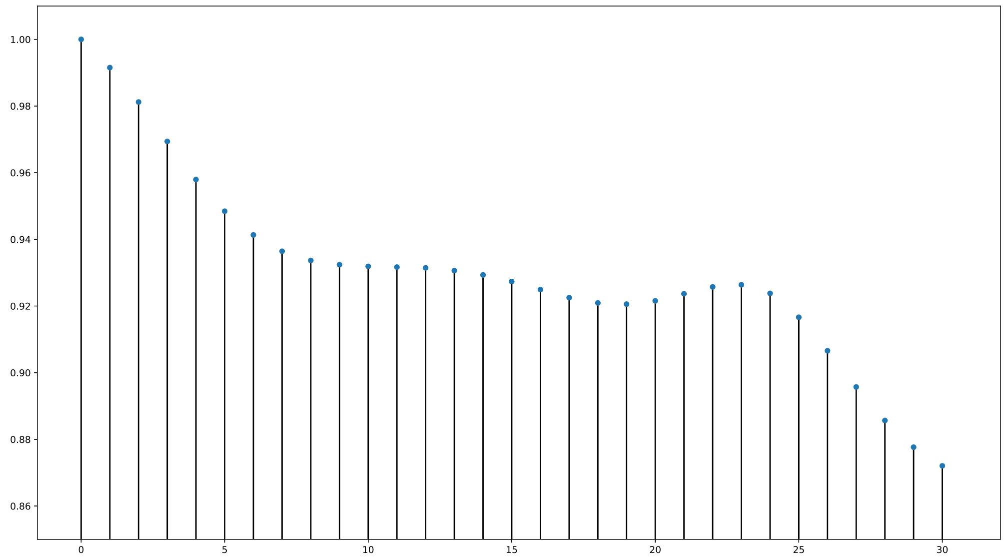

Figure B.6: Autocorrelation of load consumption of the first 30 lags for Bjønntjønn Cabin Area 2014-2018

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-

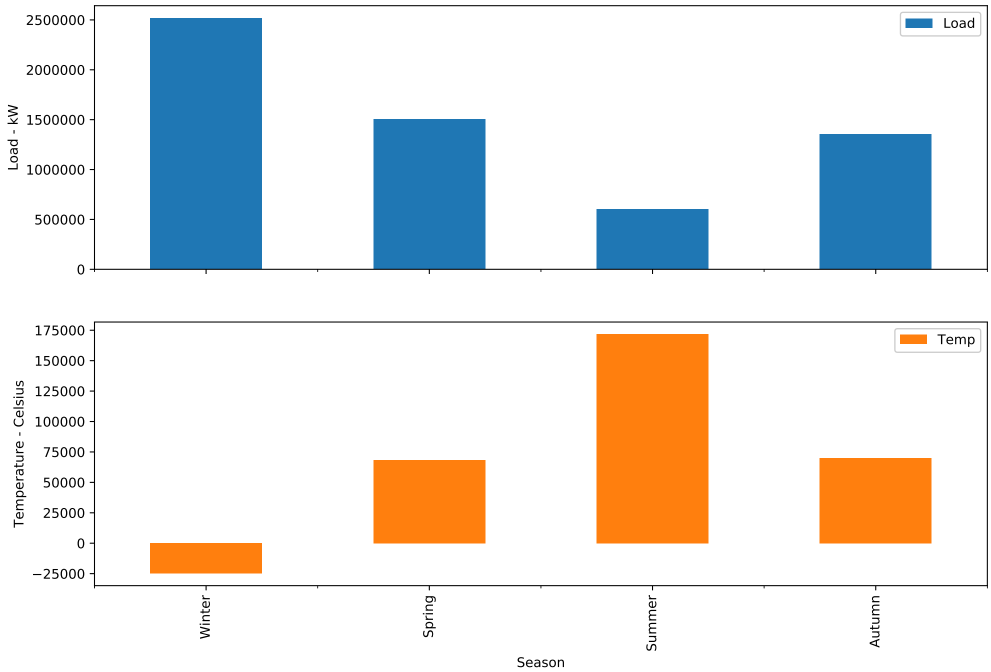

Figure B.7: Sum of load consumption and temperature on a seasonal basis

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

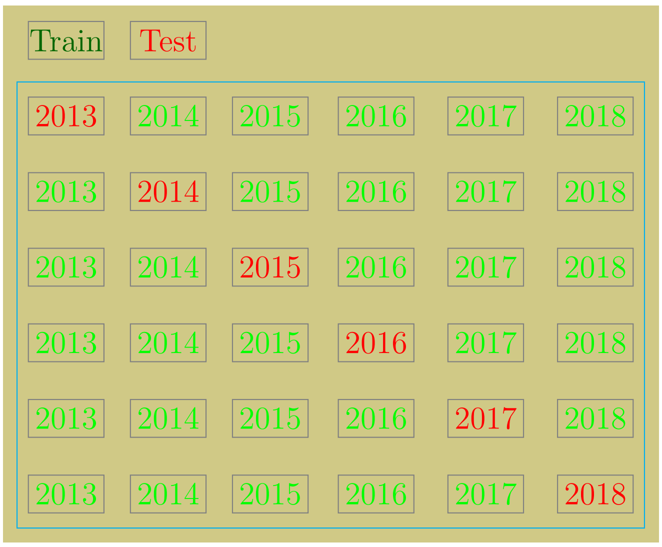

Figure B.8: Leave-one-out, or jackknife, leaves the test sample out of the training and trains the algorithm on the remaining

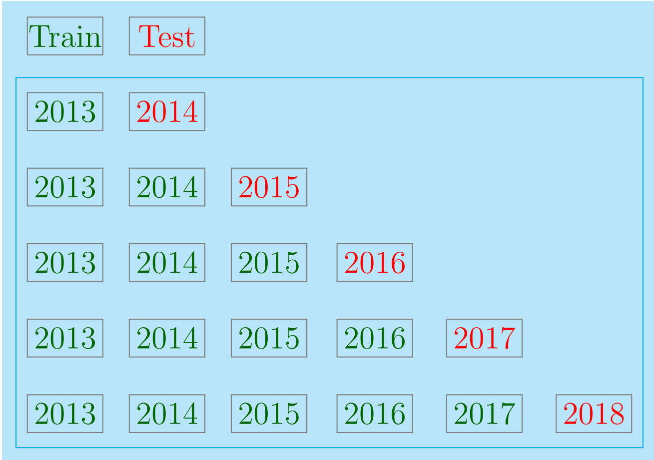

Figure B.9: Crogging combines cross validation and forecast aggregation to capture the temporal dependency of time-series

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

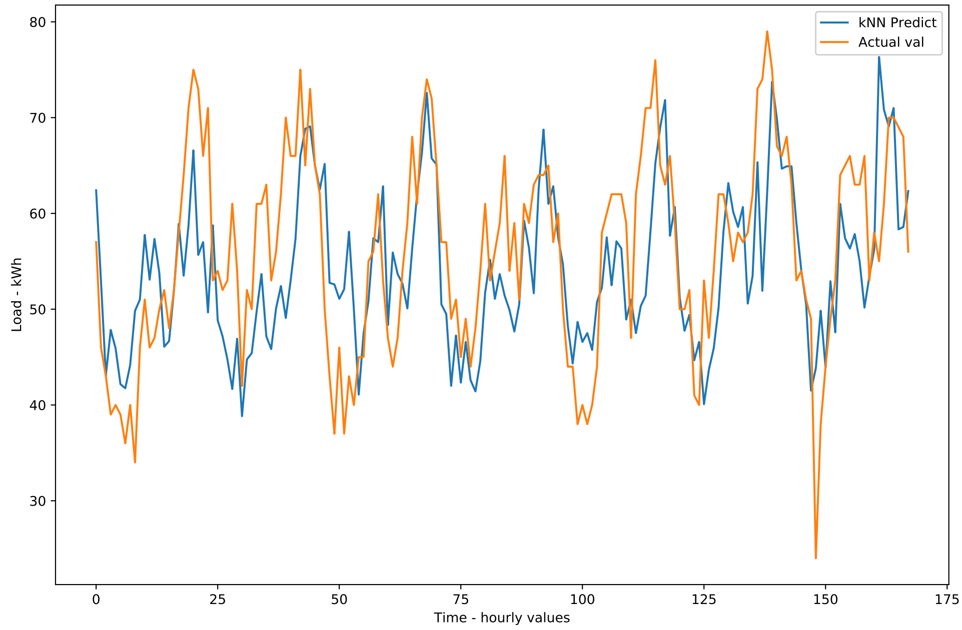

Figure B.10: Prediction for week of July 2018 scoring MAE 6.87 using vertical approach kNN-regressor only trained with time features.

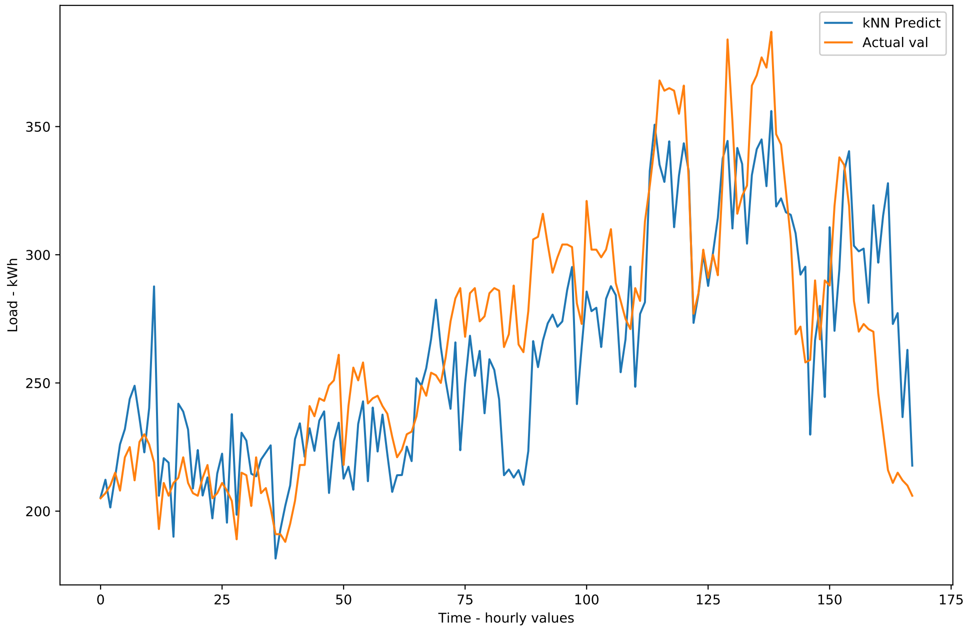

Figure B.11: kNN regressor with autoregressor by continuous approach scoring 8.86 SMAPE and 8.88 MAPE for a week in January 2018. Trained with time feature, weather parameters and holiday information.

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.