A Multi-Scheme Ensemble Using Coopetitive Soft-Gating With Application to Power Forecasting for Renewable Energy Generation

André Gensler, Bernhard Sick

Abstract

In this article, we propose a novel ensemble technique with a multi-scheme weighting based on a technique called coopetitive soft gating. This technique combines both, ensemble member competition and cooperation, in order to maximize the overall forecasting accuracy of the ensemble. The proposed algorithm combines the ideas of multiple ensemble paradigms (power forecasting model ensemble, weather forecasting model ensemble, and lagged ensemble) in a hierarchical structure. The technique is designed to be used in a flexible manner on single and multiple weather forecasting models, and for a variety of lead times. We compare the technique to other power forecasting models and ensemble techniques with a flexible number of weather forecasting models, which can have the same, or varying forecasting horizons. It is shown that the model is able to outperform those models on a number of publicly available data sets. The article closes with a discussion of properties of the proposed model which are relevant in its application.

Keywords: Ensemble techniques, Power forecasting, Multi model ensembles, Combining forecasts, Model selection, Time series, Data mining

1. Introduction

During the past decade, there has been a tremendous growth of the installed capacity of various forms of renewable energy generation. Wind turbines and photovoltaic powerplants contribute substantially to the new mix of energy, which consists of both non-renewable and renewable energy power plants. Most renewable energy sources have intermittent generation characteristics, i.e., the amount of generated power highly depends on the weather situation and it cannot be regulated the way it is possible with traditional power plants. In order to guarantee grid stability, the power generation and load in the grid have to be balanced, as the intermediate storage of electrical energy is both inefficient and expensive. Therefore, highly accurate algorithms have to forecast the available energy on various time horizons. Depending on the forecasting horizon, the forecast is of interest to different actors in the field, e.g., network operators, power plant operators, or electricity traders. Having an accurate power forecast, the technical and financial risks for all market participants can be reduced. The power forecasting process typically takes place in two steps:

1. A meteorological forecast for the desired area (the location of the renewable energy power plant) is computed. This forecast is called numerical weather prediction (NWP).
2. The NWP (and optionally other complementary data) is used to forecast the corresponding power generation of the renewable energy power plant using a power forecasting model.

In this article, we focus on the second step of the forecasting process, i.e., we assume the NWP as given. Naturally, the quality of the forecast typically decreases over time. Depending on the target application, the time horizons are categorized into
(very) short-term forecasts in the range up to hours, such as the intraday forecast, mid-term forecasts in the range of a few days (including the day-ahead forecast), and long-term forecasts in the range of weeks.

Traditionally, the computation of the renewable energy power generation from the NWP is performed using a physical model, i.e., with wind turbine / photovoltaic panel power curves. While these models yield good performance when the precise parameters for the power plant can be determined, they can easily exhibit systematic errors. Therefore, machine learning (ML) or statistical approaches became more important in the past decade. Machine learning models are “black box” techniques which train a model based on the historic power generation of the power plant and the respective corresponding NWP forecast. These trained ML models can then be used to perform a power forecast for future points in time using the NWP forecast for this point in time. There are a wide variety of models which have exhaustively been analyzed, e.g., in [10, 17, 43]. Typical models are neural networks, multi-linear regressions, or support vector based methods.

Research has shown that the combination of single forecasting models into a so-called forecasting ensemble can improve the forecasting accuracy (e.g., in [38]). Traditionally, ensemble techniques are based on diversity principles, such as data diversity, parameter diversity, or structure diversity. Furthermore, in the area of power forecasting, ensembles are also created using multiple weather forecasting models or lagged ensembles. Ensemble techniques typically aim at exploiting a particular model principle, namely to

- aggregate multiple weather forecasts of an ensemble prediction system (EPS), or
- aggregate multiple weather forecasting model predictions in a multi-model ensemble (MME), or
- aggregate the forecast for the same forecasting time period from different forecast origins (time-lagged ensemble).

In practice, all forecasts may benefit from a combination of those ensemble principles. As we will show in the following section, the combination of multiple ensemble principles is rarely applied. Therefore, in this article we propose a novel ensemble technique which combines all the above paradigms using a combination technique which we call coopetitive soft-gating. Coopetition, see e.g., [26] is a term originally emerging from economic research which describes the concept of competitors achieving a joint advantage by cooperating. We aim to include multiple ensemble paradigms in the way of using the strengths of the global quality of different weather forecasting models and forecasting algorithms, exploiting weather-situation dependent strengths of those models, and making use of the lead time-dependent properties of each weather and forecasting model. This enables the ensemble to be employed on a variety of forecasting time periods, i.e., day-ahead, or intraday forecasts.

The remainder of this article is structured as follows: In Section 2, we present the state of the art in forecasting algorithms and ensemble techniques. Section 3 highlights the way typical ensemble methods differ and how they are computed. In Section 4, we introduce the coopetitive soft gating principle and we show how to apply coopetitive soft gating to create an ensemble. Section 5 demonstrates the performance of the proposed approach in comparison to other forecasting and ensemble techniques for intraday and day-ahead forecasting, and using a varying number of weather forecasting models. Finally, our insights are summarized in Section 6.

The article closes by giving an outlook on future research directions.

2. Related Work

This section details the state of the art in the area of power forecasting, ensemble methods in general, and ensemble methods for power forecasting. Some
good write-ups on the area of power forecasting are, e.g., given in [10, 24, 43]. As each application in power forecasting typically is tied to a certain forecasting horizon, a categorization by the forecasting horizon does make sense. A good categorization is, e.g., laid out in [19]. Some articles deal with forecasting for particular forecasting horizons, e.g., (very) short-term [5, 33, 17], mid-term [29], and long-term [6, 22]. An alternative categorization is by the methodical category of the power forecasting model. Possible categories can be, e.g., physical models, such as wind turbine / photovoltaic power curves, statistical approaches, such as variants of the ARMA model, and machine learning models, such as artificial neural networks, all of which are discussed in the articles mentioned above.

A category of models which are related to machine learning models are ensemble forecasting models. Ensemble forecast is an umbrella term for the aggregation of multiple forecasts to an overall prediction. Explanatory reasons why ensembles work are founded in bias-variance decomposition [23], further reasons for the popularity of ensemble methods are also detailed in [8]. A recent survey gives an overview of the most popular forms of ensembles in the area of classification and regression [39]. Ensembles can be formed using a number of principles, the most important ones are:

- **Data Diversity:** The ensemble is formed by training each ensemble member on a different subset of the training data. Bagging [3], boosting [10], or random forests [4] are ensembles of this type.

- **Parameter Diversity:** The ensemble is created using different model parameters of the same forecasting model. Multiple Kernel Learning methods [18] are part of this diversity technique.

- **Structure Diversity:** Different types of forecasting models are used to create the ensemble. These models, sometimes also called heterogeneous ensembles, are detailed, e.g., in [28].

An overview of ensemble methods for regression is also given in [28].

A survey of ensemble techniques in the area of power forecasting is detailed in [35]. The literature suggests that ensemble methods can not only yield superior results to single models (e.g., [45, 44]), ensemble forecasts can also be used to create probabilistic forecasts to assess uncertainty (e.g., [35, 2, 42, 45]). This can, for instance, be exploited to estimate the required reserve energy [27, 21]. In the area of meteorological sciences, the term ensemble typically is used to characterize the form of aggregation of a number of NWPs in an ensemble. We will highlight the state of the art for the most popular ensemble forms in the following. More details on the computation of each ensemble type can be found in Section 3. All presented forms of ensembles have in common that the combination of the ensemble member forecasts to an overall point forecast is performed in the power domain, i.e., after applying the power forecasting model.

**Ensemble Prediction Systems (EPS),** sometimes also called single-model ensembles, are created using a systematic variation of the perturbation parameters of the weather forecasting model generating processes, yielding different NWP. The goal of such an EPS is to assess the possible weather outcomes by including an explicit model spread which reflects the stochastic nature of the forecasting task. It thereby in principle is a data diversity ensemble, which is, however, created through varying the parameters of the NWP generating process in the sense of a parameter diversity ensemble. These forms of ensemble forecasts are typically conducted by a weather forecasting model provider. Each of the weather forecasting models is then used for the power generation forecast using a power forecasting model. An EPS is employed by [44] to forecast electrical load using neural networks with multiple scenarios for the weather parameters. The study is performed for a number of load times up to ten days. In [35], the ensemble forecast is used to predict the forecasting skill using by evaluating the spread of the EPS, the EPS is also compared to lagged ensembles (explanation see below). [41] conduct a comparative study between two ensemble prediction systems regarding their forecasting accuracy for wind power. In [46], an EPS is used to create probabilistic forecasts to investigate extreme weather situations and ramp events.

**Multi-Model ensembles (MME),** sometimes also called “poor-man’s ensembles”, refers to the combination of (typically deterministic) point forecasts of different weather forecasting model providers. In principle, each of the single forecasts is created independently. MMEs have characteristics different from EPS, as each ensemble member yields the most likely point forecast and does not try to explicitly include a model spread. MMEs thus have data diversity characteristics. Furthermore, multi-model ensemble members can differ in their structure (e.g., different number of NWP parameters, etc.). [32]
use MME for photovoltaic forecasting, the creation of prediction intervals for uncertainty assessment is also investigated. An MME of four climate forecast systems using coupled ocean-atmosphere models is investigated in [17] with particular focus on forecast verification. The performance of MMEs is compared to EPS forecasts in [50].

Power forecasting model ensembles (PME) make the basic assumption that a single forecasting model cannot be an optimal estimator for a given data set due to a too simple model structure, varying intrinsic uncertainties depending on the data set, or poor generalization characteristics of certain models. This form of ensemble then uses the predictions of multiple (independently trained) power forecasting algorithms, typically given a single weather forecast, to combine to an overall forecast in the form of a data, parameter, or structure diversity ensemble. In [41], ensembles of neural networks are used for load forecasting using a number of different combination methods including principal component based methods. Bayesian adaptive model combination investigating a number of different neural network types is performed in [25] including a unified approach for model selection. An ensemble forecasting technique with adaptive weighting of the ensemble members using a locality assessment of the weather situation is investigated on a variety of models in [13], a variant for probabilistic forecasts is presented in [33]. “Standard” ensemble techniques, such as bagging [3] or boosting [41], also fall under the category of PME. These techniques can also easily be combined with other ensemble techniques.

Time-lagged ensembles (TLE) use a repetitive forecast of the same absolute point in time computed from different forecasting origins to aggregate to an ensemble. TLE can be computed using only a single power forecasting model and a single weather forecasting model in the form of a data diversity ensemble. [30] uses lagged ensembles for the aggregation of high-resolution forecasts to achieve the effect of spatial averaging. Lagged ensembles are also frequently used to assess the uncertainty of a forecast, e.g., using risk-indices [34, 35]. Table 1 summarizes the possible ensemble types by the number of weather and power forecasting models involved.

Table 1: Categorization of ensemble types by number of weather and power forecasting models.

| Ensemble Type | Weather Forecasting Models | Power Forecasting Models |
|---------------|---------------------------|--------------------------|
| No Ens.       | 1                         | 1                        |
| EPS           | > 1                       | ≥ 1                      |
| MME           | > 1                       | > 1                      |
| PME           | 1                         | > 1                      |
| TLE           | 1                         | 1                        |

Means that they aggregate single forecasts of ensemble members to an overall forecast. In addition to a more precise determination of a forecast, the assessment of the forecasting uncertainty is easier using ensemble techniques, e.g., using Bayesian model averaging [4].

3. Ensemble Computation Methods for Power Forecasting

A forecast is performed at forecasting time $t$ which is also called the forecasting origin. The forecast is normally conducted for a number of forecasting time steps (or lead times) with fixed time increment $\Delta$ and

minimum lead time $t + k_{\text{min}} \cdot \Delta$,
currently considered lead time $t + k \cdot \Delta$,
maximum lead time $t + k_{\text{max}} \cdot \Delta$,

where the lead times $k$ typically are a set with $k \in \{k_{\text{min}}, \ldots, k_{\text{max}}\} \in \mathbb{N}$, where $k_{\text{min}}, k_{\text{max}} \in \mathbb{N}$ and $k_{\text{min}} \leq k_{\text{max}}$. We will abbreviate the currently considered lead time as $t_k = t + k \cdot \Delta$. The minimum lead time $k_{\text{min}}$ and the forecasting horizon $k_{\text{max}}$ can be chosen depending on the application. An overview of the most important variables used throughout this article is also given in the nomenclature section at the beginning of the article. The process of creating a forecast is typically conducted using a power forecasting model which transforms the input data of a predictor $x_{t_k,t} \in \mathbb{R}^D$ ($D$-dimensional input data $x$ for forecasting time $t_k$ created at forecasting origin $t$) to a forecast of a target predictand $\hat{y}_{t_k,t}$ in the form

$$\hat{y}_{t_k,t} = f(x_{t_k,t}, \Theta), \quad (1)$$

where $\Theta$ describes the set of model parameters of model $f$. In case the power forecasting model is an ensemble, the deterministic forecast is in many cases created using a weighted sum of single fore-
casts in a post-processing step as
\[ \tilde{y}_{t|k} = \sum_{j=1}^{J} w^{(j)} \hat{y}_{t|k,j}, \]
where \( J \) is the number of ensemble members, \( w^{(j)} \) is the respective weighting coefficient of the power forecasting model, and \( \hat{y}_{t|k,j} \) is the forecast of the \( j \)-th ensemble member. For all types of ensembles, we require that the sum of weights has to fulfill
\[ \sum_{j=1}^{J} w^{(j)} = 1, \quad w^{(j)} \in \mathbb{R}_0^+ \]
in order to not over- or underestimate the forecast on average. Ensembles create the final prediction by aggregating the forecasts using Eq. (2). There are two basic approaches of setting weights:

- **Cooperation / Weighting**: One possibility to create an ensemble forecast is by letting the single ensemble members cooperate in creating the final point estimate. In the easiest case, the weights \( w^{(j)} \in [0, 1] \) can be chosen equally, i.e., \( w^{(j)} = \frac{1}{J} \). Other possibilities are, e.g., to set them proportional to their overall average forecasting quality, if known. The weight values typically are static in this technique, they do not change after being set.

- **Competition / Gating**: In this approach, in each situation one model succeeds in competing against the other models, i.e., \( w^{(j)} \in \{0, 1\} \) and \( \sum_{j=1}^{J} w^{(j)} = 1 \). The challenge of this approach consequently is in deciding which power forecasting model should win the competition for a particular forecast. The weight values are dynamic in this technique, they vary depending on some defined criterion.

In the area of power forecasting, the predictor \( x_{t|k} \) typically is an NWP forecast, the predictand is the expected power generation \( \hat{y}_{t|k} \). Typical applied ranges in the area of power forecasting are the so-called day-ahead forecast (where \( k_{\text{min}} = 24, k_{\text{max}} = 48, \Delta = 1h \)), or the intraday forecast (with \( k_{\text{min}} = 1, k_{\text{max}} = 24, \Delta = 1h \)). For some operational day-ahead forecasts, this time period may vary. The forecast \( \tilde{y}_{t|k} \) of each ensemble member can be computed in a number of ways depending on the form of the ensemble:

1. In case an ensemble prediction system (EPS) is used, a single power forecast is created with
\[ \hat{y}_{t|k} = f(x_{t|k}; \Theta). \]
For this type of ensemble, the input NWP values are changing, however, the type and parametrization of the power forecasting model normally remains the same. As for many EPS forecasts all NWP ensemble members are assumed to have an equal probability of being correct, the values of \( w^{(j)} \) are often chosen equally, i.e., \( w^{(j)} = \frac{1}{J} \), if a deterministic forecast is desired.

2. For Multi-Model Ensembles (MME), a single power forecast is created using
\[ \tilde{y}_{t|k} = f_j(x_{t|k}; \Theta^{(j)}). \]
For MMEs, the NWPs may be of a different form (and / or they may have different number of dimensions). The power forecasting models consequently have a different structure, thus, different model parameters \( \Theta^{(j)} \) have to be chosen for each ensemble member. This does not necessarily have to be the case for an EPS. MME members often have a varying overall quality due to a varying weather forecasting model quality. The corresponding weights \( w^{(j)} \) therefore typically have different values which can be set according to the expected quality of the models (e.g., by testing the model on some historic time period) in order to maximize the ensemble quality.

3. Forecasting model ensembles (PME) typically operate on the same input data in the form
\[ \tilde{y}_{t|k} = f_j(x_{t|k}^{(j)}; \Theta^{(j)}). \]
This form of ensemble can be computed on a single NWP (then, all \( x_{t|k}^{(j)} \) are equal), or on a subset of the NWP parameters (in subspace methods \([20]\), for instance). In this case, the dimensionality of the NWP is \( x_{t|k}^{(j)} \in \mathbb{R}^{D'} \) with \( D' < D \). The main differentiating factor of the ensemble members is their form of data, parameter, or structure diversity (see Section 2).

4. Finally, (time-)lagged ensembles (TLE) typically operate on the same NWP using the same power forecasting model. They use forecasts for the
forecasting time $t_k$ from different forecasting origins $t - \Delta \cdot v_j$ in the form

$$y_{j_k | t} = f(x_{i_k | t - \Delta \cdot v_j} | \Theta),$$

such as shown in Fig. 1. The value of $v_j \in \mathbb{N}_0^+$ denotes the amount of lag of the $j$-th ensemble member. Typically, the corresponding weights $w^{(j)}$ are chosen in a form that smaller values of $v_j$ have a higher weight (as the amount of time-lag is smaller for those forecasts, which typically correlates with an increased precision of the forecast).

It is being shown that each type of ensemble aims at exploiting a different aspect of the present data. As can be seen, the two basic approaches weighting and gating both have distinct advantages, however they are either not dynamic (weighting), or do not allow for cooperation (gating). In the following, we present a weighting scheme which combines the advantages of both cooperation and competition in the form of a coopetitive soft gating technique. This scheme is applied in an ensemble structure that combines multiple of the ensemble principles laid out above.

### 4. Coopetitive Soft Gating Ensemble

As outlined in Section 3, ensemble techniques typically aim to exploit a single principle for ensemble generation. The proposed technique aims at using a multitude of weighting principles. The principal structure of the proposed ensemble technique is visualized in Fig. 2. The weighting of the ensemble remains the same as in Eq. (2). However, we have a hierarchical ensemble structure: For each weather forecasting model $\psi = 1, \ldots, \Psi$ (which can be an arbitrary NWP of an EPS, MME, or TLE, e.g., of an intraday or day-ahead model, for a particular time step to be forecasted), a number of power forecasting models $\varphi = 1, \ldots, \Phi$ are used to forecast the target predictand for each weather forecasting model. The power forecasting models do not necessarily have to be the same for each weather forecasting model, but, for the sake of easier understanding, we will use the same type and number of power forecasting models for each weather forecasting model here. The overall number of ensemble participants $J$ consequently is $J = \Psi \cdot \Phi$. The individual predictions of each power forecasting model are then aggregated and fused to an overall forecast in a post-processing step according to Eq. (2). The main innovation here is the way the single weights $w^{(j)}$ are constructed. In order to clarify the origin of each weighting term with respect to the weather forecasting model $\psi$ and the power forecasting model $\varphi$, we define the weight of an ensemble member as $w^{(j)} = w^{(\psi, \varphi)}$. The idea is as follows: For each ensemble participant we construct the coopetitive soft gating (explained in Section 4.1) considering the following three aspects for both, power forecasting models and weather forecasting models, respectively (leading to 6 aspects):

1. **Global Weighting**: The ensemble weights are determined for the respective model regarding the overall observed performance of a model during ensemble training. This is a fixed weighting term. Thereby, overall strong models have more influence than weaker models. This form of weighting is described in Section 4.2.

2. **Local Soft Gating**: The ensemble members are weighted depending on the model input (the NWP forecast) $x_{i_k | t}$. This form of weighting assesses the quality of a model considering the current input, i.e., a local quality assessment is performed. The idea is that a number of models may have different strengths in a different set of weather situations (e.g., due to ensemble diversity effects). This form of weighting is described in Section 4.3.

3. **Lead time-dependent Soft Gating**: Models may have a lead time-dependent quality development. The goal of this form of soft gating is to weigh the model depending on the lead time $k$. In the case of power forecasting models, methods such as the persistence method perform very well on short time horizons, while they quickly lose their quality for longer time-horizons. Additionally, weather forecasting models such as intraday models typically perform very strong on short time horizons due to very recent weather measurements. This form of weighting is described...
Figure 2: Overview of the proposed ensemble model. As can be seen from the figure, the ensemble has a hierarchical structure, where a number of predictors from Ψ weather forecasting models (EPS, MME, or TLE) are used to forecast a common target predictand. Therein, Φ power forecasting models each create a power forecast for a particular weather forecast. The total number of $J = Ψ \cdot Φ$ forecasts are combined in a post-processing stage using coopetitive soft gating. The ensemble member weights are computed as described in Section 4.

The overall weighting term for each ensemble member can be described as

$$w^{(j)} = w^{(ψ, ϕ)} = \frac{w^{(ψ)} \cdot w^{(ϕ|ψ)}}{\sum_{ψ^*=1}^{Ψ} \sum_{ϕ^*=1}^{Φ} w^{(ψ^*)} \cdot w^{(ϕ^*|ψ^*)}},$$

(8)

where $w^{(j)} = w^{(ψ, ϕ)}$ is the overall weight for ensemble member $j$ computed using power forecasting model $ϕ$ and weather forecasting model $ψ$. The weights $w^{(ψ)}$ are the weather forecasting model dependent weights, $w^{(ϕ|ψ)}$ are the power forecasting model dependent weighting factors of power forecasting model $ϕ$ computed on weather forecasting model $ψ$. The denominator is a normalization term which ensures $\sum_{j=1}^{J} w^{(j)} = 1$, see Eq. (3). The weights of the weather forecasting model can be decomposed into

$$w^{(ψ)} = w^{(ψ)}_g \cdot w^{(ψ)}_l \cdot w^{(ψ)}_k,$$

(9)

while the weights of the power forecasting model are computed with

$$w^{(ϕ|ψ)} = w^{(ϕ|ψ)}_g \cdot w^{(ϕ|ψ)}_l \cdot w^{(ϕ|ψ)}_k.$$

(10)

The indices $g, l, k$ denote the respective weighting aspects global weighting $g$ (Section 4.2), local soft gating $l$ (Section 4.3), or lead time-dependent soft gating $k$ (Section 4.4) for both, weather forecasting model $ψ$ and power forecasting model $ϕ$. Using multiple weather forecasting models and multiple power forecasting models, the overall number of weights per weather forecasting model and power forecasting model combination consequently adds up to six. The ensemble training process is described in Section 4.5. In addition to the description of the weighting factors in the different sections, Fig. 5 illustrates an example of the functionality of the different weighting aspects and the overall proposed technique. Fig. 5a shows the development of the global weighting aspect, Fig. 5b shows the development of the local soft gating, and Fig. 5c illustrates the lead time-dependent weighting aspects. Section 4.6 gives an overall application example of the proposed coopetitive soft gating ensemble (CSGE) technique.

4.1. Coopetitive Soft Gating Principle

The goal of coopetitive soft gating is to weigh ensemble members according to their performance in a way that combines the properties of both principles, weighting and gating. The weighting technique should allow for a flexible application to all
of the three weighting principles (global, local, and lead time-dependent) proposed in the previous section. The weights are built from a quality estimate of the different ensemble members, which typically is represented in the form of an error, e.g., the root mean square error (RMSE). Having a quality estimate for each ensemble member, the proposed weighting technique should fulfill the following requirements:

1. It must return a score ordered from high weights (low error) to low weights (high error), i.e., the inverse of what the error score is initially represented in.
2. It must be able to weigh errors nonlinearly, as the optimal weighting possibly cannot be represented in a linear relationship. The amount of nonlinear weighting should be controllable by the user.
3. It must be insensitive to the value range of the error scores, i.e., it should only factor in the relative quality differences of the contributing ensemble members.
4. It must retain the value range of the ensemble prediction, i.e., it has to fulfill Eq. (3).

We fulfill criteria 1.--3. using the weighting function

\[ \varsigma_\eta(\Omega, \omega) = \frac{\sum_{n=1}^{N} \Omega_n}{\omega^\eta + \epsilon}, \quad \eta \in \mathbb{R}_0^+, \quad \] (11)

where \( \Omega \in \mathbb{R}^N \) is a tuple of values containing all \( N \) reference quality estimates, and \( \omega \) is an arbitrarily evaluated point of the function. Typically, \( \omega \) can be chosen to be an element of \( \Omega \). The real number \( \eta \geq 0 \) represents the amount of exponential weighting. The higher the value of \( \eta \), the higher the weight of the models models with low errors. \( \epsilon \) is a very small number that avoids division by 0 in the unlikely case that \( \omega \) is 0. Assuming that \( \omega \in \Omega \) and that \( \varsigma_\eta(\Omega, \omega) \) is computed for each element in \( \Omega \), we achieve criterion 4. by adjusting the weighting function by normalizing with the sum of the weights in \( \Omega \), i.e.,

\[ \varsigma_\eta(\Omega, \omega) = \frac{\varsigma_\eta(\Omega, \omega)}{\sum_{n=1}^{N} \varsigma_\eta(\Omega, \Omega_n)}. \] (12)

This form of the coopetitive soft gating formula can further be simplified to

\[ \varsigma_\eta(\Omega, \omega) = \frac{1}{(\omega^\eta + \epsilon) \sum_{n=1}^{N} (\Omega_n)^\eta + \epsilon} \] (13)

An example of how Eq. (13) works depending on the way \( \eta \) is set is shown in Fig. 3. The example uses exemplary quality estimates in the range of 0.2 -- 0.42 which represent the quality of different ensemble members. As can be seen in Fig. 3 when increasing \( \eta \), the weights of the well-performing ensemble members do increase. An advantage of this form of weighting is that it only has a single weighting parameter \( \eta \) for optimization (unlike methods derived from exponential functions in the form \( f(x) = e^{Ax+B} \), e.g., in [36]). The following sections use the coopetitive soft gating formula of Eq. (13) to compute the ensemble weights \( w^{(j)} \) for each of the three weighting aspects (global, local, and lead time-dependent).

4.2. Global Weighting

The global weights \( w^{(g)} \) are fixed weights which are determined during ensemble training. The proposed weighting technique aims at weighting the ensemble members according to their performance using coopetitive soft gating. A simple yet popular measure to assess the quality of an ensemble member can, for instance, be the root mean squared error (RMSE) computed by

\[ \psi^{(g)} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (\hat{y}_n^{(g)} - o_n)^2}, \] (14)

\[ \epsilon^{(g)} = \hat{y}_n^{(g)} - o_n \] (15)

on a data set containing \( N \) samples. The \( o_n \) are actual power measurements corresponding to a forecast \( \hat{y}_n^{(g)} \) computed on weather forecasting model \( \psi \) and power forecasting model \( \varphi \).

The global weight \( w^{(g)} \) can be computed from the coopetitive soft gating formula

\[ w^{(g)} = \varsigma_0(g^{(g)}, R^{(g)}) \] (16)

\[ g^{(g)} = (R^{(1)}, \ldots, R^{(g)}, \ldots, R^{(G)}) \] (17)

where \( \varsigma_0 \) is the coopetitive soft gating function of Eq. (13) for a power forecasting model \( \varphi \) computed on weather forecasting model \( \psi \).

The global forecasting ability of a weather forecasting model \( \psi \) can be observed only indirectly as actual weather measurements typically are not available. As an estimate, the overall quality of a weather forecasting model can be determined using the average quality of all power forecasting models.
for the particular weather forecasting model, i.e.,

\[ w^{(\psi)}_g = s_n(p, p^{(\psi)}), \]  

\[ p^{(\psi)} = \frac{1}{\Phi} \sum_{\psi=1}^{\Phi} R^{(\psi)}, \]  

\[ p = (p^{(1)}, \ldots, p^{(\psi)}, \ldots, p^{(\Psi)}). \]

An example for the influence of the global weighting term for a number of power forecasting models is given in Section 4.6. This weighting term is computed once during ensemble training and can then be reused for every forecast.

4.3. Local Soft Gating

The second weighting term depends on the values of the current NWP, i.e., the weighting is realized depending on the particular characteristics of the weather situation. The NWP forecast can be seen as a point in a feature space which characterizes the weather situation. The basic assumption is that both, weather and power forecasting models, may have strengths and weaknesses in varying areas of the feature space. This is due to the fact that different power forecasting algorithms yield different errors in certain areas of the feature space due to structure, data, or parameter diversity effects. In particular for sparsely covered areas of the feature space (e.g., storms for wind turbines, or Sahara dust for photovoltaic plants), this effect may become more prominent. Different NWP forecasts of certain weather forecasting models may also have a different precision depending on the particular situation. Using competitive soft gating, we aim to exploit the advantages of each model in particular observed situations during model training.

In order to obtain local weights \( w^{(\psi)}_l \) and \( w^{(\varphi|\psi)}_l \), the neighborhood of a weather forecast \( x_{t_k|t} \) has to be assessed. Similar historic weather situations are found with respect to a (historic) data set tuple \( X_H = (x_1, \ldots, x_N) \) with \( x_n \in \mathbb{R}^D \), which is used during ensemble training. The proposed ensemble algorithm is able to work with an arbitrary local quality assessment technique. Here we demonstrate the application with a simple nearest neighbor technique. Other techniques for assessing locality, such as multi-linear interpolation, are investigated, e.g., in [13].

A simple yet effective technique for locality assessment is a nearest neighbor algorithm. In order to assess the local quality of a forecast \( x_{t_k|t}^{(\psi)} \), its \( C \) nearest neighbors are searched in \( X_H \) in the way

\[ \alpha^{(\psi)} = \text{knn}(x_{t_k|t}^{(\psi)}, X_H, C), \quad \alpha^{(\psi)} \in \mathbb{N}^C, \]

where \( \alpha^{(\psi)} \) is a set containing the indices of the \( C \) nearest neighbors. Here, we use the Euclidean distance as distance metric on standardized input dimensions, though the use of more advanced distance metrics may further improve the local quality assessment. The average local quality can be assessed using

\[ q^{(\varphi|\psi)} = \frac{1}{C} \sum_{a \in \alpha^{(\psi)}} |e_a^{(\varphi|\psi)}|, \]

where \( e_a^{(\varphi|\psi)} \) is the error computed using Eq. (15) of the item at index \( a \) in the historic set using the
forecast of ensemble member with power forecasting model \( \varphi \) computed on weather forecasting model \( \psi \). From this local error score \( q(\varphi, \psi) \), the weight \( w_t^{(\varphi, \psi)} \) is computed for each power forecasting model using coopetitive soft gating of Eq. (13) in the form

\[
w_t^{(\varphi, \psi)} = \varsigma_t(q_t^{(\psi)}, q^{(\varphi)}), \quad (23)
\]

\[
q^{(\psi)} = (q^{(1)}(\psi), ..., q^{(\varphi)}(\psi), ..., q^{(\Phi)}(\psi)), \quad (24)
\]

where each value of \( q^{(\varphi)} \) is computed using Eq. (22). In the same fashion as for the global weighting (Eq. (18)), the relative quality for each forecasting time step using all available power forecasting models, i.e.,

\[
w_t = \varsigma_t(q_t, \hat{q}^{(\psi)}), \quad (25)
\]

\[
\hat{q}^{(\psi)} = \frac{1}{\Phi} \sum_{\varphi=1}^{\Phi} q^{(\varphi)}, \quad (26)
\]

\[
q = (q^{(1)}, ..., q^{(\varphi)}), \quad (27)
\]

Figs. 4a – 4c show the local quality of a number of power forecasting models in Voronoi diagrams, where green represents areas of low error and red color indicates areas with high error. The axes are given by the two most important principal components in order to better visualize the \( D \)-dimensional NWP feature space. Using the locality assessment technique of Eq. (23), each model is weighted depending on the position in the feature space in a way that reduces the overall error in the ensemble. Fig. 4d shows an example of the resulting ensemble error. It should be kept in mind that in the shown case, the improvement for the training data set is displayed. An example for the development of the local weights in a forecasting time period is shown in Section 4.6. This weighting term has to be computed during ensemble application for every NWP.

An advantage of the knn technique is that no model training is required (in the basic form if no feature subspace is selected or feature weighting is applied). However, as the data set \( X_H \) serves as basis for the locality assessment, it has to be searched in every iteration, which usually does not scale optimally if no search heuristics are being employed. The knn approach is therefore particularly useful for smaller data sets. In 13, a technique for locality assessment based on multi-linear interpolation is introduced which does require a training phase. However, during model application it no longer requires the data set \( X_H \). This technique is therefore well-suited for larger data sets. Regarding ensemble forecasting quality, both approaches turned out to behave similar.

4.4. Lead Time-Dependent Soft Gating

The lead time-dependent weighting components \( w_k^{(\varphi, \psi)} \) factor in the quality development of a model for each lead time \( k \). The idea is to weigh models according to their lead time-dependent performance. In the area of power forecasting, a prominent example for approaches with time step dependent performance is the persistence method, which performs well on very short lead times only.

The idea is to create a weight per lead time \( k \) by evaluating the quality differences of a number of coopetitive models. For the creation of this form of weighting a training data set for a particular lead time \( k \) – for which a number of forecasts \( \Psi (k, \varphi, \psi) \in \mathbb{R}^{N'} \) are created using weather forecasting model \( \psi \) and power forecasting model \( \varphi \) can be denoted as

\[
\Psi (k, \varphi, \psi) = \{ \hat{y}_{11}^{(\psi)}(t_1), \hat{y}_{12}^{(\psi)}(t_1), ..., \hat{y}_{kn}^{(\psi)}(t_1), ..., \hat{y}_{n_1}^{(\psi)}(t_1), ..., \hat{y}_{n_N}^{(\psi)}(t_{N'}), \}
\]

where \( N' \) is the number of evaluated elements with the currently evaluated lead time \( k \). The estimated quality for a particular lead time can then be created using an error metric, e.g., based on the RMSE

\[
R_k^{(\varphi, \psi)} = \sqrt{\frac{1}{N'} \sum_{n=1}^{N'} (\Psi_n^{(k, \varphi, \psi)} - o_n)^2}. \quad (29)
\]

The quality of the particular forecasting time step \( k \) in relation to other forecasting time steps of the same model can then denoted as

\[
r_k^{(\varphi, \psi)} = \frac{R_k^{(\varphi, \psi)}}{k_{\text{max}} - k_{\text{min}} + 1} \sum_{k'=k_{\text{min}}}^{k_{\text{max}}} R_{k'}^{(\varphi, \psi)}. \quad (30)
\]

Then, the weighting factor \( w_k^{(\varphi, \psi)} \) is computed for each forecasting time step using the generalized coopetitive soft gating formula of Eq. (13) in relation to other members of the ensemble

\[
w_k^{(\varphi, \psi)} = \varsigma_t(r_k^{(\varphi)}, r_k^{(\psi)}), \quad (31)
\]

\[
r_k^{(\psi)} = (r_k^{(1)}(\psi), ..., r_k^{(\varphi)}(\psi), ..., r_k^{(\Phi)}(\psi)). \quad (32)
\]

Again, the time-dependent weather forecasting model qualities are estimated using the overall
power forecasting models with

\[ w_k^{(\psi)} = s_\eta(\mathbf{r}_k, \tilde{r}_k^{(\psi)}), \quad \text{where} \]
\[ \tilde{r}_k^{(\psi)} = \frac{1}{\Phi} \sum_{\varphi=1}^{\Phi} r_k^{(\varphi)^{(\psi)}}, \]
\[ r_k = (r_k^{(1)}, \ldots, r_k^{(\psi)}, \ldots, r_k^{(\Psi)}). \]

In case there is little data available for the training process, a smoothing over weights in neighboring lead times can be applied in order to avoid noisy weights. An example of the effect of this form of weighting is described in Section [12]. This weighting term is computed once during ensemble training for every lead time and can then be reused for every forecast.

### 4.5. Model Fusion and Ensemble Training

As stated previously, the overall weighting \( w^{(j)} \) of each ensemble member \( j \) is computed using Eq. [8].

The main parameter of the cooperative soft gating ensemble (CSGE) algorithm is the hyperparameter \( \eta \) of the cooperative soft gating formula. Depending on the forecasting task and the data set, the appropriate value of \( \eta \) may differ. Furthermore, the value of \( \eta \) for each weighting aspect (global weighting, local soft gating, time-dependent soft gating) may vary. The value of \( \eta \) therefore should be chosen for each weighting aspect.

In principle, the total number of \( \eta \)-parameters \( S \) are the three aspects (global, local, temporal) for each weather forecasting model. Additionally, for each of the \( \Psi \) weather forecasting models, the three weighting aspects for the corresponding power forecasting models have to be computed, which adds up to

\[ S = 3 \cdot \Psi + \frac{3}{\text{power forecasting models}}. \]

Assuming the same number and types of power forecasting models for each weather forecasting model, the weighting parameter can be treated identically for each weather forecasting model. Therefore, the number of optimization parameters can be reduced to

\[ S = \frac{3}{\text{power forecasting models}} + \frac{3}{\text{weather forecasting models}} = 6 \]

parameters. The tuple of coopetitive soft gating parameters \( \eta = (\eta_1, \ldots, \eta_s, \ldots, \eta_S) \) can then be optimized using an arbitrary optimization algorithm solving the problem

\[
\begin{align*}
\text{minimize} \quad & \frac{1}{N} \sum_{n=1}^{N} [o_n - f_{\text{CSGE}}(\mathbf{x}_n, \boldsymbol{\eta})]^2 + \zeta \cdot \sum_{s=1}^{S} \eta_s,
\end{align*}
\]

where \( f_{\text{CSGE}}(\mathbf{x}_n, \boldsymbol{\eta}) = \sum_{j=1}^{J} w^{(j)} \cdot \hat{y}^{(j)}_n \), subject to each \( \eta_s \geq 0 \),

where \( f_{\text{CSGE}} \) is the forecast of the overall CSGE forecasting function, \( N \) are the overall evaluated points of a validation data set, \( w^{(j)} \) are the weights of a particular forecast computed using Eq. [8] using the hyperparameters \( \eta \), and \( \zeta \geq 0 \) is a regularization parameter. In this case, a squared error is chosen for optimization, however, depending on the application, also other forms of error functions can be used.

An advantage of the proposed technique is that it is a post-processing technique. Therefore, while the single forecasts of each ensemble member \( j \) are weighted differently using \( w^{(j)} \), the forecasts \( \hat{y}^{(j)} \) of each of the \( j \) ensemble members remains constant, no matter what the value of \( \eta \) may be. The values of \( \hat{y}^{(j)} \) therefore only have to be computed once for the evaluated data set during ensemble training. Furthermore, the single weights change gradually when varying the parameters in \( \eta \). Consequently, we end up with a smooth (continuously differentiable) optimization function.

### 4.6. Application Examples of the CSGE Technique

This section describes two application examples of the final CSGE algorithm. The first example shows the application of the CSGE algorithm for intraday forecasting \( (k_{\text{min}} = 1, k_{\text{max}} = 24, \Delta = 1h) \) using a single weather forecasting model and an ensemble of three forecasting algorithms, namely an ANN, a linear regression, and a persistence forecast. Fig. [5a] shows the single weighting aspects and the overall weights over time. As there exists just a single weather forecasting model, the number of weighting aspects is reduced to three. The global weights are shown in Fig. [5a]. The algorithm weighs the single algorithms according to their expected quality (ANN best, persistence worst). These weights remain constant over time. The local weights are detailed in Fig. [5b]. In this particular case, all local weights are similar. The lead time-dependent weights are shown in Fig. [5c]. Note the different horizontal axis, which is the lead
(a) Development of global weighting $w_g(\psi|\phi)$ over time. ANN performs best, followed by the linear regression. The global weighting of the persistence method is small.

(b) Development of local weighting $w_l(\psi|\phi)$ over time. As the particular weather situation is a fairly standard situation (and thus, it is relatively well represented in the data set), there is little observable local weighting in this particular case.

(c) Development of lead time-dependent weighting $w_k(\psi|\phi)$ over time. Though the global performance $w_g(\psi|\phi)$ of the persistence method is low, it performs strongly on short time horizons.

(d) Combined $w(\psi|\phi)$, consisting of $w_g(\psi|\phi)$, $w_k(\psi|\phi)$, and $w_l(\psi|\phi)$, over time. As can be seen in this example, though the overall quality of the persistence method is low, it has high impact in short time horizons.

(e) Forecasting example using combined weighting of the CSGE technique. On delivery of a new NWP each 24 h, the weighting of the persistence method is high in this example.

Figure 5: Example of the weight combination of the proposed CSGE ensemble technique. In the example, an intraday forecast is performed using a single weather forecasting model and three power forecasting models. New weather data are incorporated every 24 h. For the intraday forecast, a persistence method is combined with an artificial neural network and a linear regression model. During ensemble training, the coopetitive soft-gating parameters are optimized so that the depicted weighting emerges for an example weather situation over three days. The overall weight development $w(\psi|\phi)$ (Fig. 5d) is composed of the three weighting terms $w_g(\psi|\phi)$ (Fig. 5a), $w_l(\psi|\phi)$ (Fig. 5b), and $w_k(\psi|\phi)$ (Fig. 5c). The overall forecast is shown in Fig. 5e.
time in this case. As is to be expected, the persistence method works well on very short time horizons, but quickly loses quality in comparison to the other two approaches. The combination of the three weighting aspects is visualized in Fig. 5d. As can be seen, on delivery of new NWP forecasts every 24h, the influence of the lead time-dependent persistence technique is high. An overall forecast is shown in Fig. 5e.

The second example shows a multi-model forecast using three weather forecasting models for a day-ahead forecast, each of which is predicted using two power forecasting models. This example is visualized in Fig. 6a. The overall weighting over time is shown in Fig. 6b. Regarding the weather forecasting models, the first weather forecasting model “WM1” (which is the ECMWF IFS model) has the highest influence, while the other two weather forecasting models have lower weights. This, again, meets the expectation, as can be seen from the table in Fig. 6c which shows the overall RMSE when performing the forecast on a single weather / power forecasting model combination. The model with the lowest RMSE error gets the highest weight. Also, the quality difference between the power forecasting models is reflected in the weighting. As there are no weather or power forecasting models which are designed for a different forecasting time period (unlike in the first example), the overall weighting differences over time are not as drastic as in the first example. The forecast which is created using the weights determined by the CSGE is shown in Fig. 6d.

Figure 6: Example of weighting combination for a day-ahead forecast using three weather forecasting models and the proposed CSGE technique. Each model has different forecasting quality (where it can be observed that WM1 has the highest quality and WM3 performs worst). The overall weighting for each weather / power forecasting model combination is shown in Fig. 6a. The result is in line with the expectation when performing the prediction using a single weather and power forecasting model combination, as can be seen from the table in the figure. Values denote RMSE on normalized power values. The overall forecast is shown in Fig. 6b. Depending on the particular situation, the ensemble adapts the model weights.

5. Experimental Results

This section investigates the performance of the proposed CSGE technique in comparison to a number of state-of-the-art approaches. We evaluate the algorithms on 45 data sets which are described in Section 5.1. The experimental setup is described in Section 5.2. We examine the proposed power forecasting model using a single weather forecasting model (Section 5.3), and using multiple weather forecasting models for day-ahead forecasting (Section 5.4), as well as for intraday forecasts (Section 5.5). Finally, a detailed comparison of the performance gains when using multiple weather forecasting models for both, day-ahead and intraday forecasts, is detailed in Section 5.6. A discussion of applicability is performed in Section 5.7.

5.1. Data Set Description

The data sets used for the evaluation are the publicly available data sets of the EuropeWindFarm collection [11] using multiple weather forecasting models. The 45 data sets contain the weather forecasts and power measurements of two consecutive years of both onshore and offshore wind farms. The data sets contain the following items:

- **Time Stamp** of the forecast / power measurement,
- **Lead time** from the forecasting origin,
- **Wind Speed** in 100 m height,
- **Wind Speed** in 10 m height,
- **Wind Direction (zonal)** in 100 m height,
Figure 7: Locations of the wind farms of the EuropeWindFarm collection, see Section 5.1 for a data set description.

- Wind Direction (meridional) in 100 m height,
- Air Pressure forecast,
- Air Temperature forecast, and
- Power Generation of the wind farm.

The power generation time series are normalized with respect to the nominal capacity of each power plant to enable a scale-free comparison. All weather input parameters are normalized to in the interval [0, 1]. The data has been filtered to eliminate erroneous measurements.

5.2. Experimental Setup

In the experiments, we evaluate the power forecasting performance of the data sets (see Section 5.1). As laid out in the CSGE algorithm description of Section 4 for each weather forecasting model we perform the forecast using a number of forecasting algorithms. As optimization for the CSGE technique (see Section 4.5), a simplex algorithm [31] is chosen. The parameter search is performed in greedy fashion, where the global qualities are optimized followed by the local and the temporal optimization for both, weather and power forecasting models. The regularization parameter $\zeta$ is chosen in a way that avoids overfitting of the CSGE algorithm to the validation data.

For the evaluation, each data set is split into a training and a test subset. Afterwards, the training set is further split in a 5-fold cross-validation into three data sets which are called parameter set (3/5), optimization set (1/5) and validation set (1/5) for the sake of clarity. The single power forecasting models for each weather forecasting model are trained using the parameter set. The parameter optimization of the CSGE technique is then performed on the optimization set and finally evaluated on the validation set. The parameter combination which performed best on the validation set over all folds is chosen as final model parameterization which is used to compute the final model quality on the test set.

As pointed out, e.g., in [10], error scores beyond the RMSE are of importance for investigating a forecast. For our evaluation, we therefore include the RMSE (computed using Eq. (14)), the coefficient of determination $R^2$, and the skill score for model comparison. The $R^2 \in [0, 1]$ score is the squared correlation coefficient $R$ computed by

$$R^2 = \frac{\sum_{n=1}^{N}(\hat{y}_n - \bar{\hat{y}})((o_n - \bar{o}))^2}{\sum_{n=1}^{N}(\hat{y}_n - \bar{\hat{y}})^2 \sum_{n=1}^{N}(o_n - \bar{o})^2}.$$  (39)

It describes the amount of linear correlation between a forecast and the measured values. The optimal value of $R^2$ is 1 (perfect correlation), whereas 0 represents no correlation between the forecasts and measurements. Note that $\bar{\hat{y}}$ is the mean of all issued forecasts in the data set in this case (not an ensemble prediction). The skill score describes the amount of improvement of an evaluated technique $e_{\text{eval}}$ in comparison to a baseline technique $e_{\text{base}}$. The improvement of the forecasting skill can be computed using the error scores (either RMSE or $R^2$):

$$\text{Imp} = \frac{e_{\text{base}} - e_{\text{eval}}}{e_{\text{base}}}.$$  (40)

Furthermore, the number of wins of a particular algorithm is stated.

5.3. Day-Ahead Performance on Single Weather Forecasting Model

This experiment aims at comparing the performance of the CSGE technique to other techniques using a single weather forecasting model in comparison to other forecasting algorithms and ensembles on the day-ahead forecasting horizon ($k_{\text{min}} = 24, k_{\text{max}} = 48, \Delta = 1 h$) on the data set of Section 5.1. As there is only one weather forecasting model, the number of weighting dimensions is reduced to the 3 power forecasting model based weighting factors $w(\phi|\psi) = w_y(\phi|\psi), w_I(\phi|\psi), w_k(\phi|\psi)$ for each power forecasting model. As power forecasting techniques, we include some state of the art model types, namely feed-forward artificial neural networks (ANN), linear regression (Lin. Reg.), a boosting (LSBoost)
As can be seen from Table 2, the wind farm data sets have a varying error regarding the forecasting accuracy. Some wind farms are well predictable regarding the RMSE (e.g., wf3, wf9, wf13, wf41), whereas the forecasting algorithms struggle with other wind farms (e.g., wf6, wf20, wf31), as can be derived from the respective scores and from the color coding in the table. Regarding the overall performance, all evaluated ensemble techniques perform better on average in comparison to the forecasting techniques based on a single forecasting algorithm, as can be seen from the forecasting skill (next to last row). The best ensemble techniques on the data set are the CSGE variants, where the suffixes in the table indicate locality assessment using either the nearest neighbor (CSGE-K), or interpolation technique CSGE-I from [13]. The CSGE variants are closely followed by the CE and the bagging technique. The boosting algorithm performs weaker than the other ensemble algorithms. The $R^2$ score gives similar indication as the RMSE score. While there are some differences regarding single data sets, regarding the average $R^2$ score, the overall order of the best performing algorithms remains the same. It must be mentioned that in this scenario, the CSGE techniques are not exploited to

| Skill | 4.62% | base | 4.70% | 7.01% | 6.63% | 7.29% | 7.53% | 4.14% | base | 4.31% | 6.64% | 6.26% | 7.84% | 7.25% |
|-------|-------|------|-------|-------|-------|-------|-------|-------|------|-------|-------|-------|-------|-------|
| @Wms | 2.0   | 2.5  | 5.5   | 12.5  | 12    | 10.5  | 2.1   | 1.1   | 5.5  | 11    | 12    | 12    | 11    | 12    |
their full extend, as there is only a single weather forecasting model to make use of. Therefore, only 3 of the 6 weighting parameters are applied. In the following experiment (Section 5.4), the performance using multiple weather forecasting models is evaluated, fully exploiting the advantages of the CSGE technique.

5.4. Day-Ahead Performance on Multiple Weather Forecasting Models

This experiment aims at comparing the performance of the CSGE technique to other Multi-Model-Ensemble (MME) techniques. In the experiment, we use 3 day-ahead weather forecasting models. The three models are available for 36 windfarms for a time-period of 22 months. The time periods do not overlap entirely with the data used in Section 5.3.

For this comparison, we choose the following techniques:

1. No Ens.: The power forecasting model is the overall best non-ensemble forecasting technique (i.e., an ANN model) with the single best weather forecasting model (determined using the table in Fig. 6a).

2. MME Eq.: The model is computed using the best non-ensemble forecasting technique (ANN model) with all weather forecasting models. The model forecasts are averaged, i.e., \( \eta = \frac{1}{3} \). This technique, is, e.g., utilized in 10.

3. MME Fix.: The model is computed using the best non-ensemble forecasting technique (ANN model) with all weather forecasting models. The models are weighted with respect to their global quality using Eq. (13) with \( \eta = 2 \), such
that the best performing weather forecasting models get the highest impact in the weighting. This technique is employed, for instance, in [27].

4. CSGE-S: The CSGE technique is applied using a single power forecasting model for each of the three weather forecasting models, the best non-ensemble forecasting technique (ANN model). The CSGE locality assessment is performed using the nearest neighbor technique.

5. CSGE-M: The CSGE technique using multiple power forecasting models for each weather forecasting model is applied. The CSGE locality assessment is performed using the nearest neighbor technique.

The results of the day-ahead MME are shown in Table 3. As can be seen from this table, on average, all ensemble models perform better than a model which only uses a single weather forecasting model. Therein, the CSGE technique using multiple power forecasting models (CSGE-M) performs best regarding the RMSE score, followed by the CSGE-S, the MME Fix., and the MME Eq. models. For $R^2$, both CSGE variants perform comparably well, followed by the other techniques. For the data sets, the CSGE-M performs 8.78% better than the best non-ensemble technique using the single best weather forecasting model. The other multi-model ensemble techniques also yield better results than the baseline technique. While there are some data sets with low RMSE (e.g., wf6, wf9), this error is due to a low value of the target predictands in these data sets, and not due to a good predictability regarding correlation, as can be seen from the $R^2$ score. Particular attention should be drawn to wf9,
where the CSGE-M technique has a low $R^2$ score. This may be due to an overfitting of the ensemble model, if the regularization parameter $\zeta$ is chosen too low. Still, the CSGE-M technique exceeds the performance of the other approaches regarding the RMSE on this data set.

As can be seen from the performance of the examined ensemble techniques, the use of multiple weather forecasting models clearly improves the performance, e.g., in MME Eq. The use of a global weighting (such as conducted in MME Fix.) further increases the performance. The use of local and temporal weighting in addition to the global weighting (as performed in CSGE-S) increases the performance, even if just one power forecasting model is used. The scores can further be improved when performing the weighting for both, weather and power forecasting models (as performed in CSGE-M).

5.5. Intraday Performance as Lagged Multi-Model Ensemble

This experiment evaluates the ability of the proposed method to work as a lagged multi-model ensemble. Therefore, an intraday weather forecasting model is added to the three existing weather forecasting models, and an intraday forecast is performed ($k_{\text{min}} = 1, k_{\text{max}} = 24, \Delta = 1h$). The overall forecast therefore is created from one intraday forecast and three day-ahead forecasts. As comparison techniques, we include the pure persistence method, the best non-ensemble technique (No Ens., which is an ANN), the best conventional multi-model ensemble technique MME Fix. from the previous section, and the proposed algorithm, without (CSGE-K) and additionally including the persistence method in the CSGE-K approach (CSGE+P).

The results of the experiment are shown in Table [1]. As expected, the persistence method performs weakest. All models yield better forecasting results due to the new intraday model which is included in the forecast. Within the ensemble techniques, the proposed CSGE-K technique exceeds the performance of the MME fixed technique, whereas the inclusion of the persistence method as ensemble member (CSGE+P) yields further quality improvements regarding both, RMSE and $R^2$ score. The CSGE+P technique exceeds the persistence models performance by 49.13% regarding RMSE, followed by the CSGE-K technique (48.24%), and the fixed ensemble. The $R^2$ scores show very similar results.

5.6. Performance Development Using Varying Number of Weather Forecasting Models

This experiment gives insight into a practical problem in power forecasting: Given a number of weather forecasting models available, which of those models should be included, and will the overall quality eventually even be lowered when worse performing models or models with unknown performance are added to the overall ensemble. This section therefore describes the dependence of the performance of the techniques in Sections 5.4 and 5.5 when adding additional weather forecasting models to the forecast. In order to evaluate the performance, the average RMSE of the algorithms used in Sections 5.4 and 5.5 are computed with a varying number of weather forecasting models. Figs. [Sa] and [Sc] show the dependence of the performance when the weakest model is chosen as first model, and increasingly good weather forecasting models are added. The bounds indicate the model variance on the data sets. Figs. [Sb] and [Sd] on the other hand, show the development of the performance when the best weather forecasting model is chosen first and worse models are added subsequently.

For the day-ahead variants of Figs. [Sa] and [Sb] all ensemble models benefit from increasing the number of weather forecasting models, even if the additional models perform worse in comparison to the first weather forecasting model. Fig. [Sa] shows the RMSE error when including increasingly good performing models. As can be expected, the performance is clearly higher when adding additional weather forecasting models. As can be seen from Fig. [Sb] the performance improvement drops when including a third weather forecasting model and the added model performs worse than the already included models, in this case. The approaches for comparison benefit more when including a second model. However, the CSGE-K is able to benefit more from the inclusion of a third weather forecasting model. In particular the MME Eq. technique is not able to regulate the model influence precisely enough and therefore barely benefits from the inclusion of a third model.

For intraday forecasts, the behavior is similar to the day-ahead forecast when adding increasingly well performing models, as is indicated in Fig. [Sc]. When including an intraday model as fourth model in addition to the three day-ahead forecasts, all models benefit about equally from the inclusion of the fourth model. An interesting phenomenon is visible in Fig. [Sc] i.e., when adding additional
Day-ahead forecast when adding increasingly well performing weather forecasting models.

Day-ahead forecast when adding increasingly weak performing weather forecasting models.

Intraday forecast when adding increasingly well performing weather forecasting models.

Intraday forecast when adding increasingly weak performing weather forecasting models.

Figure 8: Average development of the RMSE quality of a number of ensemble techniques when including additional weather forecasting models to the ensemble for both, day-ahead and intraday forecasts, using the EuropeWindFarm data set collection. When adding increasingly well performing weather forecasting models, all ensemble algorithms benefit from the inclusion of additional weather forecasting models (Figs. 8a and 8c). When adding increasingly weak performing models to the ensemble, the CSGE-K technique is able to better use this additional information. The comparison techniques’ performance, on the other hand, remains constant (see Fig. 8b), or even decreases (see Fig. 8d).

Day-ahead weather forecasting models to an intraday weather forecasting model. All models benefit from the inclusion of the best-performing day-ahead weather forecasting model. The maximum benefit of the inclusion of multiple weather forecasting models seems to be reached at this point, as no model is able to benefit from additional weather forecasting models. However, when including further weather forecasting models, the model performance of the comparison approaches decreases, i.e., the inclusion of additional models irritates the ensemble forecast. The CSGE-K technique, on the other hand, is able to recognize the weak performing weather forecasting models and is able to reduce the respective weights in a way that does not negatively affect the overall model performance.

5.7. Discussion of Applicability

Beside the performance of the proposed approach, there are some other model properties worth mentioning.

Failure Mode

The ensemble weights $w^{(j)}$ are computed dynamically for each forecasting time step for which a forecast is performed. The CSGE algorithm computes the weights for the respective weighting categories using all available forecasts at that particular time. Thereby, the ensemble can create a forecast even if some power forecasting models fail to create a prediction, or if the NWP forecast for a particular weather forecasting model fails to be delivered. The proposed technique is then able to retain the optimal weighting performance given the circumstances.

Ensemble of Opportunity

Given this “failure mode” of the ensemble, an applied forecasting system can then also be designed to work as an ensemble of opportunity. A number of power forecasting models and weather forecasting models can be prepared (pretrained) for application, however, not all power forecasting models (and/or not all weather forecasting models) have to be evaluated every time a prediction is created. The number of evaluated power forecasting models can be chosen dynamically. This number could depend on one or more of the following aspects:

- Estimated difficulty of the current forecasting task, e.g., derived from weather situation.
- Uncertainty of the current prediction in a probabilistic forecast, e.g., from inner-ensemble disagreement.
- Expected quality $w^{(j)}$ in the ensemble, i.e., elimination of models with low weight if too high costs are present.
- Criticality / importance of a precise forecast to the power grid operation.
- Necessity of fast delivery time.
Financial cost of a reported deviation from the true power generation.

Available computation capacity in a computing cluster.

Cost/Reward Functions
Given the availability of (hybrid) cloud computing solutions and e.g., computing concepts such as cheap preemptible virtual machines, the employment of using cost/reward functions for financial optimizations can make sense, given the trade-off of investment in computing costs and gain in quality. The above-mentioned factors can have an impact on the design of the cost/reward function.

Parameter Determination
The proposed ensemble technique has a low number of parameters only. Besides the power forecasting models to choose, the main parameter is the regularization parameter $\zeta$ which regulates the amount of coopetitive soft gating. While choosing an improper value of $\zeta$ negatively affects the model performance, it still will perform either as a static model averaging (too much regularization) or pure gating, possibly with overfitting (too little regularization).

Variants of CSGE
In the article, a nearest neighbor technique to assess the local (weather-dependent) quality is proposed (in Section 4.3). However, one can imagine other possibilities to assess the local quality, such as a multi-linear interpolation (MLI) technique, which is presented in [13]. As shown in the experiments in the referenced article, the two techniques behave very similar regarding forecasting performance. The respective technique should therefore be chosen depending on the size of the historic training data set. The MLI has a training phase, however, it does then no longer need the training data in order to operate. This approach is, therefore, well-suited for large data sets. The CSGE-K technique is able to work without model training, however, it does have to search the data set on every query. The technique is, therefore, better suited for smaller data sets.

Parameter Optimization
An advantage of the proposed CSGE technique is that the ensemble training (the determination of the weights $\eta$) is a post-processing step of the training of the ensemble members. The ensemble is trained by Eq. 38 which optimizes the weights rather than the single ensemble member forecasts. This, in turn, means that during ensemble training, the forecasts of the ensemble members do not have to be recomputed when varying $\eta$. The evaluation of the model fitness therefore is swiftly possible. Furthermore, the weights gradually change when varying the coopetitive soft gating strengths $\eta$, thus the optimization problem is smooth. In this article, the parameter optimization was performed in a greedy fashion for the sake of simplicity and speed. However, one can also think of overall optimization using techniques such as simulated annealing, stochastic gradient descent, or particle swarms, possibly leading to an even better optimization.

On-Line Weight Improvement
The proposed technique can be extended to update the weighting methods consistently “on-line” when observing novel model input data by permanently learning on novel observations. The respective power measurements have to be included in this setup in order to enable a meaningful feedback.

6. Conclusion and Outlook
In this article, we proposed a novel multi-scheme ensemble based on a technique we call coopetitive soft gating. The technique aims to adaptively exploit the strengths of various power forecasting and weather forecasting models regarding their overall global quality, their lead time dependent quality, and their weather situation-dependent local quality. The technique is able to yield superior results in comparison to other forecasting algorithms and ensemble techniques as has been shown in a number of experiments on publicly available data sets. The flexible structure of the ensemble technique enables an employment of the proposed algorithm for a number of lead times and using a flexible number of power and weather forecasting models, respectively.

In the future, we aim to implement some of the schemes to enable the ensemble technique to work as an ensemble of opportunity using some of the influencing factors described in Section 5.7. We also aim to develop a technique for assessing the uncertainty of a forecast depending on the weights computed using coopetitive soft gating. Therein, our goal is to develop an adaptive model pruning method depending on the particular situation
to forecast when the expected quality of a power forecasting model for a particular forecasting situation (e.g., weather situation or forecasting time step) is low to avoid unnecessary computations with marginal impact. Also, the inclusion of more sophisticated base predictors, such as (deep) neural network structures (e.g., analyzed in [12]), in the ensemble may be of interest.

Acknowledgment

This article results from the project BigEnergy (HA project no. 472/15-14), which is funded in the framework of Hessen ModellProjekte, financed with funds of the LOEWE Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz. Förderlinie 3: KMU-Verbundvorhaben (State Offensive for the Development of Scientific and Economic Excellence). We want to thank the Enercast GmbH for providing the data sets.

References

[1] Alessandrini, S., Sperati, S., Pinson, P., 2013. A comparison between the ECMWF and COSMO Ensemble Prediction Systems applied to short-term wind power forecasting on real data. Applied Energy 107, 271–280.

[2] Bessa, R. J., Miranda, V., Botterud, A., Zhou, Z., Wang, J., 2012. Time-adaptive quantile-copula for wind power probabilistic forecasting. Renewable Energy 40 (1), 29–39.

[3] Breiman, L., 1996. Bagging Predictors. Machine Learning 24 (241), 123–140.

[4] Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.

[5] Costa, A., Crespo, A., Navarro, J., Lizzano, G., Madsen, H., Feitosa, E., 2008. A review on the young history of the wind power short-term prediction. Renewable and Sustainable Energy Reviews 12 (6), 1725–1744.

[6] Craig, P. P., Gadgil, A., Koomey, J. G., 2002. What can history teach us? A Retrospective Examination of Long-Term Energy Forecasts for the United States. Annual Review of Energy and the Environment 27 (1), 83–118.

[7] Delle Monache, L., Eckel, F. A., Rife, D. L., Nagarajan, B., Searight, K., 2013. Probabilistic Weather Prediction with an Analog Ensemble. Monthly Weather Review 141 (10), 3492–3516.

[8] Dietterich, T. G., 2000. Ensemble Methods in Machine Learning. Multiple Classifier Systems 1857, 1–15.

[9] Duan, Q., Ajami, N. K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. Advances in Water Resources 30 (5), 1371–1386.

[10] Foley, A. M., Leahy, P. G., Marvuglia, A., McKeogh, E. J., 2012. Current methods and advances in forecasting of wind power generation. Renewable Energy 37 (1), 1–8.

[11] Gensler, A., 2016. EuropeWindFarm Data Set, last accessed 2018-03-15. URL http://ies-research.de/Software.

[12] Gensler, A., Henze, J., Sick, B., Raabe, N., 2016. Deep Learning for Solar Power Forecasting An Approach Using Autoencoder and LSTM Neural Networks. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC16). Budapest, Hungary, pp. 2858–2865.

[13] Gensler, A., Sick, B., 2016. Forecasting wind power – an ensemble technique with gradual competitive weighting based on weather situation. In: Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN16). Vancouver, Canada, pp. 4976–4984.

[14] Gensler, A., Sick, B., 2017. Probabilistic Wind Power Forecasting: A Multi-Scheme Ensemble Technique With Gradual Coopetitive Soft Gatign. In: Proceedings of the 9th IEEE Symposium Series on Computational Intelligence (SSCI17). Honolulu, USA, pp. 1803–1812.

[15] Gensler, A., Sick, B., Pankraz, V., 2016. An analog ensemble-based similarity search technique for solar power forecasting. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC16). Budapest, Hungary, pp. 2850–2857.

[16] Gensler, A., Sick, B., Vogt, S., 2016. A review of determinist error scores and normalization techniques for power forecasting algorithms. In: Proceedings of the 8th IEEE Symposium Series on Computational Intelligence (SSCI16). Athens, Greece, pp. 1–9.

[17] Giebel, G., Brownword, R., Kariniotakis, G., Denhard, M., Draxl, C., 2011. The State-Of-The-Art in Short-Term Prediction of Wind Power: A Literature Overview, 2nd Edition. Tech. rep., Lyngby, Denmark.

[18] Gönen, M., Alpaydín, E., 2011. Multiple Kernel Learning Algorithms. Journal of Machine Learning Research 12, 2211–2268.

[19] Hagedorn, R., Doblas-Reyes, F. J., Palmer, T. N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. Tellus, Series A: Dynamic Meteorology and Oceanography 57 (3), 219–233.

[20] Ho, T. K., 1998. The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (8), 832–841.

[21] Holttin, H., Milligan, M., Ela, E., Menemenlis, N., Dobschinski, J., Rawn, B., Bessa, R. J., Flynn, D., Gómez-Lázaro, E., Detlefsen, N. K., 2012. Methodologies to determine operating reserves due to increased wind power. IEEE Transactions on Sustainable Energy 3 (4), 713–723.

[22] Hong, T., Wilson, J., Xie, J., 2014. Long term probabilistic load forecasting and normalization with hourly information. IEEE Transactions on Smart Grid 5 (1), 456–462.

[23] Kohavi, R., Wolpert, D. H., 1996. Bias plus variance decomposition for zero-one loss functions. Proceedings of the 13th International Conference on Machine Learning (ICML96). 275–283.

[24] Lei, M., Shiyan, L., Chuanwen, J.,Hongling, L., Yan, Z., 2009. A review on the forecasting of wind speed and generated power. Renewable and Sustainable Energy Reviews 13 (4), 915–920.

[25] Li, G., Shi, J., Zhou, J., 2011. Bayesian adaptive combination of short-term wind speed forecasts from neural network models. Renewable Energy 36 (1), 352–359.
Loebecke, C., Van Fenema, P. C., Powell, P., 1999. Co-opetition and knowledge transfer. Database for Advances in Information Systems 30 (2), 14–25.

Matos, M. A., Bessa, R. J., 2011. Setting the operating reserve using probabilistic wind power forecasts. IEEE Transactions on Power Systems 26 (2), 594–603.

Mendes-Moreira, J., Soares, C., Jorge, A. M., Sousa, J. F. D., 2012. Ensemble approaches for regression. ACM Computing Surveys 45 (1), 1–40.

Mirasgedis, S., Sarafidis, Y., Georgopoulou, E., Lalas, D., Moschovitis, M., Karagiannis, F., Papakonstantinou, D., 2006. Models for mid-term electricity demand forecasting incorporating weather influences. Energy 31 (2-3), 208–227.

Mittermaier, M. P., 2007. Improving short-range high-resolution model precipitation forecast skill using time-lagged ensembles. Quarterly Journal of the Royal Meteorological Society 133 (October), 1487–1500.

Nelder, J., Mead, R., 1965. A simplex method for function minimization. Computer Journal 7 (4), 308–313.

Pierro, M., Bucci, F., De Felice, M., Maggioni, E., Moser, D., Perotto, A., Spada, F., Cornaro, C., 2016. Multi-Model Ensemble for day ahead prediction of photovoltaic power generation. Solar Energy 134, 132–146.

Pinson, P., Girard, R., 2012. Evaluating the quality of scenarios of short-term wind power generation. Applied Energy 96, 12–20.

Pinson, P., Kariniotakis, G., 2004. On-line assessment of prediction risk for wind power production forecasts. Wind Energy 7 (2), 119–132.

Pinson, P., Nielsen, H., Madsen, H., Kariniotakis, G., 2009. Skill forecasting from ensemble predictions of wind power. Applied Energy 86 (7-8), 1326–1334.

Platt, J., 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in large margin classifiers 10 (3), 61–74.

Rajagopalan, B., Lall, U., Zebiak, S. E., 2002. Categorical Climate Forecasts through Regularization and Optimal Combination of Multiple GCM Ensembles. Monthly Weather Review 130 (7), 1792–1811.

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

Ren, Y., Zhang, L., Suganthan, P. N., 2016. Ensemble Classification and Regression Recent Developments, Applications and Future Directions. IEEE Computational Intelligence Magazine 11 (1), 41–53.

Schapire, R. E., 1990. The Strength of Weak Learnability. Machine Learning 5 (2), 197–227.

Siwek, R., Owowski, S., Szulpulk, R., 2009. Ensemble Neural Network Approach for Accurate Load Forecasting in a Power System. International Journal of Applied Mathematics and Computer Science 19 (2), 303–315.

Sloughter, J. M., Gneiting, T., Raftery, A. E., 2010. Probabilistic Wind Speed Forecasting using Ensembles and Bayesian Model Averaging. Journal of the American Statistical Association 105 (489), 25–35.

Ziehmann, C., 2000. Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. Tellus, Series A: Dynamic Meteorology and Oceanography 52 (3), 280–299.

[45] Taylor, J. W., McSharry, P. E., Buizza, R., 2009. Wind power density forecasting using ensemble predictions and time series models. IEEE Transactions on Energy Conversion 24 (3), 775–782.

[46] Trolldborg, N., Surensten, J., 2014. A simple atmospheric boundary layer model applied to large eddy simulations of wind turbine wakes. Wind Energy 17 (April), 657–669.

[47] van Oldenborgh, G. J., Doblas-Reyes, F. J., Wouters, B., Hazeleger, W., 2012. Decadal prediction skill in a multi-model ensemble. Climate Dynamics 38 (7-8), 1263–1280.

[48] Weigel, A. P., Liniger, M. A., Appenzeller, C., 2007. Generalization of the Discrete Brier and Ranked Probability Skill Scores for Weighted Multimodel Ensemble Forecasts. Monthly Weather Review 135 (7), 2778–2785.

[49] Zhang, Y., Wang, J., Wang, X., 2014. Review on probabilistic forecasting of wind power generation. Renewable and Sustainable Energy Reviews 32, 255–270.

[50] Ziehmann, C., 2000. Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. Tellus, Series A: Dynamic Meteorology and Oceanography 52 (3), 280–299.