Combined Forecast and Quantile Regression

Jan Kloppenborg Møller, Henrik Aalborg Nielsen and Henrik Madsen
Informatics and Mathematical Modelling
Richard Pedersens Plads
DTU-Bygning 321
2800 Lyngby
Denmark

October 31, 2007

Contents

1 Introduction

This report will examine the distribution of errors from a combined forecasting system. Combined forecast use forecasts based on different meteorological forecasts to predict future wind power production. How these combined forecasts are constructed is described in [8], but basically the model is

$$p_c = w_0 + \sum_{j=1}^{J} w_j p_j \tag{1}$$

where $\sum_{j=1}^{J} w_j = 1$ and p_j is the predicted power based on forecast j. The weights w_j are based on the correlation structure between the p_j's. The scope of this report is to examine the residuals from the combined forecast. This will be done by quantile regression as presented by Koenker in [2], and in addition some of the models will be examined with the adaptive method presented in [5] and [6].

1.1 Quantile Regression

The quantile models presented in [2] are linear models, such that the τ quantile given a vector of explanatory variables $\mathbf{x}_t$ is

$$Q_\tau(\mathbf{x}_t) = \mathbf{x}_t^T \boldsymbol{\beta} \tag{2}$$

This is fitted with an asymmetrical and piecewise linear loss function

$$\rho_\tau(r) = \begin{cases} \tau r & \text{for} \quad r \geq 0 \\ (\tau - 1) r & \text{for} \quad r < 0 \end{cases} \tag{3}$$

with $r_t = \hat{Q}_\tau(\mathbf{x}_t) - y_t$, minimizing the sum of these loss functions give the τ quantile. So the estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ given N observations is

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \rho_\tau(r_i) \tag{4}$$

In addition to the combined forecasts and the forecasts that form this, horizon and weigths will also be considered as explanatory variables. In this presentation the 25% and 75% quantiles will be used to evaluate different modelstructures.

1.2 Non-Linear Relations with Linear Regression

Quantiles of the residuals as functions of forecasted power is not expected to be a linear function. To get to the setting of linear quantile regression these functions are therefore approximated using spline basis function (see e.g. [1]), these are combined in an additive model, such that we get the model

$$\hat{Q}_\tau(x_1, ..., x_K) = \hat{\alpha}(\tau) + \sum_{j=1}^{K} \hat{f}_j(x_j) \tag{5}$$

where the functions $\hat{f}_j$ in this presentation are either linear combinations of natural spline basis functions or linear functions. x_j is the explanatory variables or some function of these. The functions must be fixed somehow to ensure uniqueness, this is done by forcing all spline basis functions to be zero at the left most boundary knot. $\hat{\alpha}$ is a common intercept for the functions.

The functions f_j are modeled with

$$\hat{f}_j(x_j) = \sum_{l=1}^{n_j} b_{j,l}(x_j)\hat{\beta}_{j,l} \tag{6}$$

where $b_{j,l}$ is the spline basis function no. l for the explanatory variable x_j and $\hat{\beta}_{j,l}$ is the coefficient to estimate.

This presentation will now go on and present the data and give a short introduction to performance measures of quantiles and then go to the presentation of some static models. The last part of the presentation give the result for some adaptive models.

2 The Data

The data set consist of hourly measurements of power production from Klim wind power plant, the predicted power from WPPT (see [3] and [4]) based on three different weather forecasting systems and the combined forecast. The prediction horizons are between 1 and 24 hours. There are 7272 data points in the data set, these span the period from February 2nd 2003 to December 12th 2003. For the analysis of the static models the dataset is divided into two part, a training set and a test set of approximately equal sizes, the training set have 3648 data points and span the period from February 2nd 2003 to July 4th 2003. The test period is the rest of the data points.

The explanatory variables in the data set is the horizon, the combined forecast and the three different forecasts used by the combined forecast. The forecasts based on different meteorological data are correlated and these are of course also correlated to the combined forecast. Therefore the objective is to model the uncertainty as a function or risk index of the combined forecast and a some function of the three forecasts, further it will be examined if the horizon should be included. The models will be additive models, and we use quantile regression to model this, the explanatory variables will be denoted

hor: The prediction horizon from the meteorological data

p_c: The combined forecast based on the three different meteorological forecasts

p_{DWD}: Predicted power based on the meteorological forecast from "Deutscher Wetterdienst"

p_{HIR}: Predicted power based on the meteorological forecast from "DMI-Hirlam"

p_{MM5}: Predicted power based on the meteorological forecast from "MM5"

3 The performance of Quantiles

The subject of giving performance measures to quantiles is not easy and there does not exist a generally accepted measure of performance of quantiles. We will present a number of different performance measures here, these methods are discussed in [9] and [5]. The performance measures presented here should all be considered on the test set.

3.1 Reliability

The reliability is simply the relative number of observations below the model. A good model should have reliability close to the required reliability, but it is not clear how close this should be. [9] go through different measures of quantiles and mention that reliability should always be considered before any other performance is considered.

3.2 Skill score

The objective function for quantile regression is also a performance parameter that will be considered, this is shown to be a skill score in [9]. The loss function for a quantile is the skill score for this quantile and the sum of two symmetric (around 50%) quantiles is a skill score for this interval. The skill score for an interval give one number and is therefore practical for comparing different models. This number should be minimized.

3.3 Reliability distance

In addition to reliability the reliability distance $d_q(x_j; tau) = d(q(y; \tau, w))$ as defined in [5] will also be considered, w is a smothing parameter and will be set to 0.1 in the presentation. Reliability distance is a performance parameter based on local reliability, also defined in [5]. Reliability distance basically measure the distance between reliability as a function of some variable and the required reliability. Reliability distance is one number while local reliability is a curve (see figure 7 and 8). The reliability distance is considered for the combined forecast, the prediction horizon, and time, it is considered as $d_q(x_j)^2 = d_q(x_j, 0.25)^2 + d_q(x_j, 0.75)^2$.

3.4 Crossings

Because quantiles are fitted individually crossings between quantile curves can (and will) occur, the number of such crossings ($\sum I(IQR < 0)$, with IQR=Inter Quantile Range, i.e $IQR = Q_{0.75} - Q_{0.25}$) on the test set is considered as a performance parameter. In addition the mean ($E(IQR|IQR < 0)$) and maximum size $min(IQR)$ of these crossings is considered. These parameters should of course be as small as possible.

3.5 Sharpness

Sharpness is a measure that tells how far quantiles are separated, given that other performance parameters are good then this should be small. Two different measures are used here. Namely the mean of IQR (Inter Quartile Range, i.e. $IQR = Q_{0.75} - Q_{0.25}$) and the median ($\tilde{E}$) of IQR. Crossings, which are undesirable will contribute positively to sharpness, therefore only IRQ's larger than zero are considered here.

3.6 Resolution

Resolution measure how well a model distinguish between different situations, this is given by measures that award variation in IQR. Here two measures are used, namely the standard deviation of IQR and the Median Absolute Deviation (MAD), i.e. the median of the absolute deviation from the median, this is multiplied with 1.44826 for normal consistency. Only IQR's larger than zero are considered.

4 Some Static Quantile Models

The main assumption is that the combined forecast is the most important explanatory variable. In addition we will examine how the forecasts which form the combined forecast can contribute to the model. When comparing models in this study, the skill score for the interval will be considered the most important performance parameter.

4.1 A Risk Index

The first hypothesis that quantiles can be modeled as

$$Q_\tau(p_c, r_m, hor) = \alpha(\tau) + f_1(p_c) + f_2(r_m) + f_3(hor) \tag{7}$$

where r_m is a risk index defined as

$$r_m^2 = \frac{1}{3}\sum_{i=j}^{3}(p_j - \overline{p})^2 \tag{8}$$

and p_j is the three forecast described above. Figure 1 show histograms of the combined forecast and the risk index defined above. The figure illustrate that there are very few observations with high values of the risk index, this can result in difficulties for the model to explain what is going on in these areas of the data space.

Figure 2 top row show the model defined in Eq. (7) with 10 degrees of freedom in each direction. The knots for the spline basis functions are placed at 10% quantiles of observed data, the knots are marked with the rugs on the first axis. The function in the direction of the combined forecast behave as expected, with large uncertainties for moderate values of predicted power.

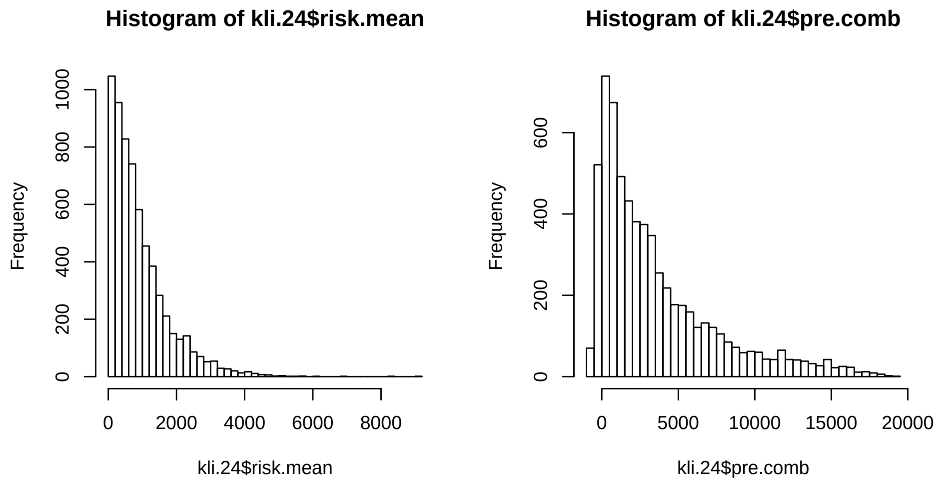

Figure 1: The figure show histograms of combined forecast and the risk index calculated by Ep. (8). Notice that there are very few observations with high values of the risk index.

The curves in the direction of the risk index show large distance between the two quantiles for large values of of risk index, this is what would be expected, but the curves only varies in areas where there are very few data so this curve should not be trusted too much. The horizon does not seem to explain any variation in the data and this will therefore not be considered further.

Table 1 give the performance of Model 1 and some other models based on the combined forecast and the risk index defined in Eq. (8). Two of these are also displayed in Figure 2. The two bottom rows in Figure 2 show models without horizon and with only 5 knots in the direction of the combined forecast. The curve in the direction of combined forecast show the same qualitative behavior as with 10 knots, with the most significant difference being that the difference between the quantiles is much larger for large forecasts. It should however again be noted that the number of observations in this area of the data space is very small.

The curve in the direction of the risk index behave quite different when there are 5 knots instead of 10, this support the point that the curve should not be trusted. When the risk index is fitted with a linear function it still

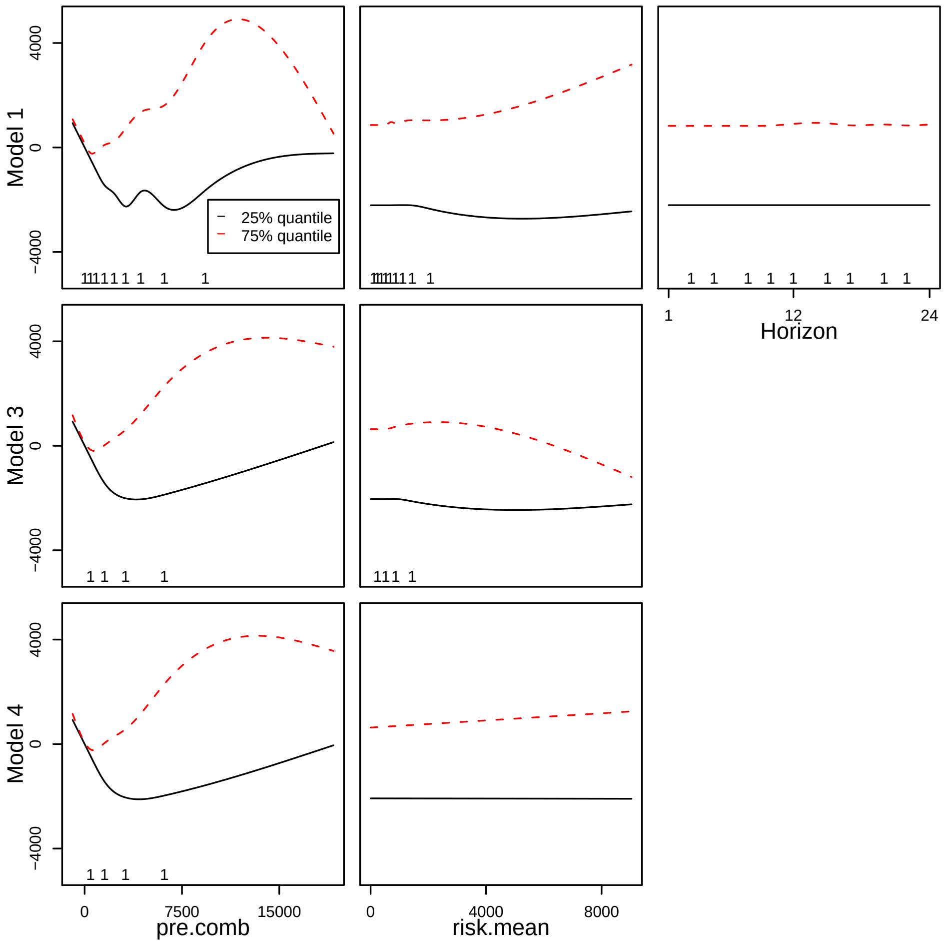

Figure 2: The figure show the effects of different components of some of the additive models presented in Table 1.

Model	1	2	3	4	5
df hor	10	0	0	0	0
df p_c	10	10	5	5	5
df r_m	10	10	5	1	0
Relai 25%	20.7	20.6	**21.8**	21.4	21.6
Relai 75%	73.5	73.6	73.3	**74.2**	73.6
Relai Interval	53.3	53.5	**52.4**	54.0	52.5
$\bar{\rho}_{0.25}(\mathbf{r})$	458.6	458.6	**454.9**	455.0	455.0
$\bar{\rho}_{0.75}(\mathbf{r})$	556.8	557.6	557.9	**556.4**	557.0
$\bar{\rho}_{0.25}(\mathbf{r}) + \bar{\rho}_{0.75}(\mathbf{r})$	1015.4	1016.1	1012.8	**1011.4**	1011.6
$d_q(p_c)$	0.068	0.068	0.070	**0.065**	0.069
$d_q(\text{hor})$	0.042	0.042	**0.037**	0.039	0.038
$d_q(t)$	0.037	0.037	**0.033**	0.036	0.035
$\sum I(IQR < 0)$	42	**24**	75	53	71
min(IQR)	-9.2	-1.4	-6.4	-8.2	**-1.2**
$E(IQR \mid IQR < 0)$	-1.4	**-0.1**	-1.6	-3.5	-0.9
$E(IQR \mid IQR > 0)$	2317.9	**2314.0**	2334.2	2324.8	2333.0
$\tilde{E}(IQR \mid IQR > 0)$	2099.0	**2093.3**	2230.3	2220.8	2226.5
$SD(IQR \mid IQR > 0)$	1869.0	**1889.8**	1801.8	1800.6	1800.2
$MAD(IQR \mid IQR > 0)$	2196.5	2249.6	**2520.7**	2493.9	2505.6

Table 1: The table show different performance parameters for different models, knots are placed at appropriate quantiles of data. The best performer in each row is marked in bold face letters.

show some dependence even though it is weak.

Table 1 show that a model (Model 4) with the risk included as a linear function give the best performance. It does however also show that the improvement in the performance is very small compared to a model with only the combined forecast included. Note also that Model 4 is the only model that perform better than a model without any risk included.

The risk index defined above use the deviation from the average as the risk factor. Another measure is based on the deviation from the combined forecast this is calculated by replacing $\bar{p}$ with p_c in Eq. (8). The result of this is given in Table 2, where model 8 and 9 are such models. The table further give the result of defining the risk index as the absolute deviation between p_{DWD} and p_{HIR}. The last index is examined because the weather forecast from MM5 and HIRLAM are known to be very correlated and the

Model	6	7	8	9
df p_c	5	5	5	5
df r_{DH}	5	1	0	0
df r_c	0	0	5	1
Relai 25%	22.0	20.5	**21.5**	20.6
Relai 75%	73.5	73.8	72.1	**74.4**
Relai Interval	51.8	53.3	**50.5**	54.6
$\bar{\rho}_{0.25}(\mathbf{r})$	455.6	**454.9**	455.3	455.0
$\bar{\rho}_{0.75}(\mathbf{r})$	558.6	557.2	**555.3**	555.8
$\bar{\rho}_{0.25}(\mathbf{r}) + \bar{\rho}_{0.75}(\mathbf{r})$	1014.2	1012.1	**1010.6**	1010.8
$d_q(p_c)$	**0.062**	0.068	0.070	0.064
$d_q(\text{hor})$	**0.034**	0.043	0.044	0.043
$d_q(t)$	**0.029**	0.041	0.037	0.040
$\sum I(IQR < 0)$	67	48	98	**34**
min(IQR)	**-2.7**	-61.9	-16.9	-5.0
$E(IQR\|IQR < 0)$	**-1.2**	-6.6	-7.1	-2.7
$E(IQR\|IQR > 0)$	2334.2	2314.4	2353.2	**2303.3**
$\tilde{E}(IQR\|IQR > 0)$	2241.2	**2192.9**	2222.4	2198.3
$SD(IQR\|IQR > 0)$	1786.7	1803.5	**1814.7**	1796.7
$MAD(IQR\|IQR > 0)$	**2495.5**	2489.1	2488.5	2493.9

Table 2: The table show different performance parameters for different models, knots are placed at appropriate quantiles of data. The best performer in each row is marked in bold face letters.

hypothesis is therefore that the deviation between one of these and p_{DWD} alone can explain the uncertainty.

Table 2 show that the risk index obtained from the deviation from the combined forecast perform better than those obtained with the mean of the forecasts, but the improvements is not seen in all performance parameters. The overall reliability is quite poor for the 25% quantile in all models while it perform better for the 75% quantile.

The risk index considered so far does not take into account that the weight change over time. It is therefore tested if this should be included. The weights sum to one as should be the case, they are however not all positive. To ensure that the risk index is greater than zero the risk index is defined such that the weights used for the risk index are proportional to the absolute size of the weights and that they sum to one. This is ensured with

Model	10	11	12	13
df p_c	5	5	5	5
df $r_{c,w}$	5	1	0	0
df $r_{m,w}$	0	0	5	1
Relai 25%	21.8	21.4	**22.2**	20.9
Relai 75%	71.9	**75.0**	73.3	74.6
Relai Interval	**50.8**	54.3	52.1	54.6
$\bar{\rho}_{0.25}(\mathbf{r})$	456.2	**455.0**	455.4	**455.0**
$\bar{\rho}_{0.75}(\mathbf{r})$	557.5	**555.6**	557.2	556.5
$\bar{\rho}_{0.25}(\mathbf{r}) + \bar{\rho}_{0.75}(\mathbf{r})$	1013.7	**1010.5**	1012.6	1011.5
$d_q(p_c)$	0.080	0.065	0.074	**0.064**
$d_q(\text{hor})$	0.045	0.037	**0.036**	0.042
$d_q(t)$	0.038	0.036	**0.030**	0.038
$\sum I(IQR < 0)$	123	**31**	79	40
min(IQR)	-24.4	-13.3	-14.4	**-11.3**
$E(IQR\|IQR < 0)$	-11.4	-6.4	**-2.7**	-5.1
$E(IQR\|IQR > 0)$	2368.9	**2310.2**	2350.8	2317.3
$\tilde{E}(IQR\|IQR > 0)$	2242.9	**2171.7**	2215.4	2199.5
$SD(IQR\|IQR > 0)$	1823.2	1796.2	**1843.4**	1807.4
$MAD(IQR\|IQR > 0)$	2454.4	2450.3	2467.9	**2492.3**

Table 3: The table show different performance parameters for different models, knots are placed at appropriate quantiles of data. The best performer in each row is marked in bold face letters.

the following construction

$$r_{m,w}^2 = \sum_{j=1}^{J} \frac{|w_j|}{\sum_{j=1}^{J} |w_j|} (p_j - \bar{p})^2 \tag{9}$$

a risk index $r_{c,w}$ based on the combined forecast is also constructed, the result of this is given in Table 3. The results is quite similar to what have been seen so far. There is a very small improvement for Model 11 which uses the weighted risk index with the combined forecast.

The risk index considered so far does not seem to give significant improvements compared to a model with only the combined forecast included.

4.2 Other Models

The approach with a risk index did not seem to produce significant improvements compared to models with only the combined forecast included. Models that take the individual forecasts into account is therefore tested.

The first approach is to set up models with the forecasts included as nonlinear functions. Table 4 give the results for combinations of this, all models have the combined forecast included. The table show large improvements for some of the models, especially Model 20 where all the forecasts are included. The forecasts and the combined forecast are connected through the formula given by Eq. (1), therefore a models that include all the forecasts might be expected to be too correlated to do a good estimation. This is partly supported by Figure 3 where the change in the function for the combined forecast seems to be compensated for by the three other functions.

The models does not break down completely because the nonlinear functions created by the spline basis functions are defined by the knots, which depends on the observations, and further the weights changes over time. If the functions had been linear and the weights constant then the construction in Model 20 would not have been possible.

In conclusion the performance parameters support Model 20, while Figure 3 indicate that the components are too correlated.

The result from Model 20 support that the forecasts based on different meteorological forecasts contain information on the uncertainty of the combined forecast. Figure 3 does however indicate that the information is not used in the right way.

The uncertainty is expected to depend on some kind of deviation between the forecasts. The next step is now to let the functions depend on such differences, the next models are therefore functions of the differences between the forecasts and the combined forecast.

Performance parameters for models of this type is given in Table 5, the functions are assumed to be linear. Model 27 perform very well, at least in terms of interval skill score. Figure 4 show the components of some of the models from Table 5, it is seen that there are clear trends, these are however not as we would expect since we have forced a linear function through and we would expect large differences between quantiles for large absolute numbers of the differences. It is also noted that there are large crossings in the plots, these are however not seen in the performance parameter. A point is again

Model	14	15	16	17	18	19	20
df p_c	5	5	5	5	5	5	5
df p_{DWD}	5	0	0	5	5	0	5
df p_{HIR}	0	5	0	5	0	5	5
df p_{MM5}	0	0	5	0	5	5	5
Relai 25%	22.7	22.6	22.6	23.6	24.3	23.6	**25.0**
Relai 75%	74.2	77.2	73.3	77.3	**74.6**	76.8	76.9
Relai Interval	51.9	55.4	52.5	54.2	**51.3**	54.1	52.8
$\bar{\rho}_{0.25}(\mathbf{r})$	452.8	455.0	458.4	**452.5**	455.8	455.9	453.4
$\bar{\rho}_{0.75}(\mathbf{r})$	554.8	560.1	558.8	551.1	555.5	557.7	**545.2**
$\bar{\rho}_{0.25}(\mathbf{r}) + \bar{\rho}_{0.75}(\mathbf{r})$	1007.6	1015.1	1017.1	1003.6	1011.3	1013.7	**998.6**
$d_q(p_c)$	0.062	0.060	0.069	0.058	**0.052**	0.054	**0.052**
$d_q(\text{hor})$	0.031	0.035	0.026	0.032	**0.016**	0.024	0.024
$d_q(t)$	0.028	0.031	0.030	0.028	0.028	**0.025**	0.030
$\sum I(IQR < 0)$	55	36	95	**32**	65	40	38
min(IQR)	-5.0	**-3.7**	-12.8	-5.7	-13.6	-14.0	-13.5
$E(IQR\|IQR < 0)$	-2.8	**-2.3**	-6.4	-3.5	-4.3	-5.6	-4.7
$E(IQR\|IQR > 0)$	2295.2	2417.8	2335.1	2374.3	**2278.4**	2379.7	2361.4
$\tilde{E}(IQR\|IQR > 0)$	2177.4	2286.1	2192.1	2268.5	**2161.4**	2234.4	2237.5
$SD(IQR\|IQR > 0)$	1779.8	**1887.8**	1803.9	1808.8	1737.2	1858.9	1753.2
$MAD(IQR\|IQR > 0)$	2352.7	**2461.2**	2320.3	2363.3	2251.0	2276.8	2267.6

Table 4: The table show different performance parameters for different models, knots are placed at appropriate quantiles of data. The best performer in each row is marked in bold face letters.

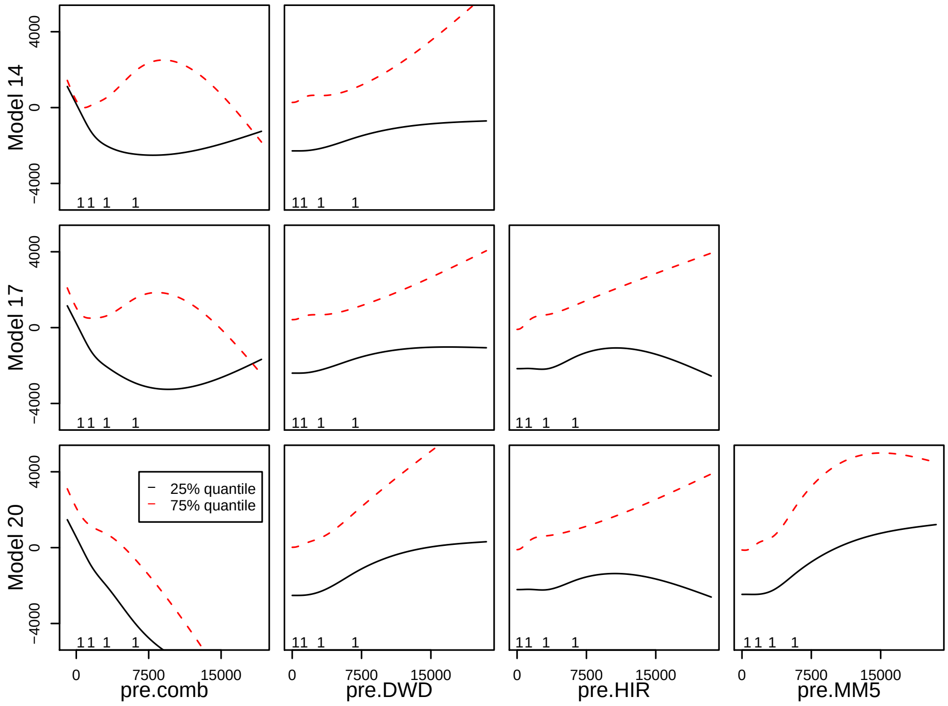

Figure 3: The figure show the effects of different components of some of the additive models presented in Table 4, the top rows is model 1 and second row is model 2.

Model	21	22	23	24	25	26	27
df p_c	5	5	5	5	5	5	5
df $p_c - p_{DWD}$	1	0	0	1	1	0	1
df $p_c - p_{HIR}$	0	1	0	1	0	1	1
df $p_c - p_{MM5}$	0	0	1	0	1	1	1
Relai 25%	**24.4**	29.2	22.6	29.1	25.7	28.6	29.2
Relai 75%	**75.4**	77.4	74.2	77.7	76.5	77.8	78.3
Relai Interval	51.0	48.2	51.6	48.7	**50.8**	**49.2**	49.1
$\bar{\rho}_{0.25}(\mathbf{r})$	455.0	454.8	454.9	454.3	454.1	454.5	**452.8**
$\bar{\rho}_{0.75}(\mathbf{r})$	557.0	557.1	549.3	552.5	549.1	549.7	**541.8**
$\bar{\rho}_{0.25}(\mathbf{r}) + \bar{\rho}_{0.75}(\mathbf{r})$	1013.0	1011.8	1004.2	1006.8	1003.3	1009.4	**994.8**
$d_q(p_c)$	**0.063**	0.094	**0.063**	0.091	0.064	0.087	0.084
$d_q(\text{hor})$	**0.023**	0.044	0.028	0.043	0.030	0.042	0.047
$d_q(t)$	**0.025**	0.041	0.030	0.041	0.038	0.040	0.049
$\sum I(IQR < 0)$	34	27	51	20	15	20	**13**
min(IQR)	-9.4	**-3.2**	**-3.2**	-11.3	-6.5	-7.1	-16.7
$E(IQR\|IQR < 0)$	-5.3	-2.5	**-2.3**	-6.2	-4.5	-4.0	-7.4
$E(IQR\|IQR > 0)$	2326.0	2402.8	**2278.9**	2375.9	2335.3	2392.6	2365.7
$\tilde{E}(IQR\|IQR > 0)$	2169.8	2197.5	2170.7	2203.1	**2153.6**	2188.0	2188.2
$SD(IQR\|IQR > 0)$	1804.2	**1887.8**	1740.5	1837.3	1759.6	1858.6	1748.6
$MAD(IQR\|IQR > 0)$	2431.1	**2518.5**	2378.4	2467.8	2273.6	2441.0	2281.3

Table 5: The table show different performance parameters for different models, knots are placed at appropriate quantiles of data. The best performer in each row is marked in bold face letters.

that there are very few observations in these areas of the data.

The models from Figure 4 are now extended to be nonlinear functions, the point that there are very few observations at high absolute values, does however mean that the spline functions give some quite large crossings. E.g. if the linear functions given above are replaced with spline basis functions, then the worst case crossing have a value of -2847. Therefore values in the tail of the distribution are removed by setting all value above the 95% empirical quantile equal to the 95% empirical quantile and all values below the 5% empirical quantile equal to the 5% empirical quantile.

The result of this is given in Table 6 and Figure 5. The table confirm the point that all three forecasts have to be considered and that we then get a quite large improvement compared to the models with one risk index, but

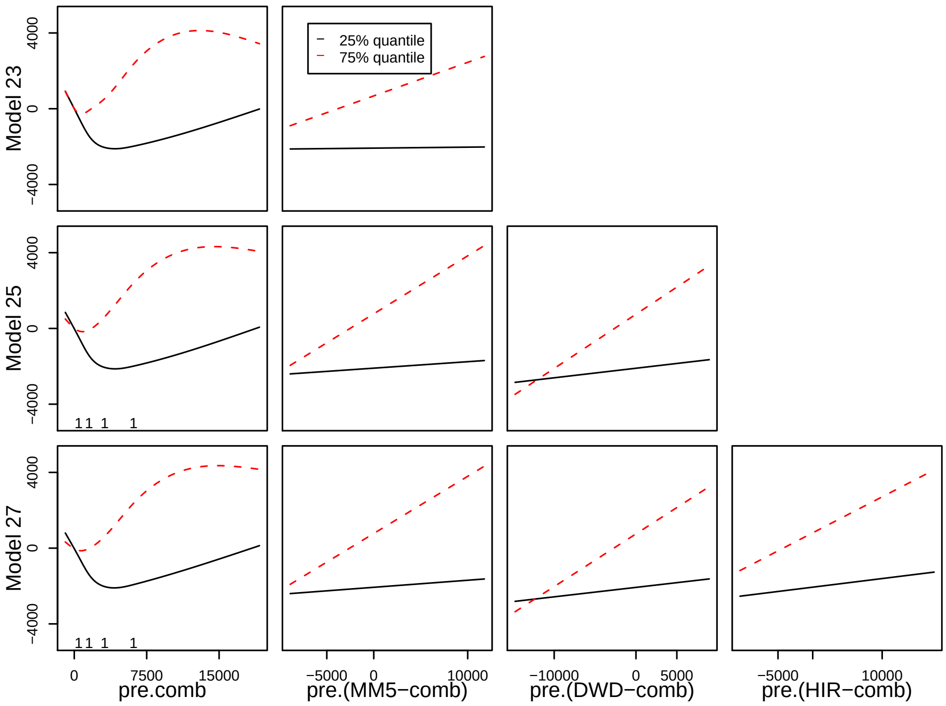

Figure 4: The figure show the effects of different components of some of the additive models explained in Table 5.

Model	28	29	30	31	32	33	34
df p_c	5	5	5	5	5	5	5
df $p_{DWD} - p_c$	5	0	0	5	5	0	5
df $p_{HIR} - p_c$	0	5	0	5	0	5	5
df $p_{MM5} - p_c$	0	0	5	0	5	5	5
Relai 25%	22.7	22.8	21.6	**25.1**	23.5	23.3	25.6
Relai 75%	**75.1**	78.9	74.4	79.0	79.0	79.3	79.9
Relai Interval	**52.5**	56.1	52.8	54.0	54.8	56.1	54.4
$\bar{\rho}_{0.25}(\mathbf{r})$	453.8	454.4	454.5	452.3	450.8	453.2	**450.3**
$\bar{\rho}_{0.75}(\mathbf{r})$	559.3	559.5	556.6	557.9	559.6	558.9	**551.6**
$\bar{\rho}_{0.25}(\mathbf{r}) + \bar{\rho}_{0.75}(\mathbf{r})$	1013.1	1013.9	1011.2	1010.3	1010.4	1012.2	**1001.9**
$d_q(p_c)$	0.062	0.070	0.055	0.061	**0.050**	0.057	0.057
$d_q(\text{hor})$	0.036	0.043	0.031	0.046	**0.024**	0.040	0.040
$d_q(t)$	**0.030**	0.036	0.033	0.033	0.031	0.038	0.041
$\sum I(IQR < 0)$	42	**4**	47	17	38	17	12
min(IQR)	-4.7	**-0.28**	-24.9	-9.6	-17.7	-19.5	-32.3
$E(IQR\|IQR < 0)$	-2.8	**-0.26**	-8.4	-4.2	-5.2	-8.7	-15.0
$E(IQR\|IQR > 0)$	**2311.1**	2410.6	2353.2	2424.8	2395.6	2459.6	2425.4
$\tilde{E}(IQR\|IQR > 0)$	**2179.8**	2220.9	2191.1	2213.8	2212.4	2332.8	2296.7
$SD(IQR\|IQR > 0)$	1841.1	1847.5	1812.7	**1863.0**	1860.0	1833.1	1795.3
$MAD(IQR\|IQR > 0)$	**2471.0**	2434.4	2398.4	2412.9	2378.8	2454.6	2361.6

Table 6: The table show different performance parameters for different models, knots are placed at appropriate quantiles of data. The best performer in each row is marked in bold face letters.

the performance was better with the linear hypothesis on the differences.

As have been discussed above, information seems to be reused as the different forecasts are taken into account. We therefore propose a hierarchic structure, where information used in one level is subtracted in the next level, this looks like

$$\begin{array}{lllll} 1 & p_c & p_1 & p_2 & p_3 \\ 2 & & p_1 - p_c & p_2 - p_c & p_3 - p_c \\ 3 & & & p_2 - p_1 & p_3 - p_1 \\ 4 & & & & p_3 - p_2 \end{array}$$

The idea is now to use one variable from each level. In the first row this is the combined forecast, in the second row p_1 is chosen as the forecast that

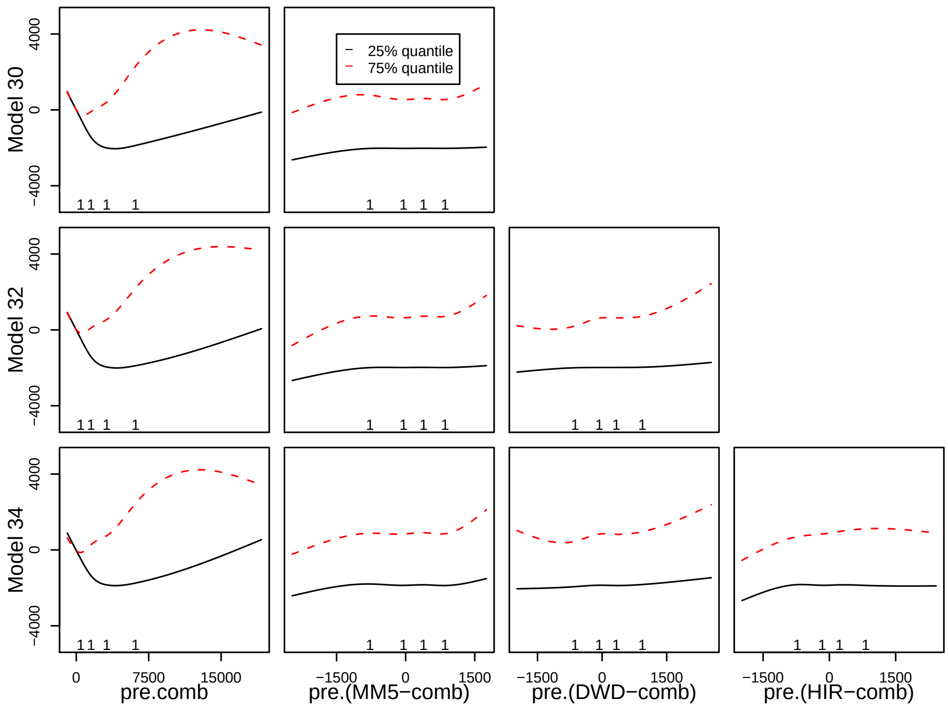

Figure 5: The figure show the effects of different components of some of the additive models presented in Table 6.

Model	30	36	37	38
df p_c	5	5	5	5
df $p_{MM5} - p_c$	5	5	5	5
df $p_{DWD} - p_{MM5}$	0	5	0	0
df $p_{HIR} - p_{MM5}$	0	0	5	5
df $p_{DWD} - p_{HIR}$	0	0	0	5
Relai 25%	21.6	**24.4**	22.7	29.7
Relai 75%	**74.4**	76.2	77.7	78.2
Relai Interval	52.8	51.8	55.3	**48.5**
$\bar{\rho}_{0.25}(\mathbf{r})$	454.5	453.1	452.3	**449.2**
$\bar{\rho}_{0.75}(\mathbf{r})$	556.6	557.8	553.7	**550.6**
$\bar{\rho}_{0.25}(\mathbf{r}) + \bar{\rho}_{0.75}(\mathbf{r})$	1011.2	1010.9	1006.1	**999.8**
$d_q(p_c)$	0.055	0.054	**0.052**	0.079
$d_q(\text{hor})$	0.031	**0.026**	0.029	0.046
$d_q(t)$	0.033	0.034	**0.030**	0.047
$\sum I(IQR < 0)$	47	36	43	**7**
min(IQR)	-24.9	**-5.0**	-52.1	-28.3
$E(IQR\|IQR < 0)$	-8.4	**-3.0**	-16.7	-9.5
$E(IQR\|IQR > 0)$	2353.2	2356.8	2435.9	**2308.7**
$\tilde{E}(IQR\|IQR > 0)$	2191.1	2161.4	2236.0	**2147.0**
$SD(IQR\|IQR > 0)$	1812.7	1806.4	**1866.6**	1752.5
$MAD(IQR\|IQR > 0)$	2398.4	2317.2	**2450.7**	2319.2

Table 7: The table show different performance parameters for different models, knots are placed at appropriate quantiles of data. The best performer in each row is marked in bold face letters.

gives the best result, etc. Model 28-30 presented in Table 6 give the result we need to chose p_1, this is the forecast based on MM5 (Model 30). The next rows is chosen in the same way, the results is presented in Figure 6 and Table 7. The performance is about the same for models that include all forecast, expect for Model 27 which perform somewhat better.

It is worth to note that that skill score have improved in the model presented in Table 7 at the same time as the reliability have moved away from the required reliability. The same tendency is seen in the reliability distance, this stress the difficulty of measuring the performance of quantiles.

This section have discussed different models for quantiles of combined forecasts as proposed in [8], the conclusion is that the individual forecasts

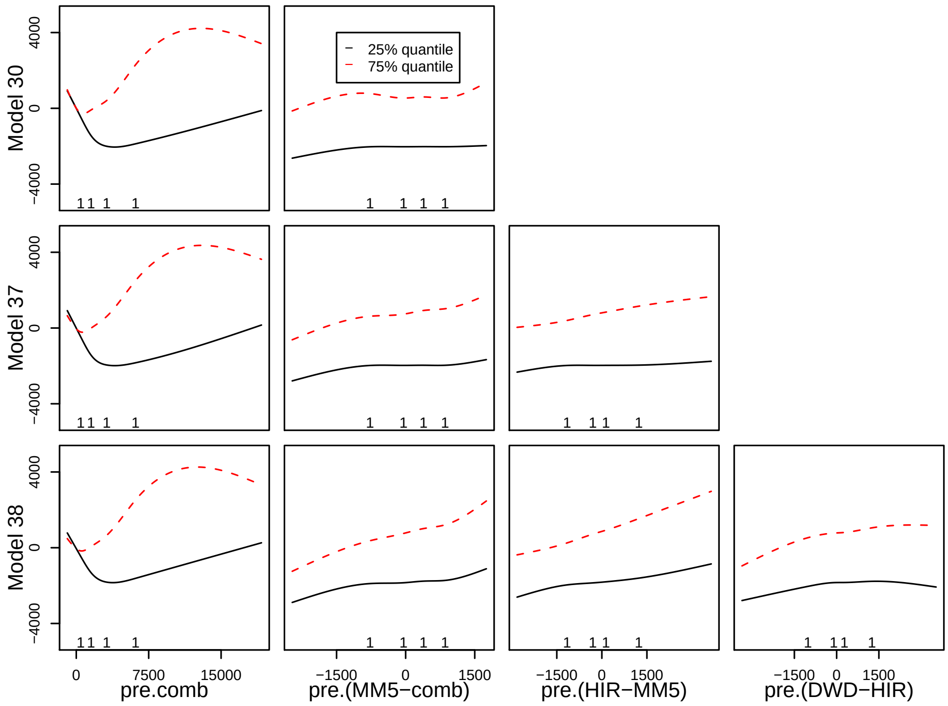

Figure 6: The figure show the effects of different components of some of the additive models presented in Table 7.

contain information on the uncertainty of these predictions. The treatment also point to the difficulties of measuring performance of quantiles. Since there is not one generally accepted method a number of different performance parameter are considered, and these does not always point to the same model.

Figure 7 show local reliability for some of the models presented in this section. The figures show that the overall reliability cover at times very large local deviation from the required reliability. This is especially clear in the direction of the combined forecast.

The next section will test some of the static models in an adaptive setting. The models which will be tested in this setting are some of the simple models and some of the best performing (and more complicated) models.

5 Adaptive Models

The adaptive procedures used here are presented in [5] where a description of the algorithms are given along with a treatment of a data set from a wind power plant. [6] give a technical description of the algorithms, and finally [7] outline some of the results from [5].

The adaptive method need an updating procedure, the one chosen here is the same as used in [5] and [7], where new observations are classified as belonging to some bin and the oldest observation from this bin is then removed. The number of observations allowed in each bin is 300 and the bins are defined by the knots for the spline basis functions in the direction of the combined forecast.

Table 8 and 9 give the results from the runs of the adaptive procedure. It is seen that the performance parameters are improved for most of the models.

The skill score for the interval have improved for most of the models, but the skill score for the 25% quantile is worse for most models while the 75% quantile have improved for all models. So in this sense the conclusion should be that models for the 75% quantile should be adaptive while models for the 25% quantile should not be.

The reliability distance have improved for most of the models and for most of the variables, the improvements are largest in the direction of the combined forecast. This point is also illustrated in Figure 8 where the local

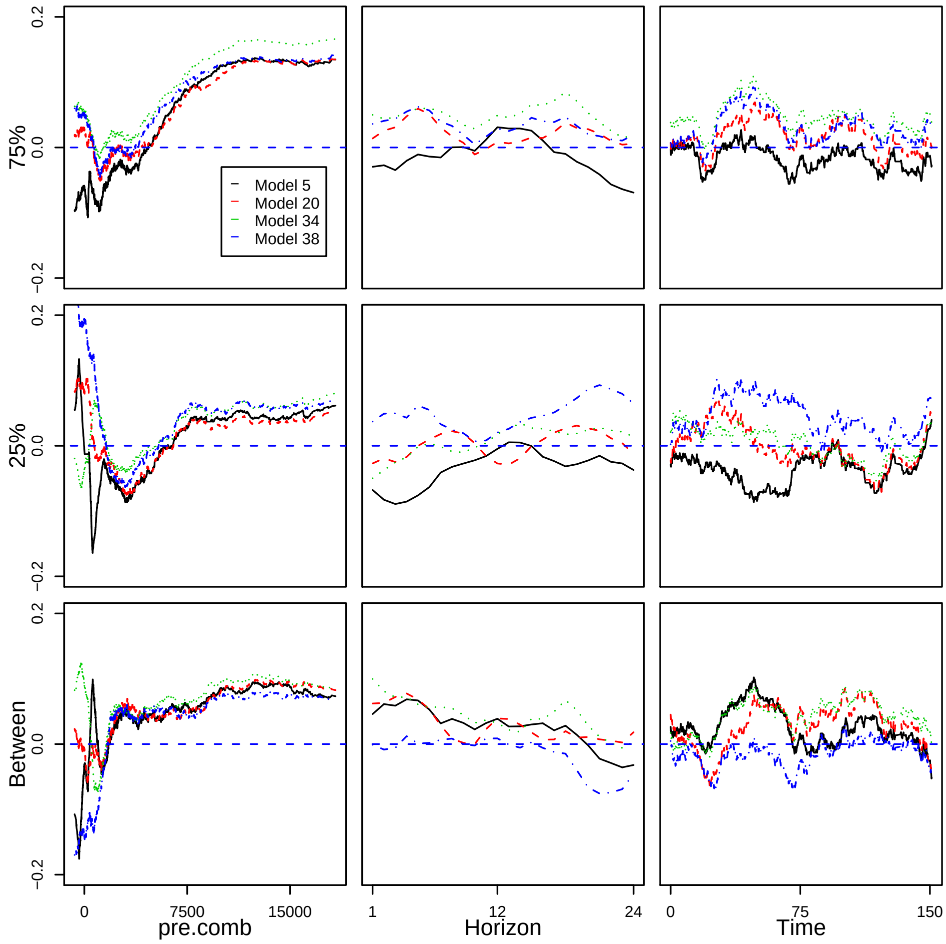

Figure 7: The figure show local reliability in the direction of forecasted power, horizon and time for some of the models presented in this section.

reliability is plotted as a function of the combined forecast, horizon and time, for the adaptive versions of the models in Figure 7. Comparing Figure 7 with Figure 8 show the improvements very clearly.

The performance of the adaptive models is worse with respect to crossings than the static models. For models with many degrees of freedom the extreme crossings can be very large for the adaptive models.

Sharpness have improved for all models, while resolution becomes worse for all models. The relative improvements for sharpness from the static to the adaptive models are between 5% and 12%, this means that the interval that cover the central 50% of data is smaller at the same time as the reliability and skill score improved.

For adaptive models the cpu time used for a step forward is also a performance parameter. Therefore the mean cpu time used for one step in the models is also given here. Not surprisingly the time increase with the number of degrees of freedoms in the models. It is also worth to notice that the the 25% quantile models are more time consuming than the 75% quantile models, at the same time as the most significant improvements was on the 75% quantiles.

6 Conclusion

The conclusion is that the input for the combined forecast can be used to explain some of the uncertainty in the quantile models. It is still not clear exactly how this should be done, the best performing model is linear in the differences between the combined forecast and the input forecasts, while the hypothesis would be that the inter quartile range should increase with the distance from zero. It does however seems that the input forecasts should be used individually rather than through risk indices.

The adaptive method generally give good results in the performance parameters. There are however big differences between the two examined quantiles, the 25% quantile models does not improve and most of the models get a little worse w.r.t. skill score, while all 75% quantile models improve by going to an adaptive setting. The reliability measure seems to improve, but this have only been recorded as one common number for both quantiles.

Model	A5	A4	A9	A11	A14	A17	A20
df	6	7	7	7	11	16	21
Relai 25%	22.0	22.8	22.0	22.7	**25.5**	26.0	25.6
Relai 75%	73.4	73.1	73.2	73.7	73.5	**75.1**	73.8
Relai Interval	51.4	**50.5**	51.5	51.5	49.3	49.5	48.6
Skill 25%	**454.4**	457.8	457.4	456.1	459.0	461.4	463.2
Skill 75%	546.2	546.0	543.5	544.5	549.0	547.0	**543.5**
Skill Interval	**1000.6**	1003.7	1000.9	1000.7	1008.0	1008.4	1006.7
dq p_c	0.037	0.038	0.039	0.036	0.030	0.028	**0.026**
dq hor	0.036	0.034	0.036	**0.030**	0.032	**0.030**	0.049
dq t	0.032	0.038	0.040	0.039	0.024	**0.020**	0.032
sum(IQR$<$ 0)	**33**	34	52	48	57	44	75
min(IQR)	**-1.8**	-20.5	-47.8	-54.1	-97.2	-17.5	-479.1
mean(IQR$<$ 0)	**-0.8**	-7.7	-14.4	-12.7	-4.2	-4.4	-25.2
mean(IQR)	2158.1	2158.7	2175.7	2168.9	2161.9	2103.6	**2083.3**
median(IQR)	2110.3	2110.6	2100.6	2114.0	2022.9	1973.4	**1897.7**
sd(IQR)	**1648.9**	1639.0	1626.9	1617.5	1614.2	1562.5	1533.3
mad(IQR)	**2062.8**	2055.3	2037.4	2031.5	2022.9	1930.6	1940.5
Time 25%	**0.024**	0.036	0.035	0.032	0.112	0.174	0.222
Time 75%	**0.017**	0.023	0.021	0.021	0.035	0.063	0.130

Table 8: The table show different performance parameters for adaptive versions of some of the models presented in section 4. The best performer in each row is given in bold face letters.

Model	A27	A34	A35	A36	A37	A38
df	9	21	11	16	16	21
Relai 25%	27.3	25.9	23.8	**24.4**	26.0	26.4
Relai 75%	**75.4**	74.3	73.2	73.0	73.5	74.0
Relai Interval	48.2	48.7	**49.5**	48.9	47.7	47.7
Skill 25%	453.2	**452.6**	456.8	456.3	456.7	452.9
Skill 75%	**531.5**	538.5	550.9	552.1	549.6	542.9
Skill Interval	**984.7**	991.1	1007.7	1008.4	1006.2	995.8
dq p_c	0.038	**0.028**	0.030	0.033	0.030	0.029
dq hor	0.051	0.031	**0.023**	**0.023**	**0.023**	0.029
dq t	0.042	**0.026**	0.029	0.028	0.030	0.032
sum(IQR<0)	**1**	70	44	63	47	17
min(IQR)	**-21.1**	-230.5	-626.8	-372.4	-428.7	-243.9
mean(IQR<0)	**-21.1**	-42.2	-48.8	-34.1	-79.5	-88.0
mean(IQR)	2094.5	2120.1	2173.2	2136.4	2182.8	**2084.5**
median(IQR)	1909.0	1946.4	2051.2	2026.9	1907.4	**1888.5**
sd(IQR)	1554.5	1584.7	1657.0	1595.0	**1676.0**	1587.0
mad(IQR)	1971.7	1977.6	2070.3	1992.6	**2098.5**	1989.5
Time 25%	**0.040**	0.216	0.085	0.131	0.097	0.152
Time 75%	**0.031**	0.159	0.037	0.082	0.078	0.144

Table 9: The table show different performance parameters for adaptive versions of some of the models presented in section 4. The best performer in each row is given in bold face letters.

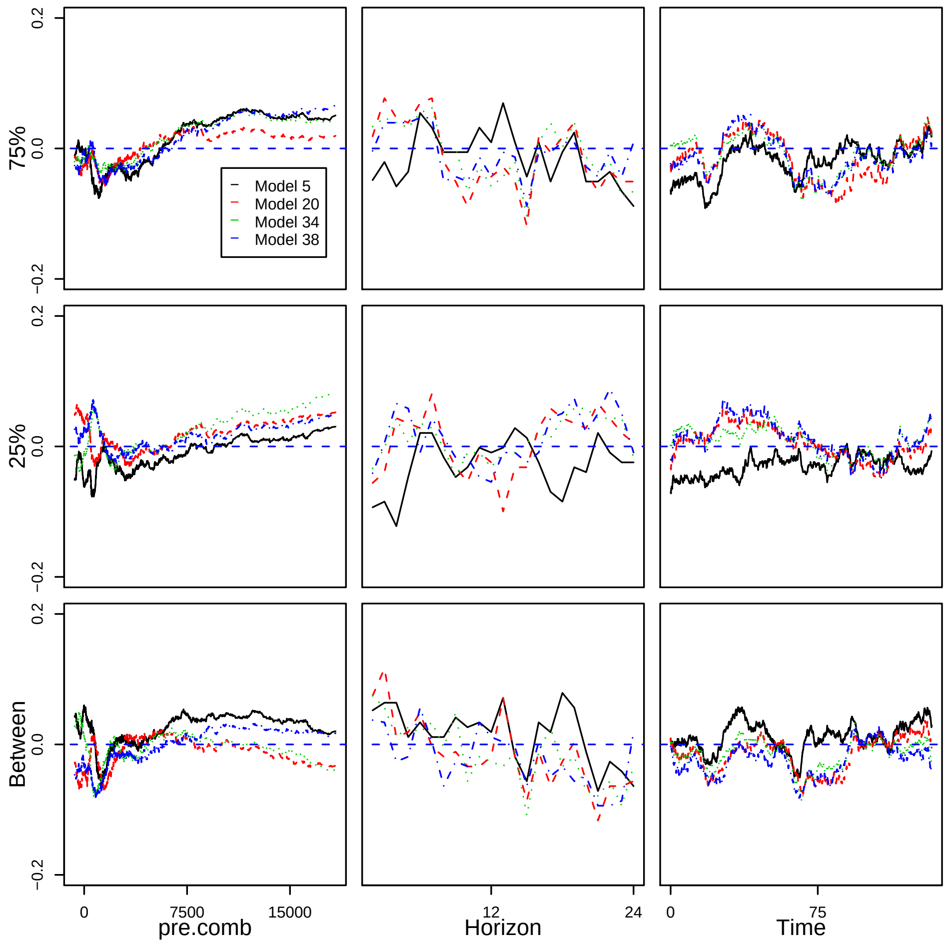

Figure 8: The figure show local reliability in the direction of forecasted power, horizon and time, for some of the adaptive models presented in Table 8 and 9.

References

[1] Carl de Boor (1978) *A practical Guide to Splines*. Applied mathematical sciences; 27. Springer-Verlag.

[2] Roger Koenker and Gilbert Bassett, Jr. (1978) *Regression Quantile*. Econometrica, Vol. 46, No 1, (33-50).

[3] Henrik Madsen (ed.) (1996) *Models and Methods for Predicting Wind Power.* IMM/ELSAM, ISBN 87-87090-29-5. 88 pp.

[4] Henrik Madsen, Henrik Aa. Nielsen and Torben S. Nielsen (2005) *A Tool for Predicting the Wind Power Production of Off-shore Wind Plants.* In Proceeding of the Copenhagen Offshore Wind Conference & Exhibition, Copenhagen October 2005. Danish Wind Industry Association. URL:`http://www.windpower.org/en/core.htm`

[5] Jan K. Møller (2006) *Modeling of Uncertainty in Wind Energy Forecast*. Master Thesis Informatics and Mathematical Modelling, Technical University of Denmark. URL:`http://www.imm.dtu.dk/pubdb/p.php?4428.`

[6] Jan K. Møller, Henrik Aa. Nielsen and Henrik Madsen (2006) *Algorithms for an Adaptive Quantile Regression Method.*. Technical Report 2006-08, Informatics and Mathematical Modelling, Technical University of Denmark URL:`http://www.imm.dtu.dk/Forskning/Soeg_publikationer.aspx` . To appear.

[7] Jan K. Møller, Henrik Aa. Nielsen and Henrik Madsen (2006) *Time Adaptive Quantile Regression*. Submitted.

[8] Henrik Aa. Nielsen, Torben S. Nielsen, Henrik Madsen and Maria J. San Isidro (2006) *Optimal Combination of Wind Power Forecasts*. Under preparation.

[9] Pinson P., Kariniotakis G., H. Aa. Nielsen, Nielsen T. S. and Madsen H. (2006) *Properties of Quantiles and Their Forecasts of Wind Generation and their Evaluation* . In Procedings of the Eouropean Wind Energy Conference & Exhibition, Athens, March 2006. EWEA. `http://www.ewea.org`, to apear.