RESEARCH

Net load forecasting using different aggregation levels

Maximilian Beichter[1†], Kaleb Phipps[1*†], Martha Maria Frysztacki[1], Ralf Mikut[1], Veit Hagenmeyer[1] and Nicole Ludwig[2]

[†]Maximilian Beichter and Kaleb Phipps contributed equally to this work

*Correspondence: kaleb.phipps@kit.edu

[1] Karlsruhe Institute of Technology, Institute for Automation and Applied Informatics, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
Full list of author information is available at the end of the article

Abstract

In the electricity grid, constantly balancing the supply and demand is critical for the network's stability and any expected deviations require balancing efforts. This balancing becomes more challenging in future energy systems characterised by a high proportion of renewable generation due to the increased volatility of these renewables. In order to know when any balancing efforts are required, it is essential to predict the so-called net load, the difference between forecast energy demand and renewable supply. Although various forecasting approaches exist for both the individual components of the net load and the net load itself, it is unclear if it is more beneficial to aggregate several specialised forecasts to obtain the net load or to aggregate the input data to forecast the net load with one approach directly. Therefore, the present paper compares three net load forecasting approaches that exploit different levels of aggregation. We compare an aggregated strategy that directly forecasts the net load, a partially aggregated strategy that forecasts demand and supply separately, and a disaggregated strategy that forecasts demand and supply from each generator separately. We evaluate the forecast performance of all strategies with a simple and a complex forecasting model, both for deterministic and probabilistic forecasts, using one year of data from a simulated realistic future energy system characterised by a high share of renewable energy sources. We find that the partially aggregated strategy performs best, suggesting that a balance between specifically tailored forecasting models and aggregation is advantageous.

Keywords: Net load forecasting, Probabilistic forecasting, Aggregation

Introduction

Balancing energy supply and demand in energy systems is essential to maintain grid frequency stability (Schmietendorf et al. 2017). Any imbalance problems are managed in today's energy systems by including enough conventional fossil fuel-based generators to react to emergencies and keep the electricity system stable. However, this controllable reserve capacity will likely not exist in future energy systems characterised by a higher

Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 2 of 21

penetration of renewable energy sources. In such renewable energy systems, it will be more difficult to react to grid imbalances spontaneously (Schmietendorf et al. 2017; Kroposki et al. 2017), through controllable generation. Furthermore, storage systems that will assist with grid stability must be efficiently scheduled to ensure enough reserve capacity is available in times of need (Zachary et al. 2021). Therefore it is essential to accurately forecast the difference between expected demand and weather-dependent renewable supply, i.e. the so-called net load (Kaur et al. 2016). Accurate net load forecasts are crucial to enable efficient grid operation, such as demand side management (Barth et al. 2018), ensure grid stability can be maintained by scheduling storage systems (Zachary et al. 2021), and allow energy suppliers and network operators to communicate and coordinate dispatch plans in a smart grid (Zhang et al. 2016).

Numerous researchers have focused on the problem of net load forecasting. Garcia and Kirschen (2006) analyse net load forecasting in detail. As well as performing exploratory analysis of net load time series to determine valuable features, they compare the forecasting performance across various forecast horizons for traditional time series methods and various neural network architectures. However, all of their experiments are deterministic, and they do not consider probabilistic forecasts.

Regarding probabilistic net load forecasts, Taylor (2006) and Salem et al. (2019) both consider probabilistic net load forecasts with quantile-based approaches. Whilst Taylor generates density forecasts based on a combination of autoregressive integrated moving average (ARIMA) point forecasts and a volatility model (Taylor 2006), Salem et al. directly generate prediction intervals using quantile regression forests (Salem et al. 2019).

In recent years increasingly complex approaches have been considered. For example, Persio et al. (2017) compare variations of ARIMA models for net load forecasting based on data from the Italian energy system, and Sreekumar et al. (2020) propose grey systems theory-based methods for net load forecasting. Stratigakos et al. (2021) develop probabilistic net load forecasts using singular spectrum analysis and long short-term memory (LSTM) neural networks. Despite these increasingly complex models, all approaches focus on directly forecasting the net load.

However, the net load is a difference between multiple time series, i.e. the demand time series on the one hand and multiple supply time series from different energy sources on the other hand. To the best of our knowledge, no previous work has considered using this underlying net load structure to investigate different forecasting strategies. Therefore, in the present paper, we investigate the performance of three forecasting strategies for the net load. One strategy where we directly forecast the net load, one where we forecast the demand and supply separately, and one where we forecast the demand and the supply from each renewable generator separately. Instead of developing new complex forecasting models, we take existing models and compare their performance using these three forecasting strategies. Thereby, we investigate different assumptions about the qualities required for the best performing forecast associated with these strategies. For example, directly forecasting the net load assumes that aggregating all involved quantities will cause a smoothing effect and, therefore, a time series that is easier to forecast. On the other hand, separately forecasting each generator and the demand assumes that it is beneficial to use tailored forecasting models. We compare these strategies for

Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 3 of 21

deterministic approaches and probabilistic forecasts by applying a copula approach already tested in the context of energy systems by Li et al. (2020). Given these strategies, deterministic and probabilistic forecasts, we consider simulated data from a realistic future energy system with enough renewable electricity generation to meet demand. With this future energy system, we evaluate the forecasting strategies using a simple (Neural Network (NN)) and a complex (Convolutional Neural Network (CNN)) model and compare their deterministic and probabilistic forecast performance.

The remainder of the present paper is structured as follows. First, we define net load and introduce our proposed forecasting strategies, including the extension to perform probabilistic forecasts in "Net load forecasting". In the "Evaluation" we describe the simulated energy system data and meteorological data used, the implementation of our forecasting models, the evaluation metrics, and we present our results. We critically discuss these results and highlight key observations in "Discussion", before concluding and presenting future research directions in "Conclusion".

Net load forecasting

The present paper investigates three strategies based on different aggregation levels for net load forecasting. In the following, we define net load before introducing our proposed forecasting strategies in detail. We then discuss the extension of our strategies to probabilistic net load forecasting.

Definition

We consider the difference between the electricity demand (i.e. load) and the electricity supply from variable renewable energy sources as the net load. Formally this net load is then represented by simple subtraction, i.e.

$$NL = D - S, \tag{1}$$

where NL is the net load, D the electricity demand in the system, and S the electricity supply from renewable sources.

The expected renewable energy supply is, however, an aggregated value made up of all the individual renewable energy generators in the system, i.e.

$$S = \sum_{g \in G} s_g, \tag{2}$$

where s_g is the energy supplied from generator g, given G different renewable generation sources in the energy system. Based on this observation, the definition of net load in Eq. (1) can also be represented by

$$NL = D - \sum_{g \in G} s_g. \tag{3}$$

We can identify different time series aggregation levels based on these net load definitions. For example, we could consider the supply from each generator separately or the total supply found by aggregating all of the generator time series together. These different levels of aggregation form the foundation for our forecasting strategies, which we describe in detail in the following.
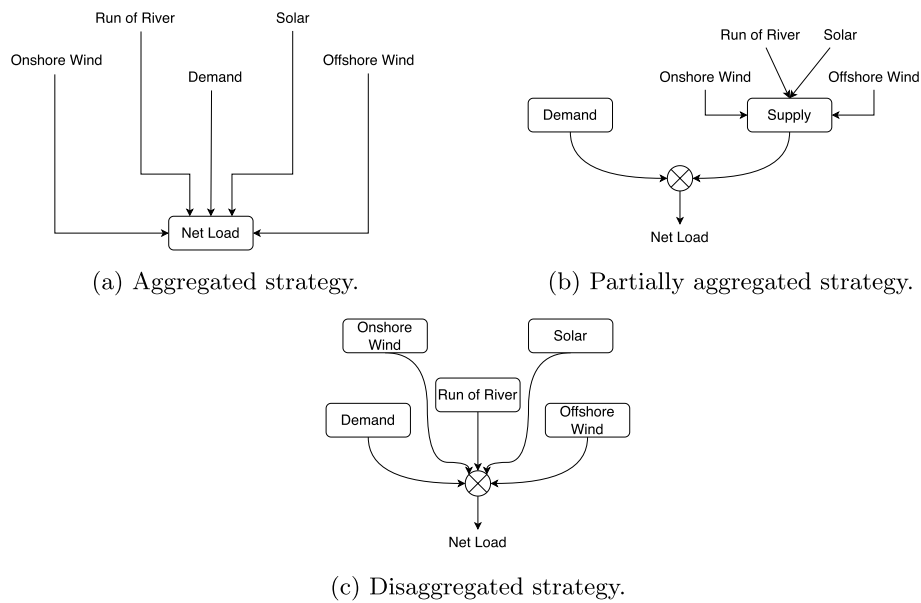
(a) Aggregated strategy.

(b) Partially aggregated strategy.

(c) Disaggregated strategy.

Fig. 1 Representation of the different strategies for forecasting net load, where each box represents a forecasting model and $\otimes$ represents the combination of forecasting models

Forecasting strategies

The three forecasting strategies, which we investigate in the present paper, differ in their level of aggregation used for the available time series. Based on the aforementioned definition of net load, we either use an *aggregated strategy*, a *partially aggregated strategy* or a *disaggregated strategy*. Figure 1 shows these strategies and we explain them in more detail in the following.

Aggregated strategy: For the aggregated strategy, we directly forecast the net load, i.e. the forecast is performed on the time series after subtracting the supply from the demand. The assumed advantage of the strategy is that more aggregation leads to a time series that is simpler to forecast as it fluctuates less with fewer extreme values and more explicit recurring patterns. However, one disadvantage of this aggregated strategy is that the relationship between net load and exogenous variables, such as the weather, might not be as straightforward as the weather primarily influences the output of the generators and not the net load itself.

Partially aggregated strategy: The partially aggregated strategy aims to keep the advantage of aggregating time series to smooth them. Additionally, it aims to reduce the disadvantage mentioned above regarding exogenous variables by forecasting supply and demand separately. This partial aggregation, in which we only aggregate the renewable electricity supply time series, gives the forecasting models the chance to specialise and select e.g. weather features for the supply time series and calendar features for the demand time series.

Disaggregated strategy: We forecast each time series separately for the disaggregated strategy, thus allowing the forecasting models to specialise even more and select a

Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 5 of 21

unique subset of features for each renewable energy source and the demand. With this strategy, we assume that the advantages of individually modelling the time series outweighs any advantage aggregation could bring to smooth the time series.

Probabilistic extension

Our three strategies can be implemented for both deterministic and probabilistic forecasts. However, while individual forecasting models can be aggregated using addition and subtraction in the case of deterministic forecasts, the aggregation for probabilistic forecasting models is more complex. Therefore, this section describes how we aggregate the probabilistic forecasting models.

When we perform probabilistic forecasts, we predict a probability density function (PDF) for different random variables and therefore when aggregating random variables, we need to predict a new PDF. For example, given two continuous random variables X and Y with PDF $f_X, f_Y : \mathbb{R} \to \mathbb{R}_+$ and their joint PDF f_{XY}, then $X + Y$ is also a continuous random variable (Casella and Berger 2021), with a PDF given by

$$f_{X+Y}(X + Y = z) = \int_{-\infty}^{\infty} f_{XY}(w,\, z - w)dw. \tag{4}$$

This PDF can be calculated as the convolution of f_X and f_Y, if X and Y are independent (Casella and Berger 2021). However, if the random variables are not independent, we have to account for their correlation in the joint PDF f_{XY} as well. Since we consider various renewable generation sources, which are all dependent on meteorological conditions, we cannot assume independence and therefore have to account for the correlations in the joint PDF.

To account for the correlations between random variables we use a copula approach similar to Li et al. (2020). According to Sklar's theorem (Sklar 1959), any multivariate joint distribution can be written in terms of the marginal distributions of each component and a copula that describes the dependency structure between the components. Following Sklar (1959), with F_x and F_Y being the marginal cumulative density functions (CDFs) for X and Y, the joint CDF F_{XY} can then be expressed as

$$F_{XY}(X \leq x,\, Y \leq y) = C(u,\, v), \tag{5}$$

with $u = F_X(X \leq x), v = F_Y(Y \leq y)$, and the copula $C(\cdot) : [0, 1]^2 \to \mathbb{R}_+$ being a continuous real valued function. Based on this definition, we can find an expression for the joint PDF by firstly deriving $C(\cdot) : [0, 1]^2 \to \mathbb{R}_+$ as,

$$c(u,\, v) = \frac{\partial^2 C}{\partial u \partial v}, \tag{6}$$

and using this to express f_{XY} via

$$f_{XY} = c(u,\, v) \cdot f_X \cdot f_Y. \tag{7}$$

Thus, we obtain an expression for the joint PDF f_{XY}, without having to explicitly calculate the unknown joint distribution and the integral in Eq. (4) can be expressed as

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} c\Big(u(w), v(z-w)\Big) \cdot f_X(w) \cdot f_Y(z-w)dw. \tag{8}$$

For our various forecasting strategies, we firstly generate a probabilistic forecast for the given quantity, either net load, demand, supply, or the electricity generated from one of the generators in the system. Depending on the level of aggregation considered, we then apply the copula method described above to determine the joint PDF. In the case of the aggregated strategies, no copula combination is required. For the partially aggregated strategy, we use the copula combination to calculate the combined PDF of supply and demand. Finally, for the disaggregated strategy, we apply the copula combination iteratively multiple times to calculate the final joint PDF. Thereby, multiple steps are required: first we apply the copula combination to calculate the combined PDF between demand and solar. Second, we use the copula combination on the resulting PDF and onshore wind to calculate the combined PDF of demand, solar, and onshore wind. Third, we repeat this process with the resulting PDF and the copula combination, including wind offshore in the next step, and finally, run of river. Following this iterative process, we can calculate the combined PDFs of supply and demand in a disaggregated manner.

Evaluation

To evaluate our three proposed forecasting strategies for net load forecasting based on different aggregation levels, we consider a hypothetical future scenario with enough renewable generation to meet demand and two exemplary forecasting models. In the following we first describe the data used for our evaluation before introducing the evaluation metrics used for feature selection and evaluating the quality of our forecasts. We then describe the simple (NN) and complex (CNN) forecasting models, as well as our feature selection, hyperparameter optimisation, and extensions for probabilistic forecasting. Lastly, we report the forecasting performance for each forecasting strategy for both the simple and complex forecasting models.

Data

To evaluate our forecasting strategies, we require data for an energy system where renewable energy generation meets demand and weather data to perform accurate forecasts. In this section, we thus briefly describe both the simulated energy system data and the weather data used.

Energy system data: We use simulated data from a realistic future energy scenario to obtain energy system data with enough renewable generation to meet electricity demand. This data is generated using the open energy system model PyPSA-Eur (Hörsch et al. 2018) and models Germany with a high spatial resolution of 100 nodes and an hourly temporal resolution. The aggregation is well described in a previous publication and details can be found in the methodology section of Frysztacki et al. (2021). The model comprises energy generated from onshore and offshore wind, solar, run of river, nuclear, oil, lignite and biomass. To ensure a realistic simulated energy system similar to the current system but with a higher share of renewable sources, the generation fleet of conventional carriers is set to the historical operation capacity of 2011 and is non-extendable, whilst the capacity of renewable energy sources is subject to optimisation.
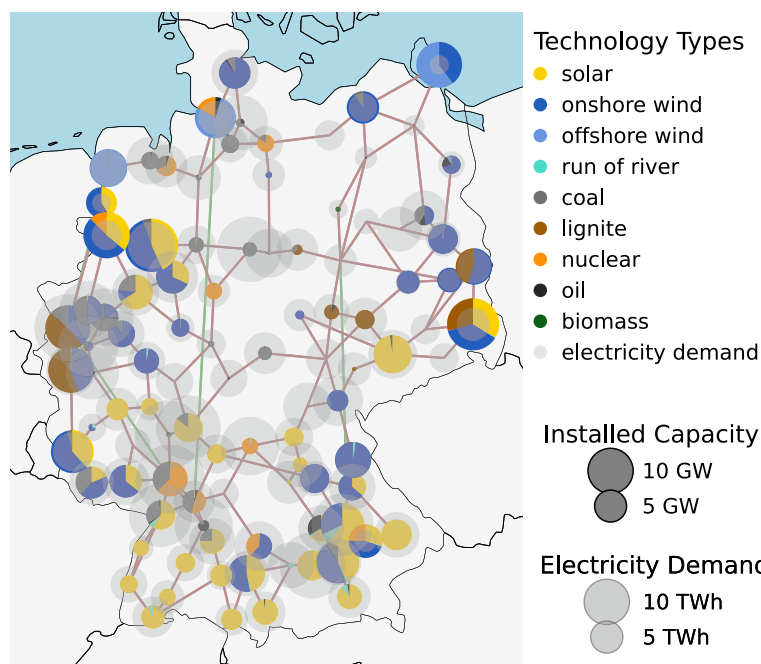
Fig. 2 The PyPSA model of Germany displaying optimal capacity and electricity demand for our future scenario

This ensures that the resulting system is realistic regarding the generation mix but optimised for a renewable future. As a result, the capacity is determined by minimising the total annual system costs and guaranteeing that electricity demand is met at all times and in all locations. The optimisation respects the physics of the system, including Kirchhoff's circuit laws and accounts for the capacity limitations of transmission lines and generators. To simulate weather conditions and weather variability, we embed historical weather data of 2010 to the investment optimisation. The problem is further subject to reducing 80% of the carbon emissions of 1990 to guarantee a significant proportion of renewable generation that is capable of meeting the electricity demand. Some nuclear power plants are included in the scenario to account for future imports from neighbouring countries, particularly France. This scenario is a possible pathway to a future fully renewable electricity system. The resulting capacity investments are displayed in Fig. 2 together with the total electricity demand of 2010. Finally, to retrieve the generation profiles, we solve an operational problem where no capacity expansion is allowed for 2010–2019, again using historical weather data. An exemplary generation profile of the model is visualised in Fig. 3, where the net load is highlighted with a red dashed line. Note that load shedding is below 0.01% of the annual demand for every simulated scenario. Using the PyPSA-EUR model at a high spatio-temporal resolution has proven to reproduce historical results, particularly the amount of renewable curtailment (Frysztacki and Brown 2020) and conventional power generation patterns as well as wind and solar production (Unnewehr et al. 2022). Therefore, the model provides a sound basis for investigating net load.
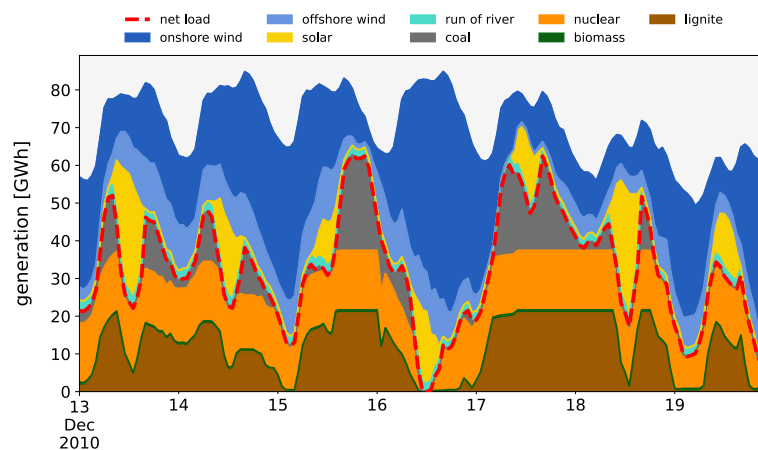
Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 8 of 21

Fig. 3 The generation profile, showing how the different suppliers make up total energy supply, for an exemplary week in our reference future electricity system scenario with a high share of renewable generation. Here, net load is displayed with a dashed red line

Given this simulated data, we use the years 2010 to 2018 for feature selection, hyperparameter optimisation, and training and validation of our forecasting models. The year 2019 is withheld from all training and used purely as test data for our final evaluation.

Weather data: Additionally to the energy system data, we also use historical weather data matching the temporal and spatial resolution of the simulated energy system as inputs for our forecasting models. More specifically, we use the ERA5 reanalysis data (Hersbach et al. 2018), available via the Copernicus Climate Data Store.[1] We acquire historical weather data in an hourly resolution for the geographical region of Germany, i.e. a grid from 45°–55° N by 5°–15° E, with a resolution of 0.25°. The variables we consider are those with the highest impact on renewable energy generation, i.e. surface net solar radiation (ssr), surface net thermal radiation (str), temperature at two meters above ground (t2m), as well as the eastward (u100) and northward (v100) component of wind at 100 m above ground. In order to map the weather data to the simulated energy data we cluster the weather information into three onshore regions i.e. north Germany (51°–54° N, 6°–14° E), south Germany (47°–50° N, 7°–13° E), and central Germany (50°–51° N, 6°–14° E) and additionally cluster the data from the Baltic sea (53.5°–55° N, 9.5°–14° E) and North sea (53.5°–55° N, 4°–9° E) to provide weather information for the offshore forecast. Given these clusters, we calculate the mean weather information across the region and use these mean values as input features for our forecasting models.

Metrics

Given the above mentioned data, we require metrics for our feature selection and the evaluation of the probabilistic and deterministic forecast performance of our different forecasting strategies. Furthermore, we perform significance tests to determine if our results are significant. This section briefly describes the metrics used and the significance test performed.

---

[1] https://cds.climate.copernicus.eu/cdsapp#!/home.

Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 9 of 21

Root Mean Squared Error: For feature selection and to evaluate the deterministic forecasts we use the Root Mean Squared Error (RMSE). The RMSE represents the square root of the second sample moment of the differences between predicted values and observed values. Therefore, given predictions $\hat{y}_t$, and observations y_t for all samples $t = 1, \ldots, N$ in the test set, the RMSE is given by

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (\hat{y}_t - y_t)^2}. \tag{9}$$

Continuous Ranked Probability Score: To evaluate our probabilistic forecasts we use the Continuous Ranked Probability Score (CRPS) (Gneiting and Katzfuss 2014). The CRPS is a measure for calibration and sharpness of a predictive cumulative distribution function F and for all forecast time steps is given by

$$\text{CRPS} = \frac{1}{N} \sum_{t=1}^{N} \int_{\mathbb{R}} \left(F(z) - \mathbb{1}\{z \leq y_t\} \right)^2 dz, \tag{10}$$

with z the integration variable, y the verifying observation, and $\mathbb{1}$ denoting an indicator function (Gneiting et al. 2005).

Significance test: To determine whether our results are significant we perform a Diebold Mariano (DM) test. The DM test determines whether a calculated loss differential d_t is significantly larger than zero (Diebold et al. 1995). Given the mean of this loss differential

$$\bar{d} = \frac{1}{N} \sum_{t=1}^{N} d_t, \tag{11}$$

with $\mu = \mathbb{E}[d_t]$, and the autocorrelation of the loss differential

$$\gamma_k = \frac{1}{N} \sum_{t=k+1}^{N} (d_t - \bar{d})(d_{t-k} - \bar{d}), \tag{12}$$

the DM test statistic for $h \geq 1$ is

$$\text{DM} = \frac{\bar{d}}{[\gamma_0 + 2 \cdot \sum_{k=1}^{h-1} \gamma_k]/N}, \tag{13}$$

where h is the order for the DM test statistic (Diebold et al. 1995). Under the null hypothesis, that the loss differential is zero, i.e. $\mu = 0$, the test statistic follows a standard normal distribution

$$\text{DM} \sim \mathcal{N}(0, 1), \tag{14}$$

and therefore the null hypothesis can be rejected using the two-tailed critical value for the standard normal distribution.

For our application of the DM test, we select the difference between the calculated mean squared error for each time step as the loss differential to compare deterministic models and the difference between the CRPS for each time step as the loss differential

Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 10 of 21

when comparing the probabilistic models (Gneiting and Katzfuss 2014; Gneiting and Raftery 2007).

Forecasting models

To evaluate our proposed strategies for forecasting net load, we implement a simple (NN) and a complex (CNN) model to show our results are not model dependent.[2] We perform a careful feature selection and hyperparameter optimisation for both models. Therefore, in the following, we explain the feature selection and hyperparameter optimisation since this is the same for both forecasting models. We then describe the implementation of the forecast models before detailing the extensions required to generate probabilistic forecasts.

Feature selection: Our applied feature selection is the same for all forecast models implemented.[3] This feature selection is based on decision trees, which are valuable for feature selection because they are accurate and can explain when a feature is essential (Grömping 2009). This explanation is based on the measure of impurity, which is the variance in regression tasks. When training a decision tree, it is possible to measure how much each feature decreases the impurity, i.e. variance. Therefore, features that decrease this impurity are considered essential features (Grömping 2009; Chen et al. 2020).

We use decision tree regression in our feature selection and measure the RMSE based on different feature sets. These feature sets are chosen to include variables relevant for the model considered, i.e. for wind forecasts, we only consider features that are relevant for wind. Within these feature sets, we consider features such as hours of the day and day of the week as categorical to ensure that either all known temporal information from one category is considered or none. We provide the decision tree with all available features and iteratively decrease the number of features. Hence, the decision tree can initially select all features; however, we restrict the decision tree to only selecting a subset in subsequent runs. In each run, we decrease the number of features the decision tree is allowed to select and calculate the RMSE to measure performance. Through this, we can generate multiple feature sets, each containing the k most essential features, as shown in Table 1. In Table 1, the number in each row indicates the smallest subset of features that still included the given feature, e.g., a value of 15 implies that the given feature was last selected in a subset of 15 features, and no longer selected when the subset was reduced to 14. These feature sets form the basis for the hyperparameter optimisation described in the following.

Hyperparameter optimisation: For the hyperparameter optimisation, we perform Bayesian optimisation. We perform 100 runs of Bayesian optimisation (Snoek et al. 2012) for each model, with a batch size of 128, using an individual hyperparameter space for each model, i.e. the number of neurons for the simple NN model, and filter and window sizes for each CNN layer in the complex CNN model. In each of these optimisation runs, we train four individual models and assess the performance based on the mean performance of these four models. These four models are obtained via fourfold time

---

[2] The implementation is available via https://github.com/KIT-IAI/NetLoadForecasting.

[3] Note that due to the small portion of run of river installed in the simulated energy system, no feature selection was performed for these models.

Table 1 An overview of the progressive feature selection for each target time series

	Target time series to be forecast					
	Onshore wind	Offshore wind	Solar	Demand	Net load	Supply
Surface net solar radiation north Germany			14		34	32
Surface net solar radiation central Germany			9		12	14
Surface net solar radiation south Germany			2		2	2
Surface net thermal radiation north Germany			6		17	17
Surface net thermal radiation central Germany			3		27	20
Surface net thermal radiation south Germany			7		13	9
Temperature 2 m over ground north Germany			10		28	8
Temperature 2 m over ground central Germany			12		14	26
Temperature 2 m over ground south Germany			8		10	7
Eastward wind velocity Baltic sea		6			24	11
Eastward wind velocity North sea		4			31	31
Northward wind velocity Baltic sea		3			32	27
Northward wind velocity North sea		8			30	29
Resulting wind velocity Baltic sea		2			9	6
Resulting wind velocity North sea		2			6	2
Eastward wind velocity north Germany	5				35	24
Eastward wind velocity central Germany	13				18	18
Eastward wind velocity south Germany	4				11	13
Northward wind velocity north Germany	8				36	12
Northward wind velocity central Germany	9				20	23
Northward wind velocity south Germany	6				26	19
Resulting wind velocity north Germany	2				2	4
Resulting wind velocity central Germany	2				3	3
Resulting wind velocity south Germany	3				8	5
One hot encoded weekdays				5	5	
One hot encoded hours	18	15	5	11	7	16
Public holiday				3	15	
Partial public holiday of Germany				12	37	
Summertime	19	16	4	13	33	15
Historical values (lags) for the last 0 h	7[a]	5[b]	2[c]	2[d]	4[e]	10[f]
Historical values (lags) for the last 24 h	15[a]	12[b]	13[c]	10[d]	21[e]	33[f]
Historical values (lags) for the last 48 h	17[a]	9[b]	19[c]	8[d]	29[e]	25[f]
Historical values (lags) for the last 72 h	10[a]	11[b]	15[c]	6[d]	23[e]	34[f]
Historical values (lags) for the last 96 h	14[a]	14[b]	17[c]	4[d]	19[e]	22[f]

Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 12 of 21

Table 1 (continued)

	Target time series to be forecast					
	Onshore wind	Offshore wind	Solar	Demand	Net load	Supply
Historical values (lags) for the last 120 h	12[a]	7[b]	11[c]	7[d]	25[e]	30[f]
Historical values (lags) for the last 144 h	16[a]	10[b]	16[c]	2[d]	16[e]	28[f]
Historical values (lags) for the last 168 h	11[a]	13[b]	18[c]	9[d]	22[e]	21[f]

Starting with a large number of features we iteratively reduce the number of features that can be selected. The number in each row indicates the smallest subset of features that still included the given feature. We perform feature selection until a minimal feature set size of 2

[a] For onshore wind, the lags correspond to historical onshore wind values

[b] For offshore wind, the lags correspond to historical offshore wind values

[c] For solar, the lags correspond to historical solar values

[d] For demand the lags correspond to historical demand values

[e] For net load, the lags correspond to historical net load values

[f] For supply, the lags correspond to historical supply values

series cross-validation (Bergmeir and Benítez 2012). For this cross-validation, we always consider 8760 samples (1 year) for validation and begin the training with our training data in 2010.

We begin with the feature sets obtained through our initial feature selection to determine the optimal features. Starting with a first subset of feature sets (see Table 2), we then optimise each model with a decreasing number of features until no performance improvement is observed on the validation data for at least three feature set decreases in a row. Finally, the resulting feature set is selected as the best performing feature set for further model evaluation.

Simple neural network model: To evaluate the forecasting strategies, we select a feed-forward neural network (NN) as a simple model. In the following, we describe the structure of our NN and refer to Goodfellow et al. (2016) for a detailed theoretical background. As inputs, the NN receives the lagged features as well as all selected features for 24 h in the future. The NN comprises six hidden layers for the deterministic model and five hidden layers for the probabilistic extension. In both cases, the final hidden layer uses a linear activation function whilst all layers before use exponential linear unit (ELU) activation functions. The output layer for the deterministic model uses a linear activation function, whilst the activation function for the output layer of the probabilistic extension depends on the distribution selected (see "Probabilistic extension"). The number of neurons in each layer is determined through hyperparameter optimisation, where the hyperparameter space ranges from 64 to 256 neurons for the first layer, 32 to 128 neurons for the second layer, 16 to 64 neurons for the third layer, and 8 to 32 neurons for the fourth layer. The final two layers are fixed with 8 and 2 neurons, respectively. The networks are trained with a batch size of 128 using kernel and bias regularisation. We use the Adam optimiser with an adaptive learning rate and early stopping for the training process. In each training step, fourfold cross-validation is applied, and the final model performance is, as with the hyperparameter optimisation, calculated as the

Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 13 of 21

Table 2 Summary of feature sets considered for the optimisation and the resulting best performing feature set selected for each forecasting model

Forecast type	Model	Complex		Simple	
		Feature sets considered	Feature set selected	Feature sets considered	Feature set selected
Deterministic forecast	Net load	37–29	37	37–29	36
	Supply	34–26	32	34–28	32
	Demand	13–10	13	13–10	13
	Offshore wind	16–5	9	16–5	9
	Onshore wind	19–7	14	19–12	18
	Solar	19–9	17	19–10	16
	Run of river	1	1	1	1
Probabilistic forecast	Net load	37–29	37	37–32	36
	Supply	34–29	32	34–27	32
	Demand	13–10	13	13–9	13
	Offshore wind	16–5	8	16–5	8
	Onshore wind	19–8	14	19–15	19
	Solar	19–10	18	19–10	14
	Run of river	1	1	1	1

Table 3 Configuration of the best performing simple neural network forecasting models used

	Model	Number of neurons			
		Layer 1	Layer 2	Layer 3	Layer 4
Deterministic forecast	Net load	192	128	12	32
	Supply	96	128	64	18
	Demand	64	64	48	32
	Offshore wind	96	32	64	8
	Onshore wind	160	80	64	8
	Solar	64	80	64	32
	Run of river	224	128	24	32
Probabilistic forecast	Net load	64	32	48	24
	Supply	64	32	48	22
	Demand	64	32	48	8
	Offshore wind	96	32	16	32
	Onshore wind	64	32	16	18
	Solar	256	128	48	32
	Run of river	64	32	64	8

mean performance of these four models. The best performing configuration is shown in Table 3.

Complex CNN model: To better account for the recurring patterns in energy time series and generate more accurate forecasts, we also develop a complex CNN model. We select a CNN since it has proven high performance when forecasting energy time series (Heidrich et al. 2020) and is relatively robust and simple to train in comparison to other machine learning methods, such as long short-term memory networks (Goodfellow et al. 2016). In the following, we again describe the structure of our CNN model, referring to Goodfellow et al. (2016) for detailed theoretical information. Our complex model
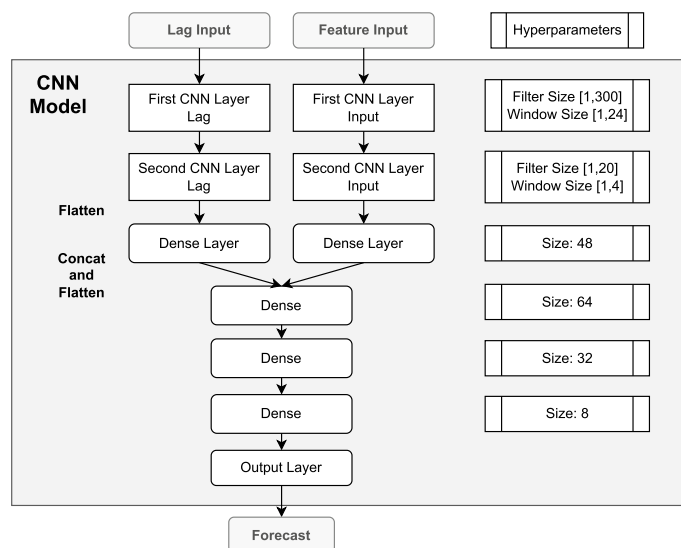
Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 14 of 21

Fig. 4 The architecture of the complex CNN model used for net load forecasting. The model takes two inputs, lagged features and non-lagged features for the upcoming 24 h. The hyperparameter space used for hyperparameter optimisation is shown on the right

CNN has two separate inputs, namely lagged features and all non-lagged features for the upcoming 24 h, as shown in Fig. 4. These two inputs are passed through two CNN layers, with batch normalisation, ELU activation functions, and max pooling between each layer. The outputs are flattened and passed into dense layers with batch normalisation and an ELU activation function following the second CNN layer. The two outputs are then concatenated and flattened before being passed to three more dense layers, with batch normalisation and ELU activation functions. The final output layer contains a linear activation function. The hyperparameter space for the complex model includes both the filter size and the window size for the CNN layers. This hyperparameter space is shown in Fig. 4. Given the optimal hyperparameters, the complex model is trained with a batch size of 128 using four-fold cross-validation and kernel and bias regularisation. The model is then evaluated on the mean of the folds. During training, we use the mean squared error (MSE) as the loss function and again the Adam optimiser with early stopping. The best performing configuration for the complex model is shown in Table 4.

Probabilistic extension: To extend our approach to probabilistic forecasts, we adapt the forecasting models in the following ways. Firstly, the output is no longer a single point forecast but rather multiple neurons designed to estimate the distribution parameters of our assumed parametric distribution. Thereby, depending on the forecast parametric distribution, we adjust the output activation function. For example, if we are forecasting a quantity with an assumed Beta or Gamma distribution we use the Softplus activation function. In contrast, we use a linear activation function for an assumed Gaussian distribution. Finally, during training, we use a proper scoring rule as a loss function, namely, the logarithmic score (Gneiting and Raftery 2007), which is a robust training metric.

To aggregate the multiple probabilistic forecasts for the three forecasting strategies, we use copulas, as described in "Probabilistic extension". We select a Frank copula for the

Table 4 Configuration of the best performing complex CNN forecasting models used

	Model	Lag input		Feature input	
		Filter size	**Window size**	**Filter size**	**Window size**
Deterministic forecast	Net load	(1, 1)	(24, 1)	(300, 1)	(1, 4)
	Supply	(218, 17)	(24, 4)	(300, 5)	(1, 4)
	Demand	(171, 20)	(1, 4)	(137, 18)	(1, 1)
	Offshore wind	(1, 1)	(1, 4)	(300, 1)	(7, 4)
	Onshore wind	(1, 20)	(7, 4)	(300, 1)	(8, 3)
	Solar	(153, 7)	(17, 3)	(78, 12)	(2, 3)
	Run of river	(128, 1)	(7, 4)	(300, 9)	(9, 4)
Probabilistic forecast	Net load	(1, 20)	(5, 1)	(99, 10)	(1, 3)
	Supply	(1, 20)	(1, 2)	(108, 5)	(1, 4)
	Demand	(171, 5)	(6, 2)	(100, 20)	(1, 3)
	Offshore wind	(1, 20)	(17, 3)	(147, 1)	(8, 4)
	Onshore wind	(1, 16)	(7, 1)	(114, 1)	(6, 4)
	Solar	(164, 8)	(18, 2)	(62, 9)	(3, 4)
	Run of river	(234, 1)	(1, 3)	(224, 20)	(24, 1)

The brackets indicate the layers, i.e. (layer one, layer two)

aggregation since the analysis of Li et al. (2020) shows that the Frank copula performs the best for aggregating energy forecasts. For our probabilistic forecast, we select parametric distributions with a data-driven approach. Therefore, we analyse the data and select a parametric distribution that approximates the underlying empirical distribution, as shown in Fig. 5. In this process, we only select distributions that are differentiable to allow their integration into our gradient-based forecasting models. Based on this data-driven approach, we thus assume a Gaussian distribution for the demand and a beta distribution for the supply, solar generation, offshore wind generation, run of river and net load. For the onshore wind generation, we assume a gamma distribution.

Results

Based on the simulated data from the realistic future scenario and our implemented forecast models, we evaluate the forecasting performance of our three net load forecasting strategies. For this evaluation, we only consider the test year of 2019, and generate forecasts based on historical weather data, i.e. a now-cast with a forecast horizon of 0 h.[4] In this section, we, therefore, report the evaluation results, firstly focusing on deterministic forecasts before also discussing probabilistic forecasts.

Deterministic forecasts: Table 5 shows the RMSE on the test data for all three forecasting strategies and deterministic forecasts. These results show that using the complex CNN model and the partially aggregated strategy significantly ($\alpha = 0.01$) performs best, with an RMSE of 4.018. The second best performing strategy is the disaggregated strategy using the complex CNN model, with an RMSE of 4.107, whilst the aggregated strategy performs worst. Additionally, the simple model ranks the forecasting strategies in the same order as the complex model, with the partially aggregated strategy again

---

[4] Since our forecast models rely on exogenous inputs, given weather forecasts of a certain forecast horizon, we can easily integrate this weather forecast information to generate supply, demand, and net load forecasts of the same forecast horizon.
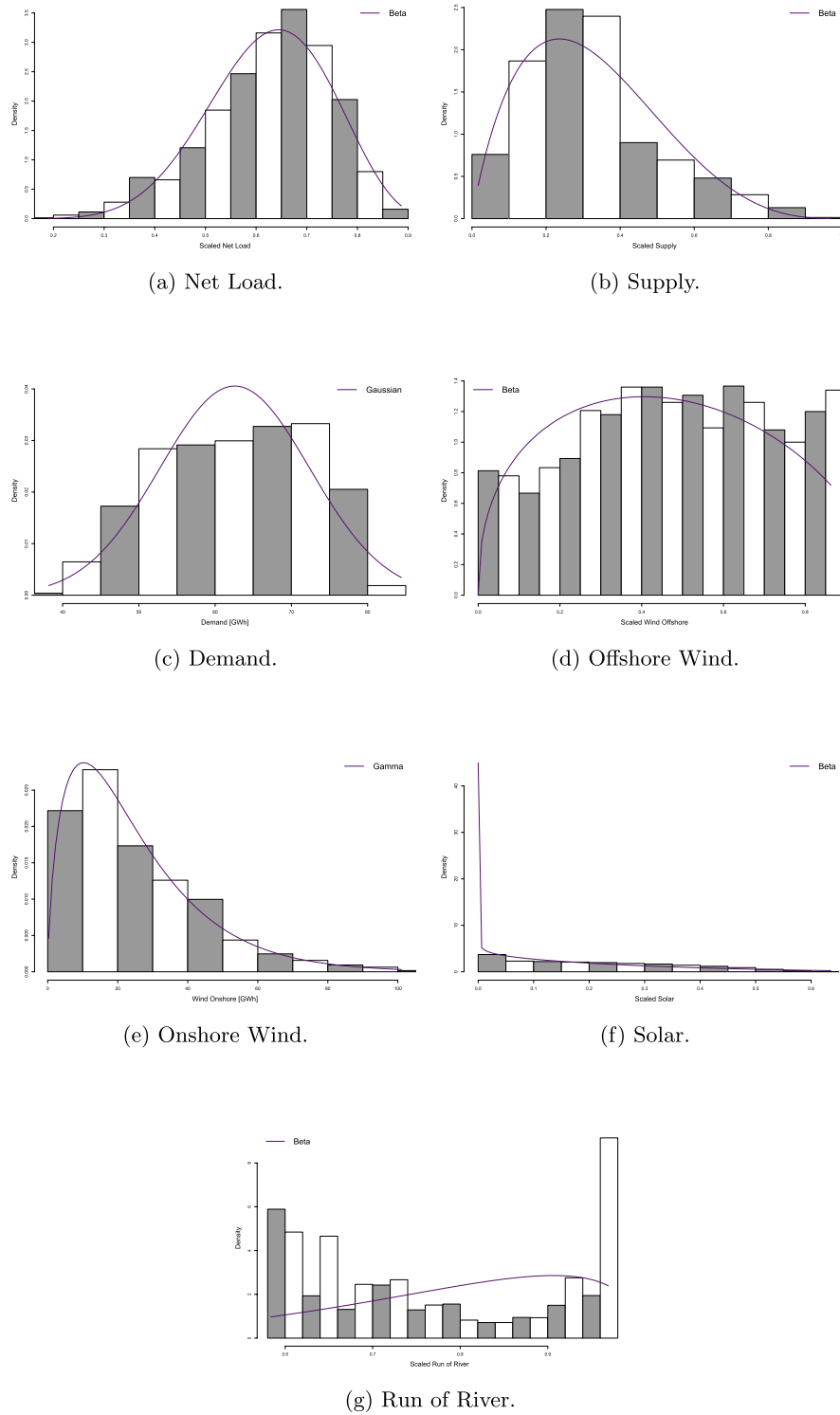
Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 16 of 21

(a) Net Load.

(b) Supply.

(c) Demand.

(d) Offshore Wind.

(e) Onshore Wind.

(f) Solar.

(g) Run of River.

Fig. 5 The predicted distribution function and the observed empirical distribution function for all considered variables. We observe, that the selected distributions, in general, are a good approximation for the data. Note that if the selected distribution required values between 0 and 1, e.g. for the Beta distribution, the data was scaled in advance

Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 17 of 21

Table 5 RMSE calculated on the test set for the different deterministic forecasting strategies

Strategy	Quantity	Complex model	Simple model
Aggregated	**Net load**	5.068	5.448
Partially aggregated	Supply	3.744	3.790
	Demand	1.408	1.558
	Net load	**4.018****	**4.132*****
Disaggregated	Offshore wind	1.186	1.281
	Onshore wind	3.508	3.764
	Solar	1.095	1.189
	Run of river	0.089	0.100
	Demand	1.408	1.558
	Net load	4.107	4.481

The best performing strategy is highlighted in bold. Significance levels are based on the DM test and show the significance between the two best performing forecasting strategies for the same forecasting model

Significance levels *** $\alpha = 0.001$, ** $\alpha = 0.01$, * $\alpha = 0.1$

performing best (RMSE = 4.132). However, all simple NN forecast models are outperformed by the complex CNN models.

Probabilistic forecasts: Table 6 shows the CRPS on the test data set for all three forecasting strategies and probabilistic forecasts. The probabilistic evaluation results mirror the deterministic ones. Again, the best performing strategy for net load forecasting is the partially aggregated strategy with a CRPS of 2.130. This difference is also significant ($\alpha = 0.01$) compared to the second best strategy, the disaggregated strategy. As before, the simple NN model ranks all strategies in the same order as the complex CNN model, with the partially aggregated strategy performing significantly ($\alpha = 0.001$) better than the other strategies. The complex model outperforms the simple NN for all forecasts apart from the run of river forecast in the disaggregated strategy, where the simple model performs slightly better.

Discussion

This section discusses the evaluation of our proposed strategies for net load forecasting, first on the results of the deterministic evaluation before considering the probabilistic forecasts.

The evaluation of our forecasting strategies for deterministic forecasts shows that the partially aggregated strategy delivers the best performance. This result suggests that combining both aggregation and tailored forecasting models is advantageous. The advantage of the aggregation is probably due to the supply consisting of multiple renewable energy sources that are geographically dispersed throughout the simulated German energy system. This dispersion and variety in renewable energy sources lead to a smoothing effect in the resulting supply time series when aggregating all generation sources. This smoothing effect may counteract the large fluctuations present in single generation time series. Such fluctuations could, for example, be the extreme difference

Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 18 of 21

Table 6 CRPS calculated on the test set for the different probabilistic forecasting strategies

Strategy	Quantity	Complex model	Simple model
Aggregated	**Net load**	2.676	2.987
Partially aggregated	Supply	1.953	2.053
	Demand	0.751	0.794
	Net load	**2.130*****	**2.226*****
Disaggregated	Offshore wind	0.602	0.657
	Onshore wind	1.810	1.911
	Solar	0.343	0.379
	Run of river	0.048	0.045
	Demand	0.751	0.794
	Net load	2.203	2.363

The best performing strategy is highlighted in bold. Significance levels are based on the DM test and show the significance between the two best performing forecasting strategies for the same forecasting model

Significance levels *** $\alpha = 0.001$, ** $\alpha = 0.01$, * $\alpha = 0.1$

between high solar generation at midday and zero solar generation at midnight. On the other hand, the advantage of tailored forecasting models may be due to the varying characteristics of supply and demand considered in the partially aggregated approach. Recurring daily and weekly patterns resulting from human behaviour characterise the demand. However, meteorological inputs primarily influence the renewable supply, and it is not likely that they follow the same patterns as the demand. Therefore, the partially aggregated approach allows the development of two separate models emphasising these individual components instead of accounting for the interaction between them as in the aggregated approach. The observed ranking of forecasting strategies can be observed for both the simple (NN) and complex (CNN) model. This observation suggests that the partially aggregated strategy is advantageous, regardless of the selected forecasting model. This information is crucial for network operators and energy suppliers. Whilst more advanced forecasting models will improve forecast performance in the future, our results show that the fundamental strategy behind these forecasts is also important and should not be forgotten.

The probabilistic forecast evaluation further strengthens the results of the deterministic forecasts. When considering the average CRPS, the forecasting strategies are ranked in the same order, the difference between the forecasting strategies is significant, and the simple and complex models are consistent in ranking. These results suggest that the partially aggregated strategy is not only the best strategy for performing standard deterministic point forecasts but is also the best strategy for capturing the uncertainty and forecasting the PDF of net load.

Although the selected parametric distributions are a good approximation for the empirical distributions, they are not ideal. Furthermore, the requirement for differentiable distributions for the gradient-based forecasting models eliminates some obvious distributions, i.e. the uniform distribution. Therefore, considering non-parametric distributions that better approximate the underlying empirical distribution and forecasting models capable of working with such non-parametric distributions could be interesting. Furthermore, whilst the probabilistic forecast quantifies the uncertainty and thus provides more information, it is not always simple to integrate into down stream operations,

Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 19 of 21

and they may be computationally more expensive. Therefore, the probabilistic and deterministic approaches may both be useful in future energy systems, depending on the required use case.

Finally, our approach has only considered one simulated future energy system and two different forecasting models. Whilst both the simple and complex forecasting models deliver similar results, it is possible that forecasting models optimised for each strategy may perform differently. Therefore, considering more data, optimising forecasting models for each given strategy, and comparing their performance in this situation should be considered. Furthermore, such analysis may also allow for the development of a forecasting model that uses the advantages different approaches. Such a model may be able to, for example, switch between the disaggregated and partially disaggregated approach depending on certain predetermined conditions.

Conclusion

The present paper investigates various net load forecasting strategies based on different levels of aggregation. We propose three forecasting strategies; an aggregated strategy that directly forecasts the net load, a partially aggregated strategy that forecasts the demand and supply separately, and a disaggregated strategy that forecasts the demand and the supply from each renewable generator in the system separately. We evaluate these three strategies on a simulated data set representing a realistic future energy system with a higher net load than today's systems. For this evaluation, we compare the deterministic and probabilistic forecast performance of a simple (NN) and complex (CNN) model.

Our evaluation shows that the partially aggregated strategy performs best for both probabilistic and deterministic forecasts regardless of the selected forecasting model. This result can be explained by a trade-off between the positive smoothing effects due to the aggregation of all individual suppliers and considering model-specific information with separate and tailored forecasting models for supply and demand. Despite these consistent results, we only evaluate two separate forecasting models for one simulated future energy system.

Therefore, in future work, we propose first considering further forecasting models optimised for each forecasting strategy and comparing these on multiple alternative future energy scenarios. Furthermore, these scenarios should include higher contributions from other renewable sources, such as solar and run of river. Based on these scenarios, forecasting models that switch between various forecasting strategies depending on predetermined conditions or optimally combine different levels of aggregation should be investigated. Secondly, non-parametric distributions and forecasting models that can successfully work with these non-parametric distributions should be considered in future work. Thirdly, since the present paper only focused on hourly forecasts based on historical weather data and aggregated for an entire country, considering different temporal resolutions, such as 15 min, forecast horizons, and different aggregation levels, could also be of interest. Finally, to enable automation in future smart grid settings our forecasting strategies should be integrated into an automated forecasting pipeline, e.g. with pyWATTS (Heidrich et al. 2021).

Beichter *et al. Energy Informatics* 2022, **5**(Suppl 1):19

Page 20 of 21

Declarations

Author details
[1]Karlsruhe Institute of Technology, Institute for Automation and Applied Informatics, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany. [2]Cluster of Excellence Machine Learning, University of Tübingen, AI Research Building, Maria-von-Linden-Str. 6, 72076 Tübingen, Germany.

Published: 7 September 2022

References

Barth L, Ludwig N, Mengelkamp E, Staudt P (2018) A comprehensive modelling framework for demand side flexibility in smart grids. Comput Sci Res Dev 33(1):13–23

Bergmeir C, Benítez J (2012) On the use of cross-validation for time series predictor evaluation. Inf Sci 191:192–213

Casella G, Berger RL (2021) Statistical inference. Cengage Learning, Brooks/Cole Cengage Learning, Belmont

Chen R-C, Dewi C, Huang S-W, Caraka RE (2020) Selecting critical features for data classification based on machine learning methods. J Big Data 7(1):1–26

Diebold F, Mariano R, Diebold F, Mariano R (1995) Comparing predictive accuracy. J Bus Econ Stat 13(3):253–263

Frysztacki MM, Brown T (2020) Modeling curtailment in Germany: how spatial resolution impacts line congestion. In: 2020 17th international conference on the European energy market (EEM), pp 1–7

Frysztacki MM, Hörsch J, Hagenmeyer V, Brown T (2021) The strong effect of network resolution on electricity system models with high shares of wind and solar. Appl Energy 291:116726

Garcia M, Kirschen D (2006) Forecasting system imbalance volumes in competitive electricity markets. IEEE Trans Power Syst 21(1):240–248

Gneiting T, Katzfuss M (2014) Probabilistic forecasting. Annu Rev Stat Appl 1:125–151

Gneiting T, Raftery A (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102(477):359–378

Gneiting T, Raftery A, Westveld A, Goldman T (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon Weather Rev 133(5):1098–1118

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge

Grömping U (2009) Variable importance assessment in regression: linear regression versus random forest. Am Stat 63(4):308–319

Heidrich B, Turowski M, Ludwig N, Mikut R, Hagenmeyer V (2020) Forecasting energy time series with profile neural networks. In: Proceedings of the eleventh acm international conference on future energy systems, pp 220–230

Heidrich B, Bartschat A, Turowski M, Neumann O, Phipps K, Meisenbacher S, Schmieder K, Ludwig N, Mikut R, Hagenmeyer V (2021) pyWATTS: Python workflow automation tool for time series. arXiv:2106.10157

Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Rozum I, Schepers D, Simmons A, Soci C, Dee D, Thépaut J-N (2018) ERA5 hourly data on single levels from 1979 to present. Copernicus

Climate Change Service (C3S) Climate Data Store (CDS). https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview. Accessed 31 July 2020

Hörsch J, Hofmann F, Schlachtberger D, Brown T (2018) PyPSA-Eur: an open optimisation model of the European transmission system. Energy Strategy Rev 22:207–215

Kaur A, Nonnenmacher L, Coimbra CF (2016) Net load forecasting for high renewable energy penetration grids. Energy 114:1073–1084

Kroposki B, Johnson B, Zhang Y, Gevorgian V, Denholm P, Hodge B-M, Hannegan B (2017) Achieving a 100% renewable grid: operating electric power systems with extremely high levels of variable renewable energy. IEEE Power Energy Mag 15(2):61–73

Li B, Zhang J, Hobbs B (2020) A copula enhanced convolution for uncertainty aggregation. In: 2020 IEEE power energy society innovative smart grid technologies conference (ISGT), pp 1–5

Persio LD, Cecchin A, Cordoni F (2017) Novel approaches to the energy load unbalance forecasting in the Italian electricity market. J Math Ind 7(1):1–15

Salem T, Kathuria K, Ramampiaro H, Langseth H (2019) Forecasting intra-hour imbalances in electric power systems. Proc AAAI Conf Artif Intell 33:9595–9600

Schmietendorf K, Peinke J, Kamps O (2017) The impact of turbulent renewable energy production on power grid stability and quality. Eur Phys J B 90(11):1–6

Sklar A (1959) Fonctions de repartition an dimensions et leurs marges. Publ Inst Stat Univ Paris 8:229–231

Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. In: Bartlett PL, Pereira FCN, Burges CJC, Bottou L, Weinberger KQ (eds) Advances in neural information processing systems 25: 26th annual conference on neural information processing systems 2012. Proceedings of a meeting held December 3–6, 2012, Lake Tahoe, pp 2960–2968

Sreekumar S, Sharma KC, Bhakar R (2020) Grey system theory based net load forecasting for high renewable penetrated power systems. Technol Econ Smart Grids Sustain Energy 5(1):1–14

Stratigakos A, Bachoumis A, Vita V, Zafiropoulos E (2021) Short-term net load forecasting with singular spectrum analysis and LSTM neural networks. Energies 14(14):4107

Taylor J (2006) Density forecasting for the efficient balancing of the generation and consumption of electricity. Int J Forecast 22(4):707–724

Unnewehr JF, Schäfer M, Weidlich A (2022) The value of network resolution—a validation study of the European energy system model PyPSA-Eur. In: 2022 open source modelling and simulation of energy systems (OSMSES), pp 1–7

Zachary S, Tindemans SH, Evans MP, Cruise JR, Angeli D (2021) Scheduling of energy storage. Philos Trans R Soc A 379(2202):20190435

Zhang Y, Rahbari-Asr N, Duan J, Chow M-Y (2016) Day-ahead smart grid cooperative distributed energy scheduling with renewable and storage integration. IEEE Trans Sustain Energy 7(4):1739–1748

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.