PSO (FU 2102)
Styring og regulering af vindkraftanlæg
Forecasting wind speeds on the minute time-scale using up-stream information

Henrik Aalborg Nielsen (`han@imm.dtu.dk`)
Henrik Madsen (`hm@imm.dtu.dk`)
Informatics and Mathematical Modelling
Technical University of Denmark
Lyngby

November 7th, 2004

Contents

Summary

Ten second averages of wind speed measured at five masts placed $300m$ apart is considered in this report and forecasting methods in the minute range is investigated. The data consists of 12 periods of length one hour or more. Initially, the data are analyzed and it is clear that the time delay between masts is roughly determined by the current overall wind speed. Varying time delays are difficult to handle using traditional transfer models and instead a, to our knowledge, novel method is developed. The method applies a continuous time formulation together with spline basis expansions and it is shown that the problem can be reduced to a multiple linear regression problem. Hence, adaptive estimation methods are also available.

The potential prediction performance of a forecast system based on up-stream information is judged in terms of Mean Absolute Error (MAE) and Root Mean Squared Error (RMS). It is clear that the performance depends on the overall level of the wind speed for $5m/s$, $10m/s$, and $15m/s$ MAE-values of respectively $0.07m/s$, $0.33m/s$, and $0.58m/s$ are obtained. However, quite large deviations for the individual periods are evident. Furthermore, the results presented here shown that when up-stream information is available the auto correlation of the wind speeds at the individual masts is not very important.

1 Introduction

With the aim of facilitating the improvement of control strategies for wind farms this report considers forecasting of wind speed based on 10 second averages. The relevant horizons are in the range of minutes. For wind farms the individual turbines may obtain up-stream information regarding e.g. wind speeds measured by nacelle anemometers.

In this report 10 second averages of wind speed at five mast placed South-North are used. With the purpose of developing models using up-stream information to improve the forecasts, periods with Southerly wind are considered. The masts are located at the Risø wind turbine test site at Høvsøre, Denmark. The distance between the masts are approximately $300m$ and the anemometers and wind vanes are located approximately $80m$ a.g.l. (above ground level).

The report is structured as follows. The data are described in Section 2. With the aim of identifying characteristics of the impulse response of pairs of masts the data are analyzed in Section 3. Based on the analysis a model is formulated in Section 4 and it is shown how estimation in the model can be handled as a multiple linear regression problem. In Section 5 the potential prediction performance of a system based on up-stream information is considered. Finally, in Section 6 we conclude on the report.

2 Data

The data used in this study consists of 12 series of measurements of wind speed and direction approximately $80m$ a.g.l. from mast 1–5 at Høvsøre, see Figure 1. More specifically the anemometers are placed at the top of the masts ($78m$ for mast 1 and $80m$ for the remaining masts). Wind vanes are placed at booms $2m$ below the anemometers. Note that data from the meteorology mast is not used in this study. The series have been selected based on two main criteria:

- The wind direction should be Southerly.
- The series should have length one hour or more.

Periods with Southerly winds are used since this lets us study the relation between the wind at the five masts as gusts move from South to North. Table 1 shows a brief summary of each series. The raw data is 1 second averages, but in the study we focus on 10 second averages. Plots of 10 second averages of the data are shown in Appendix A.
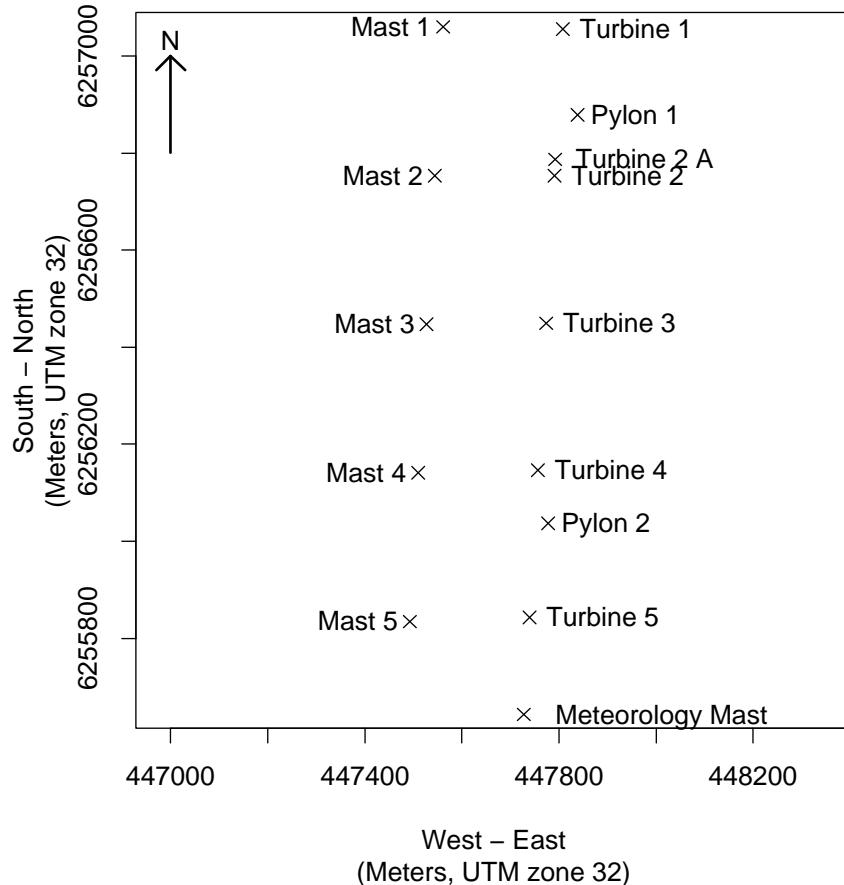
Figure 1: Locations of masts and turbines at Høvsøre.

No.	Start time	Duration (min.)	Speed (m/s)	Direction (degrees)
01	20030724 2210	60	5.5	192
02	20030801 0951	140	9.9	191
03	20030917 1641	240	10.7	193
04	20030921 0300	90	8.0	192
05	20030922 0310	90	11.9	196
06	20030922 0450	120	11.1	193
07	20030922 1510	60	13.6	193
08	20030922 1940	90	19.1	194
09	20030930 1850	80	8.2	193
10	20031001 0010	70	8.1	193
11	20031001 0150	60	8.6	192
12	20031001 2310	70	4.6	194

Table 1: Summary of the series numbered 1–12 (col. 1), start time in format yyyyddmm HHMM (col. 2), duration in minutes (col. 3), average wind speed over the five masts in m/s (col. 4), and average wind direction over the five masts in degrees (col. 5). The line of the five mast is approximately direction 183°.

3   Analysis

The primary goal of the analysis is to suggest appropriate multivariate prediction models which use both auto and cross correlation between the wind speed measurements at the five masts. To accomplish this a first step is to estimate the impulse response corresponding to all pairs of masts. To accomplish this the standard procedure of prewhitening followed by estimation of the cross correlation function is used (Chatfield, 2001); the resulting estimate is proportional to the impulse response.

However, to be able to prewhite the series an ARMA-model must be fitted to one of the series. Considering the series displayed in Appendix A this will be difficult since most of the series seems to be non-stationary on the time scale considered. More specifically the series seem to contain slow variations which are not of direct interest from an analysis point of view. For this reason a high-pass filter is designed in Section 3.1 which allows us to estimate the impulse responses of each of the periods separately in Section 3.2.

3.1   Filtering of the series

Figure 3 shows raw and smoothed periodograms (Box and Jenkins, 1976) of the 1 Hz wind speed measurements at the southern mast (No. 5), cf. Figure 1 on page 5. It is seen that in most cases the signal is dominated by frequencies below 0.1 Hz. For this reason it was decided simply to calculate the 10 second averages, i.e. not considering anti-aliasing filters. In the remaining part of this report 10 second averages are used unless otherwise stated.

The series are shown in Appendix A. It is seen that these series contain non-stationarities (trends) which must be removed before the impulse response between pairs can be calculated as outlined in the beginning of Section 3. For this purpose the series are high-pass filtered using a Butterworth filter (Hamming, 1977). The particular type of filter is chosen since its amplitude and phase characteristics are monotone. Figure 2 show characteristics for selected filters; also on the plots in the figure the relevant forecast horizons of 1 and 3 minutes are shown. While an effective elimination of the low frequencies are required it is clear that if predictions based on the filtered series are required a small phase shift is desirable. Based on this a 2nd order filter with a cutoff frequency corresponding to 12 minutes is selected. A graphical check shows that this filter is able to eliminate the apparent non-stationarities visible in the plots shown in Appendix A. The litmus test however is that an appropriate ARMA-model for prewhitening can be identified, cf. Section 3.2.
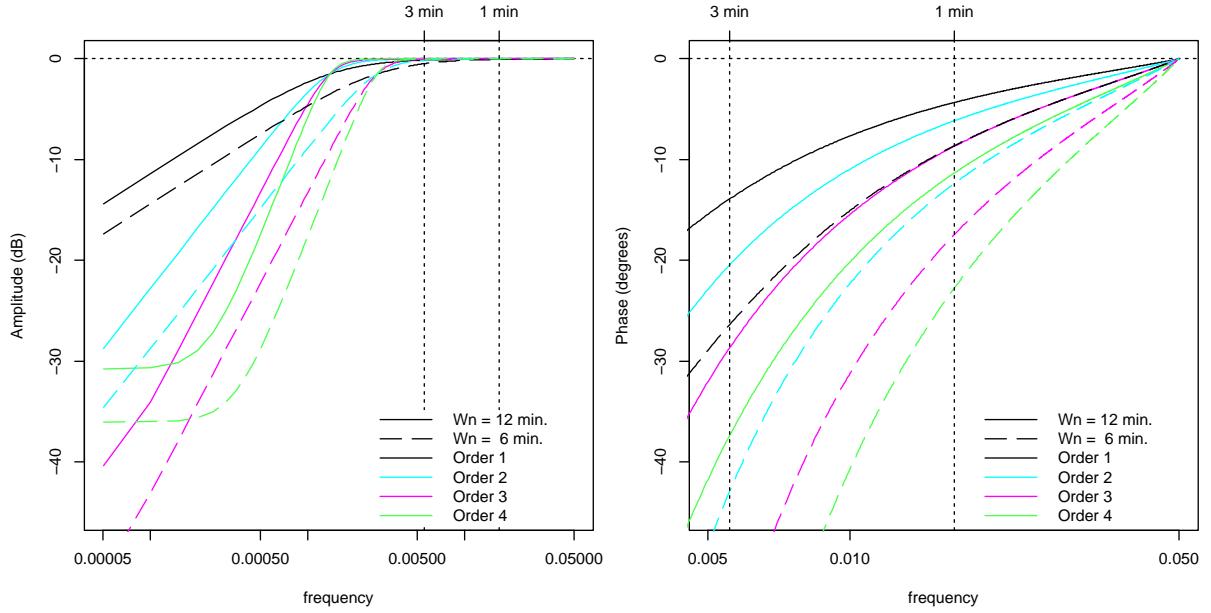
Figure 2: Filter characteristics of Butterworth filters of order 1–4 with cutoff frequencies W_n (the frequency where the magnitude response of the filter is $\sqrt{1/2}$, i.e. -1.5 dB) corresponding to 6 and 12 minutes. The vertical lines indicate 1 and 3 minutes. Note that the phase is shown for a lower range of frequencies than the amplitude.
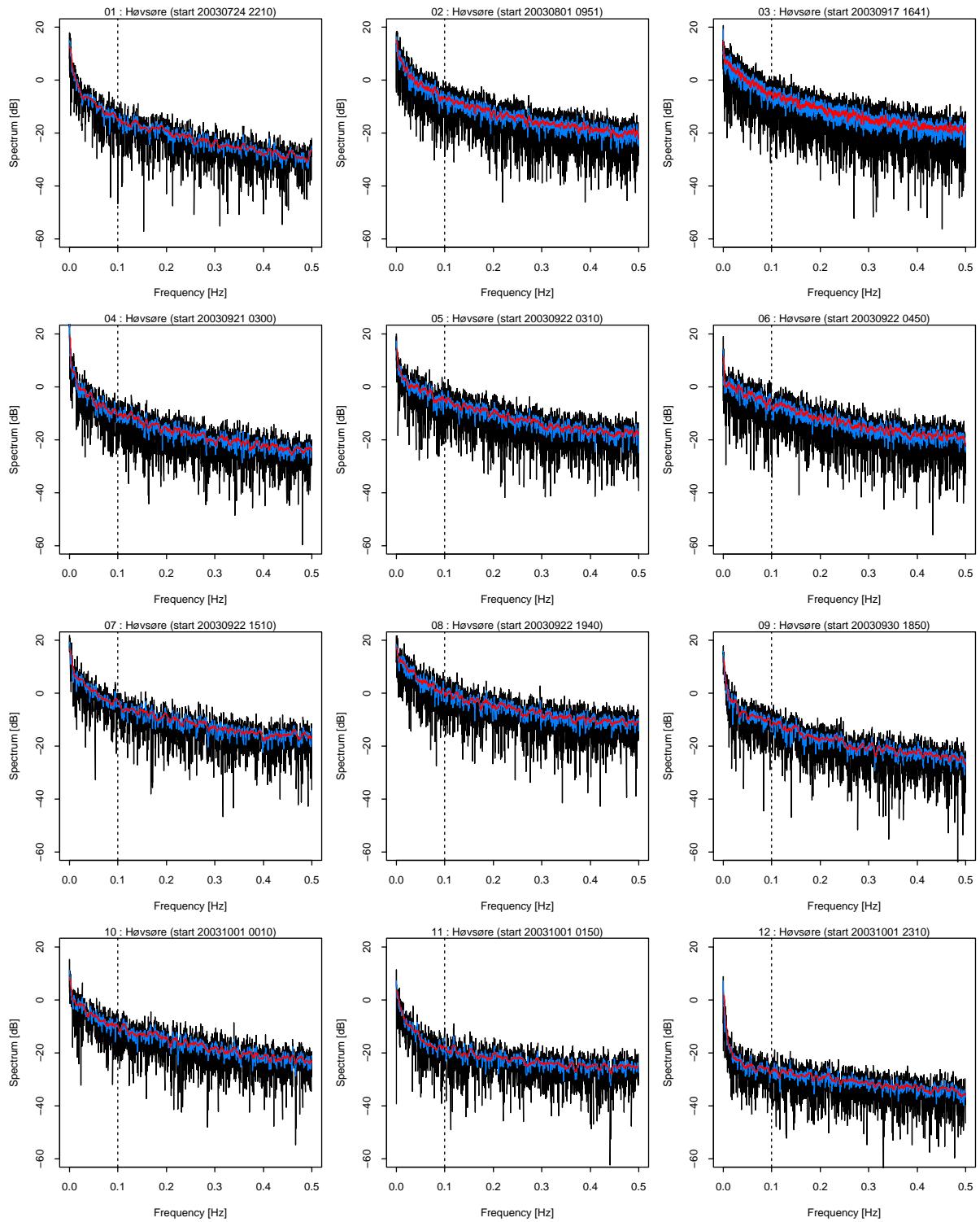
Figure 3: 1 Hz wind speed measurements at mast number 5 (south): Raw periodogram (black) and periodograms smoothed by using Daniell windows of width 7 (blue) and 31 (red). The dotted line marks a period of 10 seconds (0.1 Hz). Note that a few values exceed the dB-range on the 2nd axis.

8

3.2   Impulse response

Sample Partial Auto Correlation Function (SPACF) of the high-pass filtered series are shown in Figure 8 (page 14). From these plots it is clear that an Auto Regressive (AR) model of low order will be able to model the individual series. In order to estimate the impulse response (except for a factor of proportionality) we therefore prewhite by selecting the AR-order using Akaikes Information Criterion (AIC) (Chatfield, 2001) and estimate the parameters using Yule-Walker estimates[1] (Box and Jenkins, 1976).

The estimated impulse responses are depicted in Appendix B. Figures 4–7 (pages 10–13) show a summary of the results for (i) neighbor masts, (ii) pairs of masts with one mast in between, (iii) pairs of masts with two masts in between, and (iv) pairs of masts with three masts in between. It is seen that

- For neighbor masts (Figure 4) the peak of the impulse response is significant and near the distance traveled by the wind at the average wind speed of the series. However, for series 12 the response seems to be faster than indicated by the distance between the masts and the average wind speed.

- For pairs of masts with one mast in between (Figure 5) the relation between series is weak for most series. There is some tendency that the response is faster than indicated by the distance between the masts and the average wind speed.

- For pairs of masts with two or three mast in between (Figures 6 and 7) the relation between series is weak.

Roughly, since the distance between masts is approximately 300 m, this implies that for wind speeds around 5 m/s there is a potential benefit of using multivariate wind speed forecast models for horizons up to 2 minutes. For 10 and 15 m/s this limit drops to 1 minute and 40 seconds, respectively. However, the relation is most clear for neighbor masts and in this case the time intervals mentioned drop down to respectively 60, 30, and 20 seconds.

[1]This is the method used by the S-PLUS (http://www.splus.com) function ar.
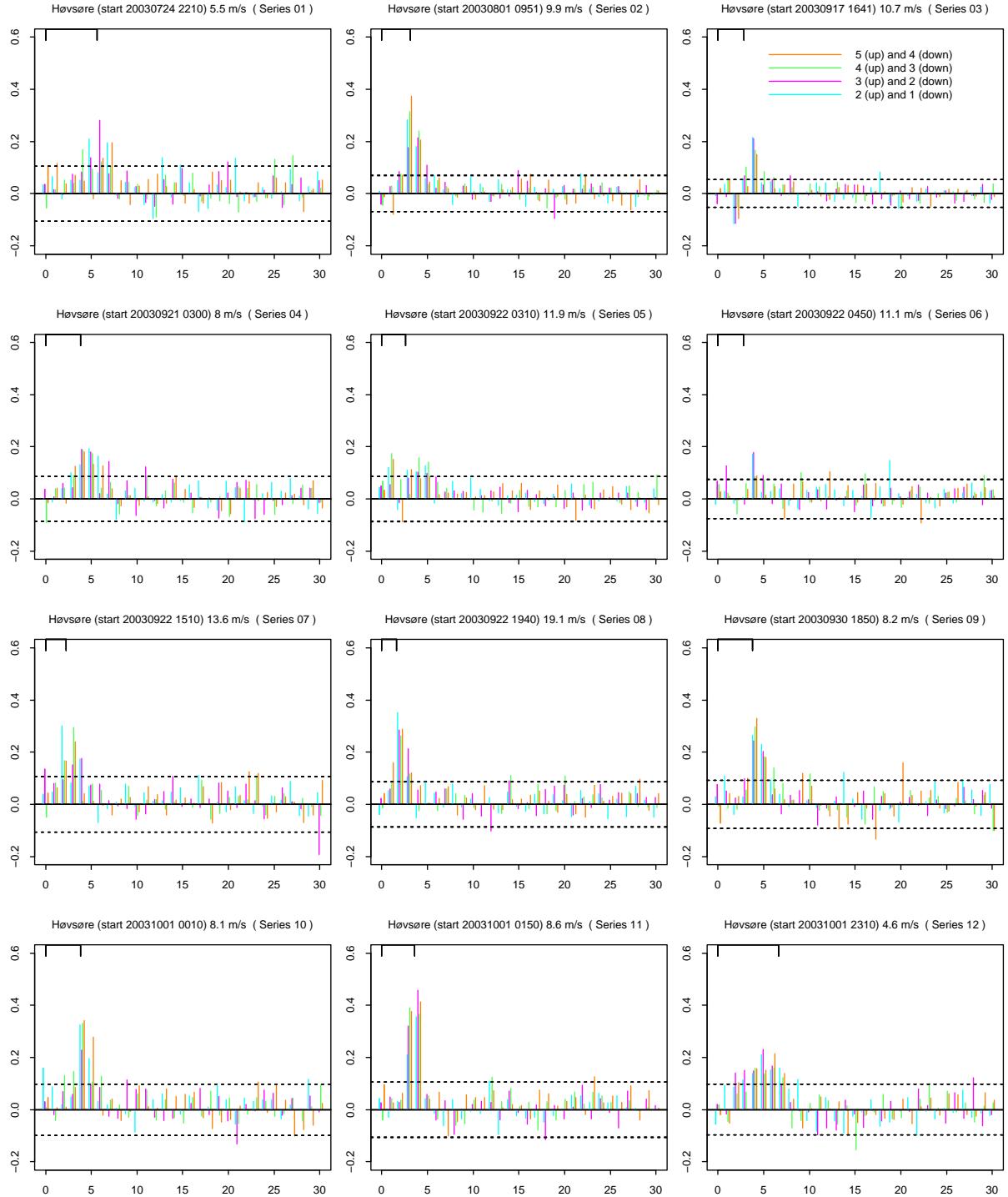
Figure 4: Estimated impulse response for neighbor masts. The ticks on top of the plots show the number of lags the wind travel at the average wind speed indicated over each plot. The two dotted lines indicate an approximate 95% confidence interval for unrelated series.
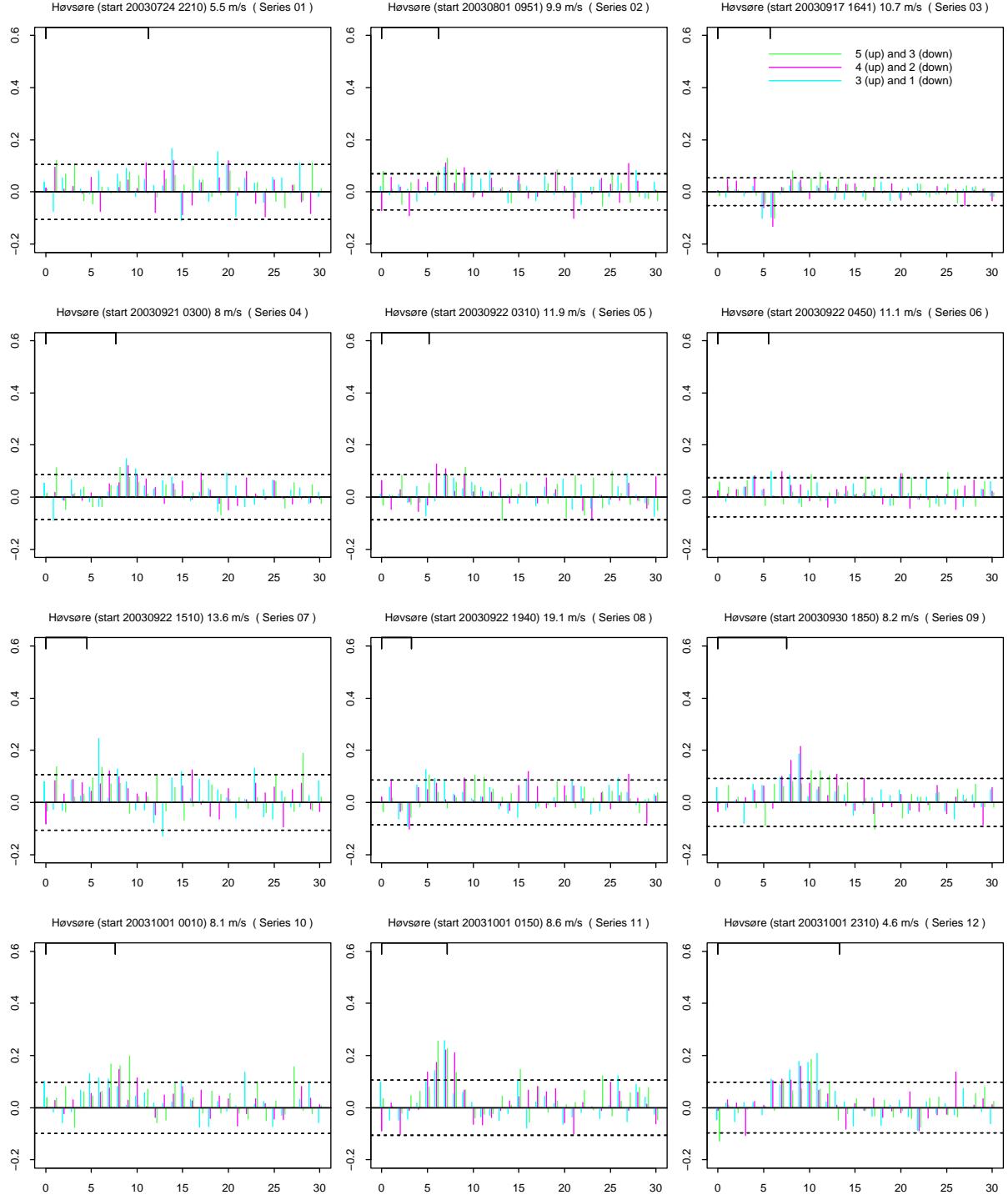
Figure 5: Estimated impulse response for pairs of masts with one mast in between. The ticks on top of the plots show the number of lags the wind travel at the average wind speed indicated over each plot. The two dotted lines indicate an approximate 95% confidence interval for unrelated series.

11
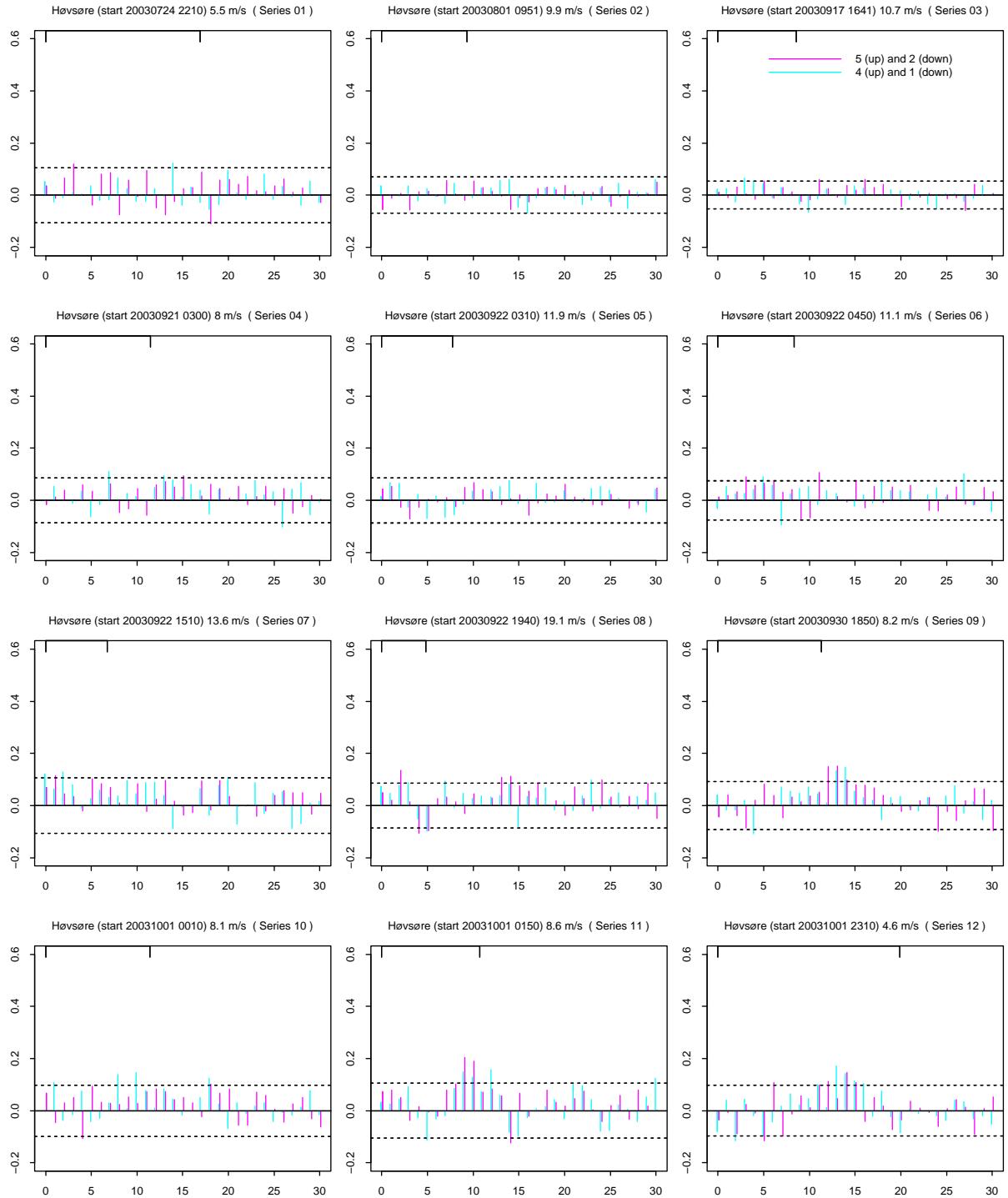
Figure 6: Estimated impulse response for pairs of masts with two mast in between. The ticks on top of the plots show the number of lags the wind travel at the average wind speed indicated over each plot. The two dotted lines indicate an approximate 95% confidence interval for unrelated series.
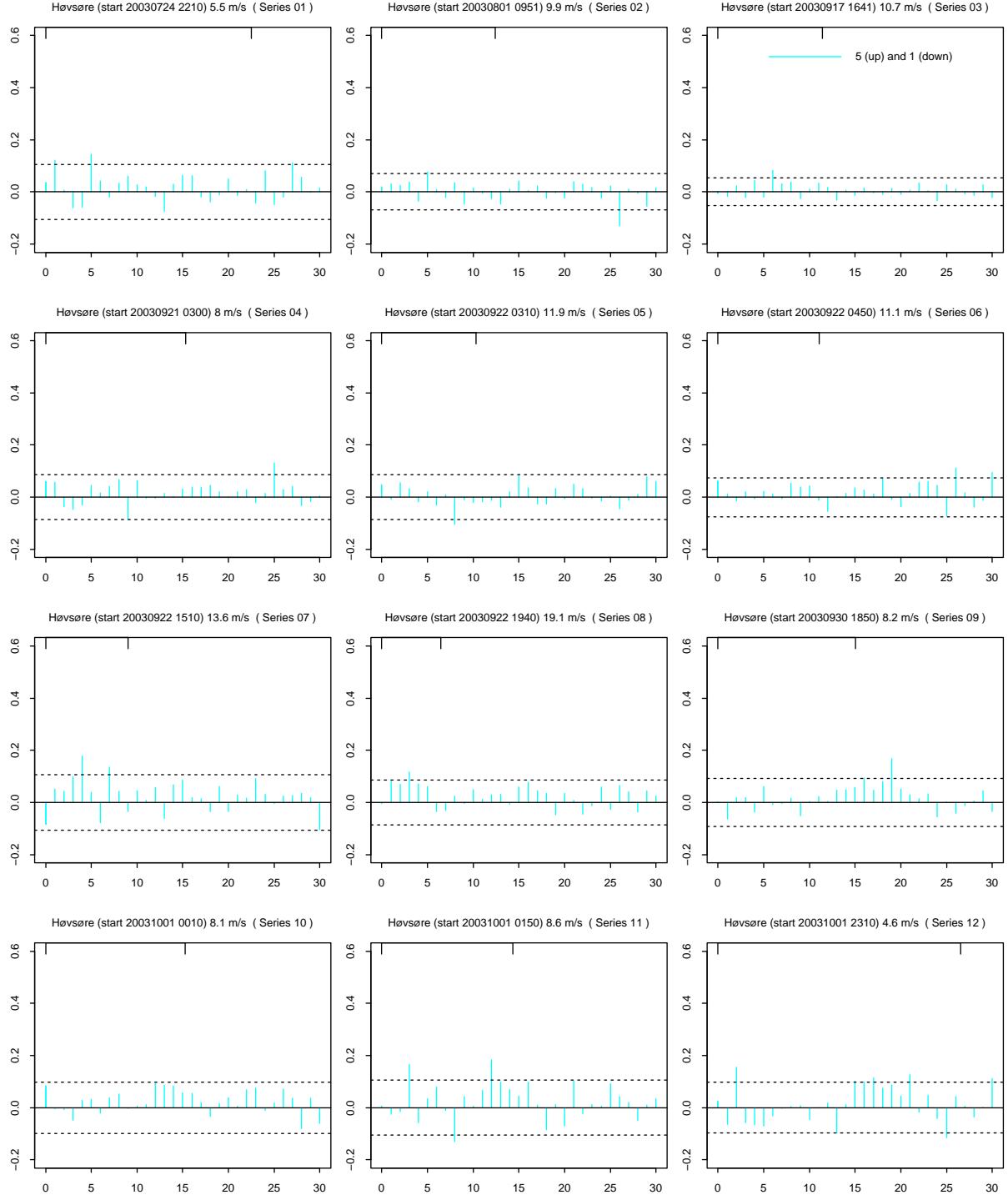
Figure 7: Estimated impulse response for pairs of masts with three mast in between. The ticks on top of the plots show the number of lags the wind travel at the average wind speed indicated over each plot. The two dotted lines indicate an approximate 95% confidence interval for unrelated series.
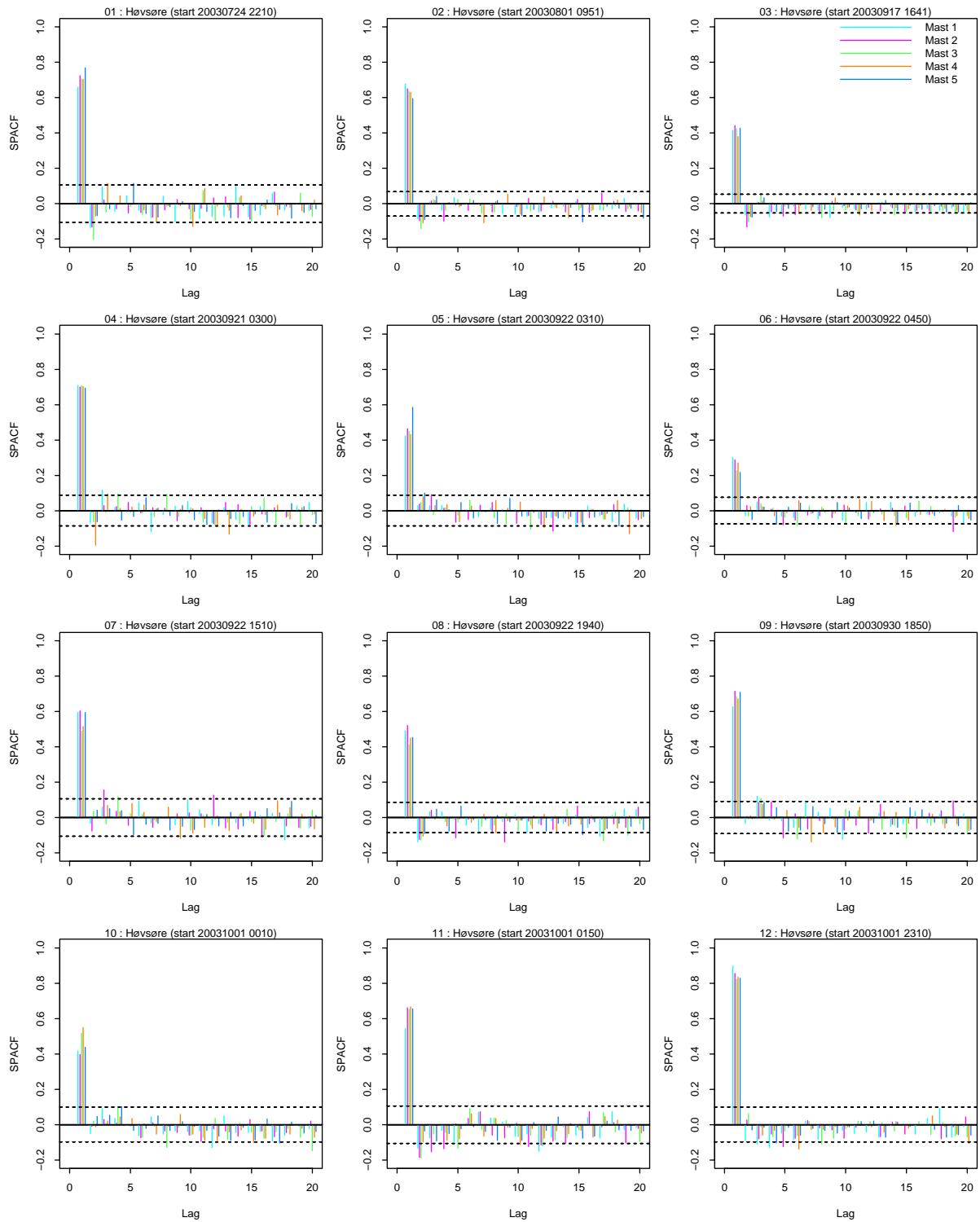
Figure 8: SPACF of the high-pass filtered 10 second averages of wind speed.

4 Model

4.1 Continuous time impulse response

4.1.1 Mathematical model

From Section 3.2 it is seen that the impulse response is roughly a smooth peak with a maximum near the distance traveled by the air at the average wind speed of the series. Impulse responses of this kind are difficult to model using traditional transfer functions[2]. Instead we formulate the problem in continuous time. Below the additive error term is excluded from the models. Let j denote a mast north of the mast denotes by i. The impact of mast i on mast j can then be characterized by the impulse response h as:

$$x_j(t) = \psi_0 + \int_0^\infty h\left(\frac{\mu(t)}{D} s\right) x_i(t-s)ds; \quad i > j, \ i = 1,\ldots,4, \tag{1}$$

where $x_j(t)$ is the wind speed at mast j at time t, ψ_0 is a constant, D is the distance between the two masts ($D = 307m$ for neighbor masts), and $\mu(t)$ is the overall level of the wind speed at time t.

It is seen that if $\mu(t)/D$ is excluded then h is a traditional impulse response in continuous time. The factor $\mu(t)/D$ models the fact that the location of the peak moves with the overall level of the wind speed as observed in Section 3.2. In order to use the model $\mu(t)$ must be known; in practice this will be an exponential smooth of the average across masts.

The data is measured with the sampling time $\Delta t = 10s$. In continuous time the data is considered constant between samples and the time stamp is related to the *end* of the sampling interval. With $t = N\Delta t$ model (1) results in

$$x_j(N\Delta t) = \psi_0 + \sum_{n=1}^\infty x_i(N\Delta t - (n-1)\Delta t) \int_{(n-1)\Delta t}^{n\Delta t} h\left(\frac{\mu(N\Delta t)}{D} s\right) ds . \tag{2}$$

With $\tau = \mu(N\Delta t)s/D$ (i.e. $ds = (D/\mu(N\Delta t))d\tau$) this can be written

$$x_j(N\Delta t) = \psi_0 + \frac{D}{\mu(N\Delta t)} \sum_{n=1}^\infty x_i(N\Delta t - (n-1)\Delta t) \int_{\tau_{N,n-1}}^{\tau_{N,n}} h(\tau)d\tau , \tag{3}$$

where $\tau_{N,n} = \mu(N\Delta t)n\Delta t/D$. Let $H(\tau)$ denote the indefinite integral of $h(\tau)$. Using this

---

[2]In principle state dependent transfer functions could be used (Madsen and Holst, 2000; Priestley, 1988) but we judge that the number of terms in the numerator will have to be large introducing too many degrees of freedom for the relative high level of noise in the system studied here. Also the alternative procedure used here leads to a very simple estimation procedure.

notation (3) can be written

$$x_j(N\Delta t) = \psi_0 + \frac{D}{\mu(N\Delta t)} \sum_{n=1}^{\infty} x_i(N\Delta t - (n-1)\Delta t) \left(H(\tau_{N,n}) - H(\tau_{N,n-1})\right) . \qquad (4)$$

In order to estimate the impulse response $h(\tau)$ the indefinite integral $H(\tau)$ of this function is approximated by a B-spline basis:

$$H(\tau) = \sum_{k=1}^{K} \psi_k B_k(\tau) , \qquad (5)$$

where $B_k(\tau)$; $k = 1, \ldots, K$ are the basis functions and ψ_k; $k = 1, \ldots, K$ is a set of coefficients to be determined from data. Some restrictions must be imposed on the coefficients in order to ensure that the integral in (1) is finite. Furthermore, restrictions can be used in order to build in prior information in the model. Here the restrictions

$$
\begin{aligned}
h(0) &= 0 \\
h'(0) &= 0 \\
h(\ell) &= 0 \\
h'(\ell) &= 0 \\
h(\tau) &= 0, \quad \tau > \ell
\end{aligned}
\qquad (6)
$$

will be used. The last restriction implies that the summation in (4) instead of running from 1 to ∞ runs to

$$n_{max}(N\Delta t) = \left[\frac{\ell D}{\Delta t \mu(N\Delta t)}\right] + 1 \qquad (7)$$

only. The square brackets indicate the integer part. Typically (7) will result in $\tau_{N,n_{max}(N\Delta t)} > \ell$, therefore $\tau_{N,n}$ is redefined as

$$\tau_{N,n} = min(\ell, \mu(N\Delta t)n\Delta t/D) . \qquad (8)$$

The remaining restrictions can be imposed on the basis in (5) by eliminating basis functions as described in Appendix D.

If (5) and (7) is used in (4), then

$$x_j(N\Delta t) = \psi_0 + \frac{D}{\mu(N\Delta t)} \sum_{k=1}^{K} \psi_k \sum_{n=1}^{n_{max}(N\Delta t)} x_i(N\Delta t - (n-1)\Delta t) \left(B_k(\tau_{N,n}) - B_k(\tau_{N,n-1})\right)$$
$$(9)$$

is obtained. It is seen that given the wind speed measurements at mast i the summation over n does not depend on unknown quantities. If we write

$$x_j(N\Delta t) = \psi_0 + \sum_{k=1}^{K} \psi_k z_{ik}(N\Delta t) , \qquad (10)$$

16

where for $k = 1, \ldots, K$

$$z_{ik}(N\Delta t) = \frac{D}{\mu(N\Delta t)} \sum_{n=1}^{n_{max}(N\Delta t)} x_i(N\Delta t - (n-1)\Delta t) \left(B_k(\tau_{N,n}) - B_k(\tau_{N,n-1}) \right) , \quad (11)$$

it is seen that $z_{ik}(N\Delta t)$; $k = 1, \ldots, K$ can be calculated for the full data set. Hereafter estimates of ψ_k; $k = 0, \ldots, K$ can be found by ordinary least squares. The impulse response $h(\tau)$ can then be found using

$$h(\tau) = \sum_{k=1}^{K} \psi_k B_k'(\tau) , \quad (12)$$

where B' is the derivative of the basis functions which can be calculated using standard software such as the function `bs` in S-PLUS.

Remark 1: The values of z_{ik} can be calculated only for $t > \ell D / \mu$.

Remark 2: The level of $H(\tau)$ can not be estimated and therefore the basis should be generated without an intercept.

Remark 3: If $H(\tau)$ is approximated by a B-spline of order 5 without intercept this can formally be considered as consisting of the basis functions τ, τ^2, τ^3, τ^4, $(\tau - \kappa_1)_+^4$, $(\tau - \kappa_2)_+^4$, ..., where κ_1, κ_2, ... are the internal knots of the basis and the subscript "+" denotes truncation of negative values. The derivative of this basis is 1, τ, τ^2, τ^3, $(\tau - \kappa_1)_+^3$, $(\tau - \kappa_2)_+^3$, ..., which corresponds to a B-spline basis of order 4. In conclusion if $h(\tau)$ is to be approximated by a cubic (i.e. 4th order) B-spline basis then $H(\tau)$ should be approximated by a 5th order B-spline basis.

4.1.2   Estimation

Estimation of the continuous time impulse responses are based on the 10s wind speed averages without high-pass filtering. Hence, the level of the model needs to be modeled as well. As a consequence model (1) is extended to include a $\mu(t)$ directly, i.e.

$$x_j(t) = \psi_{00} + \psi_{01}\mu(t) + \int_0^\infty h\left(\frac{\mu(t)}{D} s \right) x_i(t - s)ds; \quad i > j, \ i = 1, \ldots, 4, \quad (13)$$

and estimation is performed in the model

$$x_j(N\Delta t) = \psi_{00} + \psi_{01}\mu(N\Delta t) + \sum_{k=1}^{K} \psi_k z_{ik}(N\Delta t) + e_j(N\Delta t), \quad (14)$$

which is (10) extended with the term $\psi_{01}\mu(t)$ and the error term $e_j(N\Delta t)$ is included explicitly.

Since $h(\tau)$ is expected to have maximum near $\tau = 1$ the spline bases $B_k(\tau); k = 1, \ldots, K$ are constructed using cubic splines with internal knots 0.6, 1, 1.4, 2 and boundary knots 0 and 4, i.e. $\ell = 4$. Imposing the equality restrictions in (6) results in three basis functions, i.e. $K = 3$. The overall level of the wind speed $\mu(t)$ is estimated using an exponential smooth with a forgetting factor corresponding to 10 minutes (i.e. 0.9833 for for a sampling time of 10 seconds). Initialization is done using the average over the first two minutes.

The resulting estimates for neighbor masts are displayed in Figure 9. Most of the estimates behave as expected. Note that the precise placement of the peak near $\tau = 1$ will be somewhat influenced by the knot placement. This is a well known problem when using spline bases with few knots, see e.g. (Hastie and Tibshirani, 1990, p. 252-3), and as a consequence conclusions regarding the precise placement of the peaks should be avoided. A comparison with Figure 4 on page 10 shows similar behaviour as in Figure 9; also for the cases such as series 3 and 5. For series 8 the average wind speed is 19.1 m/s, corresponding to a travel time of 16 seconds between two masts; this indicates that the sampling time of 10 seconds is too high for this series and in turn this is may explain the non-minimum phase behaviour of two of the impulse responses.

Appendix C starting on page 50 shows the data and fitted values together with the estimate of $\mu(t)$ used. In general it is seen that the impulse response model indeed captures the dynamic behavior of the series. Series 3 seems to be an exception where the fitted values follow the level of the wind speed closely. The R-squared values of the fits are shown in Table 2.

Series	Mast 1	Mast 2	Mast 3	Mast 4
1	0.59	0.55	0.48	0.62
2	0.50	0.47	0.52	0.49
3	0.61	0.63	0.68	0.67
4	0.81	0.80	0.78	0.75
5	0.33	0.32	0.41	0.39
6	0.77	0.80	0.78	0.79
7	0.67	0.63	0.66	0.65
8	0.40	0.40	0.42	0.47
9	0.84	0.84	0.83	0.84
10	0.55	0.45	0.59	0.58
11	0.51	0.75	0.72	0.77
12	0.91	0.97	0.96	0.97

Table 2: R-squared values of the fits obtained using model (14) with spline bases constructed as described in Section 4.1.

Models for masts number 1, 2, and 3 using impulse responses based on two masts south of the mast under consideration are also investigated. For almost all series the estimated impulse response of the second mast is negative (not shown). The corresponding R-squared values are shown in Table 3. Note that since the extra impulse response requires a longer initialization period, the R-squared values are not based on the same observations
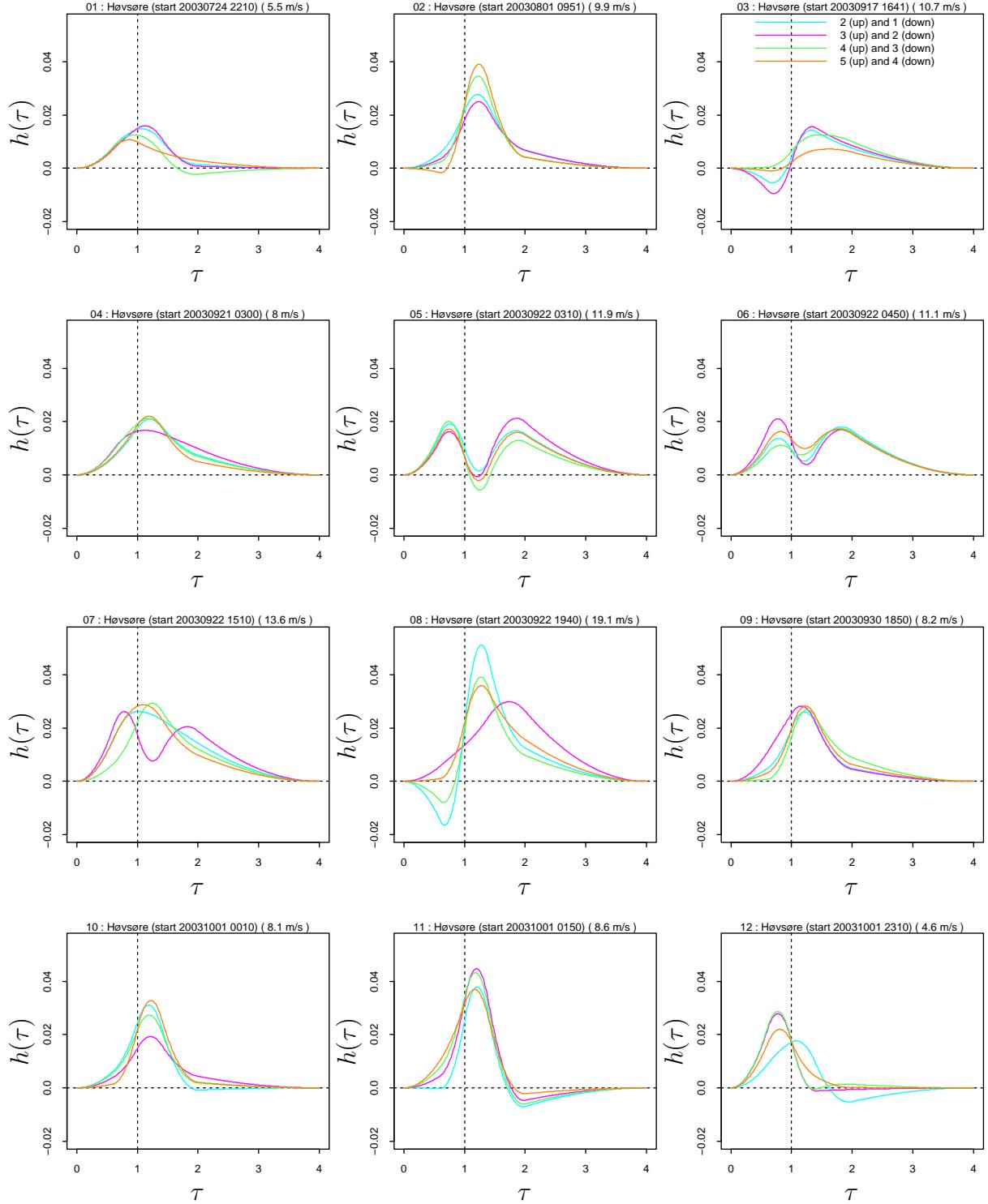
Figure 9: Estimates of continuous time impulse responses for neighbor masts. The average wind speed of each series is displayed above the individual plots.

as in Table 2. Mainly because of the negative impulse responses, and the fact that the R-squared values are comparable, it is decided not to investigate models using two souther masts any further.

Series	Mast 1	Mast 2	Mast 3
1	0.61	0.54	0.48
2	0.53	0.49	0.57
3	0.62	0.63	0.68
4	0.80	0.78	0.75
5	0.36	0.33	0.39
6	0.79	0.81	0.79
7	0.67	0.63	0.66
8	0.43	0.43	0.46
9	0.84	0.84	0.83
10	0.58	0.48	0.59
11	0.59	0.79	0.72
12	0.92	0.97	0.97

Table 3: R-squared values of the fits obtained using model (14) extended to include two masts south of the mast under consideration with spline bases constructed as described in Section 4.1.

4.2   Modelling the noise term

In this section we extend the models obtained in Section 4.1 to include a model of the autocorrelation contained in the error term, which is called $e_j(N\Delta t)$ in (14) above. Investigation of the Sample Auto Correlation Function (SACF) and the Sample Partial Auto Correlation Function (SPACF) of the residuals from the fits described in the end of Section 4.1 indicate that a low order Auto Regressive model (AR-model) is appropriate.

Table 4 shows the order of the autoregressive model selected using the AIC criterion as described in the beginning of Section 3.2. It is seen that in 32 of the 48 cases the selected order is 3 or below. Series 2 and 9 occasional show some quite high orders.

Since moderate overfitting normally does not influence the prediction performance seriously it is decided to fit a AR(4) model to the residual series. In Figure 10 P-values of the Portmanteau goodness of fit test (Box and Jenkins, 1976) are shown. Clearly, the AR(4) model is not able to describe all series. As expected from Table 4 series 9 is problematic. Furthermore series 2, (mast 4), series 11 (mast 3), and series 12 (mast 4) are problematic and minor problems are observed for series 5 (mast 2), series 6 (mast 4), and series 8 (mast 2). The SACF of the residuals from the AR(4) model of these series is investigated and shows only values marginally outside the approximate 95% confidence interval. Consequently, it is concluded that an AR(4) model is able to describe the main features of all the series.
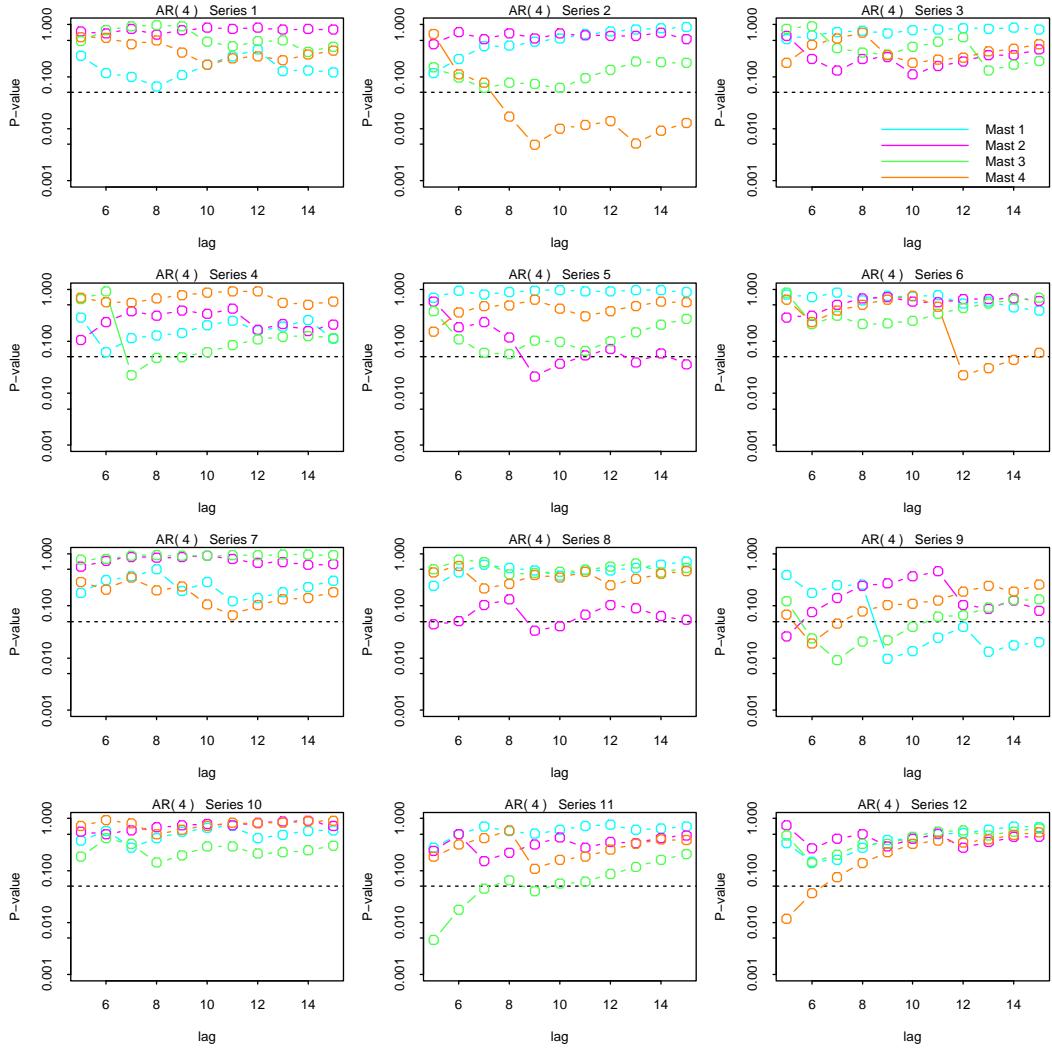
Figure 10: *P*-values versus lag for the Portmanteau goodness of fit test also called the Box-Ljung test, when fitting an AR(4) model to the residuals from the fits described in the end of Section 4.1. The horizontal line indicates 0.05, i.e. the 5% level of significance. Note the log-scale used on the 2nd axis.

Series	Mast 1	Mast 2	Mast 3	Mast 4
1	3	2	2	3
2	1	4	6	12
3	3	3	3	3
4	7	1	8	3
5	1	3	6	2
6	3	3	3	3
7	1	3	1	1
8	2	5	2	1
9	10	6	6	5
10	2	1	5	1
11	4	1	6	1
12	1	4	3	5

Table 4: Order of AR-model as suggested by the AIC criterion.

5　Prediction Performance

In this section the possible performance of a prediction system based on the previous sections are addressed. To simplify the investigations the input, i.e. z_{ik} in (14), is taken to be known. After fitting model (14) an AR(4) model is fitted to each residual series. However, the AR-model are fitted for each specific prediction horizon separately. This ensures that the estimates are optimized for each specific prediction horizon.

Figures 12 and 13 (pages 25 and 26) show the RMS and MAE of the errors obtained after the two-stage fitting outlined above. It is seen that the performance measures change with horizon up to approximately 20 seconds, i.e. for longer horizons the AR-model is unimportant (assuming the input to be known). Figure 11 shows the two performance measures plotted against the average wind speed of the series for all horizons above 60 seconds. There seems to be some dependence on the wind speed, but relatively large deviations from the line occur. If the lines are used as a rough guideline a 95% confidence interval will have length $\pm 2 \times 0.2 = \pm 0.4 m/s$ at $6m/s$, at $10m/s$ the corresponding length will be $\pm 0.8 m/s$. All these estimates are based on the assumption that the input, i.e. wind speed at the up-stream mast, is known[3]. As is seen from the impulse responses (Figure 9 on page 19) this assumption is roughly valid up to horizons defined by the average wind speed and the distance between masts ($307m$). At $5m/s$ the corresponding time interval is just over 60 seconds and at $10m/s$ just over 30 seconds. For horizons longer than these intervals the performance will gradually decrease as compared to the above as more and more wind speed values at the up-stream mast needs to be predicted, which in turn may make autoregressive error modeling more important.

---

[3]Furthermore the performance measures are based on in-sample results, but since the models contain relatively few coefficients this is a minor problem which presumably can be solved using adaptive methods for estimation.
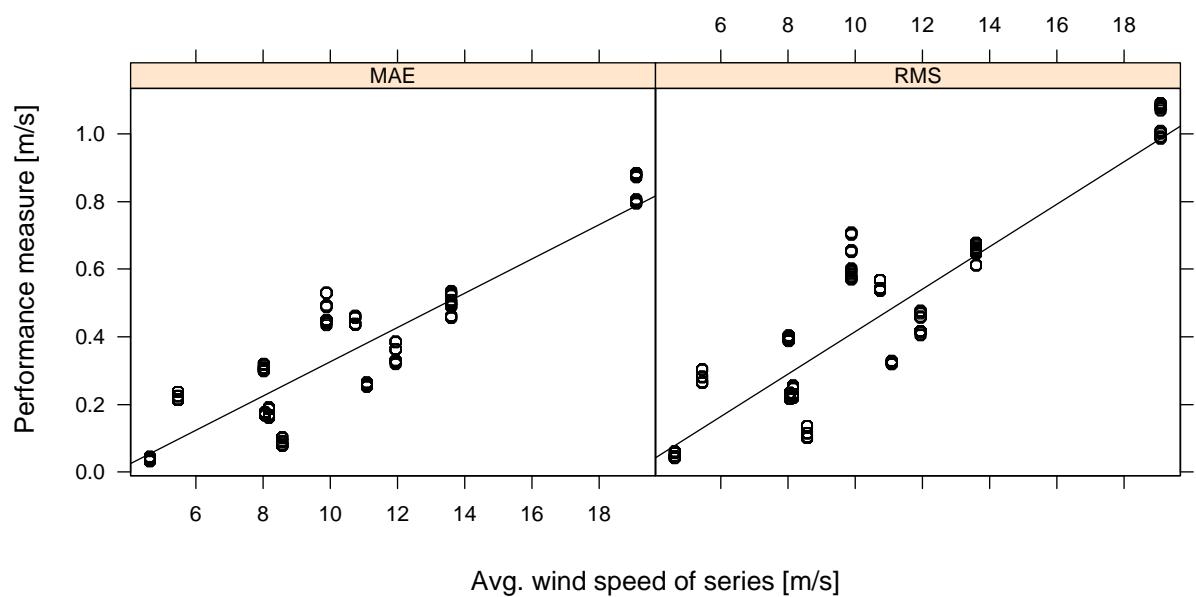
Figure 11: Performance measure for horizons 60 to 180 seconds versus average wind speed of the series. The line indicates a linear fit of a strait line.
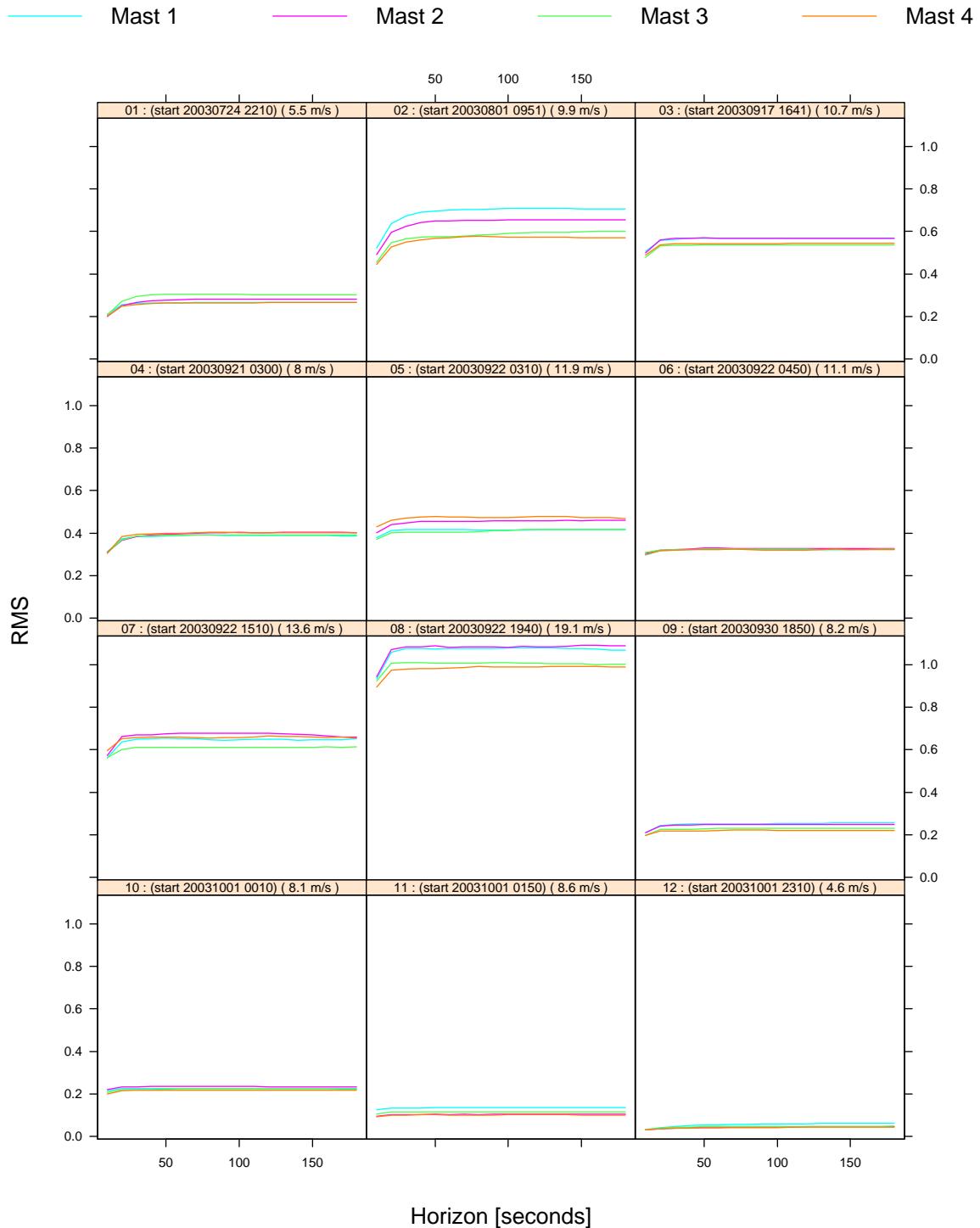
Figure 12: RMS versus prediction horizon.
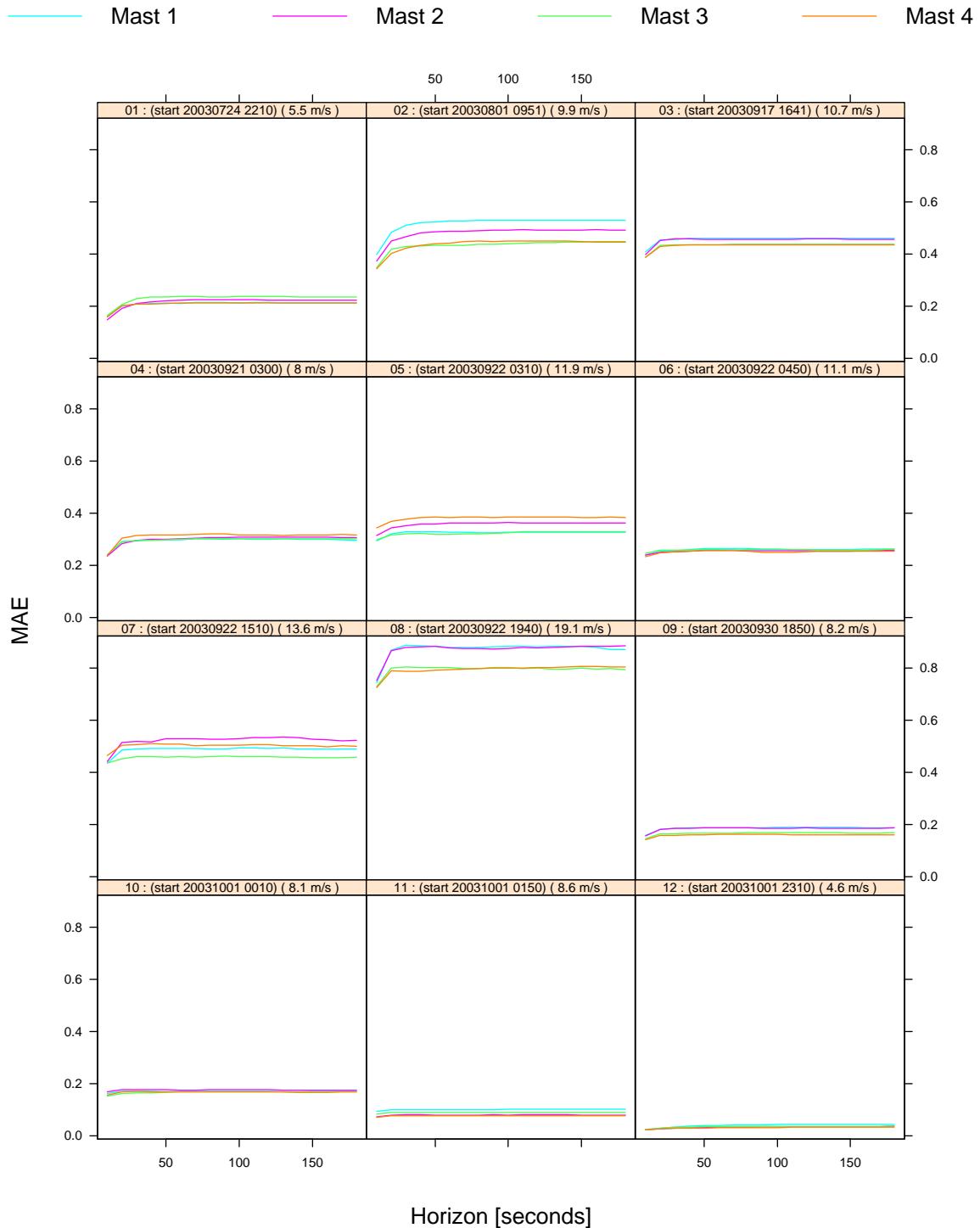
Figure 13: MAE versus prediction horizon.

26

6   Conclusion and discussion

In this report the cross correlation between wind speeds measured at five masts, located approximately North-South with a distance of approximately 300 meters, is studied. Twelve cases of length one hour or more and with Southerly winds are used after sub-sampling to 10 second averages.

After studying the data using standard methods like prefiltering, prewhitening, and estimation of the cross correlation function it is clear that the time delay between masts depend on the level of the wind speed. The correlation between masts drop with the distance between masts. Generally, a marked decrease in correlation is observed already for masts 600 meters apart. In turn this has implications for the forecast horizons at which up-stream information has the potential of improving the forecasts. When using neighbor masts these horizons are of the order one minute at $5m/s$ and half a minute at $10m/s$.

Assuming the up-stream wind speed to be known the Mean Absolute Error (MAE) range from just above $0m/s$ to $0.8m/s$, where MAE $= 0.8m/s$ occur for a series with an average wind speed of $19m/s$. For the remaining series the maximum MAE is just below $0.6m/s$. There is some tendency that the performance depend on the average wind speed (Figure 11 on page 24, which for MAE has a slope of 0.051). The strait line fitted to the data on the plot just mentioned results in MAE-values of $0.07m/s$, $0.33m/s$, and $0.58m/s$ for wind speeds of $5m/s$, $10m/s$, and $15m/s$, respectively. For horizons up to 20 seconds the performance can be slightly improved by modeling the auto correlation of the forecast error at the mast under consideration. In Appendix C a graphical representations of the performance can be seen.

In order to model the varying time delay and the smooth impulse response a, to our knowledge, new method has been developed. The problem is formulated in continuous time and the time scale used in the impulse response is controlled by the overall level of the wind speed. Hereafter the continuous time impulse response is approximated using a spline basis expansion, with appropriate restrictions imposed. In the report it is shown that the resulting problem can be transformed in to a multiple linear regression in which the independent variables are determined by the basis expansion and the overall level of the wind speed. This has the advantage that standard software can be used for estimating in the model and that it is simple to derive adaptive estimation procedures for implementation in practice.

Implementation issues includes the calculation of z_{ik} in (11) on page 17. Especially, because the most recent wind speed has little influence $(B_k(\tau_{N,0}) - B_k(\tau_{N,1}) \approx 0)$ a recursive update of $Z_{ik}((N-1)\Delta t)$ to $Z_{ik}(N\Delta t)$ seems difficult to obtain. Instead it is suggested to keep a number of recent wind speeds in a rolling buffer. The length of this buffer is determined by the minimum wind speed at which estimates in (14) on page 17 is updated, using (7) on page 16 with the values used in Section 4.1.2 and a lower level

of $3m/s$ results in a buffer length of $[4 \times 307m/(10s \times 3m/s)] + 1 = 41$.

The development of a fully functional forecast system for the nacelle anemometers in a wind farm requires more investigations than presented in this report. First of all the wind measurements at the nacelle anemometers must be transformed into the corresponding undisturbed wind. After this a model valid for all wind directions must be developed; one solution may be to use the closest wind turbine up-stream of the one considered and as a consequence the distance D used in e.g. (1) on page 15 change with the wind direction. However, the results presented in this report indicates that relation between wind speeds decrease quite drastically already when going from $300m$ to $600m$. For this reason an investigation, similar to Section 3, for the case where the wind direction is approximately orthogonal to the axis of the five masts at Høvsøre is relevant. Since the masts are placed West of the wind turbines at Høvsøre (Figure 1 on page 5) periods with westerly winds are appropriate. Such an investigation will aid in clarifying how up-stream information is most appropriately defined for wind turbines placed in a grid. Finally, for the long horizons, some of the up-stream information will not be available as measurements and instead forecasts has to be used. To be able to initiate this process pure autoregressive models for the the wind turbines placed at the border of the farm must run in parallel with models of the type described in this report.

References

Box G. and Jenkins G. *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco, 1976.

Chatfield C. *Time-series forecasting.* Chapman & Hall / CRC, 2001.

Hamming R.W. *Digital Filters.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1977.

Hastie T.J. and Tibshirani R.J. *Generalized Additive Models.* Chapman & Hall, London/New York, 1990.

Madsen H. and Holst J. *Modelling Non-Linear and Non-stationary Time Series.* Informatics and Mathematical Modelling, The Technical University of Denmark, Kongens Lyngby, Denmark, 2000.

Priestley M.B. *Non-linear and Non-stationary Time Series Analysis.* Academic Press, New York/London, 1988.

A   Plots of 10 second averages of speed and direction
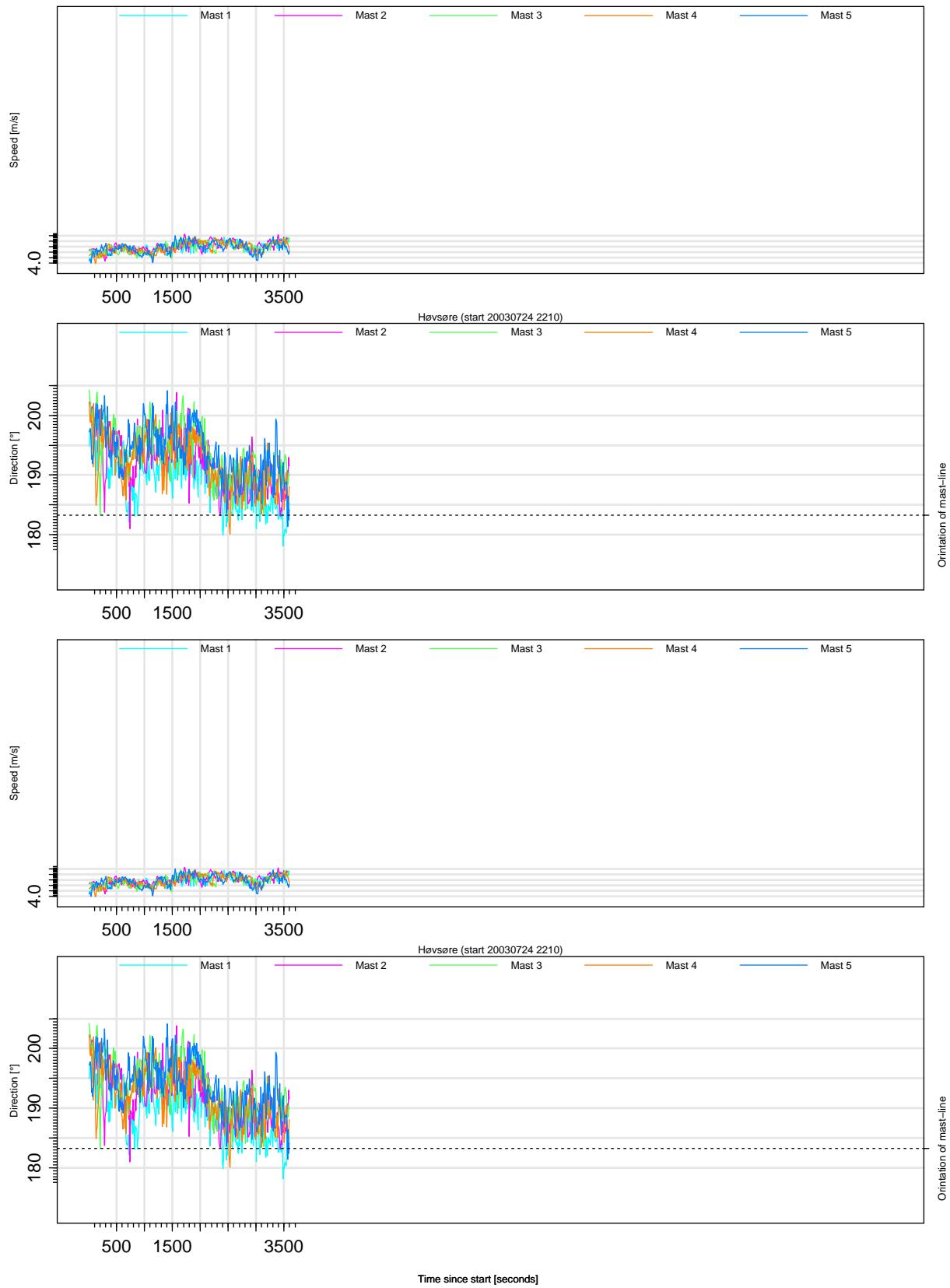
Figure 14: 10s averages of wind speed and direction, series: 1 & 2.
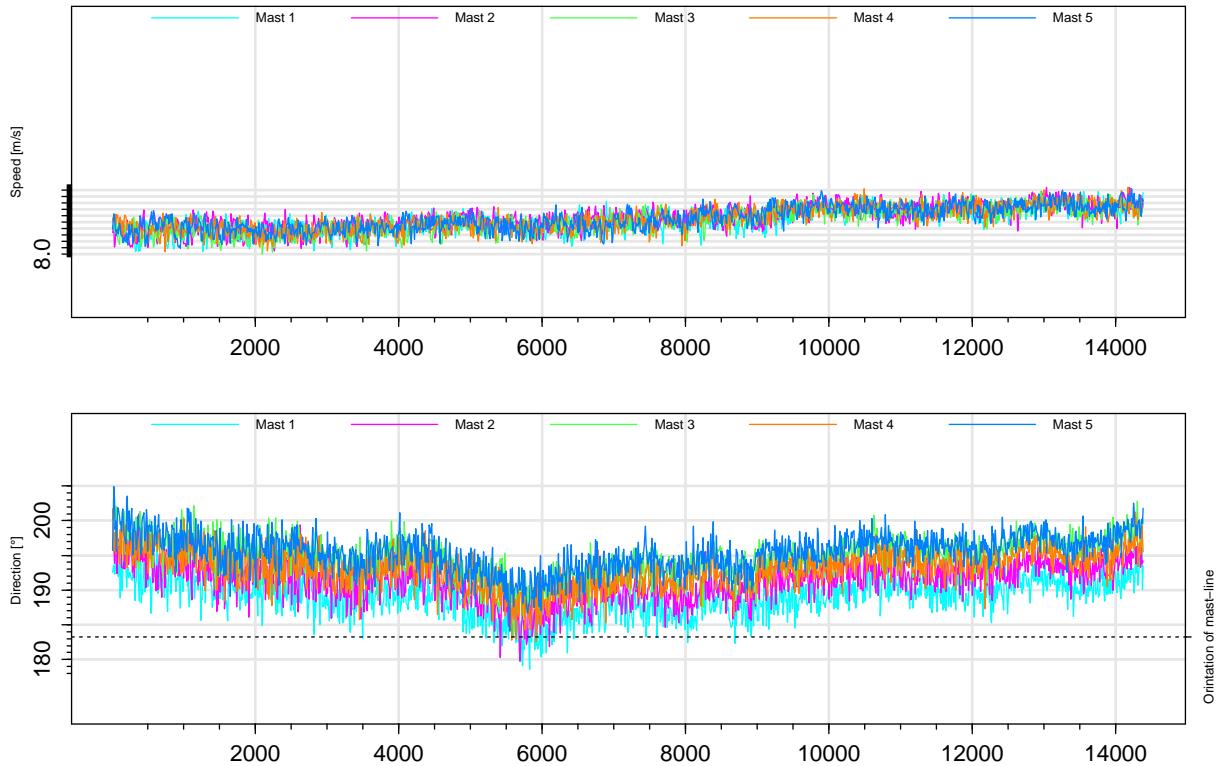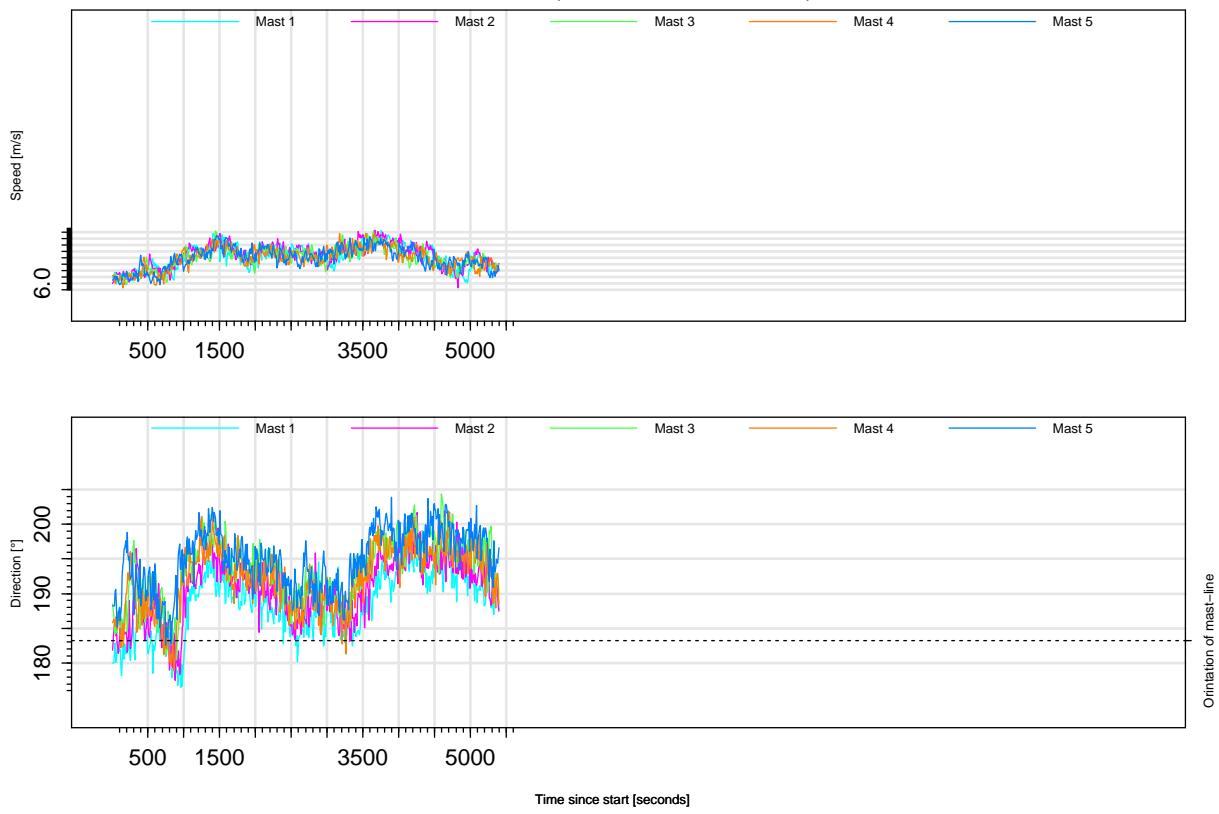
Figure 15: 10s averages of wind speed and direction, series: 3 & 4.
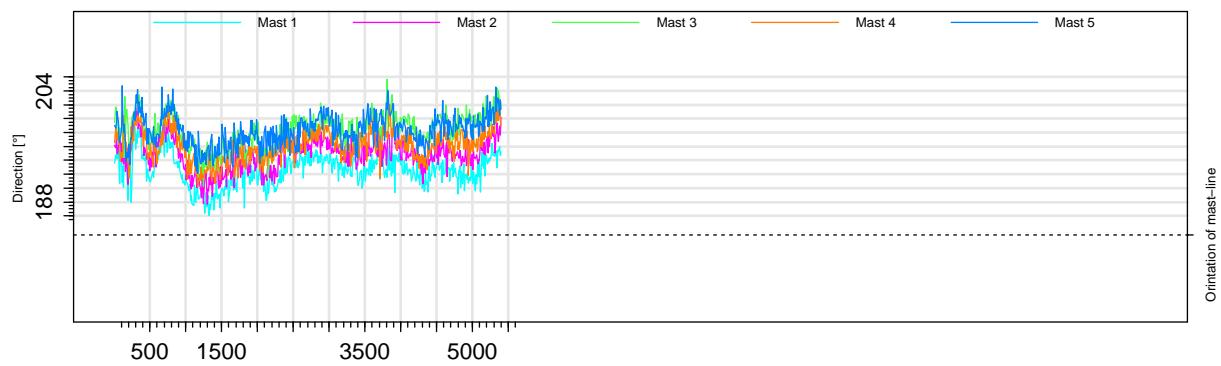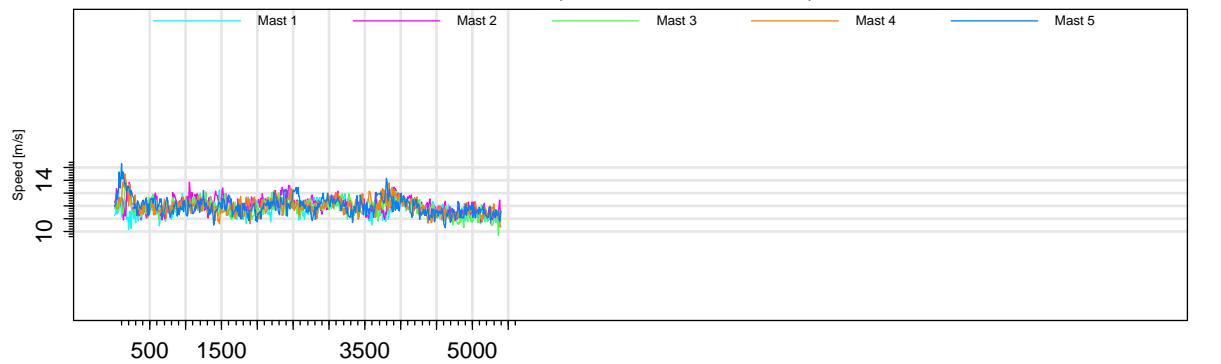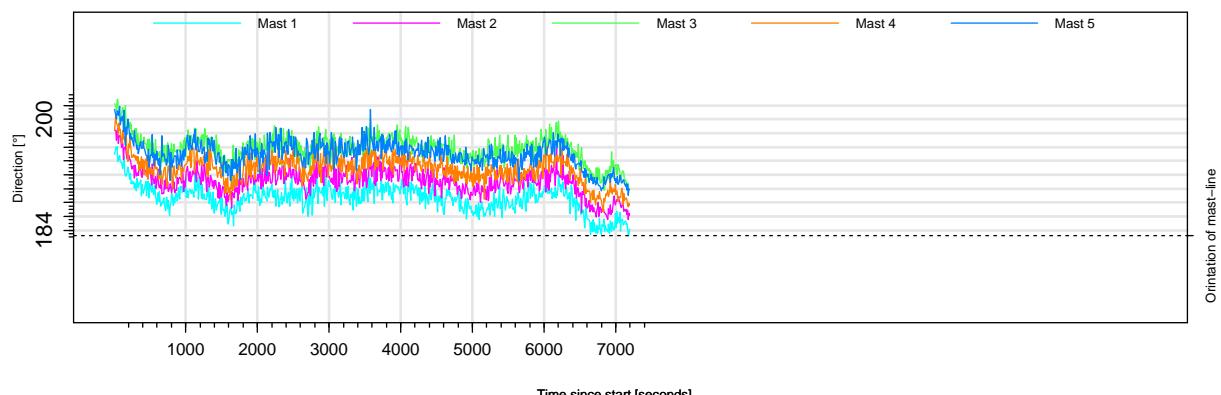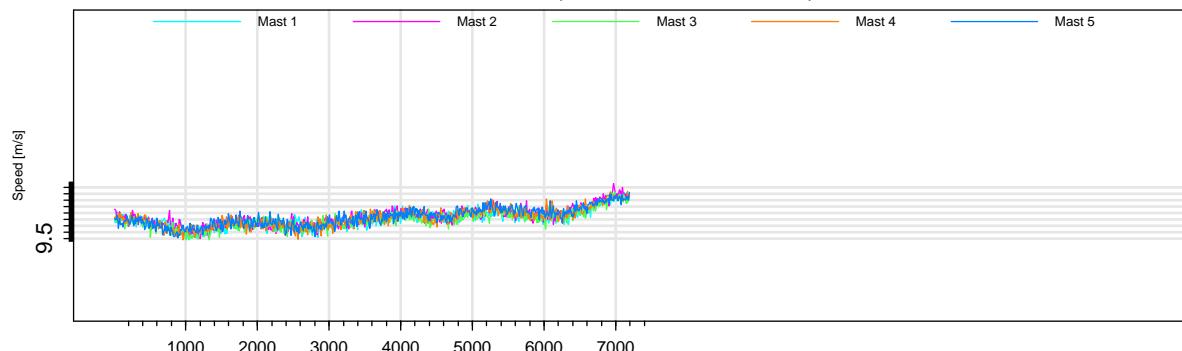
Figure 16: 10s averages of wind speed and direction, series: 5 & 6.
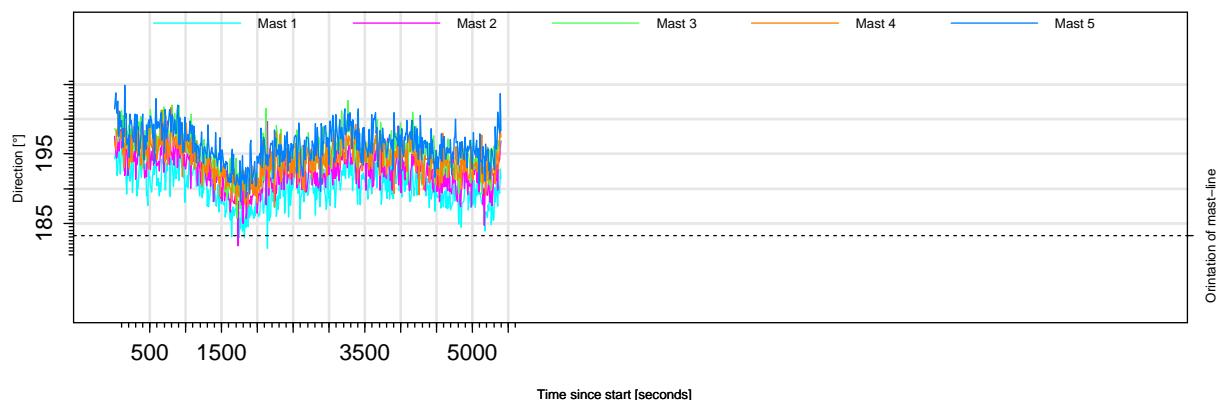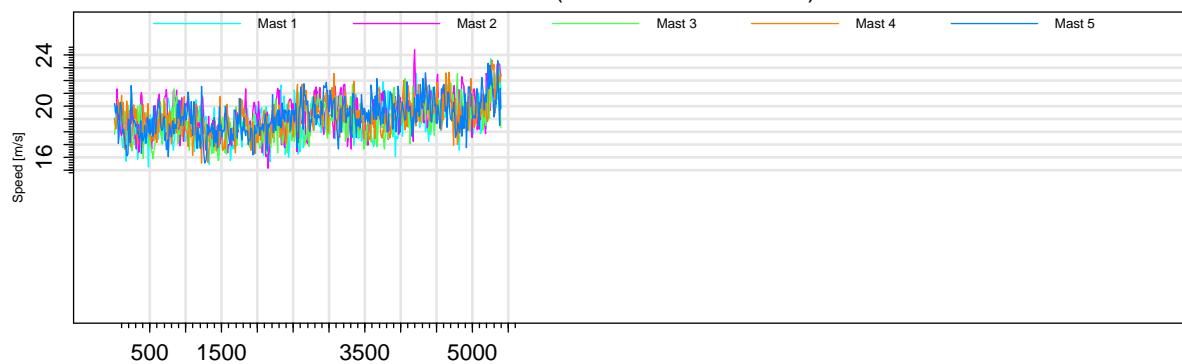
Figure 17: 10s averages of wind speed and direction, series: 7 & 8.

Figure 18: 10s averages of wind speed and direction, series: 9 & 10.
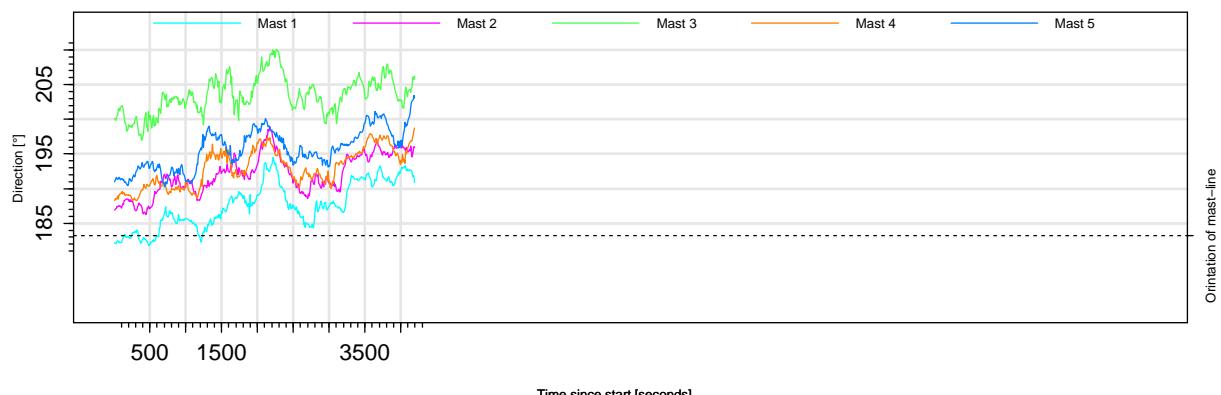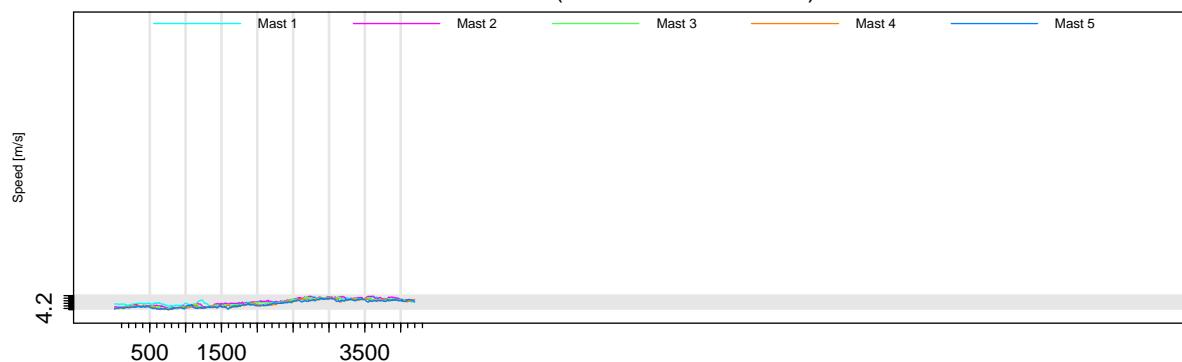
Figure 19: 10s averages of wind speed and direction, series: 11 & 12.

B Estimated impulse response based on high-pass filtered series

The horizontal line marks an approximate 95% confidence interval under the hypothesis that the two series under consideration is uncorrelated. For the Sample Partial Auto Correlation Function (SPACF) the confidence intervals are for the white noise hypothesis.

The tick marks over the plots of the estimated impulse responses marks the distance between the masts under consideration in terms of lags of 10 seconds at the average wind speed for the series.
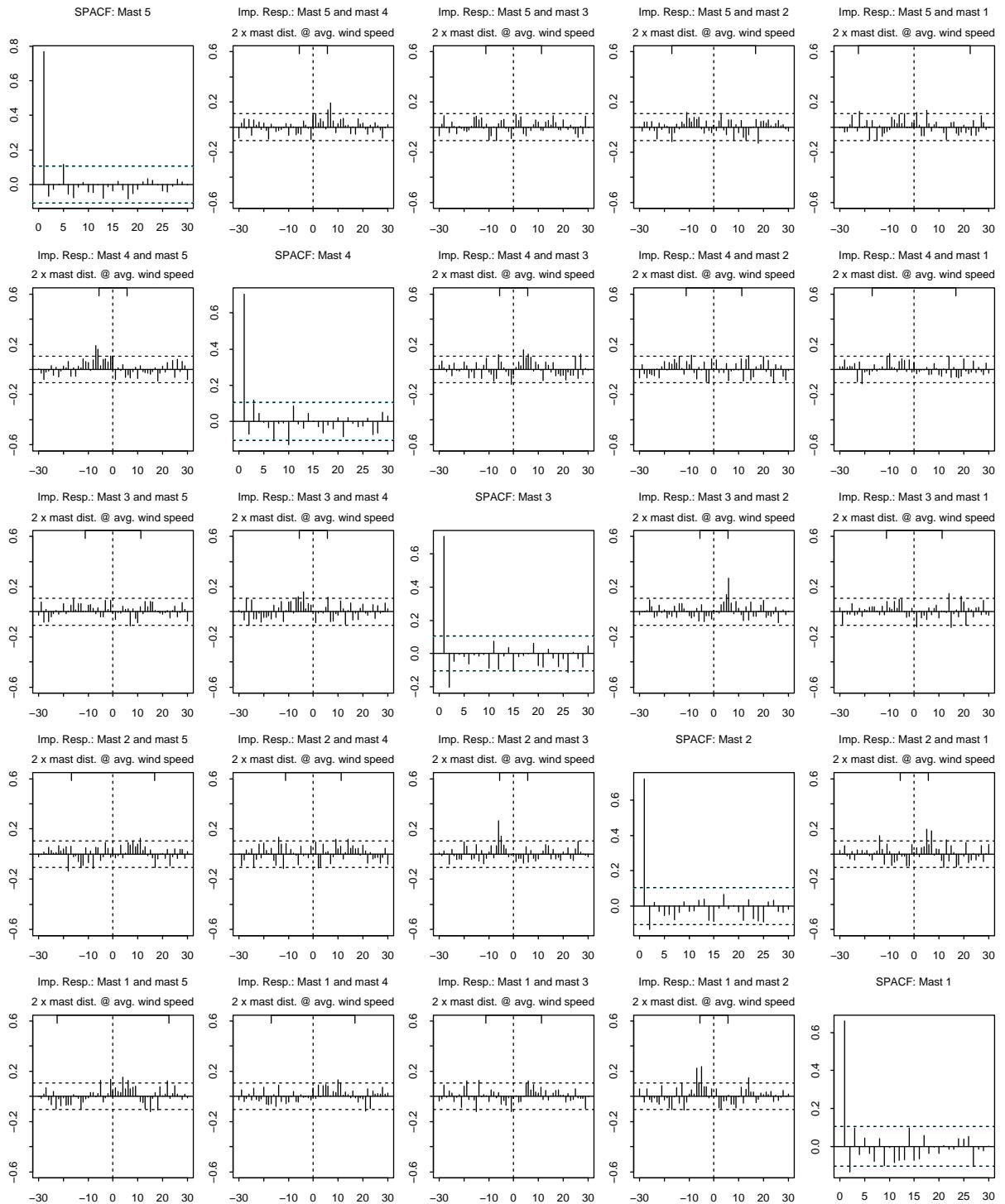
Figure 20: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 1.

Høvsøre (start 20030801 0951) 9.9 m/s (Series 02)
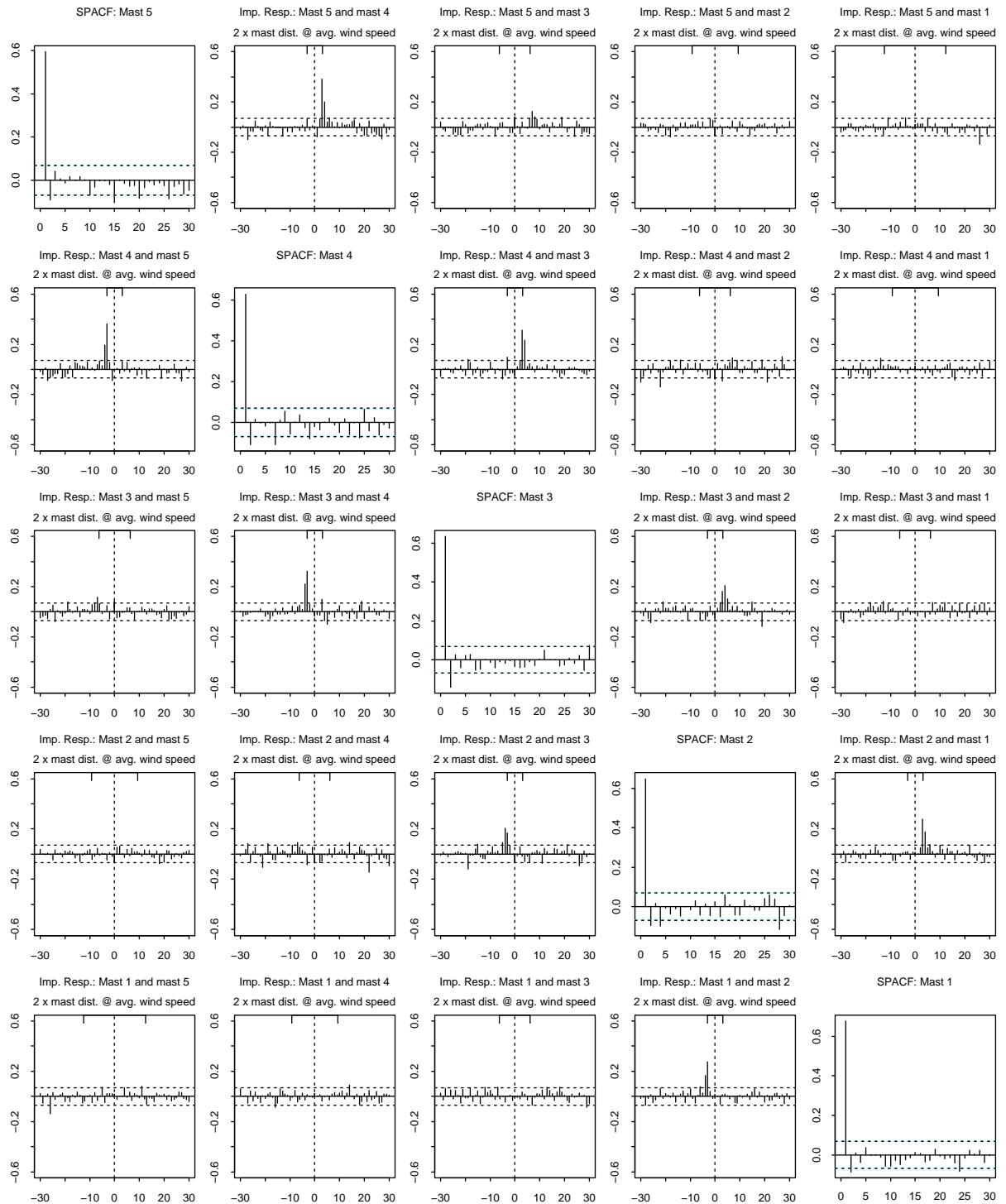
Figure 21: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 2.

39
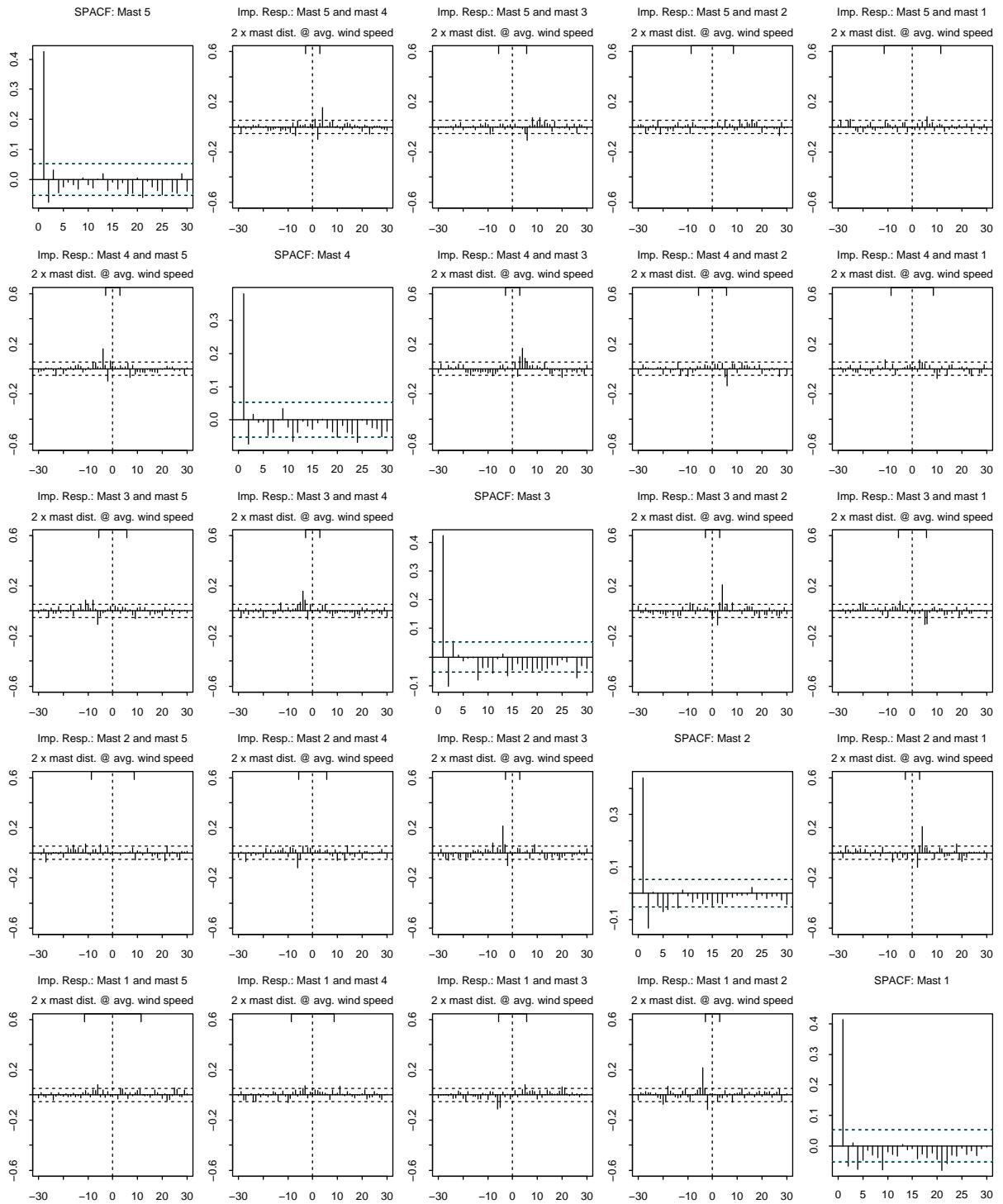
Figure 22: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 3.
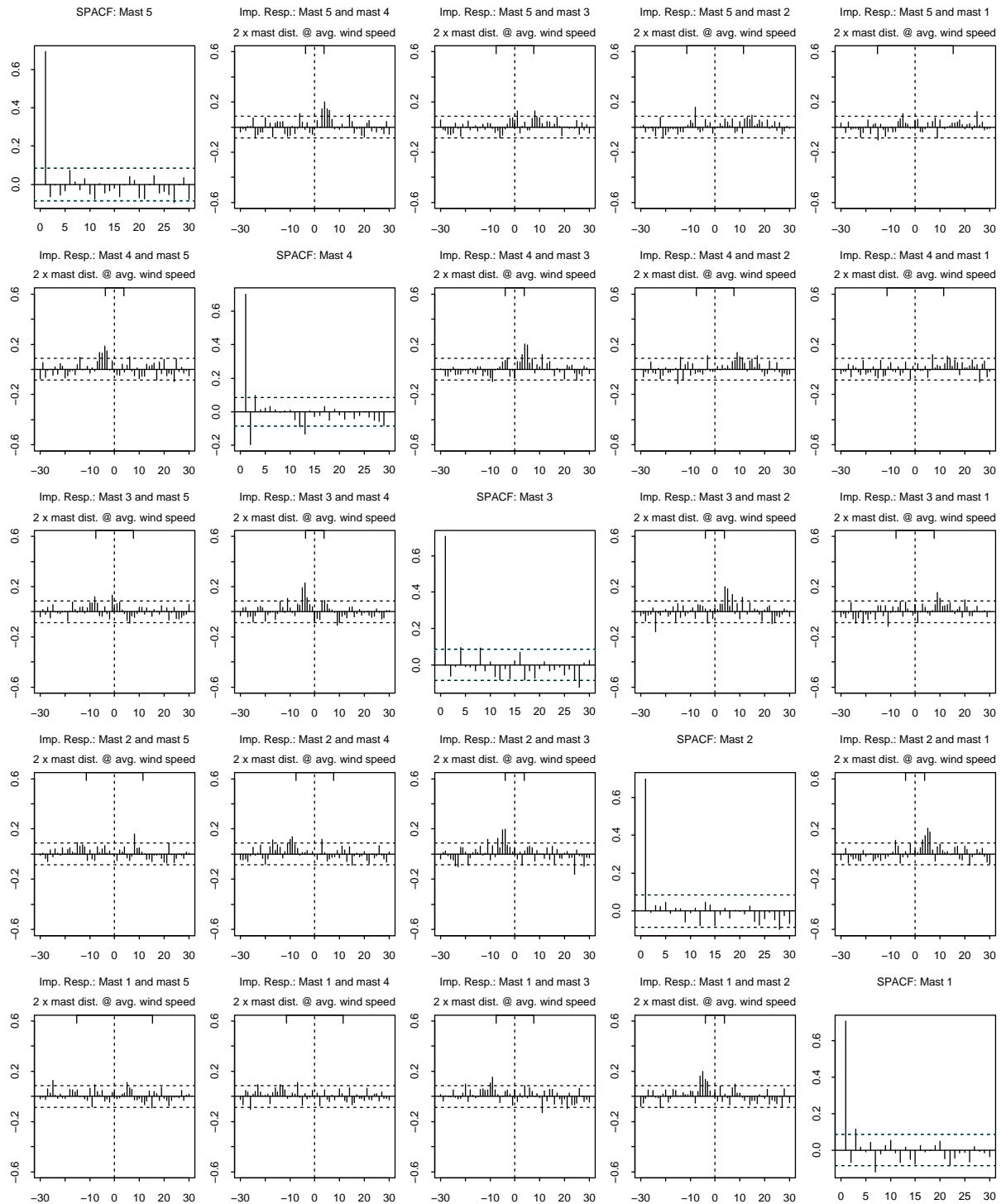
Figure 23: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 4.
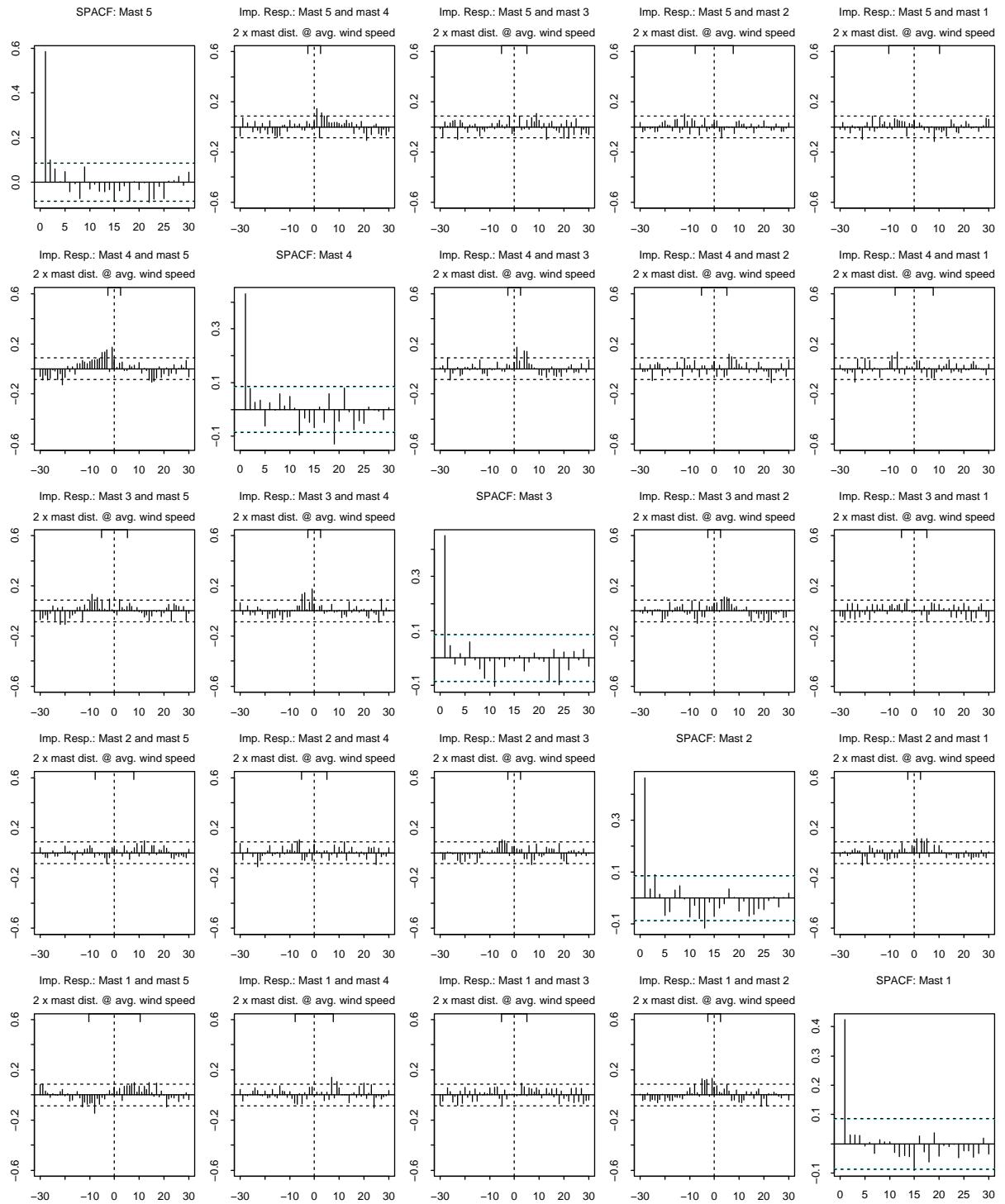
Figure 24: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 5.
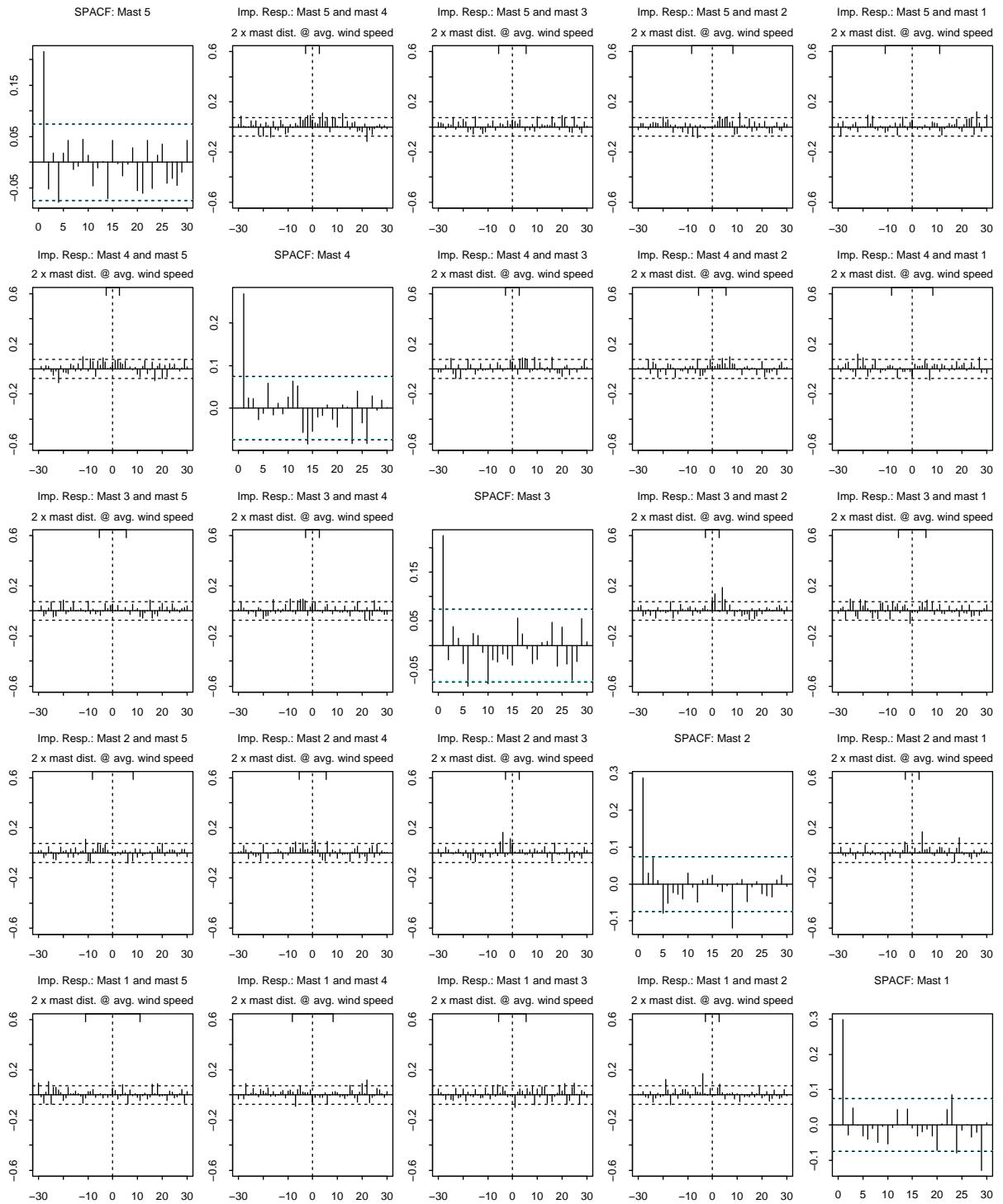
Figure 25: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 6.
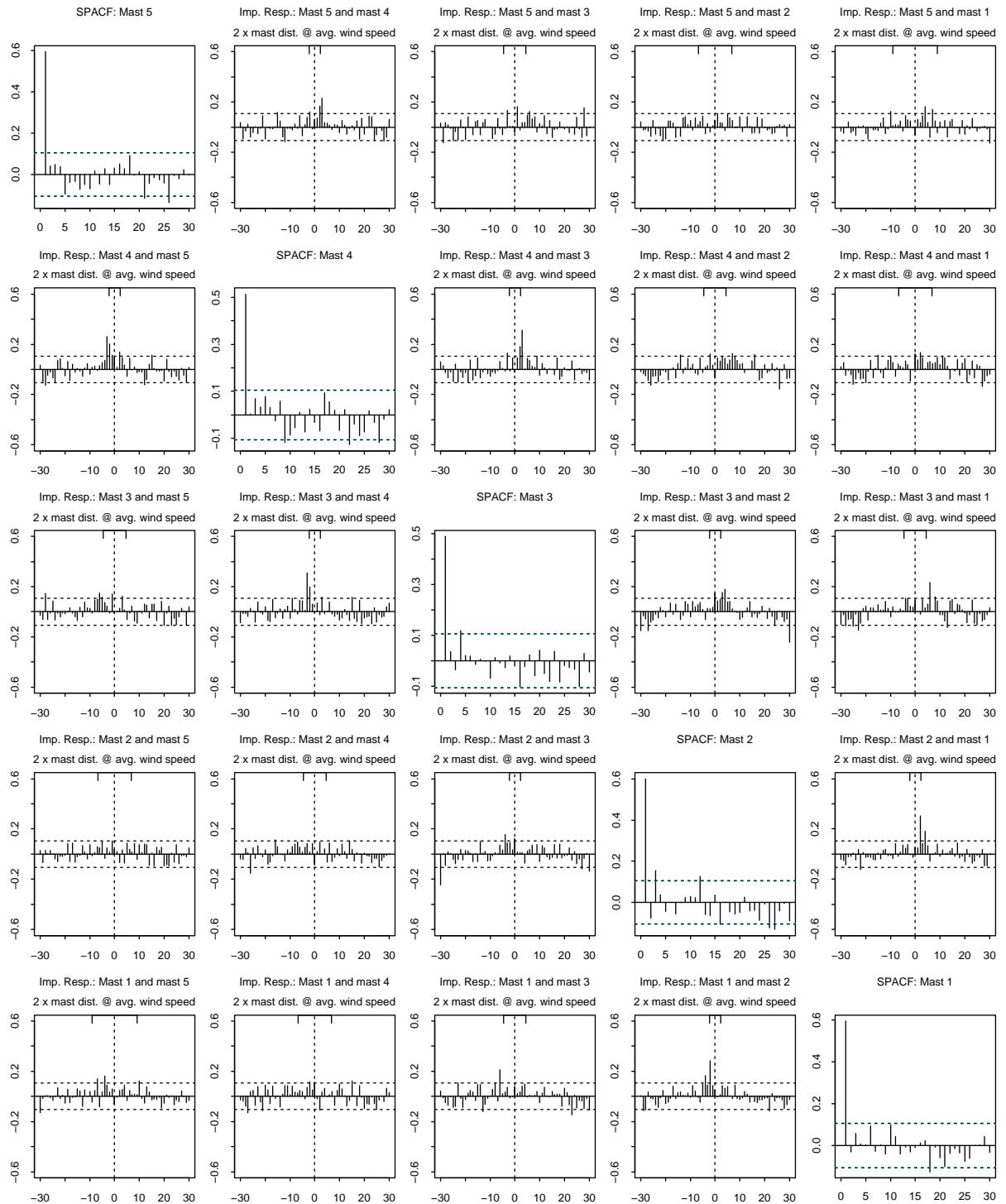
Figure 26: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 7.
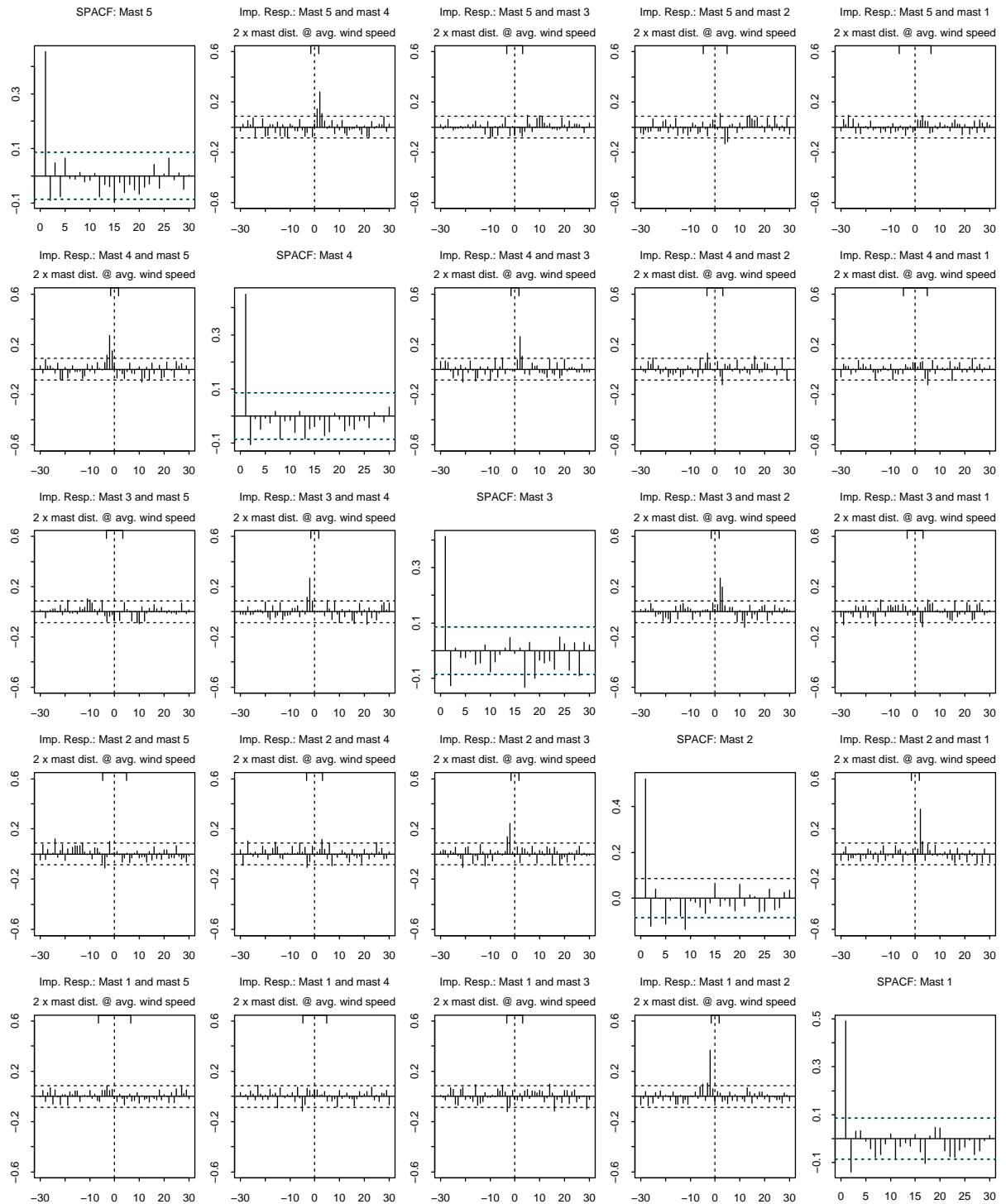
Figure 27: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 8.

Høvsøre (start 20030930 1850) 8.2 m/s (Series 09)
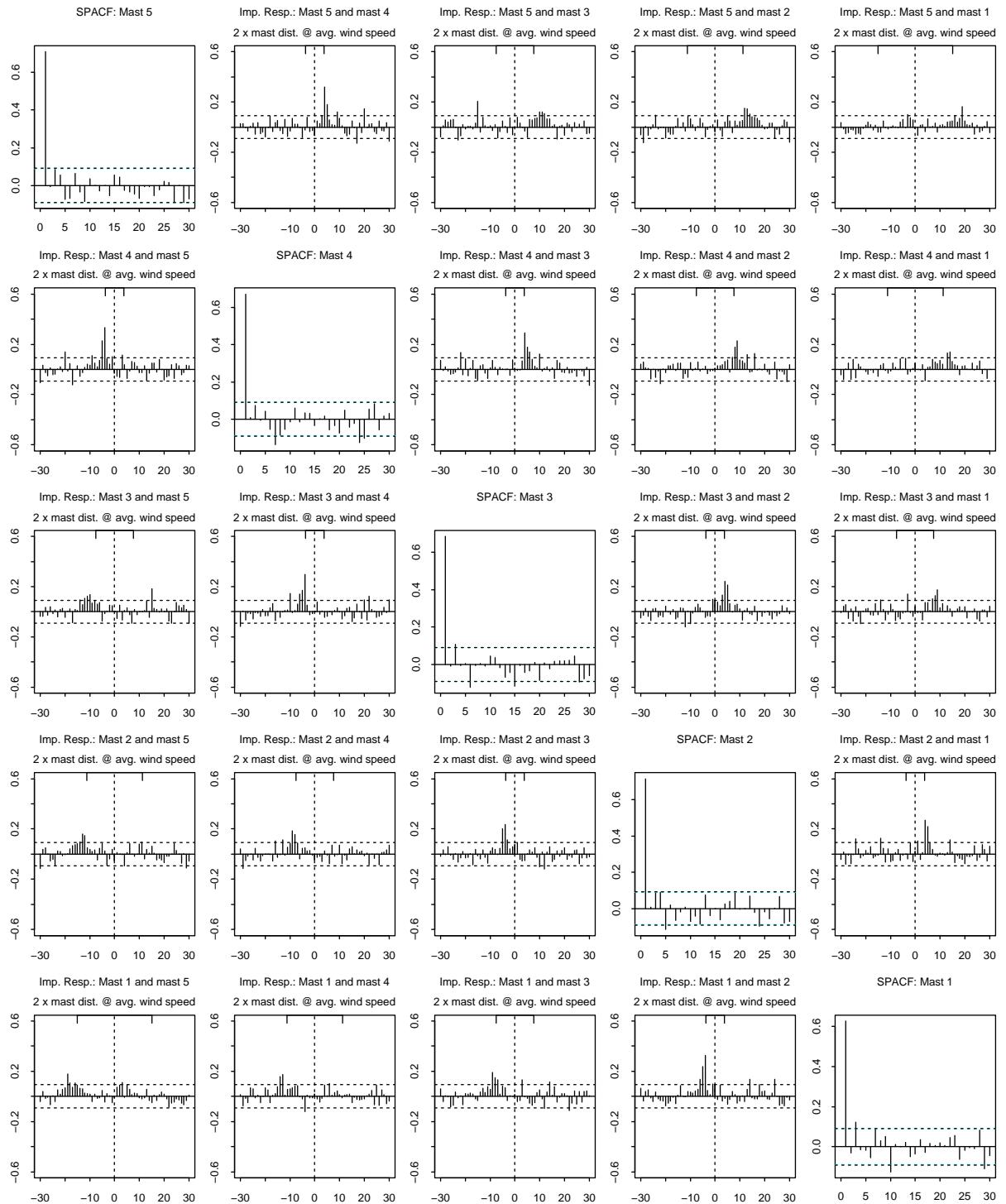
Figure 28: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 9.

46
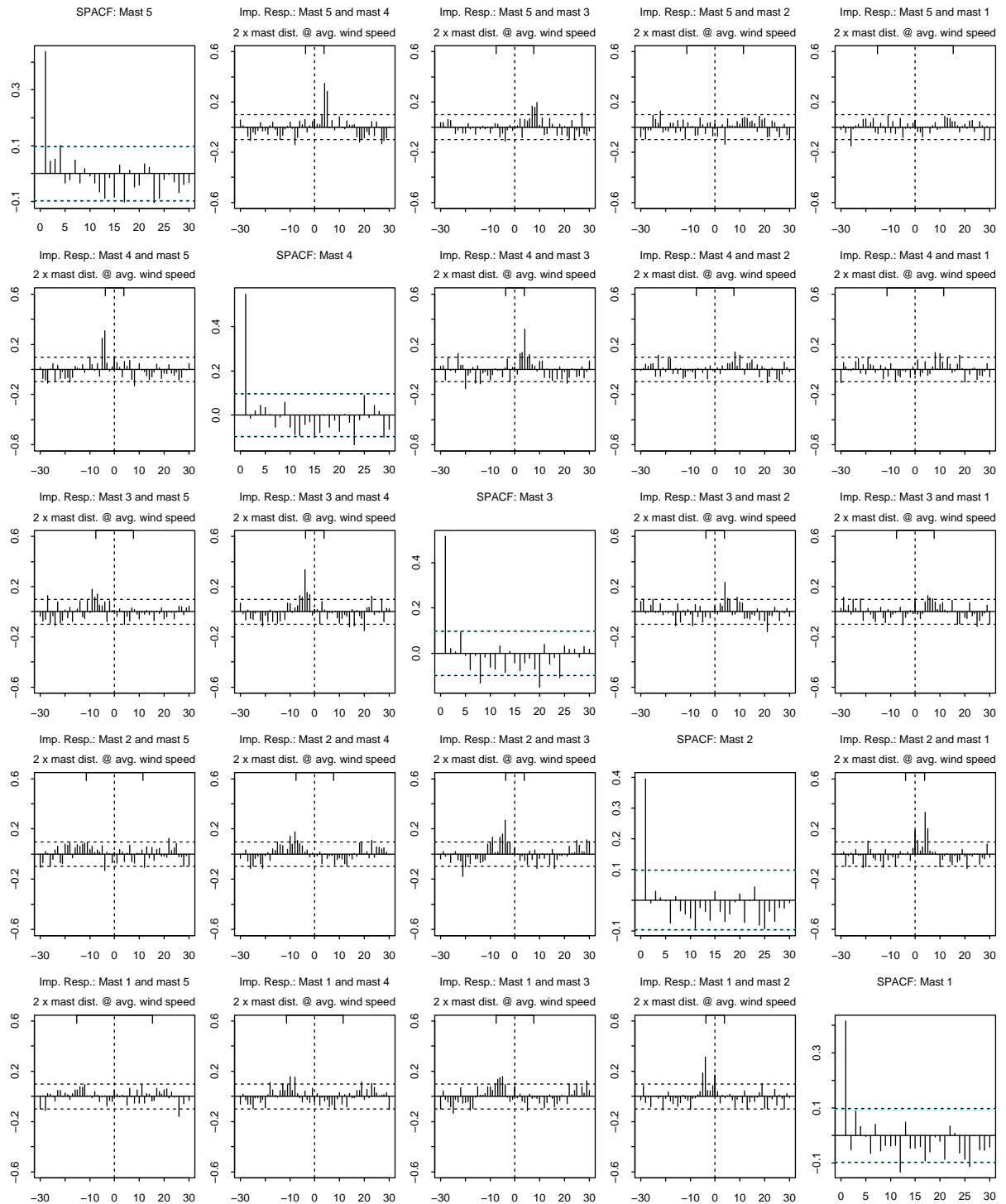
Figure 29: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 10.
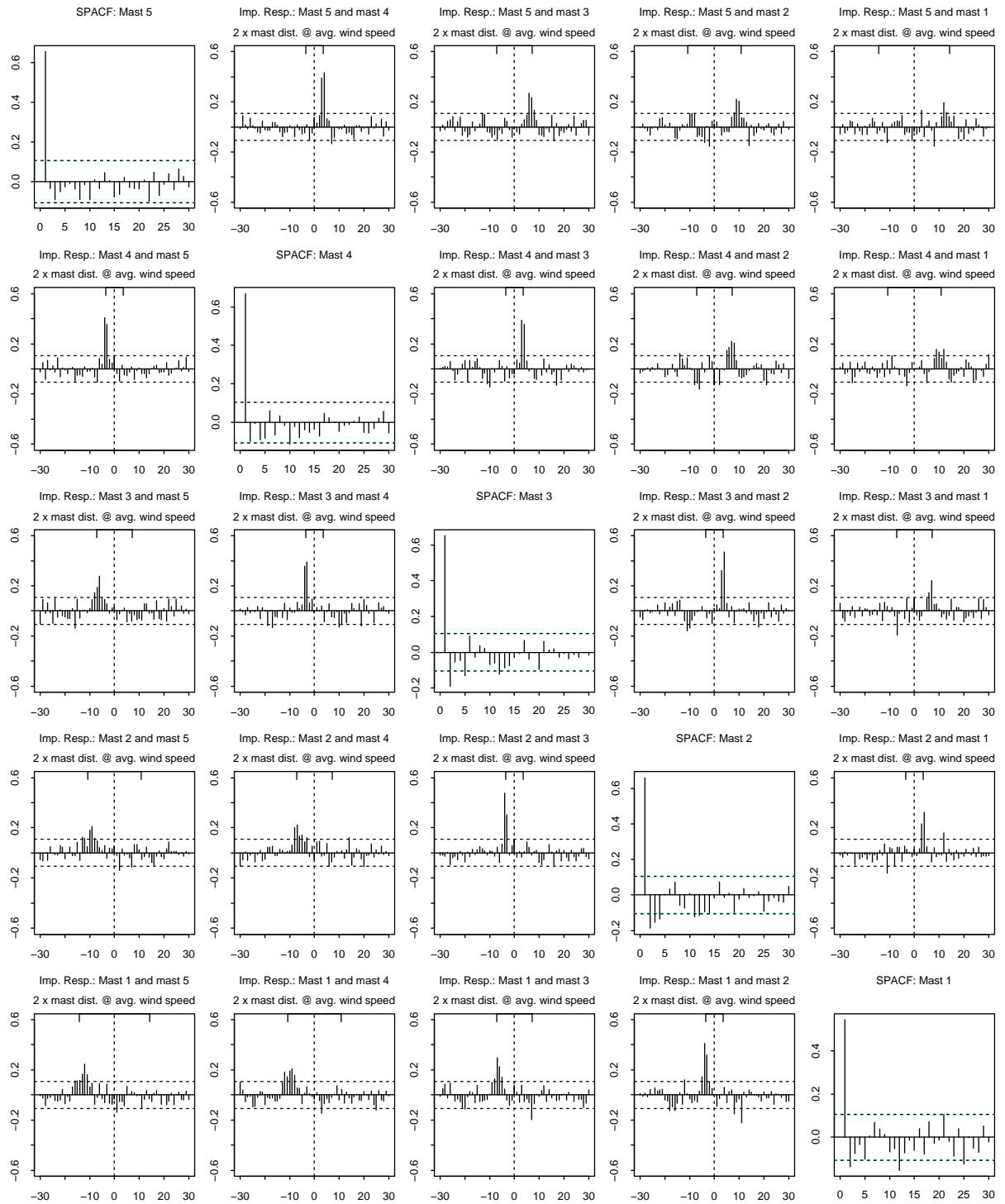
Figure 30: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 11.
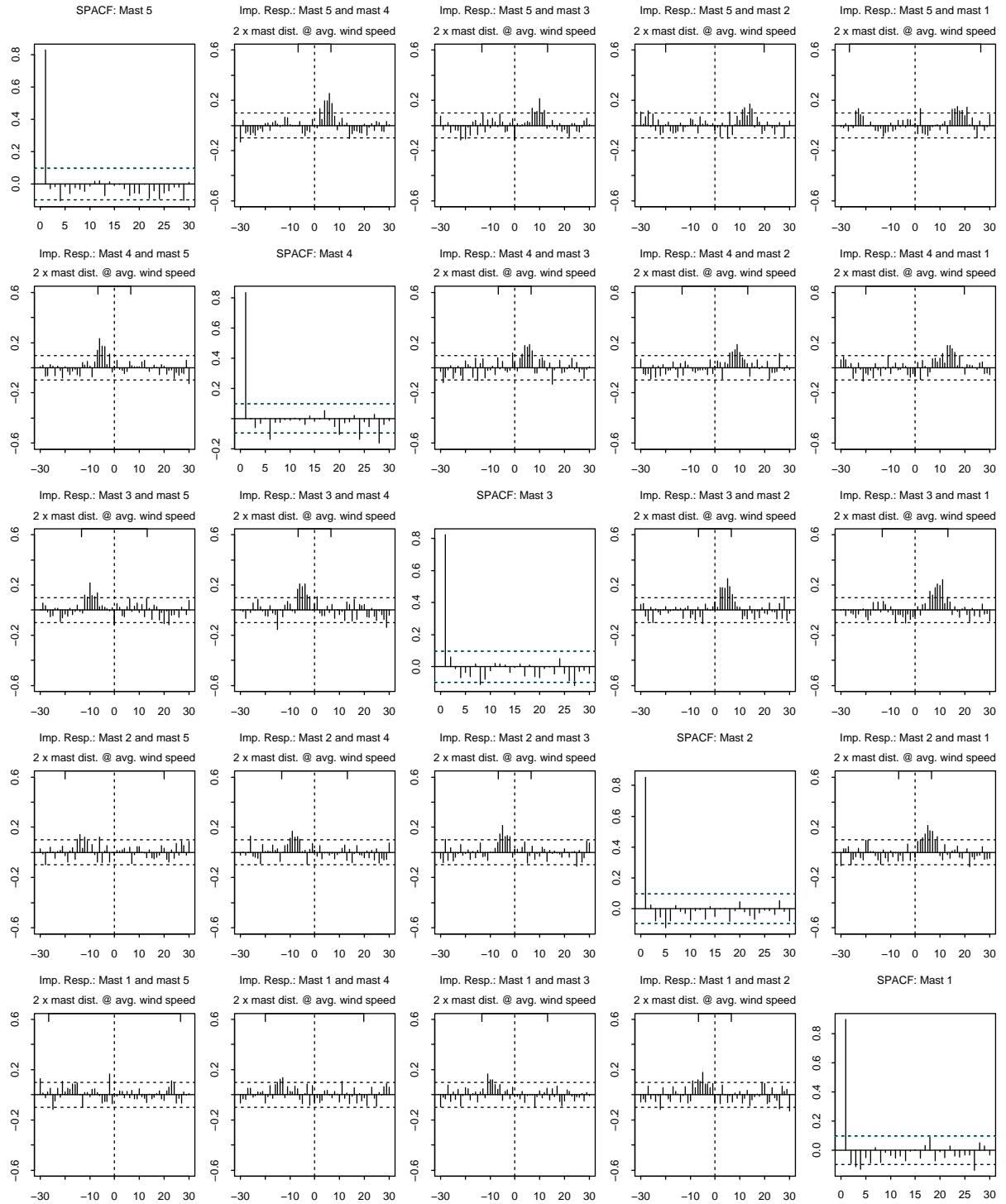
Figure 31: Estimated impulse responses (off-diagonal) and partial auto correlation function (diagonal) for high-pass filtered 10s averages, series: 12.

C   Fitted values when using continuous time impulse responses on neighbor masts
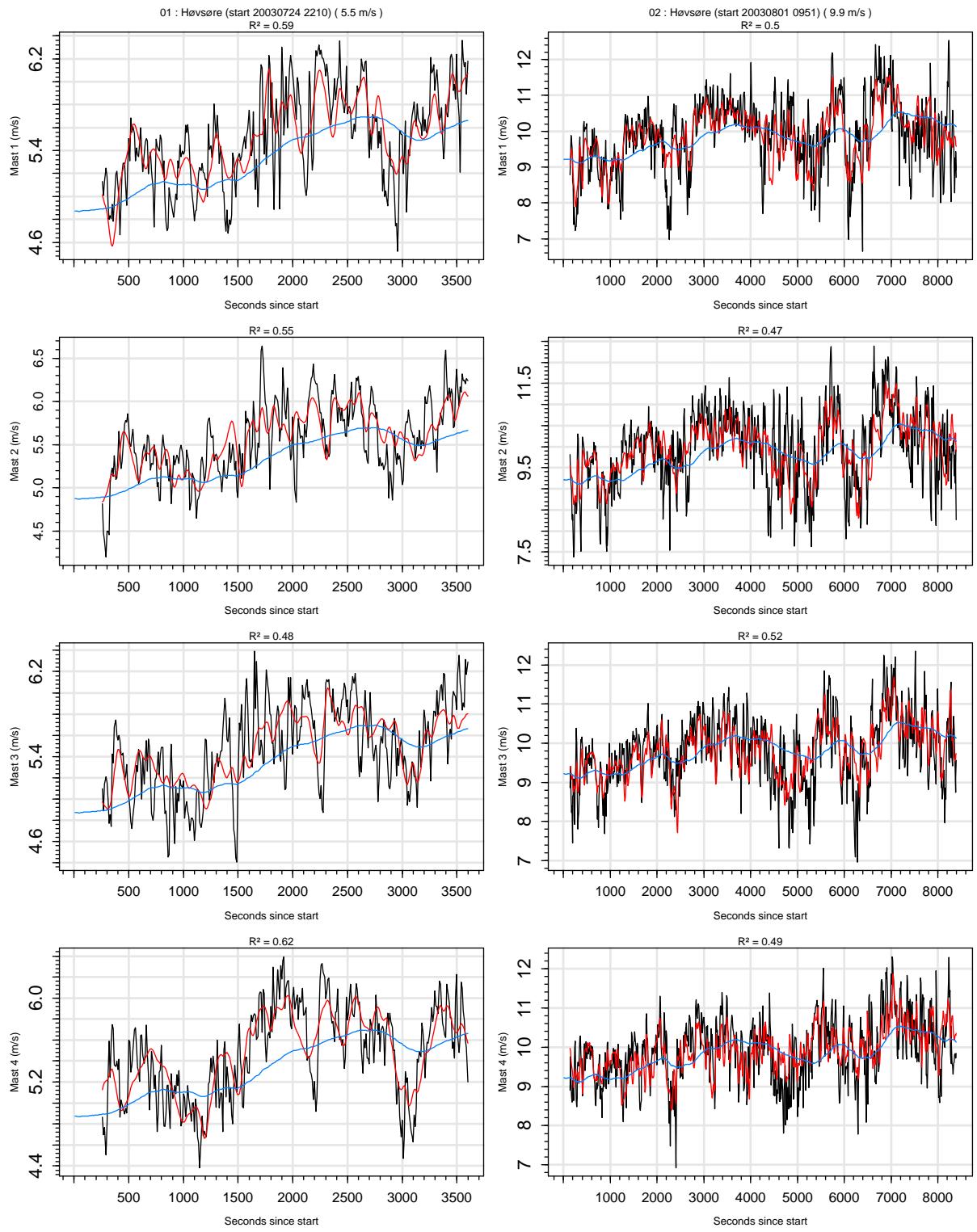
Figure 32: Series 1 and 2: Data (black) and fitted values (red) for model (14) with spline bases constructed as described in Section 4.1. The blue curve is the estimate of $\mu(t)$ used. R-squared of the fits are shown above each plot.
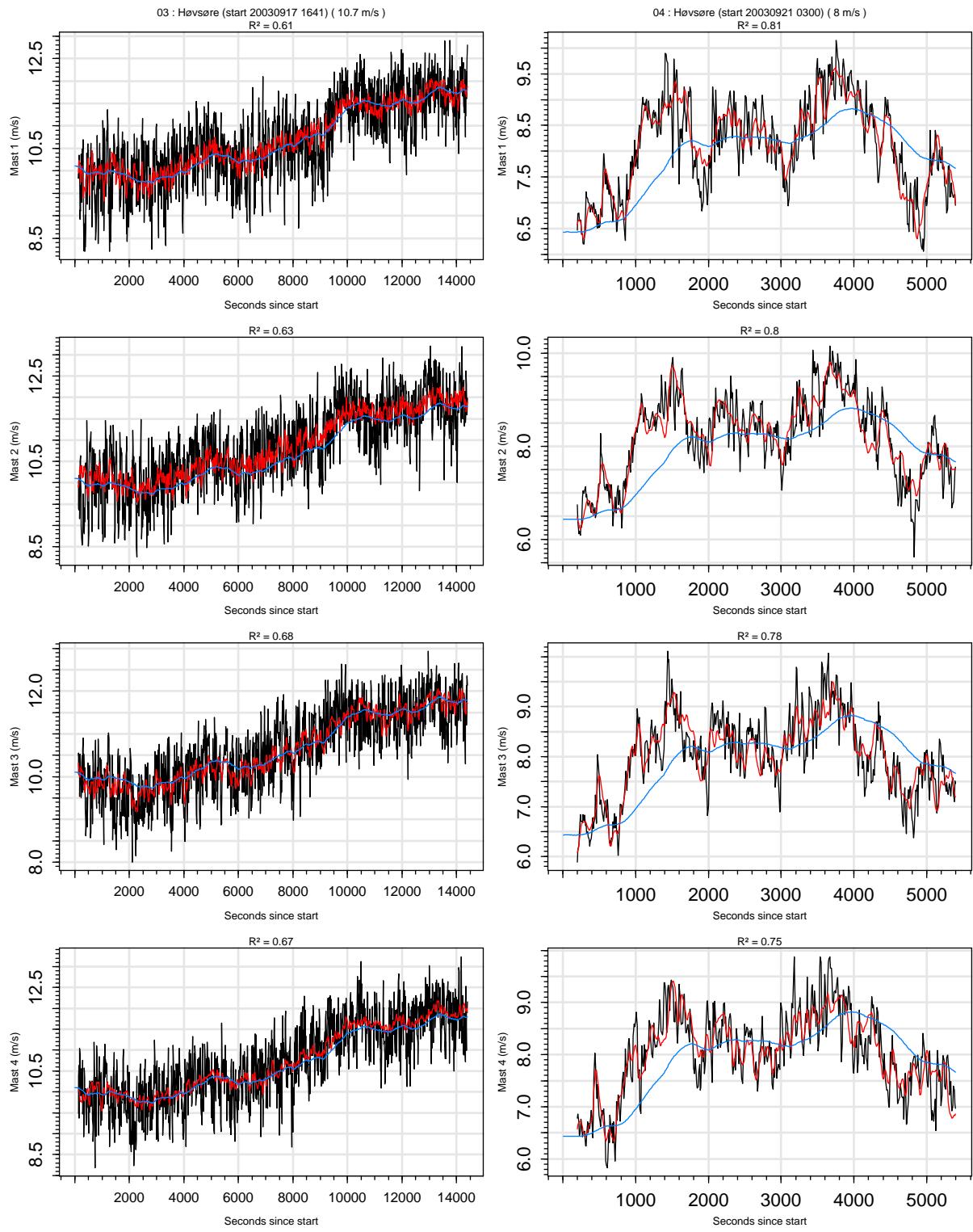
Figure 33: Series 3 and 4: Data (black) and fitted values (red) for model (14) with spline bases constructed as described in Section 4.1. The blue curve is the estimate of $\mu(t)$ used. R-squared of the fits are shown above each plot.
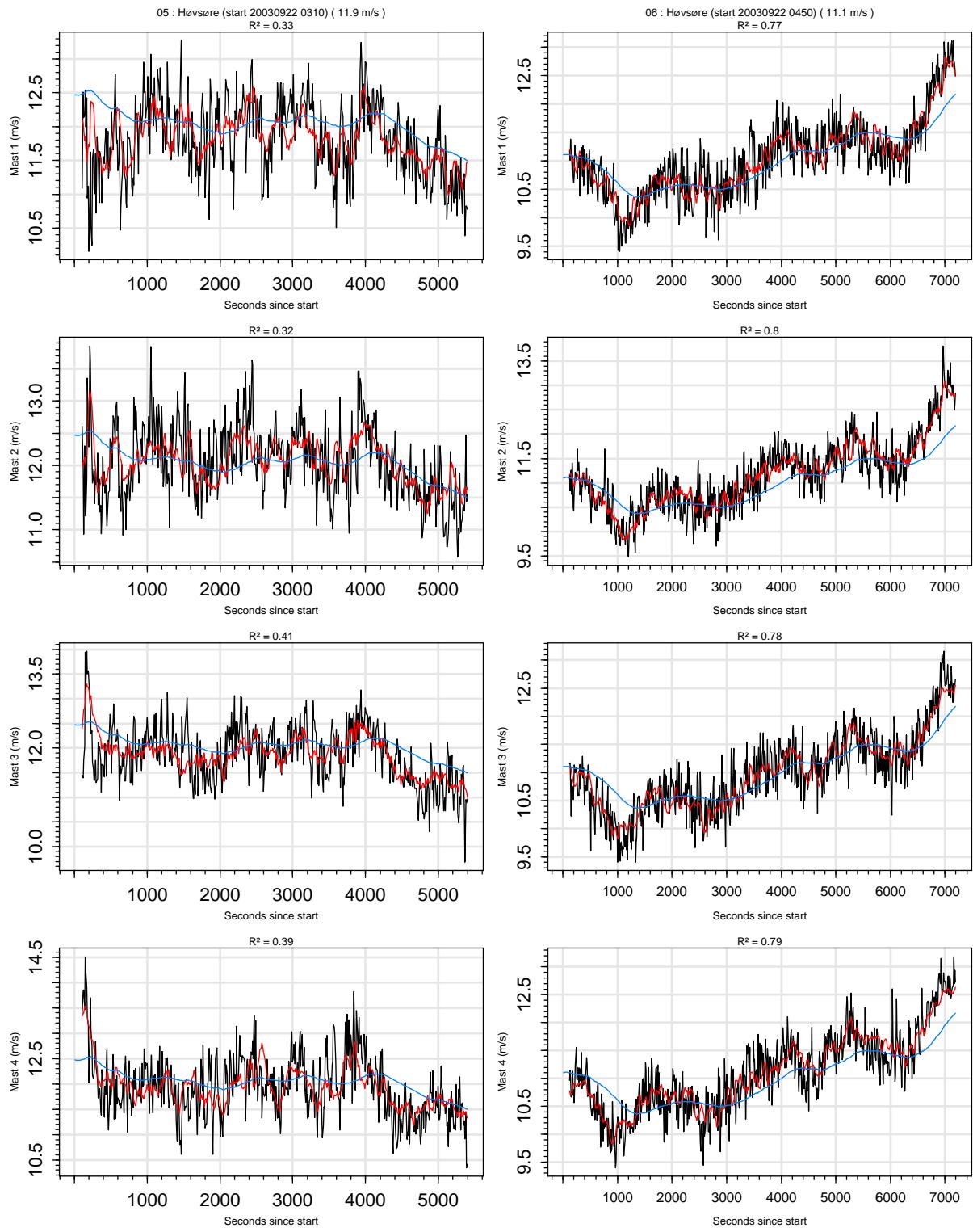
Figure 34: Series 5 and 6: Data (black) and fitted values (red) for model (14) with spline bases constructed as described in Section 4.1. The blue curve is the estimate of $\mu(t)$ used. R-squared of the fits are shown above each plot.
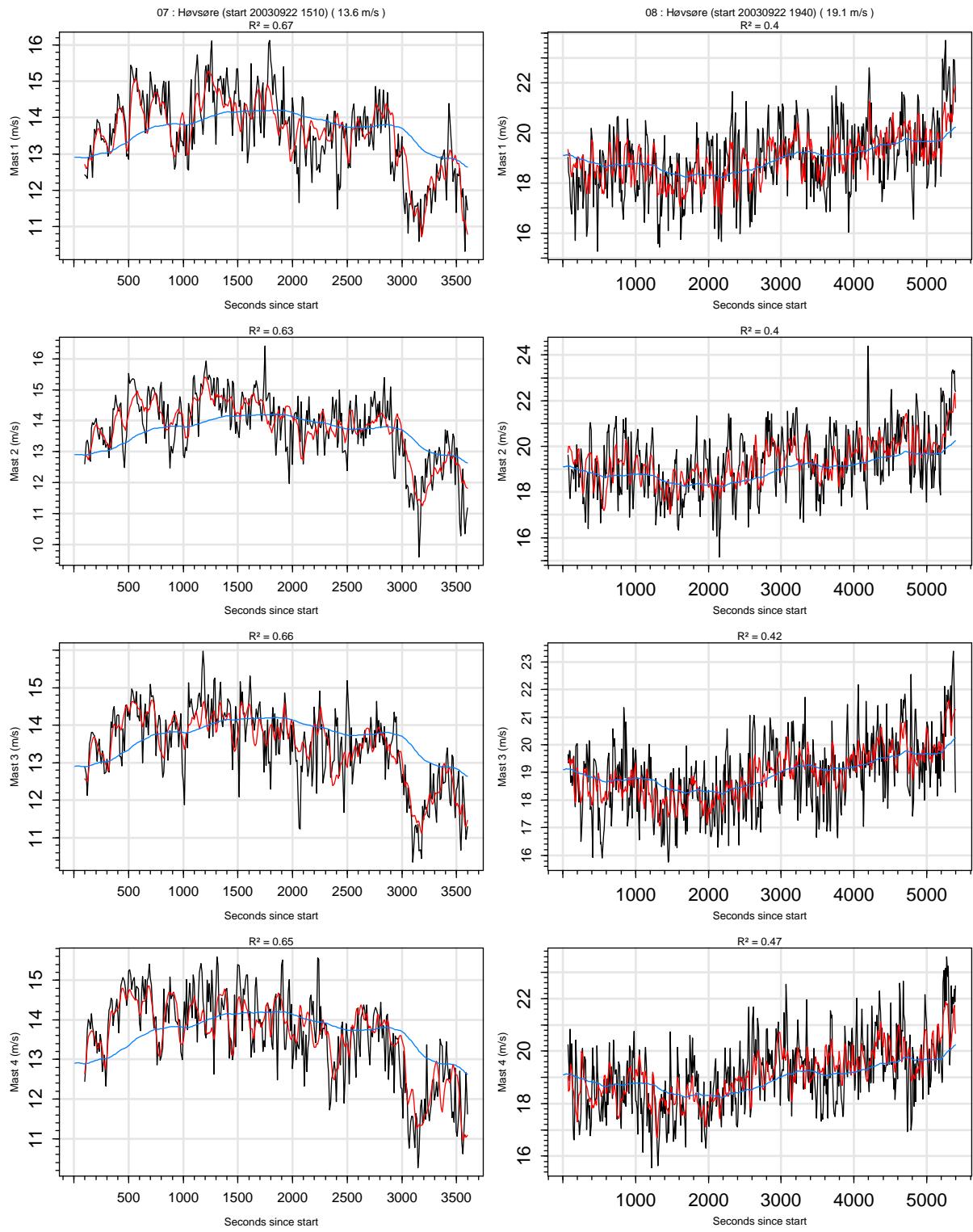
Figure 35: Series 7 and 8: Data (black) and fitted values (red) for model (14) with spline bases constructed as described in Section 4.1. The blue curve is the estimate of $\mu(t)$ used. R-squared of the fits are shown above each plot.
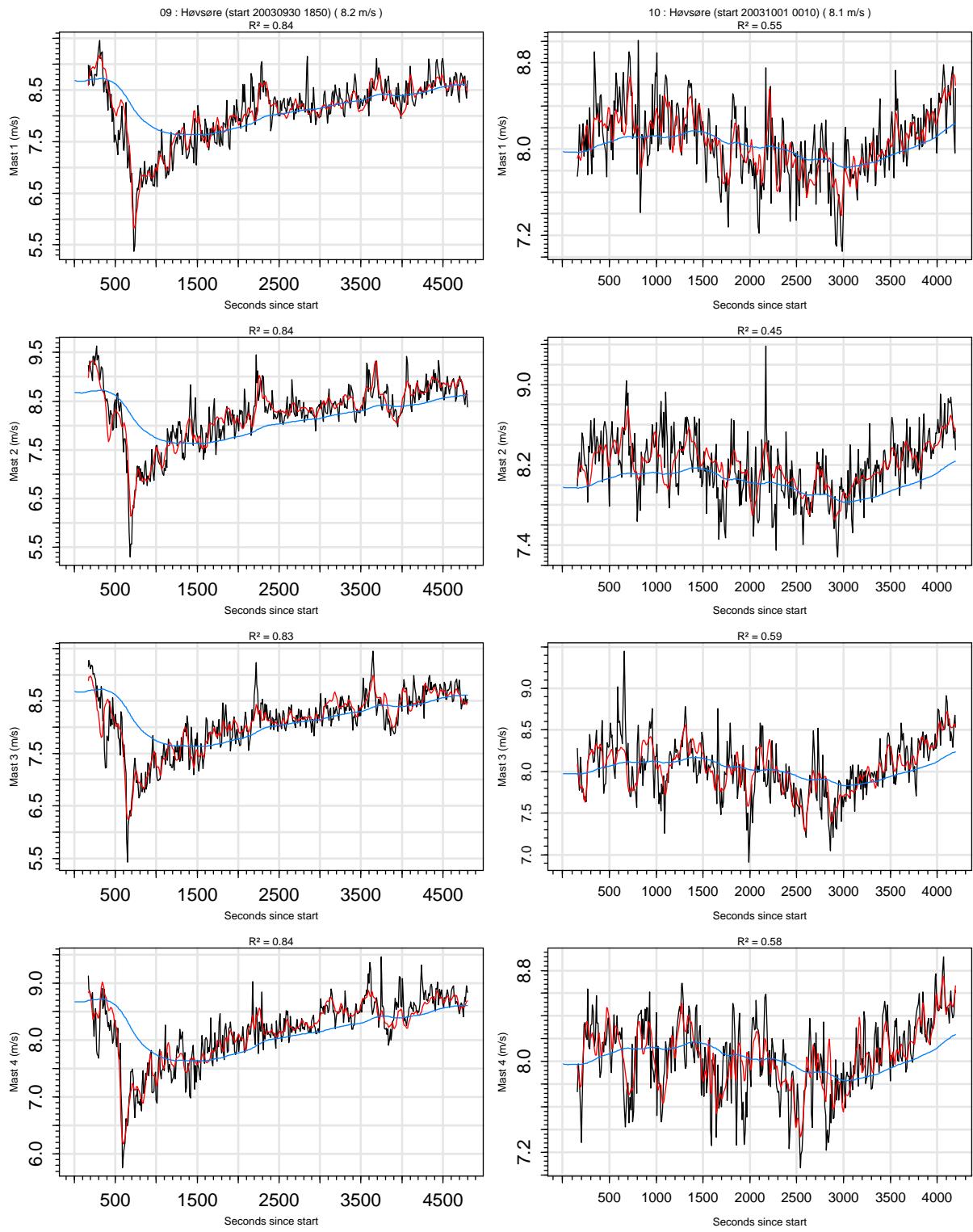
Figure 36: Series 9 and 10: Data (black) and fitted values (red) for model (14) with spline bases constructed as described in Section 4.1. The blue curve is the estimate of $\mu(t)$ used. R-squared of the fits are shown above each plot.
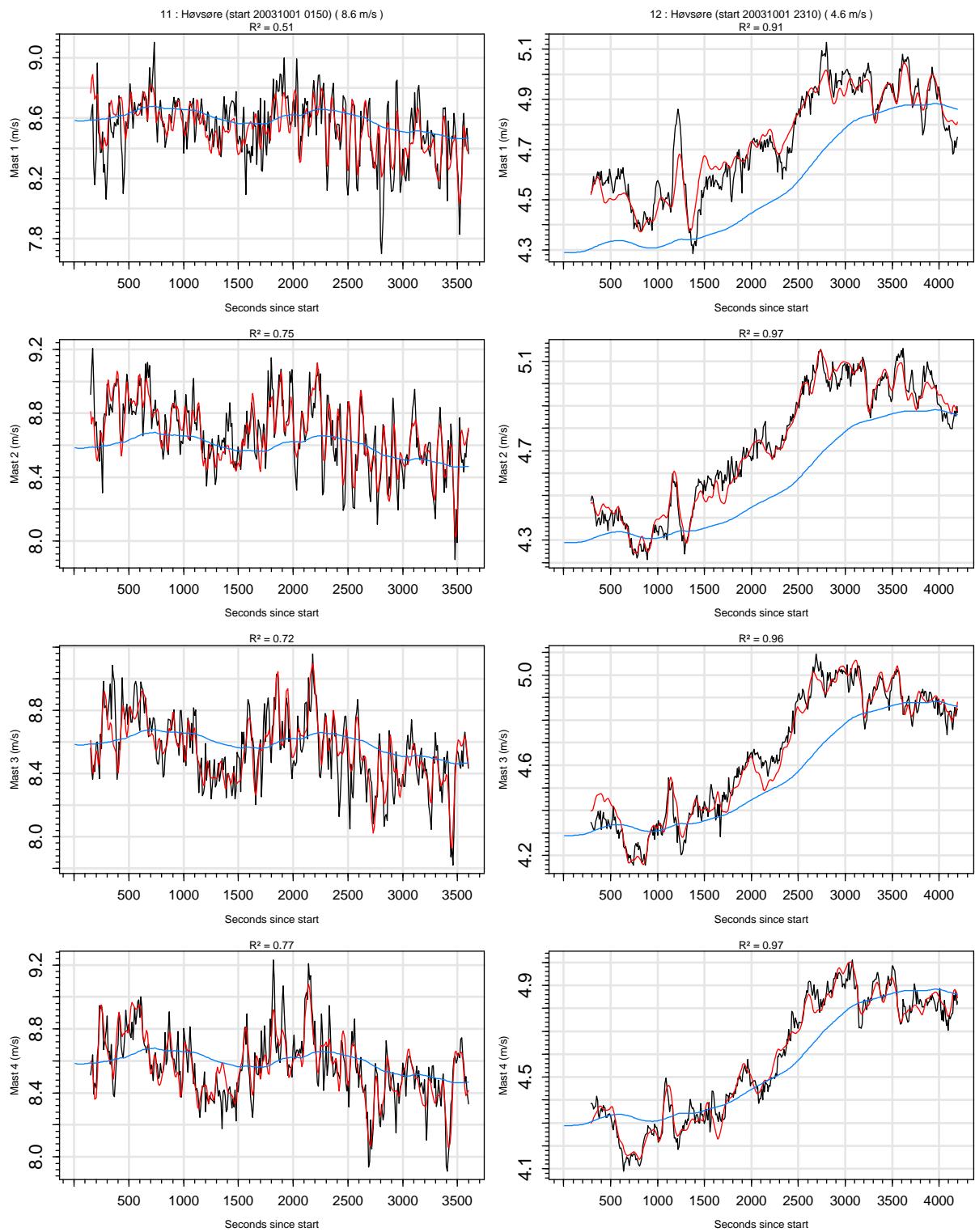
Figure 37: Series 11 and 12: Data (black) and fitted values (red) for model (14) with spline bases constructed as described in Section 4.1. The blue curve is the estimate of $\mu(t)$ used. R-squared of the fits are shown above each plot.

D Restricted spline bases

Consider a function $f(x)$ approximated by a linear combination of spline basis functions $b_k(x); \ k = 1, \ldots, K$, i.e.

$$f(x) = \sum_{k=1}^{K} a_k b_k(x) \ , \tag{15}$$

where $a_k; \ k = 1, \ldots, K$ are constants.

We want to apply *a number* of restrictions of the type

$$f^{(n_0)}(x_0) = 0 \ , \tag{16}$$

where the superscript indicates the n_0'th order derivative of f. From (15) it follows that

$$f^{(n)}(x) = \sum_{k=1}^{K} a_k b_k^{(n)}(x) \ . \tag{17}$$

This requires that the n'th order derivatives of the basis functions exists. In turn this implies that the upper bound on n is 2 below the order of the spline.

For any $j = 1, \ldots, K$ for which $b_j^{(n_0)}(x_0) \neq 0$ the restriction (16) implies

$$a_j = -\sum_{k \neq j} a_k \frac{b_k^{(n_0)}(x_0)}{b_j^{(n_0)}(x_0)} \ . \tag{18}$$

Inserting this into (17), which included (15) as a special case, shows how a new basis can be constructed by eliminating column j:

$$f^{(n)}(x) = \sum_{k \neq j} a_k \left(b_k^{(n)}(x) - \frac{b_k^{(n_0)}(x_0)}{b_j^{(n_0)}(x_0)} b_j^{(n)}(x) \right) \ . \tag{19}$$

Using the basis functions resulting from this elimination the process can be repeated for additional restrictions.

To obtain a procedure which is unambiguous the column to be eliminated j is selected as the first column corresponding to the largest $|b_j^{(n_0)}(x_0)|$.

An S-PLUS function implementing the above is shown below.

```
bfix <- function(x,
                 x0 = range(x),
                 derivFix = rep(0, length(x0)),
                 valueFix = rep(0, length(x0)),
                 FUN = bs, verbose = F, derivs = 0, intercept = T, ...)
{
  ##

  ## Generate a basis with equality constaints on the function
  ## approximated or on its derivatives.
  ##
  ## x0       : the value(s) of x for which the approx. or its derivative
  ##            must be valueFix.
  ## derivFix : The derivative which must be fixed
  ##            (max. order-2, i.e. degree-1; NOT CHECKED).
  ## valueFix : The value at which it must be fixed (ONLY ZERO IS SUPPORTED!).
  ## FUN      : the function used to generate the unrestricted basis
  ##            (bs or ns).
  ## verbose  : Print information about which columns are eliminated.
  ## derivs   : Like in bs/ns.
  ## intercept: Like in bs/ns, but the default is changed to TRUE.
  ## ...      : Additional arguments parsed to FUN.
  ##
  ## Examples : See the file 'bfixEx.S'.
  ##
  ## Non-zero valueFix is not implemented (it generates an offset and
  ## thus the output must contain both a basis and a offset to be
  ## subtracted from the response in the linear case).

  ##

  if(any(valueFix != 0)) stop("non-zero fixed values currently not supported")

  if(length(derivs) == 1)
    derivs <- rep(derivs, length(x))

  derivs <- c(derivFix, derivs)
  x <- c(x0, x)

  B <- FUN(x, derivs = derivs, intercept = intercept, ...)
  for(ll in 1:length(x0)) {
    jj <- (1:ncol(B))[abs(B[ll, ]) == max(abs(B[ll, ]))][1] # 1st col. at max
    if(verbose) cat("Step:", ll, "; eliminating column", jj, "\n")
    B <- B[, -jj] - outer(B[, jj], B[ll, -jj] / B[ll, jj])
  }

  B <- B[-(1:length(x0)), ]

  return(B)
}
```