Submitted to Wind Energy December 10, 2008.

Conditional Weighted Combination of Wind Power Forecasts

Fannar Ö. Thordarson*[1], Henrik Madsen[1], Henrik Aa. Nielsen[2], Pierre Pinson[1]

[1] **DTU Informatics, Technical University of Denmark;**
Richard Petersens Plads (bg. 321), DK-2800 Kgs. Lyngby, Denmark.

[2] **ENFOR A/S;**
Lyngsø Allé 3, DK-2970 Hørsholm, Denmark.

Abstract

The classical regression model for combined forecasting is reformulated in order to impose restrictions on the combination weights. This restricted linear combination model is then extended to the case where the weights are allowed to be a nonparametric function of some meteorological variables, yielding the so-called conditional weighted combination method. The weight functions are estimated with local regression techniques. The conditional weighted combination method is applied to a test case of a wind farm over a period of 10 months. Various combinations of two forecasts out of the three available ones are considered. Analysis of the data suggests that meteorological forecasts of air density and turbulent kinetic energy may be considered as relevant external variables in the combination scheme. A performance comparison shows that the conditional weighted combination method introduced in this paper outperforms the least-squares combination method for almost all prediction horizons, especially for larger ones. This indicates that further developments based on conditional combination methods, including adaptivity of the weight functions estimation, may significantly enhance forecast accuracy and dampen the risk of large prediction errors.

Key words: wind power, forecasting, linear combination, local regression, conditional parametric models, conditional weighted combination

---

**Corresponding author:*
F. Thordarson,
DTU Informatics, Richard Petersens Plads (bg. 321), DK-2800 Kgs. Lyngby, Denmark.
Tel: +45 4525 3349, fax: +45 4588 2673,
E-mail: ft@imm.dtu.dk, Webpage: www.imm.dtu.dk

Introduction

The wind power industry is growing rapidly, while the share of wind generation in the electricity mix of a large number of countries is getting larger every year. Research and development have improved the technology, with advances in performance as well as reduced production costs, and made wind energy competitive on the energy market. Expectations in the medium-term are that 10-20 percent of world electricity consumption will be generated by wind power [1]. However, the variability and low-predictability of wind generation are commonly seen as penalizing, since they translate to e.g. an increase in reserve requirements for Transmission System Operators (TSOs), or to a diminution of potential gains on the electricity market for the case of utilities or independent wind power producers. This makes forecasting wind power generation for time-scales ranging from a few minutes to a few days ahead paramount. Forecasting tools are helping to secure the value of wind generation on the market, while contributing to increasing security of supply. A review of state of the art wind power forecasting has been published by Giebel *et al.* [2].

The loss function of wind power forecast users, i.e. the mathematical representation of their sensitivity to consequences of forecast errors, is usually not linear. This translates to saying that their sensitivity to forecast errors changes with their range of magnitude. Indeed, errors perceived as small are easily accepted, while forecasters will always be reminded about their errors of very large magnitude and that yielded significant financial or technical consequences. Today, users of wind power predictions are increasingly provided with several forecasts for the coming period, generated from various methodologies and/or based on different meteorological input. In such a case, it can be attractive procedure to combine these forecasts. By combining the independent information included in each individual forecast, more accurate prediction can be accomplished, and the risk of large prediction errors significantly reduced [3]. This was originally suggested by Bates & Granger [4] for two individual forecasts, and since then a number of procedures have been suggested, see e.g. Clemen [5] or de Menezes *et al.* [6] for a literature review.

Combining wind power forecasts is especially appealing if several meteorological forecasts from different meteorological offices are available. Since meteorological forecasts directly input wind power prediction methods (for look-ahead times ranging from 6 to 72 hours ahead), employing different meteorological forecasts will yield power forecasts with different characteristics. Combination is traditionally performed through a linear regression model, for which the parameters represent the weights of corresponding individual forecasts in the combination. Common linear methods for forecast combination include the simple plain average [7], the Bayesian approach [8] and regression-based methods [9]. Initially the present paper employs a linear model, with restrictions on the combination weights. The main originality of our approach is that the parameters in the linear model (i.e. the combination weights) are further allowed to be dependent upon some external variables, which then leads to a nonlinear model. In other words, the combination scheme can be made conditional to the predicted meteorological conditions, if summarized by a set of predicted meteorological variables. The block diagram in Figure 1 illustrates the proposed approach to forecast combination, and more specifically how information from meteorological forecasts impact weight estimation. Owing to its specific characteristics, our newly introduced combination scheme is referred to as Conditional Weighted Combination (CWC). This proposal comprises an alternative approach to the situation-specific combination approach introduced by Lange *et al.* [10] and Lange & Focken [11], for which forecast combination is proposed to be made dependent upon the prevailing synoptic weather situation. Classification of weather situations in a number of weather 'modes' is performed through a method combining principal component analysis and cluster analyses. Within each weather class the accuracy of various prediction models is evaluated and an optimal combination determined. Such a method may be cumbersome, as it requires analysis of a huge amount of data in order to determine prevailing weather situations. Another drawback is that such a method involves the possibly demanding task of predicting, to which weather class the coming weather situation belongs to. In contrast, the situation-specific combination method originally introduced here only necessitates a limited quantity of forecast information at the level of the wind power production site considered.

——————— Figure 1 is approx. here ———————

The paper begins with a description of the mathematical framework for forecast combination by considering combination as a restricted linear model with the individual forecasts as explanatory variables. Then the CWC scheme is introduced by replacing the weights in the linear model with smooth, but otherwise unknown, functions of some meteorological variables. A method for estimating such functions is briefly presented. These methods are then evaluated and compared in a least-squares minimization fashion [9]. The test case is that of real wind farm located in the north of Jutland (Denmark), for which three different types of power forecasts are available (with three different meteorological forecast inputs), along with corresponding wind power measurements for a period of 10 months. The main evaluation criterion is the Root Mean Square Error (RMSE), which gives more weight to large prediction errors. Its use is thus consistent with the idea that forecast users may be more sensitive to consequences of large forecast errors [12]. In parallel, employing the RMSE as an error measure is also consistent with the idea that wind power point forecasts commonly inform on the mean potential generation for each look-ahead time, or in more mathematical terms, give the conditional expectation of the wind power generation random variable for every forecast horizon. The expectation is the summary statistics of a random variable that minimizes a quadratic loss function, while the RMSE is the corresponding quadratic error measure. The paper concludes with some discussion and perspectives regarding further developments related to conditional combination schemes.

Mathematical Framework for Combining Wind Power Forecasts

Since the issue of combining forecasts was first introduced, many methods have been developed, sometimes even based on advanced complex techniques. However, it is still not obvious which combination method will always perform best [6]. In this paper, the linear combination model is adopted first because of its inherent simplicity and its performance often deemed as highly acceptable. Later, however, an extension is proposed in order to allow for the coefficients to be functions of some exogenous variables, and this makes the model nonlinear.

Formulation of a Restricted Linear Combination Model

For the wind power application, the variable of interest to be predicted is the wind power production at time $t+h$, which is denoted as y_{t+h}. In parallel, let $\widehat{y}_{i,t+h|t}$ be the i^{th} individual forecast for the wind power at time $t+h$ (out of the k different forecasts available) calculated using the actual information available at time t, with h being the forecast horizon, i.e. the number of hours between the moment at which the forecast is issued (time t) and the time point for which the prediction is made. The i^{th} prediction error, that is, the difference between the measured power production and the i^{th} competing prediction is $e_{i,h}$. For the combination of k individual forecasts, a linear model of the forecasts is formulated as

$$\widehat{y}_{c,t+h|t} = \mu_h + \sum_{i=1}^{k} w_{i,h}\widehat{y}_{i,t+h|t} \tag{1}$$

where $w_{i,h} = w_i$ is the weight given to the i^{th} of the k forecasts. In the following, the forecast horizon h is omitted from the weight terms of the model in order to simplify notations. In the present study, a linear model is written for each horizon independently, and the horizon of interest can easily be deduced from notations employed for the various forecast variables. The term μ represents the constant term in the linear model, which is provided if the individual forecasts are biased, especially when the coefficients for the individual forecasts are restricted to sum to one [13, 14]. Considering an additional noise in the power production process, the combined forecast corresponds to the conditional expectation of power production at time $t + h$, given the information available at time t,

$$y_{t+h} = \widehat{y}_{c,t+h|t} + e_{c,t+h} \tag{2}$$

where $\{e_{c,t+h}\}$ is a noise sequence, which in the present case consists of a set of identically distributed realizations of a random variable with finite variance.

The parameters of interest in (1) are the weights which indicate the influence of each individual forecast in the combined forecast. The weights in the model are restricted to sum to one, since each weight defines the fraction of information provided by each constituent prediction in the combination, i.e.

$$\sum_{i=1}^{k} w_i = 1 \tag{3}$$

By applying this restriction alone, negative weights may be obtained as there is no constraint on a minimum value for each weight. Here, weights are allowed to be negative and no consideration is given to a potential lower limit, although, according to Gunter [15], a non-negativity restriction may be assumed to be more effective and robust, in addition to being more easily interpretable from a physical point of view. The constant term in the linear combination model is included to detect and correct for any global discrepancy in the combination, i.e. any systematic error still remaining after merging the forecast information provided by each individual prediction. Any potential bias in the individual forecasts is embraced in the constant term, with the same estimated weight as for the corresponding individual forecast. This is then equivalent to correcting the bias for the combined forecast as a weighted sum of bias terms originating from the individual forecasts.

Inclusion of the additive restriction can be done in a straightforward manner: the constraint is added into the linear combination model (1) for the k^{th} weight (by expressing this weight as 1 minus the others, i.e. $w_k = 1 - \sum_{i=1}^{k-1} w_i$), and the k^{th} forecast is then subtracted. This is a quite general manner to extend a combination scheme to more than two individual forecasts, as discussed in Nielsen *et al.* [16]. This yields a new formulation in the form of a restricted linear combination model,

$$y^*_{t+h} = \mu + \sum_{i=1}^{k-1} w_i \widehat{y}^*_{i,t+h|t} + e_{c,t+h} \tag{4}$$

where $\widehat{y}^*_{i,t+h|t} = \widehat{y}_{i,t+h|t} - \widehat{y}_{k,t+h|t}$ and $y^*_{t+h} = y_{t+h} - \widehat{y}_{k,t+h|t}$. Neither the prediction error $e_{c,t+h}$ nor the intercept μ are affected by this modification. The weights are not only consistently estimated, but the restricted model in (4) also gives forecast errors for the linear combination model in (1) and the corresponding intercept.

The method applied in the following to estimate the weight coefficients of the restricted linear model (4) is the least squares method [17].

The Least-Squares Method in a more General Regression Framework

By integrating the restriction into the most general form for forecast combination, i.e. in a regression model, and considering the weights to be time-varying, we obtain

$$y^*_{t+h} = g(\widehat{\boldsymbol{y}}^*_{t+h|t}, t; \boldsymbol{w}_{t+h}) + e_{c,t+h} \tag{5}$$

where $g(\widehat{\boldsymbol{y}}^*_{t+h|t}, t; \boldsymbol{w}_{t+h})$ is a known mathematical function of the independent individual forecasts $\widehat{\boldsymbol{y}}^*_{t+h|t} = (\widehat{y}^*_0, \widehat{y}^*_{1,t+h|t}, \ldots, \widehat{y}^*_{k-1,t+h|t})^\top$, but the weights $\boldsymbol{w}_{t+h} = (\mu, w_{1,t+h}, \ldots, w_{k-1,t+h})^\top$ are unknown. To properly estimate the constant μ in the following study is $\widehat{y}^*_0 = 1$ kW. The error $e_{c,t+h}$ is as defined in the combination model (2). The general linear model is a special case of the regression model where the response is defined as a linear function of individual forecasts, i.e.

$$y^*_{t+h} = \boldsymbol{w}^\top_{t+h} \widehat{\boldsymbol{y}}^*_{t+h|t} + e_{c,t+h} \tag{6}$$

In order to estimate the weights in (6), it can be chosen to minimize a loss function $\mathcal{L}$, where $\mathcal{L}$ represents the cost of deviation between combined forecast and measurement. For the combined forecast to be competitive with the individual forecasts, its loss function value has to be less than or equal to those of the individual forecasts, i.e. $\mathcal{L}(e_{c,t+h}) \leq \min_i\{\mathcal{L}(e_{i,t+h})\}$. The loss function that is the most commonly chosen for weight estimation in (6) is the Least Squares (LS), where $\mathcal{L}(e_{c,t+h}) = \mathrm{E}[(e_{c,t+h})^2]$.

The data set is used to estimate the unknown parameters, thus the design matrix is a matrix of the constituent forecasts, including all values in the data set ($\widehat{\boldsymbol{Y}}$), and the measurements form a vector of observed values ($\boldsymbol{y}$). The solution to the problem is an estimator of the weights that minimizes the sum of squared residuals, i.e.

$$\widehat{\boldsymbol{w}} = (\widehat{\boldsymbol{Y}}^\top \widehat{\boldsymbol{Y}})^{-1} \widehat{\boldsymbol{Y}}^\top \boldsymbol{y} \tag{7}$$

where the matrix $\widehat{\boldsymbol{Y}}^\top \widehat{\boldsymbol{Y}}$ is assumed to be invertible [18].

Conditional Weighted Combination

Let us consider a class of models that are linear to some regressors, and for which the coefficients are assumed to be changing smoothly as an unknown function of some other variables. These kinds of models are called varying-coefficient models [19], but when all coefficients depend on the same variable, the model is referred to as a conditional parametric model. The formulation is related to the procedure of locally weighted regression [20, 21, 22], where the function g in the regression model (5) is now a smooth function, estimated by fitting a polynomial of the dependent variables to the response. By extending the weights in the linear combination model in (6) to be function of some external variables, a new method of combining forecasts is introduced, the so-called Conditional Weighted Combination (CWC). In the following analysis the weight restriction in (3) is implemented in the CWC model. Notations similar to those employed above are also used here. However, the mathematical framework of CWC may be seen as generic and employed for forecast combination in other applications than wind power prediction.

When using a conditional parametric model to describe the response y_t^*, the explanatory variables are divided in two groups. In the case of combining forecasts, the group of individual forecasts enter globally as in the linear regression model, while the weights depend on some other group of variables, $\boldsymbol{u}_{t+h}$. In practise will $\boldsymbol{u}_{t+h}$ denote a group of forecasted variables. The combination model in (6) hence becomes

$$y_{t+h}^* = \boldsymbol{w}(\boldsymbol{u}_{t+h})^\top \widehat{\boldsymbol{y}}_{t+h|t}^* + e_{c,t+h}, \tag{8}$$

where $\boldsymbol{w}(\boldsymbol{u}_{t+h})$ is a vector of coefficient functions to be estimated at time $t+h$, and $e_{c,t+h}$ is as described in (2). The dimension of $\widehat{\boldsymbol{y}}_{t+h|t}^*$ can be quite large, but for practical purposes the dimension of $\boldsymbol{u}_{t+h}$ must be low [23]. The weight functions are estimated at a number of distinct fitting points by locally approximating these functions using polynomials, and then fitting the resulting linear model locally to each of these points. Let $\boldsymbol{u}$ denote a particular fitting point and $\boldsymbol{p}_d(\boldsymbol{u}_{t+h})$ be a column vector of terms in the corresponding d^{th} order polynomial evaluated at $\boldsymbol{u}$. The order of the polynomial can be different for the elements in the vector $\boldsymbol{w}(\boldsymbol{u}_{t+h})$. Furthermore, we have the $k-1$ individual forecasts, $\widehat{\boldsymbol{y}}_{t+h|t}^* = [\widehat{y}_{1,t+h|t}^*, \cdots, \widehat{y}_{k-1,t+h|t}^*]^\top$, then

$$\boldsymbol{z}_{t+h}^\top = \left[\widehat{y}_{1,t+h|t}^* \boldsymbol{p}_d^\top(\boldsymbol{u}_{t+h}), \cdots, \widehat{y}_{k-1,t+h|t}^* \boldsymbol{p}_d^\top(\boldsymbol{u}_{t+h})\right], \tag{9}$$

$$\boldsymbol{\phi}^\top(\boldsymbol{u}) = \left[\boldsymbol{\phi}_1^\top(\boldsymbol{u}), \cdots, \boldsymbol{\phi}_{k-1}^\top(\boldsymbol{u})\right] \tag{10}$$

where the terms in $\boldsymbol{\phi}(\boldsymbol{u})$ are column vectors of local coefficients at $\boldsymbol{u}$, corresponding to the terms in $\boldsymbol{z}_{t+h}^\top$. The linear model $y_{t+h}^* = \boldsymbol{z}_{t+h}^\top \boldsymbol{\phi}(\boldsymbol{u}) + e_{c,t+h}$ is then fitted locally at $\boldsymbol{u}$ using weighted least squares:

$$\widehat{\boldsymbol{\phi}}(\boldsymbol{u}) = \underset{\boldsymbol{\phi}(\boldsymbol{u})}{\arg\min} \sum_{t=1}^{N} \beta_{\boldsymbol{u}}(\boldsymbol{u}_t) \left(y_t^* - \boldsymbol{z}_t^\top \boldsymbol{\phi}(\boldsymbol{u})\right)^2 \tag{11}$$

with N being the number of available observations [20]. The weights $\beta_{\boldsymbol{u}}(\boldsymbol{u}_t)$ are assigned by using the tri-cube function [21, 23].

The i^{th} element of the weighting vector $\boldsymbol{w}(\boldsymbol{u})$ is then estimated as

$$\widehat{w}_i(\boldsymbol{u}) = \boldsymbol{p}_d^\top(\boldsymbol{u})\widehat{\boldsymbol{\phi}}_i(\boldsymbol{u}) \tag{12}$$

where $\widehat{\boldsymbol{\phi}}_i(\boldsymbol{u})$ is the weighted least squares estimate of $\phi_i(\boldsymbol{u})$. The procedure described above is identified as an offline procedure, since a given set of data is used for the estimation. The coefficients depend on the values of the meteorological forecasts exclusively where no time-variation is considered. When considering potential long-term variations in the coefficients, it is possible to track adaptively such coefficients with methods similar to that proposed in Nielsen *et al.* [24]. This possibility is not, however, considered in this paper.

Application Results

So far the mathematical framework for combining forecasts has been introduced, with the introduction of a new approach allowing for conditional weighted combination (CWC) of wind power production forecasts. In the following, CWC's of two individual forecasts are considered, for which the external variables are selected from the available predicted variables of one considered meteorological forecasting system. The test case and available data are introduced in a first stage, followed by a detailed description of the study performed to select the external variables to be used in the CWC scheme, and to identify the model. Actually, the study on forecast combination of wind power predictions detailed in Nielsen *et al.* [16] is based on the same data. In that study the combination weighted is not conditional, but depends on the bias of the individual forecasts and the variance-covariance matrix for the individual forecast errors.

Description of the Case Study

Power predictions are produced using WPPT (Wind Power Prediction Tool [25]) for the Klim wind farm, which is a 21MW wind farm located in north-western Jutland, Denmark. The data is available in two separate data sets: one data set includes the wind power predictions and corresponding wind power measurements, and the second data set consists of meteorological forecasts to be used as the external variables in the CWC procedure, cf. Figure 1. A more detailed description of the data is given below:

1. Using three meteorological forecasts, WPPT is used to generate a forecast of the wind power for each meteorological forecast. In the first data set the three wind power forecasts are listed along with the corresponding measured power production. The meteorological forecasts for WPPT are provided from three different global meteorological systems. First are predictions from the meteorological system Deutscher Wetterdienst, second are predictions from the HIRLAM model of the Danish Meteorological Institute [26], third there are predictions from WPPT when using meteorological forecasts based on MM5 modelling [27] and GFS (Global Forecast System) from the National Centers for Environmental Predictions, in the USA. Furthermore, the MM5 forecasts were prepared for WPPT by CENER, the National Renewable Energies Centre, in Spain, see for instance Villanueva and Martí [28]. In the following study the individual wind power forecasts will be abbreviated DWD, HIR and MM5,

respectively. Even though the three different sets of forecasts are based on different meteorological data, they are highly correlated.

2. The second data set consists of meteorological forecasts from the Danish Meteorological Institute forecasting system HIRLAM. The objective is to generate a conditional weighted model for the combined forecast, where the weights depend on one or several of the available forecast variables. For each of four different pressure levels in the atmosphere, the variables included are wind speed, wind direction and turbulent kinetic energy, along with air density, friction velocity and radiation. The wind speed and wind direction at 10 meters above ground level also available as forecast variables, but are already used as input in WPPT for providing the power predictions for HIRLAM and, therefore, of less importance in the analysis. Correlation is strong among wind speed variables, and for the various wind direction and turbulent kinetic energy variables. As a consequence, only one variable in each group is considered in the study.

Both data sets contain forecasts provided at midnight, denoted as 00Z in the figures, with an hourly temporal resolution up to 24 hours ahead. All variables span the period February 2^{nd} 2003 to December 2^{nd} 2003, which in total gives 7272 data points for each variable in the data sets.

The individual wind power forecasts are investigated to see if the conclusion drawn from the combination improves the individual predictions for certain. The RMSE values using the competing meteorological forecasts for horizons up to 24 hours are depicted in Figure 2a. It is clear that the RMSE for all three cases increases with larger prediction horizons. MM5 is the least accurate forecast, while HIR is the best prediction. DWD appears to be somewhere in between, though closer to HIR. From the figure it could be assumed that combining the two best performing forecasts would result in the most adequate combination. Figure 2b shows that the correlation between the individual forecasts is rather high, around 0.6. Further, it can be seen that the correlation with the HIR forecast is roughly the same for all prediction horizons, while the lowest correlation appears to be that between the DWD and MM5 wind power forecasts.

——— Figure 2 is approx. here ———

Results and Discussion

With the available data for this study, the weights in the combined forecasts are considered as weight functions being dependent on possibly any of the two available meteorological forecasts, then serving as explanatory variables [29]. Selection in Thordarson [29] is performed through data analysis with coplots, employed in combination with conditionally parametric fits [30]. By doing so, any relation of a weight in the combined forecast with any two meteorological forecast variables can simultaneously be studied and demonstrated. An available meteorological forecast is selected and partitioned in several bins, and for each of the bins the weights in the combination are smoothed regarding some other available meteorological forecast (with a conditionally parametric fit). The study reveals that of the available forecasts, the two most adequately influencing weights in the combination are air density and turbulent kinetic energy.

In a first stage, Figure 3 depicts coplots with the DWD weights as the response variable (w_1), while air density (u_{ad}) and turbulent kinetic energy (u_{tke}) are seen as the explanatory variables. For both Figures 3a and 3b the partitioned variable has its lowest values in the lower left panel and, first, increases to the right and, then, upwards. The partitioning is performed so that there is the same number of data points in each of the bins. For each fraction of the partitioned variable, the panels show the distribution of a combination weight as a function of the second external variable. The fitted line in each panel is the local regression for the cloud of data points.

In Figure 3a u_{tke} is partitioned. This figure shows how the slope of the local fit between weights and u_{ad} decreases with higher values in u_{tke}, as well as the variance of the weights. For low values in u_{tke} (bottom row of panels) little or no change in slope occurs, but that is due to the density of the variable for low values. When the turbulent kinetic energy increases, first the intermediate panels show some shifting in slope of the local fit, around u_{ad} =1270, and then the slope approaches zero as seen in the top row of Figure 3a. The distribution of the points in each panel also shows how the response spreads with increasing u_{ad}, but the distribution is reduced with increasing u_{tke}. By partitioning the air density and fitting the weights to the 4$^{\text{th}}$ root of u_{tke} (Figure 3b) the local regression appears to be constant for almost all panels. Closer inspection of the fit reveals that the smoothed line shifts upwards with increasing air density. The top row of panels also show that the fit has a negative slope before it forms a concave curvature. The data points in the panels verify the distribution of the weights with increasing air density, while the variance of w_1 is constant with changes in u_{tke} within the panels.

——— Figure 3 is approx. here ———

In conclusion there seems to be coherency between forecasted values of u_{ad} and u_{tke}, and the DWD weights. The u_{ad} and u_{tke} forecasts from DWD are therefore used as external variables for the CWC model. The combination forecast is then expressed as

$$\widehat{y}_{c,t+h|t} = \mu(u_{ad}, u_{tke}) + w_1(u_{ad}, u_{tke})\widehat{y}_{1,t+h|t} + w_2(u_{ad}, u_{tke})\widehat{y}_{2,t+h|t} \tag{13}$$

Using the constraint for forecast coefficients as introduced in (4), the above equation can be rewritten as

$$\widehat{y}^*_{c,t+h|t} = \mu(u_{ad}, u_{tke}) + w_1(u_{ad}, u_{tke})\widehat{y}^*_{1,t+h|t} \tag{14}$$

The scatterplot of u_{ad} and u_{tke} makes a basis for the weight estimation in (13) but observing the u_{ad}-u_{tke} plane in Figure 4a reveals that the meteorological data does not cover the whole plane. The u_{tke} forecasts are more densely distributed for low values, while the u_{ad} forecasts are closer to being Gaussian distributed. This implies that for high values of u_{tke} and either low or high values of u_{ad}, very few values exist. The area where the weights have some valid estimation on the u_{ad}-u_{tke} plane is thus defined as the area below the dashed lines in Figure 4a.

Figure 4b shows the time series for the two DWD meteorological forecast variables of interest. Air density is known to be inversely related to air temperature. Figure 4b is indeed an illustration of this effect, since air density is high through the colder months of the year but decreases during the summer period. In contrast, turbulent kinetic energy is not found to be correlated with any of the other available meteorological variables. It varies through the entire period, although less at both tails of the data set, indicating reduced variation over the winter period. Turbulent kinetic energy is a variable used to study turbulence and its evolution in the atmospheric boundary layer. When the boundary layer becomes stable, turbulent kinetic energy is suppressed. The time series plot shows that layers of air in the atmosphere are more stable over the winter period.

——————— Figure 4 is approx. here ———————

Weights for all three possible combinations of two individual forecasts, are examined on the u_{ad}-u_{tke} plane. Only one weight from each combination is displayed since the weights are restricted to sum to one. The local polynomials for the two weight functions are considered to be of zero and of first order for μ and w_1, respectively. From (12), the local models for the weight functions are then

$$\mu(u_{ad}, u_{tke}) = \mu, \tag{15}$$

$$w_1(u_{ad}, u_{tke}) = w_{10} + w_{11}u_{ad} + w_{12}u_{tke} \tag{16}$$

where the parameters are estimated with the weighted least squares method in (11). Inserting the local models directly into the restricted linear combination model in (14) would then make our combination model equivalent to the following linear model

$$\widehat{y}^*_{c,t+h|t} = \mu + w_{10}\widehat{y}^*_{1,t+h|t} + w_{11}u_{ad}\widehat{y}^*_{1,t+h|t} + w_{12}u_{tke}\widehat{y}^*_{1,t+h|t} \tag{17}$$

Combining wind power forecasts as in (17) indicates that the weights are linearly dependent on the forecasts of the two meteorological variables considered, i.e. air density and turbulent kinetic energy. If such an equivalent linear model were then considered, the parameters μ, w_{10}, w_{11} and w_{12} would be estimated as constant terms, independent from the meteorological variables. In contrast, for our proposed combination scheme, the weights are allowed to be unknown general functions of the meteorological variables, and by smoothing the weights over the u_{ad}-u_{tke} plane, the surface plots in Figure 5 are obtained. Since we only consider the combination of 2 forecasts, and the weights are constrained to sum to one, the weight surfaces in this figure represent the relative importance given to either one or the other constituent wind power forecast, depending on the forecast value of air density and turbulent kinetic energy. Figure 5 shows that the weights (w_1) on DWD, when combined with HIR or MM5, appear to have similar surfaces. For low values of u_{tke}, the DWD weights become larger as u_{ad} increases, but for higher values of turbulent kinetic energy the weights are reduced with progressing air density. Hence this defines zones where wind power predictions from DWD meteorological forecasts are more or less to be trusted in comparison with those based on HIR and MM5 meteorological forecasts. In parallel, the behavior of the HIR weight when combined with MM5 is more challenging to interpret. There is some fluctuation for the low values of u_{tke}, but with increasing turbulent kinetic energy the surfaces' variation is reduced. It can also be seen that the share of the HIR forecasts is higher in the combination for larger values of turbulent kinetic energy. Finally, the surfaces for the intercepts are similar where low u_{tke} implies high negative value for the intercept, but with increasing u_{tke} the intercepts increase as well. The intercepts depending on u_{ad} show some kind of bell-shaped structure where the high and low values of air density imply low value for the intercept, but around mean u_{ad} the intercept is at its maximum.

——————— Figure 5 is approx. here ———————

Considering two meteorological forecasts to fit the weights in the CWC model gives rather smooth surfaces, especially for the weight in the DWD/HIR combination (Figure 5b). For minimum turbulent kinetic energy, the weight appears to increases linearly with increase in air density, with some relative slope and intercept. However, with increasing u_{tke} this linear relation is maintained, but the slope is reduced and the intercept is shifted upwards. Thus, to simplify the CWC model further the shape of the surface for w_1 in the DWD/HIR combination is merged into the model. From now on, let us only focus on the DWD weight in the DWD/HIR combination in order to study how the combination scheme could be further enhanced.

From Figure 5b it can be concluded that there is a linear relationship between the weight and air density, where both intercept and slope are functions of the turbulent kinetic energy, i.e.

$$w_1(u_{ad}, u_{tke}) = v_{10}(u_{tke}) + v_{11}(u_{tke})u_{ad}, \tag{18}$$

where v_{10} and v_{11} are the intercept and the slope, respectively. The plot shows that for increase in u_{tke} the intercept increases as well, but the slope decreases and becomes negative for u_{tke} higher than 8000. Figure 5a is a surface plot of corresponding intercept in (14), but it shows a wave-like behavior for low values of u_{tke} around the mean value of u_{ad}, which is challenging to interpret. However, the influence of the variance of the intercept can be appraised by comparing the terms of the CWC model in (14), i.e. the variance of the intercept is compared to the product of the forecast weight and the forecast. A comparison of the diagonal terms in the covariance matrix for the product terms in the model shows that the second term in (14) is much greater than the first term. This indicates that the intercept term does not have much influence in the combination [29].

By omitting the deep valley on the surface of the intercept, the relationship between the weight and the air density appears to be a bell-shaped function which fades out with increasing u_{tke}. Such a function is difficult to formulate, and its implementation in the model would give a complicated interpretation. The naive assumption is that u_{ad} does not affect the intercept but the intercept is a function of u_{tke}, such as

$$\mu(u_{ad}, u_{tke}) = v_0(u_{tke}). \tag{19}$$

This assumption might be a bit crude but will make it easier to integrate the proposed functions from (18) and (19) into the conditional weighted model. This finally gives

$$\begin{aligned}\widehat{y}^{*}_{c,t+h|t} &= v_0(u_{tke}) + \left[v_{10}(u_{tke}) + v_{11}(u_{tke})u_{ad}\right]\widehat{y}^{*}_{1,t+h|t}\\ &= v_0(u_{tke}) + v_{10}(u_{tke})\widehat{y}^{*}_{1,t+h|t} + v_{11}(u_{tke})\widehat{y}^{*}_{1,t+h|t}u_{ad}\\ &= v_0(u_{tke}) + v_{10}(u_{tke})z_1 + v_{11}(u_{tke})z_2\end{aligned} \tag{20}$$

where $z_1 = \widehat{y}^{*}_{1,t+h|t}$ and $z_2 = \widehat{y}^{*}_{1,t+h|t}u_{ad}$. The model in (20) is a CWC model where the coefficients only depend on one unknown variable instead of two, namely the turbulent kinetic energy. Figure 6 shows how the weights in (20) change with u_{tke}. The same valley appears for the intercept, i.e. as in Figure 5a, and the same test is performed as before to estimate sufficiency of the variance of the intercept. The diagonal elements of the matrix reveal that the variance of the intercept is only a fraction of the other variances, and, therefore, the valley at u_{tke} of around 3000 can be neglected. The opposite behavior of the parameters v_{10} and v_{11} is not surprising since they form the weight in the original model (14). From Figure 6 it can be assumed that the parameters, v_{10} and v_{11}, are linear functions of u_{tke} which changes slope when u_{tke} is approximately 9000.

——— Figure 6 is approx. here ———

In a last stage, the quality of the forecasts obtained from the combination methods is investigated using the RMSE criterion. For a discussion on error measures in wind power forecasting, we refer to Madsen *et al.* [12]. First; if comparing the forecast accuracy of combined forecasts with that of the individual forecasts in Figure 2a, it is clear that forecast combination leads to a significant decrease in RMSE. Notice that combinations involving the least accurate forecasts (that is, MM5) are as good as those involving the most accurate ones (DWD and HIR). This shows that one should not just pick the most accurate predictions as input for combination, since their correlation may also have an impact on the quality of the combined forecasts. This point is further discussed in Nielsen *et al.* [16]. Second; in Figure 7a the CWC forecasts are compared to the LS-combination, for the three combinations of the two forecasts considered. For the shorter prediction horizons the CWC method performs similar to the LS method. However, for the intermediate and larger prediction horizons the CWC method outperforms the LS predictions. The improvement of applying the CWC is displayed in Figure 7b in percentages. It shows how the LS model is performing better for the first five to nine horizons, but for the intermediate horizons CWC outperforms LS by up six percent for the HIR/MM5 combination and four percent for DWD/HIR and DWD/MM5. For the larger horizons the improvement is reduced for all combinations, but still in favor of the CWC method. On average, the performance of CWC is 1% better then the LS combination, but considering only prediction horizons larger than 8 hours ahead this improvement is closer to 2%. By looking at each combination, the average prediction horizon in DWD/HIR, DWD/MM5 and HIR/MM5 shows improvement of 0.9%, 2.1% and 2.5%, respectively. The performance of the alternative CWC proposed in (20) (i.e. weights depending on only one external variable, but with an adapted model structure) for the specific case of the DWD/HIR combination is also depicted in Figure 7 (denoted as CWC.1 in the legend of the Figure). It appears to be approaching the performance of the LS combination method (in terms of NRMSE) with a slight improvement.

———————— Figure 7 is approx. here ————————

The results show that a first impression of the introduced CWC model is capable of improving a linear model for combining forecasts in all three combinations. The advantage of the new method over the classical approach is obvious, even though the results appear to be marginal. Conditioning the weights on the meteorological forecasts establish a new scheme for combining forecasts that can easily be extended, such that further improvement is achieved.

Conclusions

The combination of wind power forecasts, when several meteorological forecasts are available, is known to lead to an increase in forecast accuracy. This has also been confirmed by the analysis in the present paper. Various methods for combining forecasts have been presented with the classical linear regression model as a basis. This particular model has been thoroughly described, in its form including a constant term and a weight restriction, and for both offline and online cases. It has been shown that not considering a constant term to account for bias in the combination model may lead to a loss of forecast accuracy. Therefore, our advice for future developments related to combining wind power forecasts is to always account for this bias term.

In addition, an original proposal in this paper is to replace the constant weights traditionally found in combination schemes with weights that are nonparametric functions of some external variables. This new combination scheme is referred to as Conditional Weighted Combination (CWC). The weight functions are estimated using local polynomial regression. For wind power prediction, the combination scheme can then be made conditional on the evolution of some meteorological variables. This goes along with an idea originally proposed in Lange *et al.* [10] which consists of a forecast combination, conditioned to the prevailing weather situation. Here, our alternative proposal is to consider that forecasts of certain meteorological variables may directly be used to determine the weights in the combination of several wind power predictions.

A study of available forecasts of meteorological variables showed that forecasts of air density and turbulent kinetic energy are relevant candidates for conditioning forecast combinations. It is not possible to generalize application of these two variables to other wind farms, located in different terrains and subject to different wind climatologies. However, the proposed methodology can be employed for other wind farms to determine which forecasts of meteorological variables may be considered for conditioning a wind power forecast combination scheme. As a consequence, perspectives regarding future work must obviously include applications of the proposed conditional weighted combination scheme to other test cases, i.e. to wind farms located in regions with different climates, and for which a variety of meteorological forecasts are used as input to wind power prediction. This may permit validation of the importance of employing air density and turbulent kinetic energy as external variables in the combination scheme in general. In parallel, this may reveal the interest of considering different external variables depending on local wind climatologies, or on the type of meteorological forecasts employed.

From a mathematical point of view, the method proposed to estimate the weight functions in the CWC model should be made (time-) adaptive, by employing some recursive formulation of the problem, and exponential forgetting of past observations. A method similar to that described in Nielsen *et al.* [24] could be applied for this purpose. This would certainly further improve the forecast accuracy of the combined forecasts. Finally, emphasis should be put on providing uncertainty information along with combined forecasts. Even if forecast uncertainty is reduced by employing forecast combination schemes, the characteristics of the uncertainty in the combined forecasts obtained is also becoming more complex. This is due to the fact that overall uncertainty is then driven by an evolving weighting of the situation-specific uncertainty of each individual wind power forecasts used as input. Relevant nonparametric methods taking the aspects into account will be the focus of further studies.

Acknowledgements

This study was sponsored by the Danish Public Service Obligation (PSO) fund FU5766, and we would like to express out gratitude for this sponsorship. We are also grateful to the Danish Meteorological Institute for providing the meteorological data set used in the study. The HIRLAM data was kindly provided by Elsam Kraft A/S (now part of DONG Energy), the MM5 data was prepared by CENER, the Spanish National Renewable Energies Centre, with data originally provided by the National Center for Environmental Prediction (USA), and the DWD forecasts were obtained commercially via financial support from the Danish PSO fund under contract FU2101.

References

[1] Thor S-E, Weis-Taylor P. Long-term research and development needs for wind energy for the time frame 2000-2020. *Wind Energy* 2002; **5**:73–75.

[2] Giebel G, Kariniotakis G, Brownsword R. The state-of-the-art in short-term prediction of wind power. *Technical Report, Anemos Project Deliverable Report* D1.1, 2003. Available online: http://anemos.cma.fr (accessed March 1st 2008).

[3] Hibon M, Evgeniou T. To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting* 2005; **21**:15–24.

[4] Bates JM, Granger CWJ. The combination of forecasts. *Operational Research Quarterly* 1969; **20**:451–468.

[5] Clemen RT. Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting* 1989; **5**:559–589.

[6] de Menezes LM, Bunn DW, Taylor JW. Review of guidelines for the use of combined forecasts. *European Journal of Operational Research* 2000; **120**:190–204.

[7] Gunter SI. Theoretical justification of the efficiency of simple average combination. Working paper, Temple University, Philadelphia, PA; 1990.

[8] Bunn DW. A bayesian approach to the linear combination of forecasts. *Operational Research Quarterly (1970-1977)* 1975; **26**:325–329.

[9] Sánchez I. Adaptive combination of forecast with application to wind energy forecast. *International Journal of Forecasting* 2008; **24**:679–693.

[10] Lange M, Focken U, Meyer R, Denhardt M, Ernst B, Berster F. Optimal combination of different numerical weather models for improved wind power predictions. In: *Proc. of the 6th Inter-*

national Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Farms, Delft, The Netherlands, 2006.

[11] Lange M, Focken U. *Physical Approach to Short-Term Wind Power Prediction*. Springer Verlag: Berlin, 2005.

[12] Madsen H, Pinson P, Kariniotakis G, Nielsen HAa, Nielsen TS. Standardizing the Performance evaluation of short-term wind power prediction models. *Wind Engineering* 2005; **29**:475–489.

[13] Granger CWJ, Ramanathan R. Improved methods of combining forecasts. *Journal of Forecasting* 1984; **3**:197–204.

[14] Holden K, Peel DA, Thomson JL. *Economic Forecasting: An Introduction*. Cambrigde University Press: Cambridge, 1990.

[15] Gunter SI. Non-negativity restricted least squares combinations. *International Journal of Forecasting* 1992; **8**:45–59.

[16] Nielsen HAa, Nielsen TS, Madsen H, San Isidro Pindado MJ, Martí Perez I. Optimal combination of wind power forecasts. *Wind Energy* 2007; **10**:471–482.

[17] Madsen H. *Time Series Analysis*. Chapman & Hall/CRC: London, 2007.

[18] Basilevsky A. *Applied Matrix Algebra in the Statistical Sciences*. Dover Publications, Inc: Mineola, New York, 2005.

[19] Hastie T, Tibshirani R. Varying coefficient models. *Journal of the Royal Statistical Society. Series B* 1993; **55**(4):757–796.

[20] Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 1979; **74**:829–836.

[21] Cleveland WS, Devlin SJ. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* 1988; **83**:596–610.

[22] Hastie T, Loader C. Local regression: automatic kernel carpentry. *Statistical Science* 1993; **8**:120–129.

[23] Hastie T, Tibshirani RJ. *Generalized Additive Models*. Chapman & Hall: London, 1990.

[24] Nielsen HAa, Nielsen TS, Joensen AK, Madsen H, Holst J. Tracking time-varying-coefficient functions. *International Journal of Adaptive Control and Signal Processing* 2000; **14**:813–828.

[25] Madsen H, Nielsen HAa, Nielsen TS. A tool for predicting the wind power production of offshore wind plants. In: *Proc. of the Copenhagen Offshore Wind Conference & Exhibition*, Copenhagen, Denmark, 2005.

[26] Feddersen H, Sattler K. Verification of wind forecasts for a set of experimental DMI-HIRLAM ensemble experiments. DMI Scientific Report 05-01, 2005.

[27] Dudhia J, Gill D, Manning K, Wang W, Bruyere C. PSU/NCAR mesoscale modeling system tutorial class notes and users' guide (MM5 modeling system version 3). Pennsylvania State University/National Center for Atmospheric Research numerical, January 2005. `http://www.mmm.ucar.edu/mm5/`.

[28] Villanueva J, Martí I. Improving wind resource assessment and wind forecast by using a CFD model developed in CENER. *Proc. of the Conference: The Science of Making Torque from Wind*, Delft, The Netherlands, April 19–21, 2004.

[29] Thordarson F. Optimal combined wind power forecasts using exogeneous variables. Master Thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2007.

[30] Cleveland WS. Coplots, nonparametric regression and conditionally parametric fits. In: *Multivariate Analysis and its Applications* (Anderson TW, Fang KT, Olkin I, ed.). IMS Lecture Note-Monograph Series 1994; **24**:21–36.

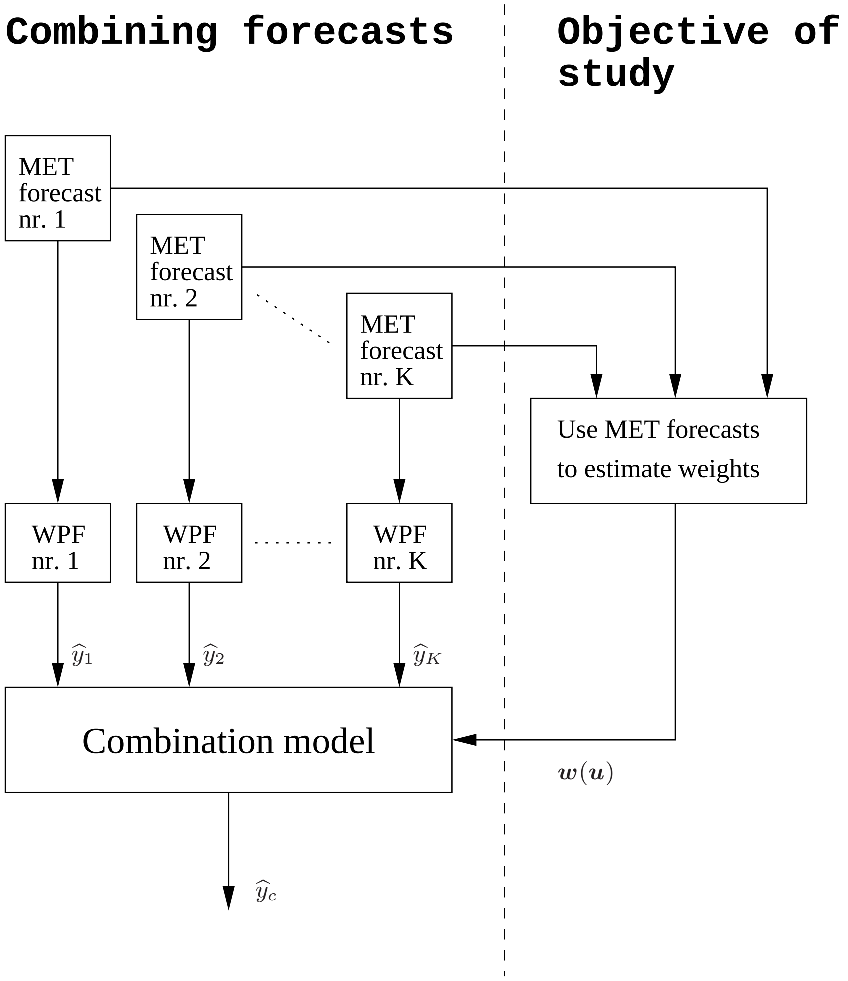
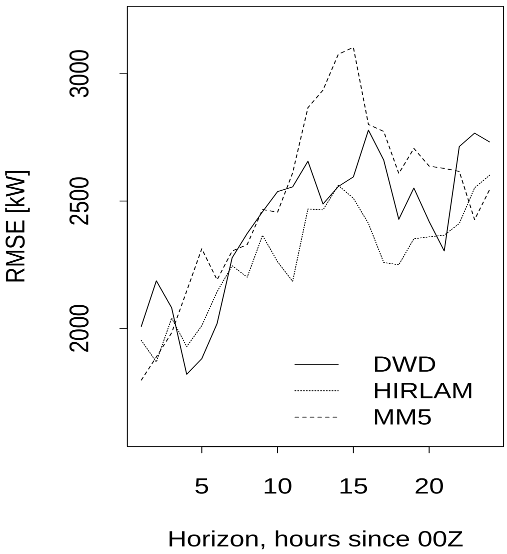
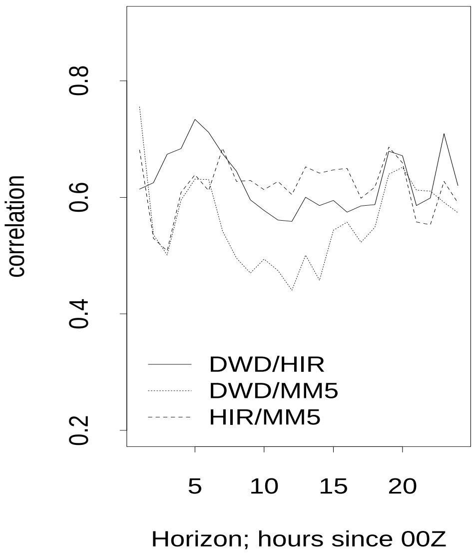


Figure 1: Block diagram of the process of combining wind power forecasts. To the left of the dashed line the flow for combining wind power forecasts (individual forecasts abbreviated WPF in diagram) is described, while the right side shows the model for the weights with the MET forecasts as input.

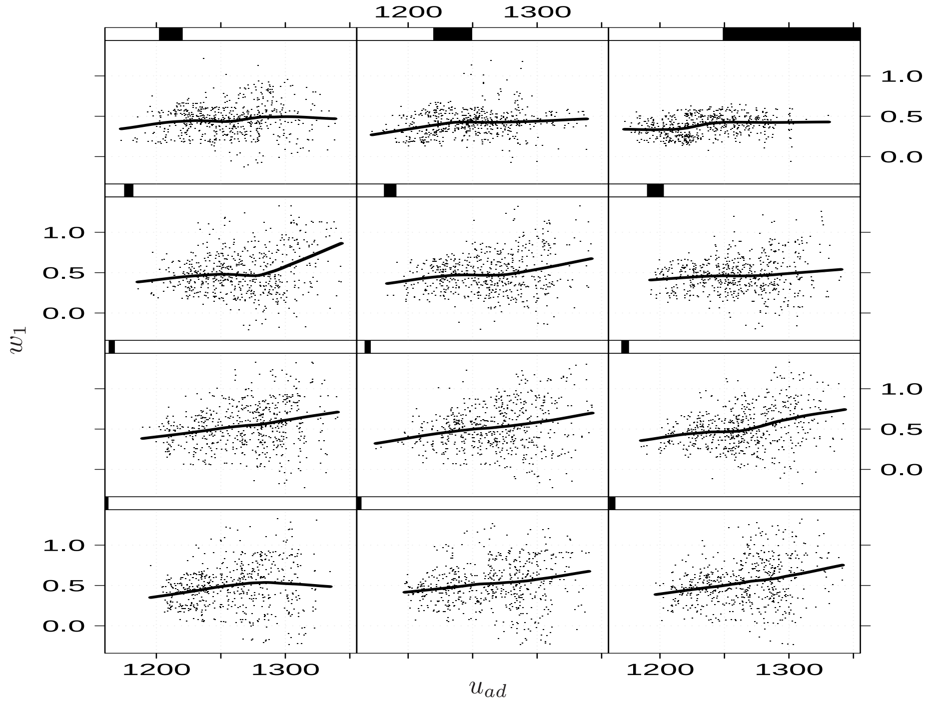


(a) Performance comparison between the three individual wind power forecasts.

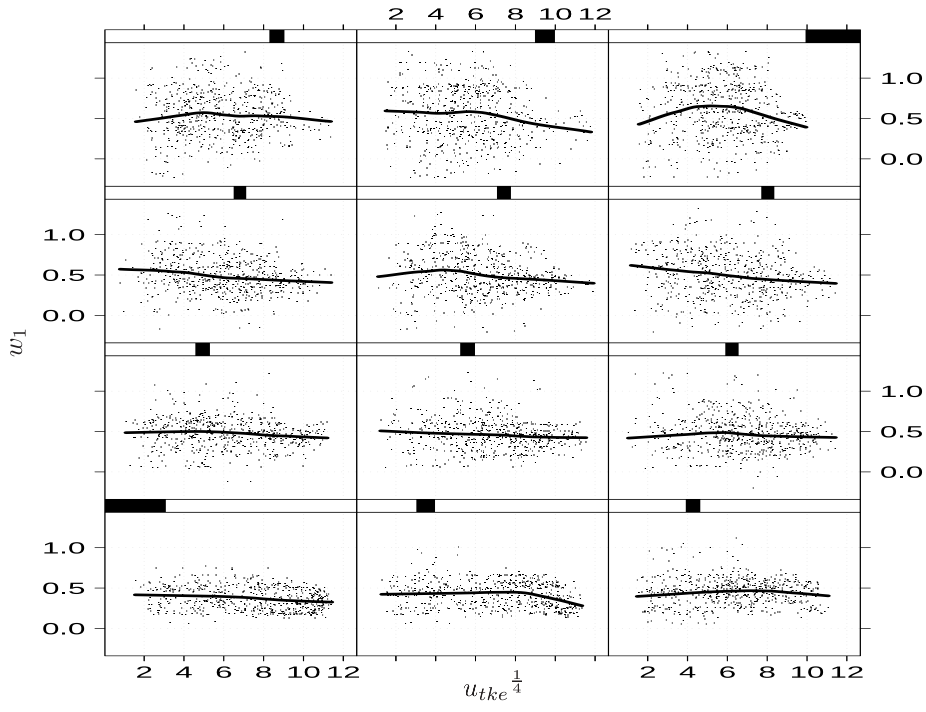


(b) Correlation between any possible combination of two wind power forecasts.

Figure 2: Comparison between the individual wind power forecasts available for the study. Left: Performance comparison between the individual forecasts used in the study. Hirlam forecast is the best, but MM5 the worst. Right: Correlation between any possible combination of two individual forecasts. the correlations are similar, though obviously the lowest for DWD/MM5 combination.

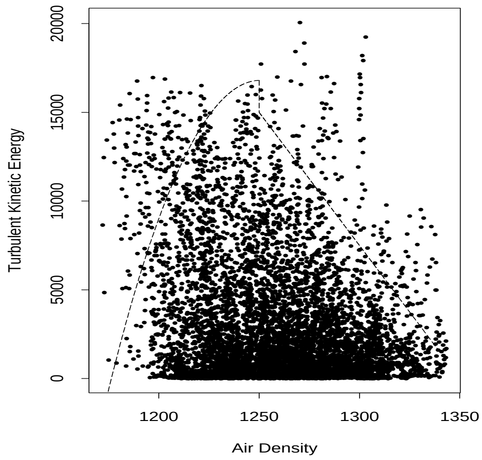


(a) Scatterplot of weights and u_{ad} where u_{tke} is partitioned.

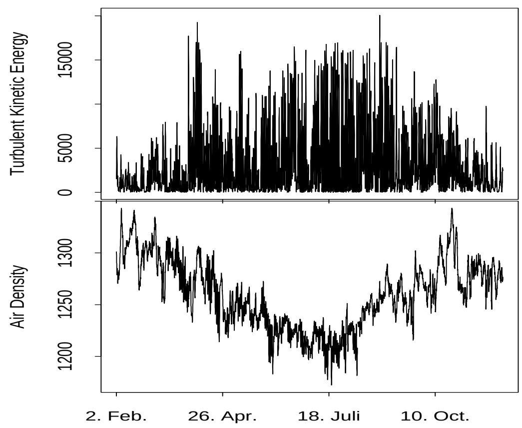


(b) Scatterplot of weights and $u_{tke}^{\frac{1}{4}}$ where u_{ad} is partitioned.

Figure 3: Coplots where DWD weights are depending on air density (g/m^2)and turbulent kinetic energy (Wm^2/s).

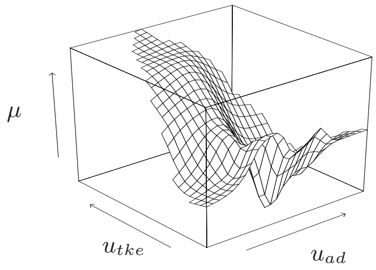


(a) Scatterplot of u_{ad} and u_{tke} variables; the u_{ad}-u_{tke} plane

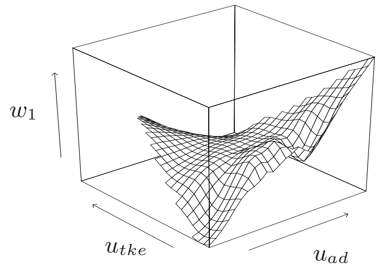


(b) Time series for u_{ad} and u_{tke}

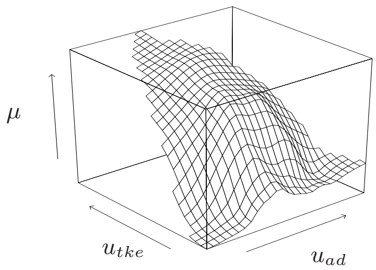
Figure 4: Scatterplot and time series plots for the air density and turbulent kinetic energy variables. Left: The relation between variables shows that main part of the data points have low values for u_{tke}, whilst u_{ad} appears to be normal distributed. Right: The time series for u_{tke} varies extremely during the warmer months, while u_{ad} is quite stable and has higher values during the cold period than in the summer period.



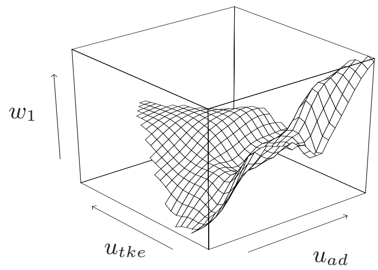
(a) Intercept in DWD/HIR combination



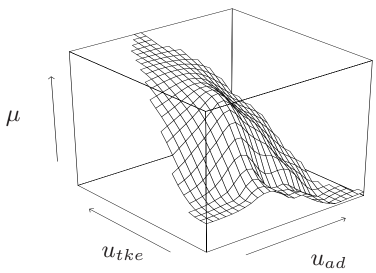
(b) DWD weights in DWD/HIR combination



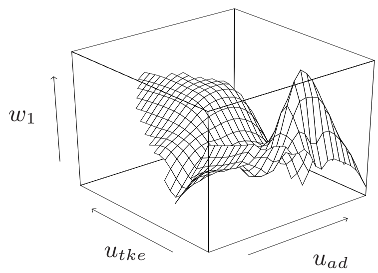
(c) Intercept in DWD/MM5 combination



(d) DWD weights in DWD/MM5 combination



(e) Intercept in HIR/MM5 combination



(f) HIRLAM weights in HIR/MM5 combination

Figure 5: Surface plots for the estimated intercepts ($\mu(u_{ad}, u_{tke})$) and the estimated weights ($w_1(u_{ad}, u_{tke})$).

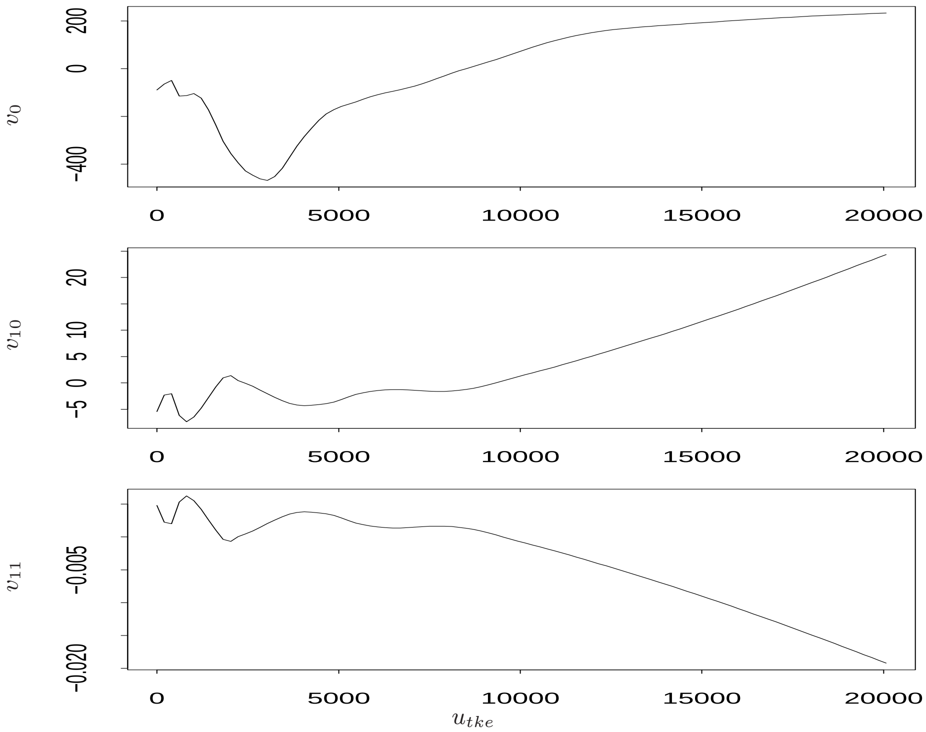
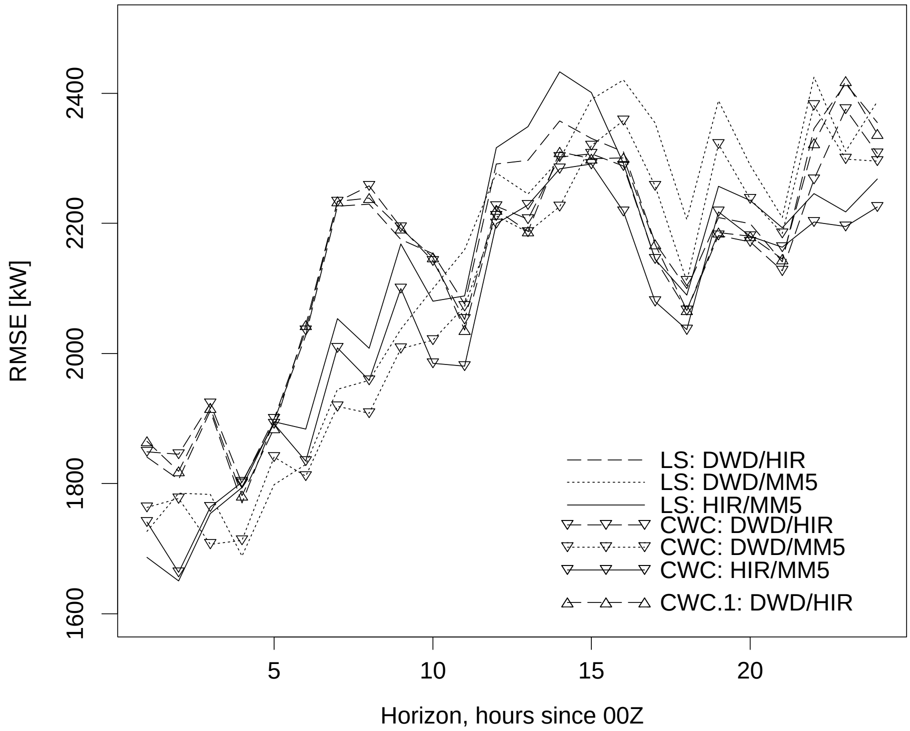
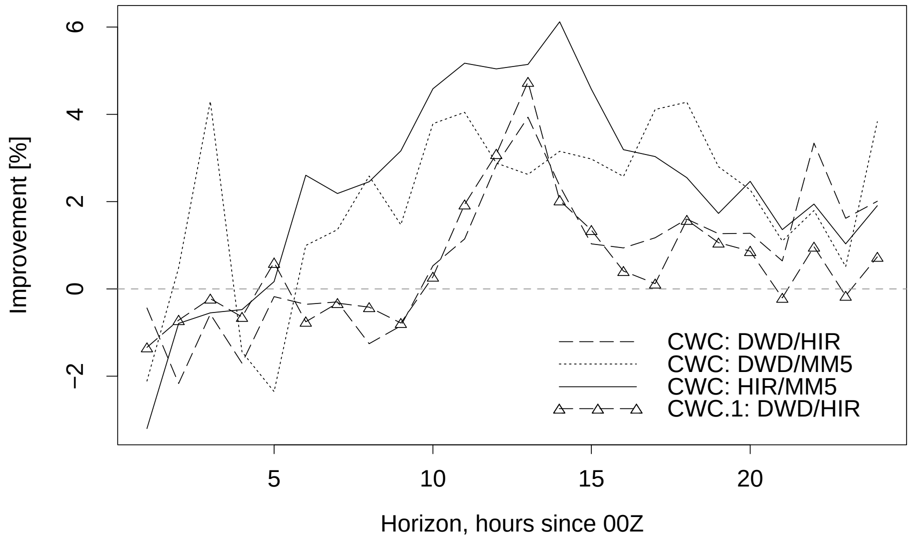


Figure 6: Coefficients in (20) *as a functions of turbulent kinetic energy. The intercept v_0 is the same as seen in Figure 5a, but through the mean value for u_{ad}. The parameter v_{11} is a slope coefficient for the weight being a function of u_{ad}, while v_{10} considers the bias between the weight and air density.*



(a) RMSE values for the LS and CWC methods.



(b) Improvements by applying the CWC scheme.

Figure 7: Comparison of the methods for all 24 prediction horizons. Overall, the CWC method is outperforming the LS-method, especially for the prediction horizons larger than 8 hours. The simplified CWC model (CWC.1) shows very limited improvement over the corresponding LS model.