Forecasting Power Grid States for Regional Energy Markets with Deep Neural Networks

1st Yujiang He
Intelligent Embedded System Group
University of Kassel
Kassel, Germany
yujiang.he@uni-kassel.de

2nd Janosch Henze
Intelligent Embedded System Group
University of Kassel
Kassel, Germany
jhenze@uni-kassel.de

3rd Bernhard Sick
Intelligent Embedded System Group
University of Kassel
Kassel, Germany
bsick@uni-kassel.de

***Abstract*—An increasing amount of volatile infrastructure, such as solar power plants, wind power plants, battery storages, or electric vehicles, will be added to the distribution grid. Such volatile infrastructure adds strain to the existing infrastructure and may lead to high transmission load on lines and corrupt power grid states. One way to resolve these issues is to establish regional energy markets. Such energy markets try to coordinate energy consumption or production regionally, based on estimates of future power grid states, thus reducing stress on the power grid. For these markets, it is crucial to know possible threatening grid states in advance to allow the market operators to keep the power grid in a healthy grid state with the help of regionally consumed or produced energy. In this work, we propose three different approaches to forecast future grid states with deep neural networks based on a learned numerical weather prediction representation. We evaluate these approaches with experimental results and discuss the remaining challenges that have to be addressed to enable regional energy markets. Furthermore, we provide the dataset to allow other researchers to reimplement our work.**

***Index Terms*—smart grids, deep learning, forecasting, renewable energy, grid states**

I. Introduction

An increasing amount of smart and renewable infrastructure is connected to the power grids. Renewable power plants, such as solar power plants or wind power plants, are producing their energy depending on volatile weather conditions. Other smart infrastructure, including energy storage, electric vehicles, or cogeneration power plants, can plan their consumption or production based on external incentives. Such incentives are often economic, e.g., the owner wants to sell energy during periods when energy prices are high, and store energy in periods the energy prices are low. Volatility in the energy source can lead to an overload of the power lines within the power grid, and therefore, can usually be seen as a regional risk for the energy supply chain. Currently, in Germany, the grid operators try to overcome this risk by shutting down individual renewable energy power plants, requiring high compensation payments to the power plant owner. Another smart way of integrating infrastructure and volatile energy producers is currently under research in Germany within the c/sells project [1] as part of the sinteg research initiative [2]. The main idea of the research project is to align the production and consumption of energy on a regional level by building a regional energy market.

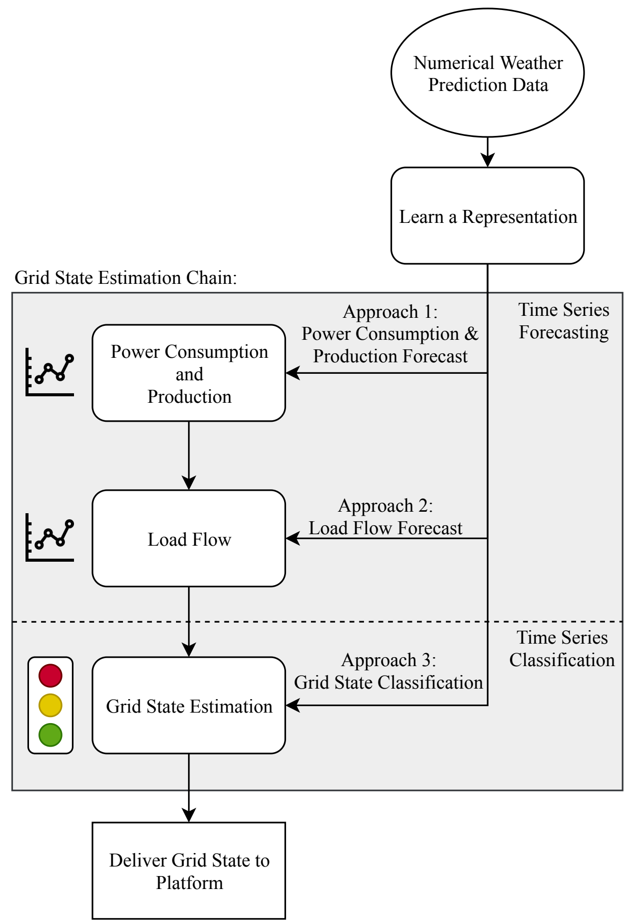


Fig. 1. This flow chart shows the three proposed approaches as well as the grid state estimation chain. Each forecast task replaces all previous tasks in the grid estimation chain.

Regional energy markets connect regional generation and consumption sites. Such spatial links between generation and consumption exist in cities with a lot of roof-top photovoltaics, battery storages, electric vehicles, or close-by wind power

This work was supported within the c/sells RegioFlexMarkt Nordhessen (03SIN119) project, funded by BMWi: Deutsches Bundesministerium für Wirtschaft und Energie/German Federal Ministry for Economic Affairs and Energy.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

plants. Usually, production as well as consumption have to coincide in time and geographically, to reduce costs due to transportation. An often discussed variant of regional energy markets is a system-as-a-service variant, as described in [3]. At these so-called regional flexibility markets, the main idea is to trade grid supporting flexibility options. Such flexibility options are implemented in a positive way, i.e., increased generation or reduced load, or in a negative way, i.e., reduced generation or increased load. The overall goal is to use these flexibility options to stabilize the local power grid at the location of the critical network state, which is in contrast to how power grids are operated in Germany today.

To be able to trade flexibility options on time, the markets need precise forecasts of power grid states. In Germany, the grid states are symbolized with a traffic light system [4]. The three phases are green, yellow, and red. The green phase indicates no problems in the power grid. The yellow phase indicates that some parts of the power grid infrastructure are in an alarm situation, and the balance between consumption and production could be corrupted soon. The red phase indicates an existing problem in the grid, and immediate actions need to be taken. Forecasting the power grid states needs to take the current as well as possible future generation and production into account. It creates a significant real-world challenge today, as in the future the amount of renewable energy sources is increasing.

Therefore, we propose three different approaches to forecast the grid states, as shown in Fig. 1.

1) The first approach forecasts energy production and consumption and then calculates the load percentage of each transmission line to determine the state of the power grid.
2) The second approach forecasts the results of the load flow calculation to estimate the grid states.
3) The third approach forecasts the grid states directly.

Every forecasting or classification task replaces the current and the preceding elements in the grid state estimation chain. Therefore, in each approach the whole network has to abstract more information about the underlying power grid. All the approaches use the same learned representation of numerical weather prediction (NWP) data. This representation is the input to a long short term memory (LSTM) network. These deep learning techniques are combined with traditional power grid operation tools for power system modeling and analysis. We detail the advantages and disadvantages of each approach, techniques to overcome these disadvantages, and an experiment showing the feasibility of the solution.

Our contributions are summarized as below:

- We propose three approaches to estimate the grid states based on a real-world power grid dataset.
- We design and implement three experiments corresponding to these approaches.
- We evaluate the results of the experiments and outline the existing challenges for implementing the regional smart energy market successfully.
- We publish the real-world power grid dataset on which we conduct the experiments.

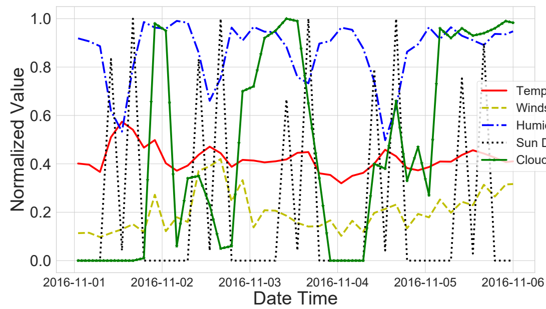
The remainder of the article starts with an overview of related work on power time series forecasting in Section II. Afterward, we give an insight into our dataset, including what it contains and how it is preprocessed, and also outline three approaches, the methods, and the experiments to forecast the grid states in Section III. The results of the experiments are evaluated and discussed in Section IV. In Section V, we conclude the article by outlining challenges in forecasting and infrastructure techniques that need to be solved to implement regional energy markets successfully.

II. Related Work

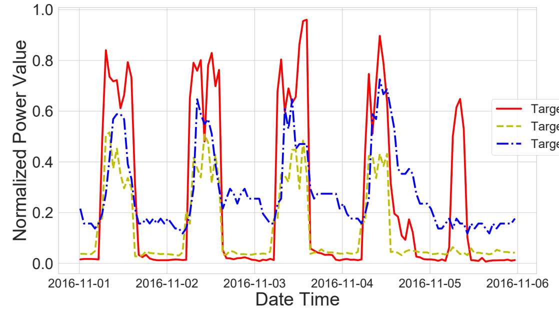
Forecasting the power of regenerative energy sources is not a new research topic. An excellent survey of prior work in solar power forecasting can be found in [5]. Other overview papers also cover the forecasting of other types of regenerative power plants [6]–[8]. One possible way to categorize different approaches is to sort them by their forecasting horizon. The area of short-term forecasting is covered in [7], [9], mid-term forecasting is described in [10], long-term forecasting techniques are described in [11]. In recent years, artificial neural network (ANN) had regained attention in research when deep learning emerged [12]. Research in deep learning focuses on different network structures, e.g., auto-encoders, deep belief networks, and LSTM, for tasks such as data encoding, information extraction, or forecasting [13]–[15]. Auto-encoders and deep belief networks are used to learn representations of the data [12], [13], [15]. LSTM networks are a subcategory of recurrent neural networks and use additional memory cells to store hidden states. Hence, they have shown to be compelling for time series prediction tasks [16], [17]. Deep learning architectures have recently been used to forecast regenerative energy sources. For example, deep belief networks are used in [18] to predict wind power, and stacked auto-encoders are used in [19] to predict short-term wind speed. The creation of forecasting ensembles can improve the overall forecasting quality. An overview of various ensemble techniques is shown in [20], a hybrid ensemble technique based on the gradual weighting of weather situations is described in [21].

III. Forecasting Grid States

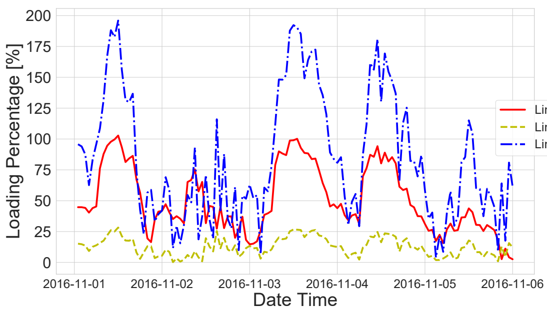
For a forecasting process, we use data of NWP models as an input. NWP models are mathematical models of the atmosphere and oceans to predict current and future weather conditions. We use deep neural networks to map NWP data to forecast target variables within the specified period. For example, in our experiments the forecast period is 6 hours into the future, NWP data from time $t + 0$h to $t + 6$h are used to forecast the output of renewable energy plants, or the states of the transmission cables from time $t + 0$h to $t + 6$h. Additional temporal features are selected as inputs, i.e., month of the year, day of the year, week of the year, day of the week, and hour of the day.



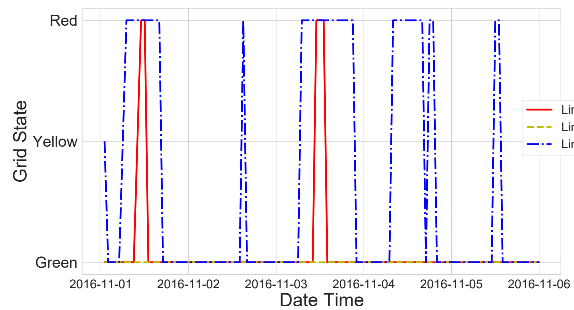
(a) NWP time series



(b) Power time series



(c) Load flow time series



(d) Grid state time series

Fig. 2. Examples of input and output time series of the three experiments, including NWP time series (input), power time series (output), load flow time series (output), and power grid state time series (output).

A. *Problem statement*

The state of the power grid depends on the capacity of each component in the power grid. Load flow calculations are used to determine these capacities. Typically, the inputs to the load flow calculation are values of the produced or consumed power at each node in the power grid. Each power line has a maximum capacity which the line can transfer before it gets damaged due to heat or too high currents. According to the load percentage on the lines, we determine the power grid states based on the information released by the German Association of Energy and Water Industries (BDEW) in [4]. As required, the grid states are defined as a traffic light system with similar thresholds as the BDEW provides:

- Green phase: $0\% \ldots 95\%$ of max. load percentage,
- Yellow phase: $95\% \ldots 98\%$ of max. load percentage,
- Red phase: over 98% of max. load percentage.

A typical workflow to determine the power grid state is shown in Fig. 1. By calculating the load flow, the power grid operators get an overview of the current state of the power grid. Fig. 2 displays the few examples of input and output time series, including 5 NWP features in Fig. 2(a), 3 power consumers in Fig. 2(b), loading percentage of 3 transmission lines in Fig. 2(c), and the corresponding grid states of the 3 lines in Fig. 2(d). To be able to forecast the power grid state, we can forecast each of the outputs of three blocks through the NWP data, seen in Fig. 1:

1) energy production and consumption,
2) load flow results, or
3) grid state estimation.

Approach I: Forecasting energy production and consumption is a straight forward solution to obtain information about the future power grid states. As mentioned in Section II, research and applications for this approach are vast. The main challenge is to forecast all production and consumption nodes precisely. The forecasts are used to calculate load flow results. Finally, the grid state can be determined using the definition described above.

Approach II: Forecasting the load flow results is a challenge similar to forecasting the energy production and consumption. Each line, transformer, and each utility that transmits power needs to be forecasted, potentially increasing the number of individual forecasts. In addition to the increased quantity of forecasts, we need to add load flow calculations to preprocess our data. According to the forecast results, the grid states can be directly determined under the definition.

Approach III: Forecasting the states of the power grid is as challenging as forecasting the load flow results. Using the traffic light system reduces the problem to a classification problem rather than a regression of a time series.

B. *Description of Dataset*

Our dataset consists of one year of data for a regional power grid in central Germany. It contains information about:

- NWP data, as features,
- power data (generation and consumption), as targets.

The raw NWP dataset contains the historic regional NWP data for one year in a six-hour temporal resolution. Twenty-seven weather features are available at each time step and the corresponding 180-hour forecast values starting from the measurement time point in a 3-hour resolution. Utilizing shifting and interpolating, we obtain NWP data in a 15-minute resolution, which is the same temporal resolution as the power data. The data is normalized between 0 and 1 using min-max normalization.

The power data includes the power generation of 11 renewable power generators, the sum of the power consumption in a residential area (44 consumers), and the power consumption of 45 local consumers for one year in a 15-minute resolution. The 11 renewable power generators consist of 3 wind parks, 6 photovoltaic generators, and 2 combined heat and power plants. The power data are the targets for the forecast task. The raw dataset, including NWP data, the power consumption data, power generation data, as well as the power grid data, is available through our homepage (www.ies.uni-kassel.de → Downloads).

The produced power of each generator is normalized using the rated capacity of the corresponding facility. The power consumption measurements are normalized between 0 and 1 using min-max normalization. The temporal features are sine cosine normalized, resulting in 10 encoded features.

In the load forecast task (Approach I), the sum of the power consumption in a residential area is referred to as the target. In Approach II and III, the values averaged by 44 residences are used to calculate the load flow and to estimate the states of the power grid. Additionally, in Approach II and III, the topology data of the power grid is essential for the load flow calculation and the estimation of the power grid states. The data describes a regional power grid, i.e., information about consumers and producers, topological information about the power grid, and connection points to higher and lower power grid levels.

C. Splitting Datasets

The dataset contains 52 weeks (354 days in all) of measurements. Due to the limited available data, traditional linear splitting, i.e., splitting the dataset into training and test dataset with a known percentage, is not applicable in the context of time series forecast. In detail, the available one-year dataset usually could be split into 2 sections, the training dataset, including the first 9 months (could cover 3 seasons), and the test dataset, including the last 3 months (could cover only one season). On the one hand, the differences in seasons result in various effects on energy consumption, which cannot easily be abstracted from the available features in the training data. On the other hand, the existing measurement data reveals that some power generators can be enabled or disabled by an operator. For example, a generator was enabled less than three months in the one-year measurement data, as shown in Fig. 3. The traditional dataset split could not report the true data distribution of the whole dataset.

A better way to split the dataset for forecasting a time series is to split the dataset into 52 sections, each split covers one week. From each week, five days are selected for training, one day for validation, and one day for testing, as shown in Fig. 4. Rolling the sequence could balance the data distribution in all three subsets.

In our work, we choose a model named Auto-LSTM as the forecast neural network, which combines an auto-encoder and an LSTM network, as proposed by [14]. An auto-encoder is a multi-layer symmetrical deep neural network with a small bottleneck layer, as shown in Fig. 5. The representations of the sophisticated high-dimensional input features, i.e., the NWP data, are extracted and compressed in the encoder and reconstructed in the decoder. Generally, all parameters of the auto-encoder are fine-tuned during the training of the auto-encoder. During the training for the forecasting task, the encoder is frozen, and the decoder is not used. The representation of the NWP data is concatenated with the sine cosine encoded temporal features. We choose to concatenate the temporal data as previous work [15] has shown that auto-encoders

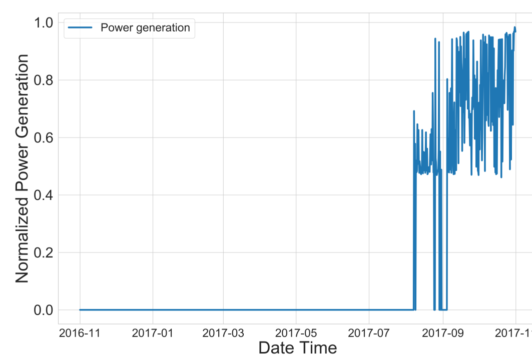


Fig. 3. The measurement of a power generator from November 2016 to October 2017. The generator was enabled only from August to October in 2017.

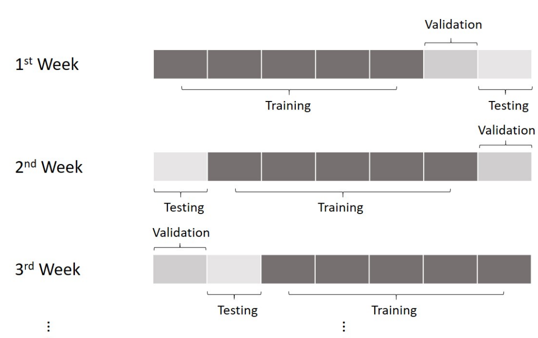


Fig. 4. An illustration of splitting the dataset into training, validation, and testing dataset.

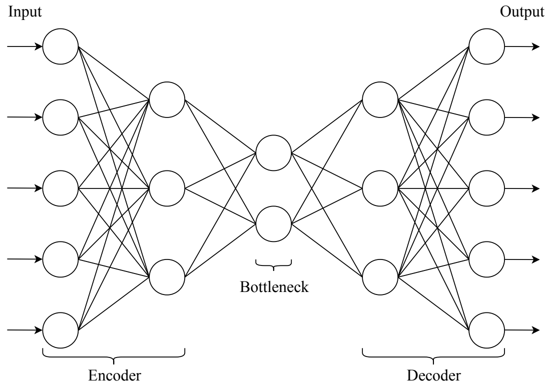


Fig. 5. An illustration of an Auto-Encoder with the different components: input and output layer, encoder and decoder part, and the bottleneck.

TABLE I
ARCHITECTURE PARAMETERS OF THE AUTOENCODER. DROPOUT LAYERS ARE USED IN THE AUTOENCODER (FROM ENCODER 1 TO ENCODER 4) WITH A DROPOUT RATE OF 0.1. THE SLOPE PARAMETER OF THE LEAKYRELU IS SET TO 0.3.

Layer	Size	Activation
Input	1×27	None
Encoder 1	27×19	Leaky ReLu
Encoder 2	19×15	Leaky ReLu
Encoder 3	15×12	Leaky ReLu
Encoder 4	12×8	Leaky ReLu
Decoder 1	8×12	Leaky ReLu
Decoder 2	12×15	Leaky ReLu
Decoder 3	15×19	Leaky ReLu
Output	19×27	None

perform worse when trying to learn temporal features. The combination of the learned representation and the temporal features is used as the input for an LSTM network. The LSTM setup, which is most commonly used, was initially described in [22] and has been one of the state-of-the-art models for solving time series problems. A standard LSTM unit consists of three gates and one cell. Compared to other multi-layer deep neural networks, the gates in the LSTM regulate the information flow into and out of the cell more efficiently, where memories can be stored over time. The main challenge is to use the learned NWP representation, which is the same for all 3 approaches, to forecast the abstract power-grid-related targets. In each approach, the LSTM network needs to abstract more knowledge about the underlying power grid. Therefore, with each step down the power grid state estimation chain, the forecast gets more complicated.

In order to put more focus on the comparison of the three approaches, the same auto-encoder structures and preprocessing methods are used in the three experiments. Parameters are selected using a grid search based on the smallest validation error. Different forecast targets are corresponding to different parameters of LSTM. The parameter selection of the auto-encoder and the grid search table for LSTM are listed in Tables I and II. The batch size is 32, and Adam is chosen as the optimizer. The training process will be stopped by early stopping if the training error does not drop within 50 iterations.

TABLE II
GRID SEARCH TABLE FOR TUNING THE HYPERPARAMETER OF LSTM NETWORK. THE SLOPE PARAMETER OF LEAKY RELU IS SET TO 0.3.

Parameter	Values
Layers of LSTM	1, 2
Activation function	Sigmoid, Tanh, Relu, Leaky relu
Dropout rate	0.1, 0.2, 0.3, 0.4
Batch size	32
Optimizer	Adam

D. Experimental Setup

The final goal of the experiments is to estimate the states of the given power grid using a 6-hour forecast horizon. This horizon is chosen in accordance with the requirements of our work within the c/sells project. Our neural network predicts the states from time $t{+}0$h to $t{+}6$h with a resolution of 1 hour. For each of the three approaches, we propose one experiment:

Experiment I: In Experiment I, the power generation or consumption of all known power plants is forecasted using the Auto-LSTM network. Afterward, the load flow is calculated, with the help of Pandapower [23] and the available power grid data. Finally, from the load flow results, the grid state is estimated.

Experiment II: In Experiment II, the Auto-LSTM network aims at forecasting the results of the load flow calculation, which we mainly refer to as line capacity. There are 372 lines in the existing power grid. One neural network model is trained for each line. The state for each line is estimated based on the results of load flow forecasting.

Experiment III: Experiment III is a multi-class classification experiment, where we forecast the power grid state of each line directly. The Auto-LSTM is extended with a classification layer and trained using a cross-entropy loss function. Similar to Experiment II, there are 372 lines as targets, and one model is trained for each target. For each line, three classes exist, namely, green, yellow, and red, which are defined in Section III-A.

IV. RESULTS

In this section, we evaluate the results of the three experiments. First, we describe our evaluation metrics before applying them to each experiment.

A. Evaluation Metrics

In the article, the forecasting is evaluated using the NRMSE metric, and the classification tasks are evaluated using an F_1 score. The root-mean-squared error (RMSE) is a common metric, which is defined as in (1):

$$\mathrm{RMSE}(\hat{\mathbf{x}}(t), \mathbf{x}(t)) = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (\hat{x}_i(t) - x_i(t))^2} \quad (1)$$

where N represents the numer of samples, $\hat{\mathbf{x}}$ is the predicted time series, $\mathbf{x}$ is the measured time series. $\hat{x}_i$ and x_i are the i_{th} sample at time step t, and t ranges from 0 to 6 in the context. The normalized-root-mean-squared error (NRMSE) is the RMSE divided by the difference of the maximum $x_{\max}$

and minimum $x_{\min}$ of the measured time series at time step t, defined as in (2):

$$\text{NRMSE}(\hat{\mathbf{x}}(t), \mathbf{x}(t)) = \frac{\text{RMSE}(\hat{\mathbf{x}}(t), \mathbf{x}(t))}{\mathrm{x}(t)_{max} - \mathrm{x}(t)_{min}} \quad (2)$$

The value range differs in the different target datatsets, i.e., generators, consumers, or transmission lines. The normalized RMSE simplifies the comparison of the results from the datasets with different scales.

All experiments are also evaluated based on their accuracy when forecasting the state of the power grid using the F_β-Score with $\beta = 1$:

$$\text{F}_\beta(t) = \frac{(1+\beta^2)\cdot \text{TP}(t)}{(1+\beta^2)\cdot \text{TP}(t) + (\beta^2)\cdot \text{FN}(t) + \text{FP}(t)} \quad (3)$$

In (3), true positive (TP) is a correctly predicted grid state, false positive (FP) is when a certain grid state is predicted even if it is not present, and false negative (FN) is a grid state that is present but has not been predicted. Here, t indicates the time step, same as above. The F_1 score is the harmonic average of precision and recall. An F_1 score reaches its best result at 1 and the worst at 0.

B. Experimental Results

Experiment I: In Experiment I, the Auto-LSTM networks forecast the power generation and consumption. The forecast results for different types of targets are scored using NRMSE and averaged by the number of power plants, as shown in Fig. 6 and Table III. Forecasting the photovoltaic generators yields the lowest NRMSE and standard deviation. The prediction results of combined heat and power plants (HPs) and consumers (Loads) have a higher standard deviation. This higher standard deviation indicates that the targets are affected not only by weather conditions but also substantially by economic or human factors, which are hardly predictable based on the existing dataset.

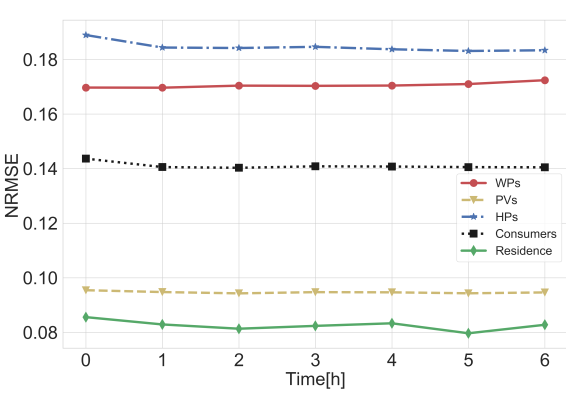


Fig. 6. Experiment I: short-term (0-6h) power forecasting results of NRMSE averaged over 3 wind parks (WPs), 6 photovoltaic generators (PVs), 2 combined heat and power plants (HPs), 45 consumers (Loads), and 1 residential area (44 consumers in all).

With the help of the power grid analysis tool pandapower [23], the load flow results for the transmission lines are calculated based on the power forecasts. The load flow results allow for an estimation of the power grid states. Fig. 7 presents the F_1 score for the three power grid states in this experiment. The definition of the traffic light system in Section III-A leads to an imbalance problem in the grid states. The minority class, i.e., yellow state, which makes up only about 0.4% of the available samples, is the most difficult to predict.

Experiment II: In Experiment II, the forecast targets are the load flow calculations on the 372 transmission lines. In Table IV, we present the RMSE of the load flow result forecasting from Experiment I and II. The table shows that using the Auto-LSTM model to predict the load flow calculations directly (Experiment II) is a more reliable method in comparison to Experiment I. However, the 0.01 improvement is not significant for forecasting power grid states. In Fig. 8 we can observe that when estimating the power grid states, the minority classes, i.e., yellow and red states, are not detected anymore at the +0h and +1h forecast horizon in the results of Experiment II. Even the lower error could still result in misclassification due to the strict definition of the traffic light system. Similarly, to Experiment II, the yellow power grid state is hardly predictable as well.

TABLE III
EXPERIMENT I: COMPARISON OF THE FORECAST ERROR (NRMSE) OF THE POWER GENERATION AND CONSUMPTION

	WPs	PVs	HPs	Loads	Residential
Mean	0.17	0.09	0.18	0.14	0.08
Std	±0.028	±0.004	±0.095	±0.064	–

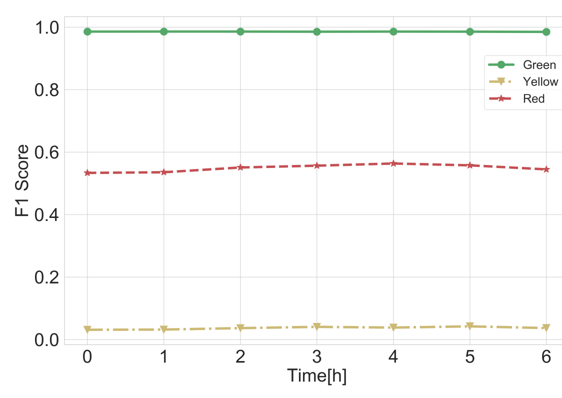


Fig. 7. Experiment I: F_1 score of short-term (0-6h) grid state forecast.

TABLE IV
COMPARISON OF THE FORECAST ERROR (RMSE) OF THE LOAD FLOW RESULTS BETWEEN EXPERIMENTS I AND II

	Mean RMSE (Std)
Experiment I	0.129 (±0.117)
Experiment II	0.119 (±0.110)

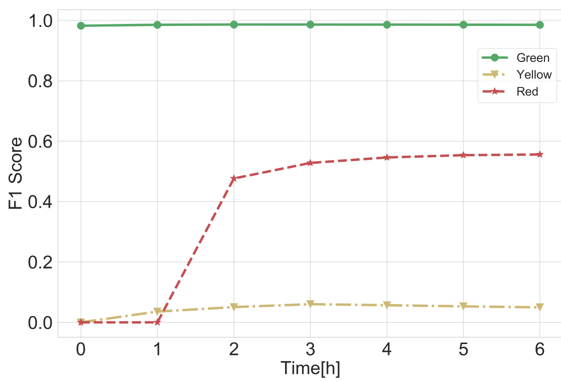


Fig. 8. Experiment II: F_1 score of short-term (0-6h) grid state forecast.

Experiment III: As mentioned earlier, during the third experiment, the analyzed problem transforms into a classification problem. The classification labels are the three power grid states for the 372 transmission lines. Fig. 9 again shows the unpredictability of the yellow power grid state, due to data imbalance.

Table V shows a comparison of all F_1 scores of the three experiments. First of all, all three approaches can predict the green power grid state precisely but fail to predict the yellow power grid state. This can mainly be accounted to the data imbalance problem caused by the strict definition of the traffic light system. In Experiment I, the entire information about the power grid and the involved energy plants is essential. It could lead to a massive challenge for dataset collection. In comparison, the power grid could is a black box in Approach II and III, where the internal structure of the power grid and the complicated electrical computation are skippable, and we can put more focus on tuning the neural network parameters. Moreover, using Experiment III, a higher F_1 score and lower standard deviation for green and red power grid states can be achieved. However, the yellow power state has been ignored

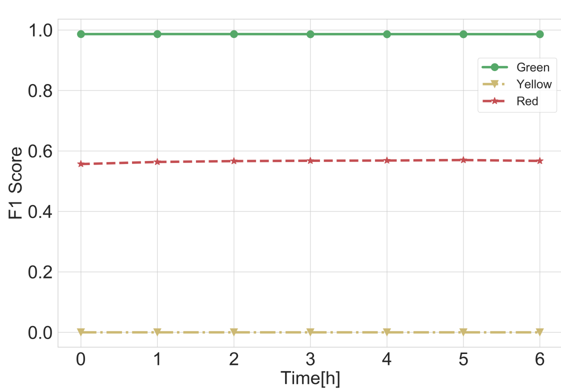


Fig. 9. Experiment III: F_1 score of short-term (0-6h) grid state forecast.

TABLE V
COMPARISON OF F_1 SCORE OF THE GRID STATE FORECAST FOR THE 3 EXPERIMENTS

	Experiment I	Experiment II	Experiment III
Green	0.9853 ± 0.0004	0.9852 ± 0.0013	0.9862 ± 0.0002
Yellow	0.0367 ± 0.0038	0.0437 ± 0.0192	0.0000 ± 0.0000
Red	0.5487 ± 0.0106	0.3800 ± 0.2416	0.5654 ± 0.0041

due to the class imbalance. Strategies for solving the imbalance problem have to be applied directly during the predicting of the power grid states in future research. For example, future research could include sampling strategies to change the data distribution or designing a novel loss function to weigh the minority classes higher during training.

V. Discussion

With our work, we want to provide techniques that allow establishing a regional energy market. Such markets can forecast the potential risks in the power grid and regulate energy supply and consumption regionally based on the forecasted information. With this forecast knowledge, the power grid operator can trigger energy auctions at the regional energy markets to solve instability issues on the power grid.

We proposed three methods to forecast power grid states using deep learning techniques. Our results in Section IV show that the trained predictors perform great on forecasting safe power grid states (green), but very poorly on the less frequent power grid states such as yellow and red.

The results indicate that the proposed approaches indeed contribute to the development of the smart power grid significantly, e.g., more reasonable re-allocation of personal responsibility, ownership, expertise, and decision concerning energy supply. However, they leave significant room for improvement in forecasting these minority classes. We plan to implement over- and undersampling methods to adjust the proportion of minority classes to achieve better class balance [24], [25]. Another approaches are to design a loss function to emphasize the importance of the minority classes during training [26], or to generate additional data for the minority classes using Generative Adversarial Networks (GANs) [27]. Other future research might include working with attention-based transformer network models to improve forecast accuracy [28].

One of the next steps is to apply the forecast models in a real-world scenario for regional energy markets, e.g., using real-time data, applying model updates to new training data [29], and combining several models to an ensemble for increased forecast quality [30].

Acknowledgment

We want to thank EnergieNetz Mitte for providing and allowing us to publish the power grid data.

References

[1] "C/sells Website," http://www.csells.net, accessed: 2020-01-20.
[2] "sinteg website," https://www.sinteg.de/en/, accessed: 2020-01-20.
[3] D. B. Fey and R. Ag, "Regional Flexibility Markets," VDE/ETG, Tech. Rep., 2014.

[4] BDEW Bundesverband der Energie- und Wasserwirtschaft e.V., “Smart Grid Traffic Light Concept,” p. 14, 2015.

[5] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, “Solar forecasting methods for renewable energy integration,” *Progress in Energy and Combustion Science*, vol. 39, no. 6, pp. 535–576, 2013.

[6] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, “Current methods and advances in forecasting of wind power generation,” *Renewable Energy*, vol. 37, no. 1, pp. 1–8, 2012.

[7] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, “The State-Of-The-Art in Short-Term Prediction of Wind Power A Literature Overview,” *Technical Repor, ANEMOS.plus*, pp. 1–109, 2011.

[8] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, “A review on the forecasting of wind speed and generated power,” *Renewable and Sustainable Energy Reviews*, vol. 13, no. 4, pp. 915–920, 2009.

[9] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, “A review on the young history of the wind power short-term prediction,” *Renewable and Sustainable Energy Reviews*, vol. 12, no. 6, pp. 1725–1744, 2008.

[10] S. Mirasgedis, Y. Sarafidis, E. Georgopoulou, D. Lalas, M. Moschovits, F. Karagiannis, and D. Papakonstantinou, “Models for mid-term electricity demand forecasting incorporating weather influences,” *Energy*, vol. 31, no. 2-3, pp. 208–227, 2006.

[11] T. Hong, J. Wilson, and J. Xie, “Long term probabilistic load forecasting and normalization with hourly information,” *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 456–462, jan 2014.

[12] Y. Bengio *et al.*, “Learning deep architectures for AI,” *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[13] L. Deng and D. Yu, “Deep learning: Methods and applications,” *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[14] A. Gensler, J. Henze, B. Sick, and N. Raabe, “Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks,” in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2016, pp. 002 858–002 865.

[15] J. Henze, J. Schreiber, and B. Sick, *Representation Learning in Power Time Series Forecasting*. Cham: Springer International Publishing, 2020, pp. 67–101. [Online]. Available: https://doi.org/10.1007/978-3-030-31760-7_3

[16] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[17] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, “LSTM: A Search Space Odyssey,” *arXiv*, p. 10, 2015. [Online]. Available: http://arxiv.org/abs/1503.04069

[18] C. Q. Y Tao H Chen, “Wind Power Prediction and Pattern Feature Based on Deep Learning Method,” pp. 1–4, 2014.

[19] K. Kimura and T. Kimura, “Neural networks approach for wind-solar energy system with complex networks,” in *Power Electronics and Drive Systems (PEDS), 2013 IEEE 10th International Conference on*. IEEE, 2013, pp. 1–5.

[20] Y. Ren, P. Suganthan, and N. Srikanth, “Ensemble methods for wind and solar power forecasting - A state-of-the-art review,” *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 82–91, 2015.

[21] A. Gensler and B. Sick, “Forecasting Wind Power - An Ensemble Technique With Gradual Coopetitive Weighting Based on Weather Situation,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN16)*, Vancouver, Canada, 2016, pp. 1–9.

[22] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional lstm and other neural network architectures,” *Neural networks*, vol. 18, no. 5-6, pp. 602–610, 2005.

[23] L. Thurner, A. Scheidler, F. Schäfer, J. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun, “pandapower — an open-source python tool for convenient modeling, analysis, and optimization of electric power systems,” *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510–6521, Nov 2018.

[24] M. Buda, A. Maki, and M. A. Mazurowski, “A systematic study of the class imbalance problem in convolutional neural networks,” *Neural Networks*, vol. 106, pp. 249–259, 2018.

[25] S. Pouyanfar, Y. Tao, A. Mohan, H. Tian, A. S. Kaseb, K. Gauen, R. Dailey, S. Aghajanzadeh, Y.-H. Lu, S.-C. Chen *et al.*, “Dynamic sampling in convolutional neural networks for imbalanced data classification,” in *2018 IEEE conference on multimedia information processing and retrieval (MIPR)*. IEEE, 2018, pp. 112–117.

[26] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, and P. J. Kennedy, “Training deep neural networks on imbalanced data sets,” in *2016 international joint conference on neural networks (IJCNN)*. IEEE, 2016, pp. 4368–4374.

[27] G. Douzas and F. Bacao, “Effective data generation for imbalanced learning using conditional generative adversarial networks,” *Expert Systems with applications*, vol. 91, pp. 464–471, 2018.

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[29] Y. He, J. Henze, and B. Sick, “Continuous Learning of Deep Neural Networks to Improve Forecasts for Regional Energy Markets,” 2019, unpublished.

[30] A. Gensler, *Wind Power Ensemble Forecasting: Performance Measures and Ensemble Architectures for Deterministic and Probabilistic Forecasts*. kassel university press GmbH, 2019, vol. 12.