Multi-Timescale Forecasting of Battery Energy Storage State-of-Charge under Frequency Containment Reserve for Normal Operation

Aleksei Mashlakov (iD)
School of Energy Systems
LUT University
Lappeenranta, Finland
aleksei.mashlakov@lut.fi

Samuli Honkapuro (iD)
School of Energy Systems
LUT University
Lappeenranta, Finland
samuli.honkapuro@lut.fi

Ville Tikka (iD)
School of Energy Systems
LUT University
Lappeenranta, Finland
ville.tikka@lut.fi

Arto Kaarna (iD)
School of Engineering Science
LUT University
Lappeenranta, Finland
arto.kaarna@lut.fi

Lasse Lensu (iD)
School of Engineering Science
LUT University
Lappeenranta, Finland
lasse.lensu@lut.fi

Abstract—Forecasting the state-of-charge changes of battery energy storage, anticipated from a provision of different services, can facilitate planning of its market participation strategy and leverage the maximum potential of its energy capacity. This paper provides a performance comparison study of multiple decision-tree and data-driven machine learning methods for point forecasts of the state-of-charge of battery energy storage under frequency containment reserve for normal operation on day-, hour-, and 15-minute-ahead basis. The battery state-of-charge data for the performance evaluation were simulated with a droop curve battery model based on the historical frequency data in the northern Europe synchronous area. The results show that the data-driven methods outperform the decision-tree based methods on the 15-minute- and day-ahead time scales while demonstrating a comparable performance for the hour-ahead time scale.

Index Terms—battery state-of-charge, point forecasting, decision trees, machine learning, frequency containment reserve

I. INTRODUCTION

The European Union has recently revised its 2030 climate and energy framework. An important change is hardening its key target for the share of renewable energy on the grid to 32% in pursuit of low-carbon economy and its commitments under the Paris Agreement [1]. The implementation of these directives will demand more flexibility to guarantee the stability of power system operation in the conditions of gradual replacement of dispatchable fossil-fuel based power plants by intermittent renewable power generation.

Battery energy storage systems (BESSs) are one of the flexibility elements of power systems that are able to handle the foreseeable stability challenges provoked by the enlarged share of renewables [2]. In fact, the control efficiency of a BESS as a frequency containment reserve for normal operation

(FCR-N) surpasses conventional power plants in ramping capabilities [3]. In this scenario, frequency regulation is the most common primary application for battery storage globally [4]. However, in addition to FCR-N, technical capabilities of BESSs enable provision of other services [5], and hence, its efficient utilization is bound to various decision making activities for finding the trade-off between the benefits of different services and markets [6]. These decisions are dependent on the forecast uncertainties of coupled resources and corresponding stochastic processes that predetermine the development of BESS state-of-charge (SOC) in time. Consequently, accurate forecasting that would facilitate the economic dispatch of BESSs and unit commitments is essential.

The goal of this study is to analyze and compare the performance of several traditional and state-of-the-art machine learning methods for point forecasting of BESS SOC in the condition of BESS operating as a FCR-N in the northern Europe synchronous area. The results of the study include the performance evaluation of a decision tree (DT), random forest (RF), Light Gradient Boosting Machine (LightGBM), dilated convolutional neural network (DCNN), and multi-attention recurrent neural network (MARNN) on day-, hour-, and 15-minute-ahead time scales. Moreover, a data analysis of the real frequency measurements utilized to simulate a BESS SOC dataset and feature engineering for the dataset are provided.

The rest of this paper is organized as follows: Section II describes the general principles of the machine learning methods applied in this study for point forecasting. In Section III, the case study is introduced including frequency data analysis in northern Europe, feature engineering for the BESS SOC dataset, implementation details, and evaluation criterion for point forecasting. The results are presented in Section IV. Finally, the conclusions are drawn in Section V.

II. METHODS

A. Decision Tree

Decision tree represents a non-parametric supervised learning method utilized for classification and regression tasks [7]. A fitting model of the method is developed by recursively partitioning a dataset into smaller subsets by using tests applied to the feature or sample values. Each consecutive splitting point is defined by a greedy algorithm with cost function minimization as the criterion. The final structure of the model corresponds to a binary tree where the topmost node represents the root node, the intermediary nodes constitute the binary decision rules, and each leaf node can be regarded as the classification or regression outcome. The advantages of DTs include the ability to capture non-linear patterns in the data, while sensitivity to noisy data is one of their drawbacks.

B. Random Forest

An ensemble of multiple DTs is known as the random forest method [8]. In this method, the final output of multiple DT models is combined to stabilize the performance of individual DTs for both regression and classification tasks. The aggregation of the individual models is implemented with a bagging technique that involves training each DT model on a different dataset with replacement. In addition to the bagging, random feature selection is utilized. Then, an average of all the predictions from the individual trees is computed to define the final outcome. As a result, RFs provide more robust performance for high-dimensional data and reduce the variance of the DTs.

C. Light Gradient Boosting Machine

Light Gradient Boosting Machine is a modern and fast gradient boosting framework that is also based on the DT algorithm [9]. Gradient boosting decision trees follow the principle of consecutively fitting an ensemble of DT models to the data and weighting the model outcomes based on the previous outcome with the gradient descent algorithm. Eventually, the prediction is generated by adding the weights of individual models. The main advantage of this approach is that LightGBM splits the tree in a leaf-wise manner with the best fit in contrast to the depth-wise or level-wise algorithms for tree splitting used in other boosting algorithms. This split approach enables construction of more complex trees and leads to better loss reduction than in the level-wise algorithms.

D. Dilated Convolutional Neural Network

Convolutional neural networks are state-of-the-art deep learning approaches for image classification problems [10]. They have recently gained popularity also in time-series forecasting [11]. In this study, we adopt DCNN that connects the past inputs to future time step outputs in a causal structure. Moreover, dilated convolutions presented in [12] enable the exponential relationship between the dilated layer depth and input size to handle longer sequences. Such relationship is realized by skipping constant dilation rate inputs in between each of the inputs.

E. Multi-Attention Recurrent Neural Network

Recurrent neural networks (RNNs) have proven to be a robust method for prediction of sequentially dependent data [13]. An attention mechanism introduced in [14] is a recent advancement in the recurrent deep learning that further improves the results of vanilla RNN in memorizing long source sequences. The main difference with the RNN is that the attention develops an aggregate context vector that is filtered specifically for each output time step and memorized in the decoder layer [15]. In this study, we utilize MARNN model described in [16] that deploys lag sample values from previous input sequences at the decoding time.

III. CASE STUDY

A. Description

The case study corresponds to a BESS operating as a FCR-N in the northern Europe synchronous area. The FCR-N market in this area is individually organized by the transmission system operators (TSOs) that also set the service rules. The reserve procurement occurs on the basis of competitive bidding at the yearly market and at the hourly market arranged once a day for additional procurement. The forecasting time scales for the performance evaluation were selected considering the battery operation in a multi-objective environment with current procurement practices and possible future developments towards shorter procurement intervals.

B. Data

In this study, the real frequency measurements for the period of four years from 2015 to 2018 served as an input for the BESS model to generate the variations of BESS SOC. The droop curve parameters set by the TSO service rules in the Finnish bidding area for the corresponding period were used [17]. The details about the assumptions and BESS model characteristics can be found in [16]. Moreover, to support BESS SOC forecasting, several time, market, frequency, and modified SOC features were retrieved and their correlations with the simulated hourly SOC data are illustrated in Fig. A.1 of Appendix A. The dataset is publicly available in [18] for examination, and its features are described below.

The frequency measurements utilized for the simulation were retrieved from open data [19]. The frequency distribution patterns for hourly mean frequency values are graphically represented as a boxplot in Fig. 1. Following the medians of the boxes, several over-frequency (0, 9, 21, and 23) as well as under-frequency (2, 3, 4, 11, and 18) hourly trends can be identified. If compared to the other hours, the morning hours (6, 7, and 8) are highlighted by the smallest interquantile range and variations between the maximum and minimum frequency values. Also, the frequency distribution is close to being symmetrical in most of the hours. In general, the sequential distribution pattern of the frequency data resembles wave-like structure with different wavelengths and amplitudes.

Fig. 2 illustrates the relationships between daily sets of 15-minute mean frequency data with 4 main clusters. To obtain the clusters, a hierarchical dendrogram was built with
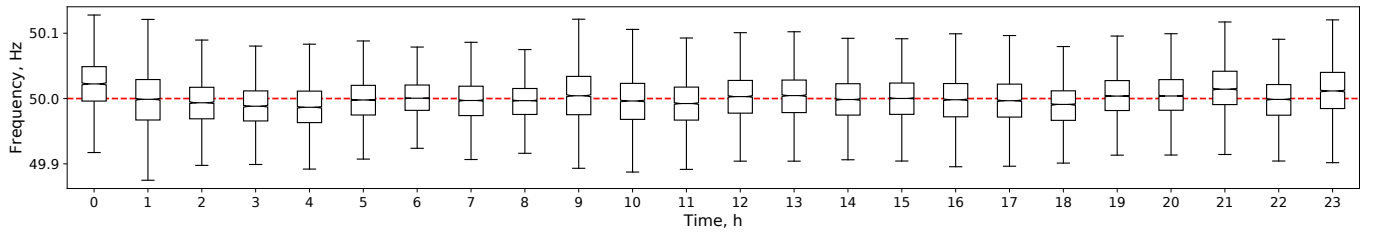
Fig. 1: Distribution of hourly mean frequency values in the northern Europe synchronous area (2015 – 2018). The outliers are not shown.
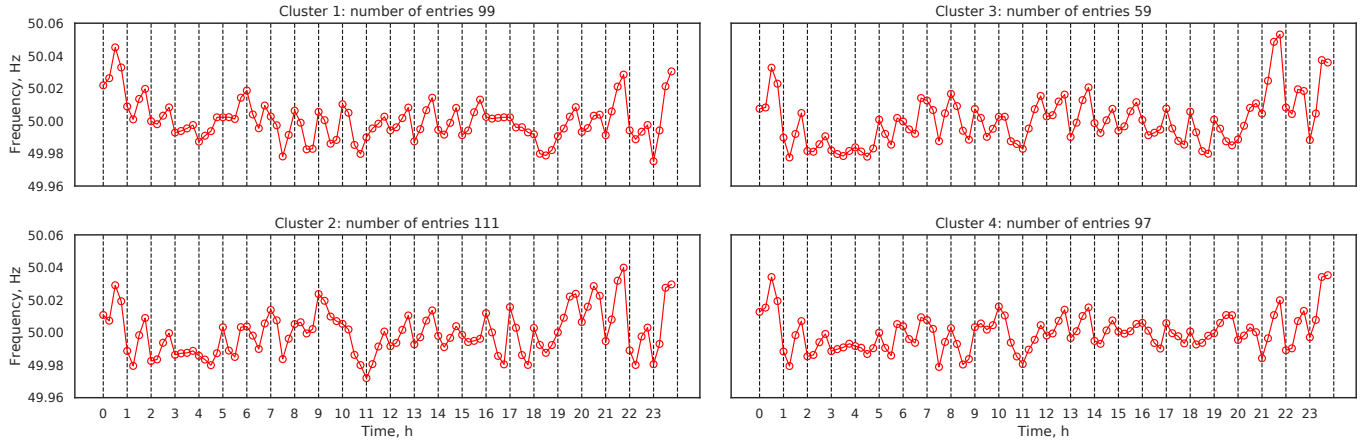

Fig. 2: Clusters of daily 15-minute mean frequency values in the northern Europe synchronous area (2015 – 2018).

(a) Before autocorrelation removal

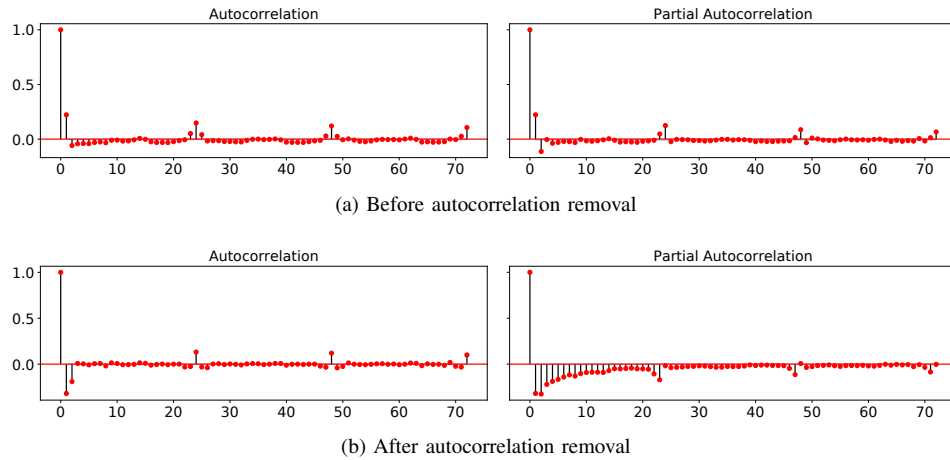

(b) After autocorrelation removal

Fig. 3: Examples of autocorrelation and partial autocorrelation plots for the hourly BESS SOC data (a) before and (b) after the removal of hourly correlation.

Ward's minimum variance method [20]. From the clusters, the hourly dips and spikes identified by the medians in Fig. 1 can be clearly seen, and, in addition, intra-hour variations are distinguishable. For instance, one of the intra-hour properties of such variations is in the frequency rise at the beginning of the peak morning hours (Clusters 1 – 4) and, in some cases, peak evening hours (Clusters 2 – 3) followed by the frequency dips. The opposite trend is illustrated by all the clusters during the daily hours (11 – 15) and late evening hours (20, 21, and

23). This power system dynamics can be potentially explained by the trading model that includes hourly market with an uneven distribution of load during the hours and primarily stable power output of the generation units dispatched in the market [21]. In the markets with high volume of intermittent renewable generation, mistakes in the forecasted power output of the renewables can be an additional factor contributing to the mismatch between the market equilibrium and the real power balance situation.

The time features in Fig. A.1a incorporate weekly (as w_day) and monthly (m_day) day values for the data time range, as well as sine and cosine functions of hourly and monthly values that were obtained as follows:

$$hr\_sin(h) = \sin\left(\frac{2\pi}{24}h\right),$$
$$mn\_sin(m) = \sin\left(\frac{2\pi}{12}(m-1)\right),$$
$$hr\_cos(h) = \cos\left(\frac{2\pi}{24}h\right),$$
$$mn\_cos(m) = \cos\left(\frac{2\pi}{12}(m-1)\right).$$

(1)

The market data in Fig. A.1b were retrieved from open data in [22] and incorporate a hourly sum of wind power generation in Finland, Sweden, and Denmark (wind), as well as hourly automatic activated reserve (up_aar, dn_aar) and regulation volumes (up_reg, dn_reg) summarized for all nordic bidding areas. Also, the frequency data (freq) were added as an hourly mean to Fig. A.1b. The additional statistical frequency data were retrieved from the frequency measurements and include hourly sums of the number of seconds when the frequency was above or below daily single (1s_up, 2s_up) or double (1s_dn, 2s_dn) the standard deviation value, respectively. Moreover, the means of these features related to the positive (sm_up) or negative (sm_dn) deviation were obtained. Furthermore, the SOC data modified by shifting (sh_1, sh_24, sh_48) and applying difference (df_1, df_24, df_48) with steps 1, 24, and 48 were investigated because of the highest correlation at these time steps as seen in Fig. 3a.

The heatmaps indicate that the correlations of the time features with the mean hourly SOC are accounting for less than 8%. A stronger correlation can be retrieved from the market and frequency data. For example, up- and down-regulation volumes have about 10% correlation with the SOC while for the automatic activated reserve volumes this metric is reaching 20%. A non-linear conversion of the frequency measurements led to 92% correlation with the SOC data. The lowest correlation value among the market features is for the wind data that is slightly exceeding 3%. The artificially created frequency data features have the correlation varying from 60% to 80%. The maximal correlation for the modified SOC data is achieved by the difference values and reaches almost 70% while the correlation of the shifted values is declining from 22% along with the difference in time.

C. Feature Selection

Feature selection is a part of machine learning model engineering that incorporates finding a subset of relevant and informative features from the initial dataset. The benefits of proper selection can decrease training time, minimize the risk of overfitting, potentially improve model performance, and reduce model complexity [23]. A review of the state-of-art methods for the feature selection can be found in [24]. In this study, two types of feature selection methods were applied to the dataset that are filter and embedded methods.
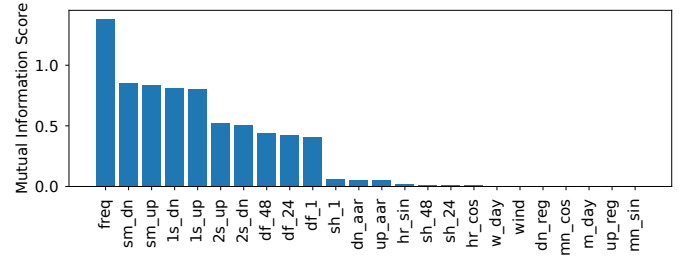
(a) Filter method
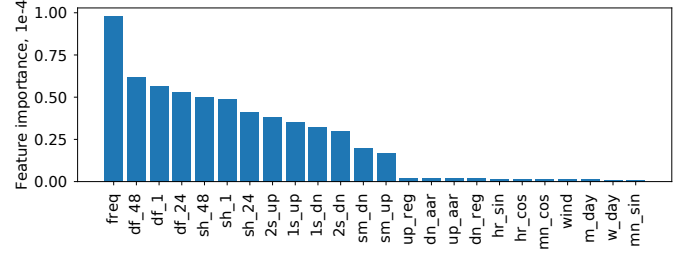
(b) Embedded method

Fig. 4: The results of feature selection for BESS SOC dataset: (a) filter method and (b) embedded method.

For the filter method that enables model-independent feature selection, Mutual information (MI) score is utilized. This score evaluates the linear and non-linear dependence of the target variable of each feature by quantifying the amount of information that can be obtained about the target observing the feature. As a result, a high MI score indicates dependence between the target and the feature. According to the MI scores presented in Fig. 4a, the most important features for the SOC data are the frequency measurements, its derivatives such as sm_dn, sm_up, and modified SOC data features such as df_48, df_24, df_1. However, since the highly correlated frequency derivative features are redundant, only sm_dn and sm_up features can be be kept without loss of information. In contrast to SOC and frequency data, the market and time features have low importance with only up_aar and dn_aar accounting for distinguishable MI score values.

In the embedded methods, feature selection is integrated into the model algorithms at the training stage. Here, LightGBM was used as the model algorithm to perform the feature selection. The feature importance metric according to the embedded method is presented in Fig. 4b. The results are correlated with the MI scores and highlight the importance of the frequency measurement, its derivatives, and modified SOC data features.

Thus, the final set of features utilized in the experiment included frequency mean values, all modified SOC data features, sm_up and sm_dn features from the frequency data, and automatically activated reserves from the market data.

D. Training, Validation, and Testing Sets

The input data for the experiments were obtained from the BESS model simulation. The data were consisted of a

15-minute BESS SOC data in percent of a nominal battery capacity and corresponding frequency data features. For the hour- and day-ahead predictions, these data were transformed to an hourly resolution and enriched by the selected features from the hourly BESS SOC dataset presented in the previous subsection. For the 15-minute forecasts, the hourly market features were resampled to 15 minute resolution using forward filling, and the SOC shifting and difference procedures were applied with a coefficient factor equal to 4. The target variable for the forecasts was the difference between the consecutive hours (df_1 and df_4 for the hourly and 15-minute datasets, respectively) to prevent the persistence model relying only on the value of the previous step in its predictions. The distinction between the autocorrelation and partial autocorrelation before and after the removal of hourly correlation can be seen in Fig. 3. Moreover, MinMax scaling with the range from 0 to 1 was utilized for the datasets.

The testing was done using hold-out validation, in which the BESS SOC datasets were split for training and testing sets in proportions of 75% and 25%. Also, 15% of the training set was reserved for a parameter validation during the testing of the MARNN and DCNN models.

E. Performance Index

The error of the point forecast was estimated by the deviation between the actual observation and the prediction $\varepsilon_t = y_t - f(x_t)$. In this study, the widely used Root Mean Squared Error (RMSE) error metric was employed to quantify the statistical quality of the forecasts for the testing data points:

$$\text{RMSE} \;=\; \sqrt{\frac{1}{N}\sum_{t=1}^{N}\varepsilon_t^2}. \tag{2}$$

F. Implementation Details

The data-driven models were implemented using Keras 2.0.2 high-level neural networks API [25] with Tensorflow 1.0.1 [26] as the backend in the Python 3.6 environment. The DCNN was based on the model presented in [27], and the MARNN model was developed based on [28] with the Monte Carlo dropout added to the model as a regularization method with the astroNN package [29]. The random forest and decision tree models were implemented with the Scikit-learn library [30], and the LightGBM model was built based on its library [31].

IV. Experimental Results

The experimental results are presented in Table B.I and Fig. B.1 of Appendix B. According to the performance evaluation, the MARNN model achieved the best score in point forecasting of BESS SOC for the hour-ahead and day-ahead time scales while DCNN demonstrated the lowest loss for the 15-minute-ahead forecast. However, the difference between the performance of the RF, LightGBM, DCNN, and MARNN models in hour-ahead forecasting is relatively small and can be neglected.

The efficiency of tree based methods has an expected distribution from DT to RF and LightGBM in the descending order of their loss score, and this disposition is in correlation with known simulation results. The data-driven models have noticeable improvement in particular time horizons if compared to each other. These time scales are 15-minute-ahead and day-ahead for the DCNN model and the MARNN model, respectively. This forecast quality can be explained by their structures, in which the input of the DCNN model is a sequence of consecutive values from the past, while for the MARNN model, it is lag sample values from several previous input sequences. Thus, the DCNN model has more restricted scope that can be beneficial for the short-term forecasts but the MARNN model can leverage lag sample values to predict long-term dependencies.

The heatmaps of prediction and actual values in Fig. B.1 show forecast properties of the models. For example, the RF, LightGBM, and DCNN models were unable to catch the BESS SOC patterns in day-ahead forecasting, and this is illustrated in Fig. B.1 by a low amplitude of their predictions. Better performance of the MARNN, DCNN, RF, and LightGBM models on hour-ahead time scale compared with the DT model is demonstrated by a location of their predictions closer to the imaginary diagonal line that corresponds to an absolutely accurate prediction. In general, the forecast distributions of BESS SOC imply the presents of errors that together with the drawbacks of point forecasts makes the predictions too uncertain and ambiguous to utilize them in decision making.

V. Conclusions

In this study, a comparison between the several traditional and state-of-the-art machine learning algorithms was conducted for the forecasting of BESS SOC in the northern Europe synchronous area. The performance evaluation of the algorithms was carried out on 15-minute-ahead, hour-ahead, and day-ahead time scales. The results demonstrate that the MARNN model outperforms DT, RF, LightGBM, and DCNN on the day-ahead scale, DCNN has better loss for the 15-minute-ahead forecasting, while comparable results can be obtained for the hour-ahead forecasting by RF, LightGBM, DCNN, and MARNN models.

Moreover, cross-dependency of BESS SOC data with time, market, and frequency data features was analyzed, and the most important features were chosen with the filter and embedded feature selection methods. The findings suggest that the BESS SOC data has low correlation with the market and time features while frequency features as well as applying shifting and difference to the SOC data can be more beneficial to support the forecasting.

Forecasting of BESS SOC can potentially provide an efficient and meaningful tool for decision making applications in multi-objective smart grid environment. However, taking into account inevitable errors of point forecasts related to stochastic processes such as the frequency regulation, probabilistic forecasts can be more informative and should be investigated for such applications in the future research.

REFERENCES

[1] E. Union, "Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources," *Official Journal of the European Union*, vol. 5, pp. 82–209, 2018.

[2] A. Ulbig, T. S. Borsche, and G. Andersson, "Impact of low rotational inertia on power system stability and operation," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 7290–7297, 2014.

[3] Z. A. Obaid, L. M. Cipcigan, L. Abrahim, and M. T. Muhssin, "Frequency control of future power systems: reviewing and evaluating challenges and new control methods," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 1, pp. 9–25, 2019.

[4] "Supercharged: Challenges and opportunities in global battery storage markets," Report, The Deloitte Center for Energy Solutions, March 2018.

[5] R. H. Byrne, T. A. Nguyen, D. A. Copp, B. R. Chalamala, and I. Gyuk, "Energy management and optimization methods for grid energy storage systems," *IEEE Access*, vol. 6, pp. 13 231–13 260, 2017.

[6] N. Belonogova, V. Tikka, S. Honkapuro, J. Lassila, J. Haakana, A. Lana, A. Romanenko, J. Haapaniemi, A. Narayanan, T. Kaipia *et al.*, "Multi-objective role of battery energy storages in an energy system," *LUT Scientific and Expertise Publications/Tutkimusraportit–Research Reports*, 2018.

[7] L. Breiman, *Classification and regression trees*. Routledge, 2017.

[8] ——, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[9] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, 2017, pp. 3146–3154.

[10] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner *et al.*, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[11] A. Borovykh, S. Bohte, and C. W. Oosterlee, "Conditional time series forecasting with convolutional neural networks," *arXiv preprint arXiv:1703.04691*, 2017.

[12] A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio." *SSW*, vol. 125, 2016.

[13] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[14] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[15] A. Mashlakov, V. Tikka, L. Lensu, A. Romanenko, and S. Honkapuro, "Hyper-parameter optimization of multi-attention recurrent neural net-work for battery state-of-charge forecasting," in *EPIA Conference on Artificial Intelligence*. Springer, 2019, in press.

[16] A. Mashlakov, L. Lensu, A. Kaarna, V. Tikka, and S. Honkapuro, "Probabilistic forecasting of battery energy storage state-of-charge under primary frequency control," Manuscript submitted for publication.

[17] "Terms and conditions for providers of Frequency Containment Reserves (FCR)," https://www.fingrid. fi/globalassets/dokumentit/en/electricity-market/reserves/ appendix2---terms-and-conditions-for-providers-of-fcr.pdf/, 2018.

[18] A. Mashlakov, "BESS SOC forecasting," https://github.com/ aleksei-mashlakov/BESS-SOC-forecasting/, 2019.

[19] Fingrid, "Frequency - historical data," 2019, data retrieved from Fingrid, https://data.fingrid.fi/en/dataset/frequency-historical-data.

[20] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.

[21] B. Schäfer, C. Beck, K. Aihara, D. Witthaut, and M. Timme, "Non-gaussian power grid frequency fluctuations characterized by lévy-stable laws and superstatistics," *Nature Energy*, vol. 3, no. 2, p. 119, 2018.

[22] Nordpool, "Historical market data." 2019–, [Online; accessed 15-06-2019]. [Online]. Available: https://www.nordpoolgroup.com/ Market-data1/#/nordic/table/

[23] I. Koprinska, M. Rana, and V. G. Agelidis, "Correlation and instance based feature selection for electricity load forecasting," *Knowledge-Based Systems*, vol. 82, pp. 29–40, 2015.

[24] J. C. Ang, A. Mirzal, H. Haron, and H. N. A. Hamed, "Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 13, no. 5, pp. 971–989, 2015.

[25] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[26] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016, pp. 265–283.

[27] J. Eddy, "Timeseries seq2seq," 2019–, [Online; accessed 15-06-2019]. [Online]. Available: https://github.com/JEddy92/TimeSeries_Seq2Seq/

[28] M. Ratsimbazafy, "Mckinsey smartcities traffic prediction," https:// github.com/mratsim/McKinsey-SmartCities-Traffic-Prediction, 2018.

[29] H. Leung, "AstroNN," https://github.com/henrysky/astroNN/tree/master/ astroNN, 2019.

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

[31] LightGBM, "Light gradient boosting machine." 2019–, [Online; accessed 15-06-2019]. [Online]. Available: https://github.com/microsoft/ LightGBM/
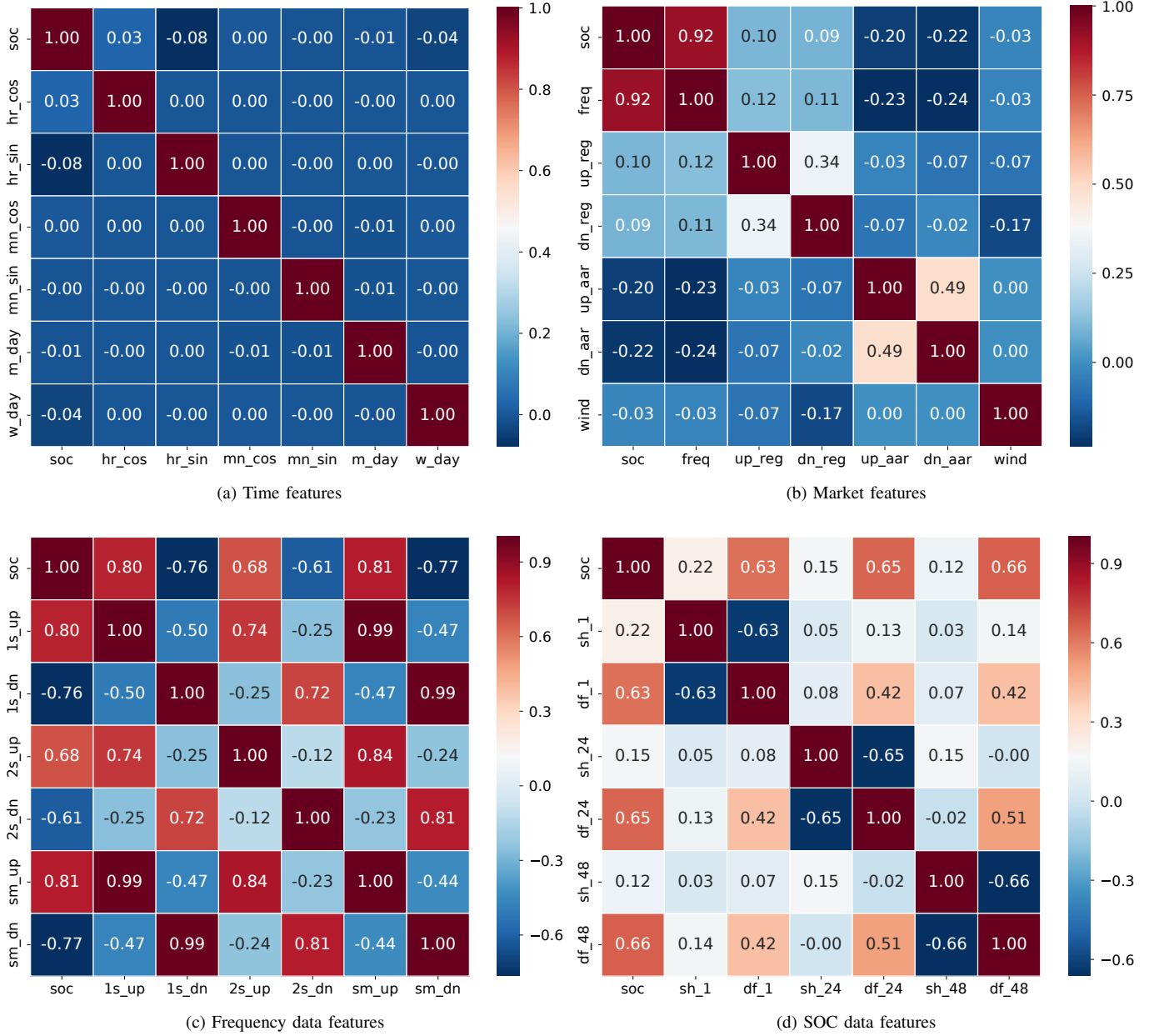
Fig. A.1: Correlation of simulated hourly state-of-charge (SOC) values in northern Europe synchronous area (2015 to 2018): (a) time, (b) market, (c) frequency data, and (d) modified SOC data features.

TABLE B.I: Performance evaluation of the point forecasts.

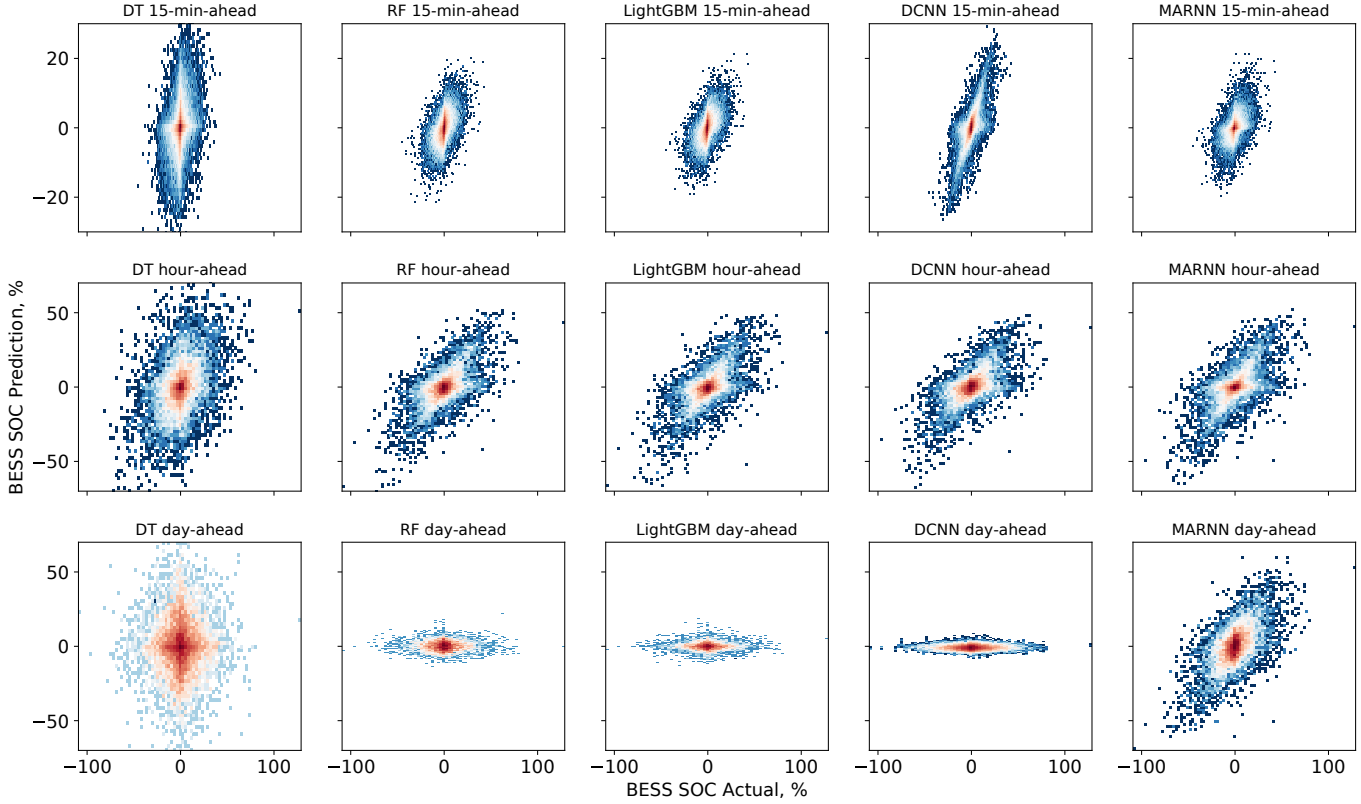| Timescale	Metric	DT	RF	LightGBM	DCNN	MARNN
15-min-ahead	RMSE, [%]	7.35	5.17	5.07	**3.98**[a]	5.18
Hour-ahead	RMSE, [%]	19.30	13.06	12.95	13.01	**12.93**
Day-ahead	RMSE, [%]	25.51	17.43	17.26	17.12	**13.80**

[a]The best RMSE loss is marked in bold.

Fig. B.1: Performance of a decision tree (DT), random forest (RF), Light Gradient Boosting Machine (LightGBM), dilated convolutional neural network (DCNN), and multi-attention recurrent neural network (MARNN) on day-, hour-, and 15-minute-ahead time scales of the BESS SOC testing set.