Direct spectral verification of a mesoscale ensemble

Jesper Nissen, Claire Vincent

Risø National Laboratory for Sustainable Energy – Technical University of Denmark

Project report for the PSO project: High Resolution Ensemble for Horns Rev

18 February 2010

Objective and abstract

The motivation for spectral verification of atmospheric meteorological models lies in the need to assess model skills in capturing meteorological phenomena on a variety of spatial and temporal scales. These skills might not show up properly in conventional statistical validation metrics such as root mean square error and mean absolute error, since small timing inaccuracies in the simulations can result in poor model performance scores due to phase errors in atmospheric phenomena that were otherwise well captured by the model.

In this work, we aim to verify the shorter time scales found in the Multi Scheme Ensemble Prediction System (MSEPS). The variability in the forecasts and observations will be directly verified using the Ensemble Empirical Mode Decomposition (EEMD). The EEMD represents a significant development on the earlier Empirical Mode Decomposition, since timescales are not mixed between the components. This allows direct comparison of the modes (components) of the forecast and observed time series. Therefore, we can see directly the scales at which the model starts to produce realistic fluctuations.

The methodology is developed using both a 2 month segment of the continuous 'warm run' of the MSEPS, and using a 20 day period of individual 24 hour forecasts. The key results are that the model starts to produce variance with a realistic amplitude for time scales longer than about 2-3 hours, that some faster scale variability events are captured by the model, but that in general, realistic variability is not found in the model on time scales shorter than about 2-3 hours.

Introduction

The development of suitable verification methodologies is essential for the assessment and comparison of mesoscale models. As discussed in Vincent et al. (2009), it is impossible for a mesoscale model to capture the exact timing of the peaks and troughs of wind fluctuations on time scales of around an hour or less. A successful mesoscale forecast is one that captures the statistics of the wind speed variability (that is, the amplitude and period of the fluctuations), rather than one that minimises a global verification score such as RMSE or bias. This is in contrast to large scale variability (such as the synoptic or diurnal cycles), where the phase as well as the amplitude are important aspects of a good forecast.

Previous work in this project used the Hilbert-Huang transform to separate scales in MSEPS forecasts, and verify the model performance over different time scales (Vincent et al. 2009). It was found that certain MSEPS members consistently scored well on faster time scales, while other MSEPS members were more effective on slower time scales. The relative performance of all 75 ensemble members on a range of different time scales was presented. This previous study used a special 'warm run' version of the MSEPS model, which was forced with the large scale weather patterns from the boundaries, but which was not reinitialized with a new set of initial conditions every six hours.

During the past two years, there have been significant developments in the Hilbert Huang Transform methodology which address some of its previous limitations and which potentially make it a simpler and more useful tool for mesoscale verification. In this report, the limitations of the methodology in Vincent et al(2009) will be discussed, the new developments to the Hilbert Huang Transform will be described. Preliminary results of applying the new methodology will be presented.

Section 1: The Ensemble Empirical Mode Decomposition

The Hilbert Huang transform is a two step process for frequency-time analysis of non-stationary time series. In its original form (Huang et al. 1998), the time series was first decomposed into a set of components (called Intrinsic Mode Functions or IMFs) using an empirical sifting technique, and then each component was transformed using the Hilbert Transform to find its instantaneous frequencies and amplitudes. There were two serious limitations of the empirical sifting technique, which was based on recursively subtracting the mean of the time series. Firstly, the decomposition was not unique because it depended on the choice of convergence criteria. Secondly, a signal of a particular frequency could be split between more than one IMF, a problem known as 'mode mixing'. This meant that it was not possible to attach a physical interpretation to a single IMF.

These two limitations were addressed in Wu and Huang (2009) with the development of the Ensemble Empirical Mode Decomposition (EEMD). In the EEMD, the time series is pre-whitened N times with N different realisations of white noise. Each of the N pre-whitened versions of the time series is decomposed using the original empirical sifting technique, and the N realisations of each IMF are then averaged to find the 'true' decomposition. The development addresses both the problems of uniqueness and mode-mixing. Conceptually, the original decomposition technique and the EEMD are compared in figures 1 and 2.

Figure 1: Original methodology for decomposing the time series

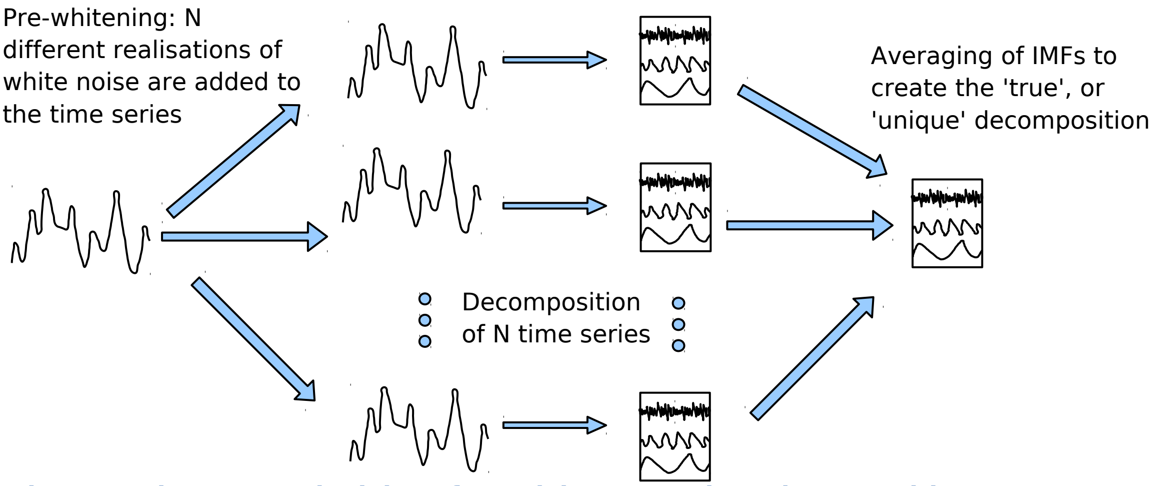

Figure 2: The new methodology for arriving at a unique decomposition of the time series, consisting of pre-whitening, decomposition and averaging.

For verification purposes, the important aspect of the updated methodology is that it may be possible to directly compare the IMFs from the forecast with the IMFs from the model. This relies on the assumption that since all frequencies should be represented in the pre-whitened time series, there should be no missing frequencies in the IMFs. In other words, if a feature appears in a particular IMF in the observations, then it should appear in exactly the same IMF in the forecast, allowing for direct comparison of IMFs. It may be possible to abandon the need to calculate instantaneous frequencies and amplitudes in the IMFs using the Hilbert transform, and instead directly compare IMFs. Whether or not it is in fact possible to directly compare the IMFs from the forecast with the IMFs from the model will be discussed further below.

Section 2: Application of the EEMD to the MSEPS warm run

The EEMD was applied to a 61 day segment for two ensemble members of the MSEPS warm run, which is described in Vincent et al. (2009). The 61 days corresponded to the months of February and March 2004. The forecast and observed time series are shown in figure 3. By eye, there are clearly periods of large amplitude wind variability in the observations that are not captured by the model. For example, the wind behaviour between days 50 and 60 is poorly forecast in terms of the wind variability.

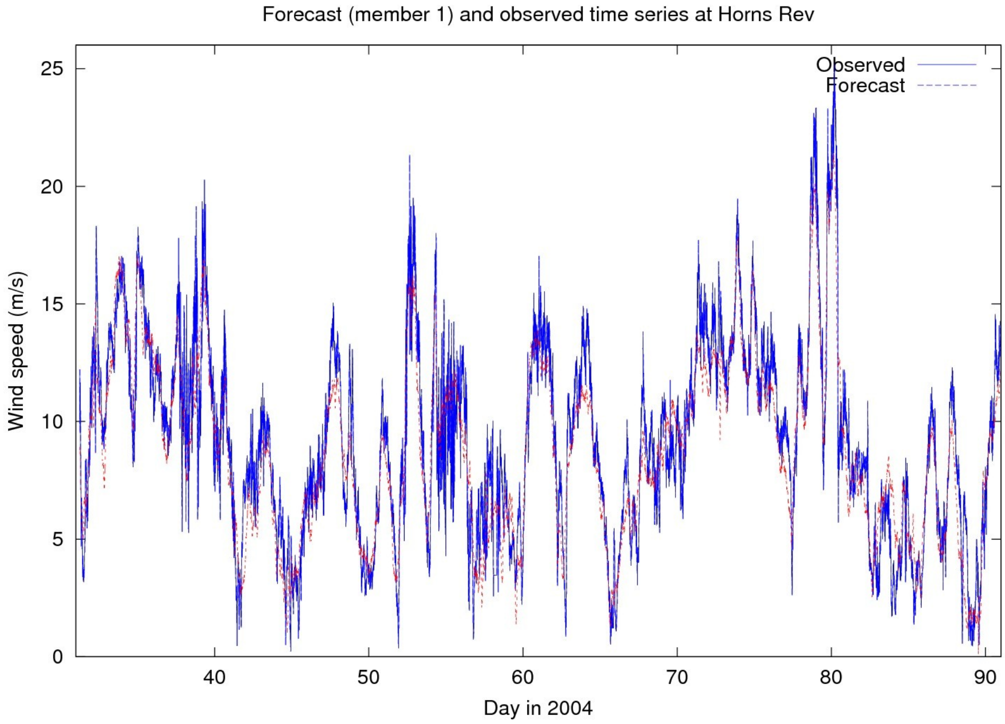

Figure 3: Observed (blue) and forecast (red) time series for the 61 day test period.

The forecast time series had a time step close to 6 minutes. They were interpolated to a resolution of exactly 6 minutes for consistency with observations which were also interpolated to 6 minutes. The two EEMD parameters that need to be chosen for such a decomposition are the ratio of the standard deviation of the added white noise to that of the original time series, and the number of pre-whitened time series that should be created to form the 'average' decomposition. In this case, the two parameters were chosen to be 0.15 and 40 respectively. These two parameter choices were somewhat arbitrary. The decomposition of the forecast and observed time series are given in figure 4.

Inspection of the decomposed time series (figure 4) shows that the technique is much more satisfactory than the methodology used in previous studies. The IMFs can now be directly compared, and it seems unnecessary to transfer them to frequency space via the Hilbert transform. It can be immediately seen that the model captures very little of the variance in the first four IMFs, and that for the last four IMFs the fluctuations in the model are nearly perfect. For IMFs 5 and 6, the model captures some of the variability. In particular, the large fluctuations on day 80 appear to be captured by the model. The right hand plot in figure 4 is zoomed in around the event on day 80.

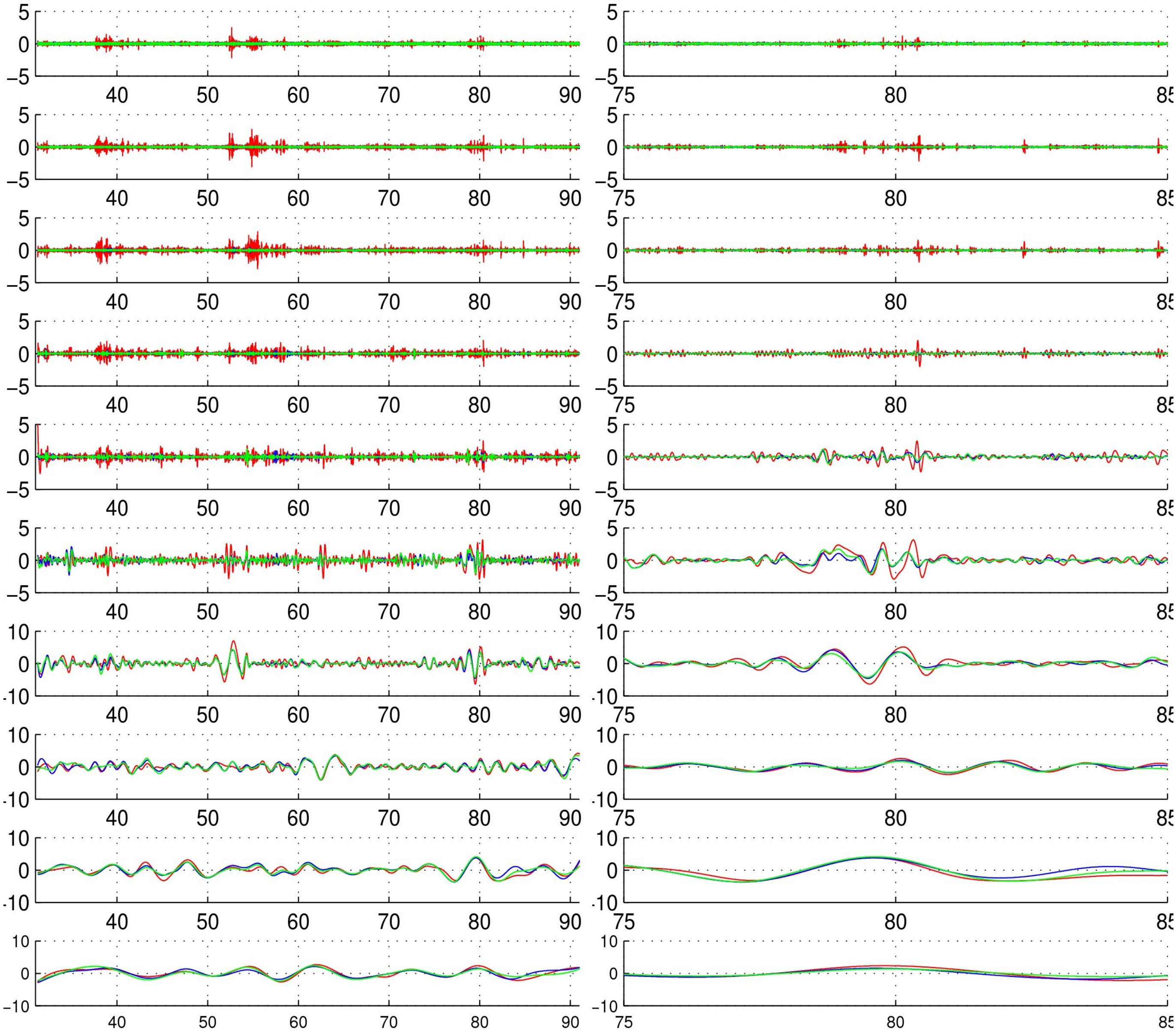

Figure 4: The first 10 IMFs for the observed time series (RED) and the first two MSEPS ensemble members (BLUE and GREEN). Left: The 61 day study period. Right: 10 days surrounding the variability event on day 80.

The average periods in each IMF for the observations and for the two ensemble members are tabulated in table 1. It is seen that they are nearly identical, which suggests that directly comparing the IMFs is indeed reasonable for at least the first 9 IMFs.

IMF number	Period of observed IMFs (hours)	Period of forecast IMFs – member 1 (hours)	Period of forecast IMFs – member 2 (hours)
1	0.15	0.14	0.14
2	0.33	0.29	0.29
3	0.68	0.61	0.60
4	1.4	1.3	1.3
5	3.1	2.9	2.9
6	7.1	7.0	7.2
7	18	18	19
8	38	40	42
9	86	81	92
10	133	183	183

Table 1: Average periods in the first 10 IMFs for the observed time series and the first two ensemble members.

The average periods of the forecast and observed IMFs are shown on a log-log axis in figure 5. It is seen that they agree very well for the first 9 IMFs. Further, it can be noted that the periods are approximately evenly spaced on a log-log axis, which agrees with the results of Wu and Huang (2004) and Flandrin et al. (2004).

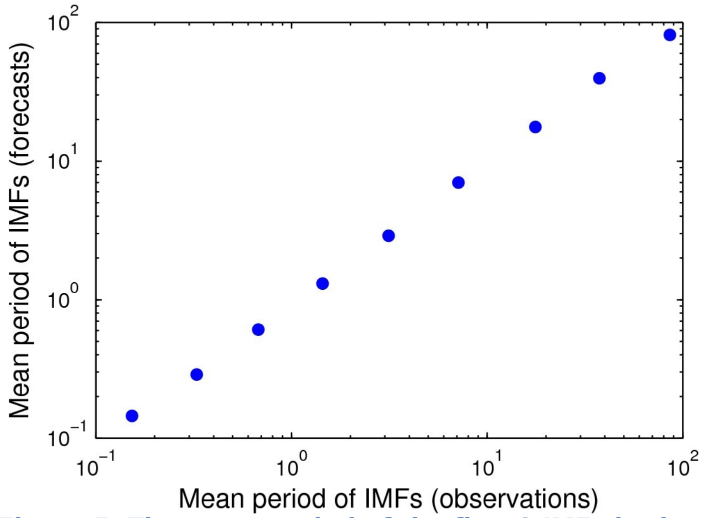

Figure 5: The mean period of the first 9 IMFs in the observations and in the MSEPS ensemble member 1, on a log-log axis. The periods of the IMFs are nearly identical, which suggests that the amplitudes can be directly compared.

Since the IMFs in the forecasts and the observations seem to be capturing nearly identical sets of frequencies, it is reasonable to use them as a direct verification tool. In figure 6, the average periods of the forecast and observed IMFs are shown. The average periods were calculated as the average of the absolute value of the maxima and minima of the IMFs. The plot shows that for the first 5 IMFs (time scales shorter than 3 hours), the amplitudes are dramatically suppressed. For IMFs 6 and 7 (time scales 7 and 18 hours), the amplitudes of the forecasts start to improve, and for IMFs 8 and 9 (time scales 40 and 80 hours), the forecast amplitudes are nearly correct.

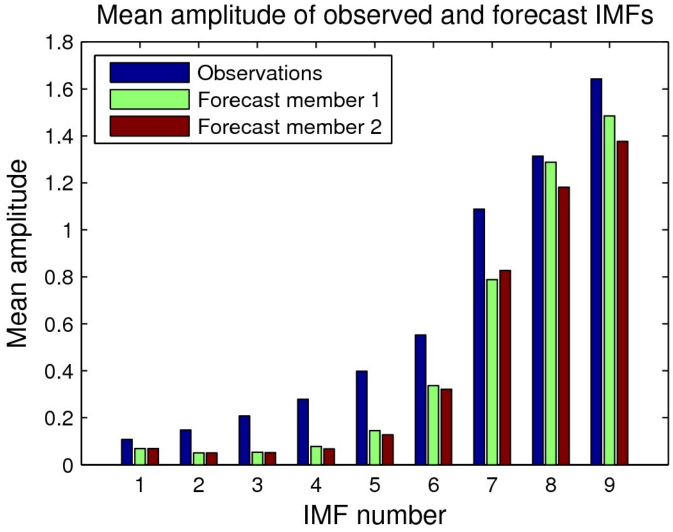

Figure 6: The average amplitude for each of the first 9 IMFs for the observations and the first two MSEPS ensemble members.

Section 3: Application of the EEMD to individual 24 hour forecasts

The analysis in section 2 related to a continuous 3 month period of the warm run experiment. This was ideal for EEMD analysis because it did not have any gaps in it and represented a wide range of scales. However, it is also of interest to analyse individual 24 hour forecasts, because in many cases this is how forecasts are actually provided to end users. The EEMD was applied to forecasts from once per day in a 20 day period in January 2004.

The main frequencies that appeared in each IMF were analysed by taking the Fourier transforms of the IMFs. The corresponding observations were analysed in the same manner, and compared with the forecasts for all 75 ensemble members. The days were validated individually, and the time scales that existed in the observed IMFs on each day were compared with the timescales reproduced by MSEPS. That is, what were the dominant time scales on each day, and how often were those dominant time scales represented in MSEPS?

To quantify the number of times that the correct dominant frequencies were represented in MSEPS, a score parameter was calculated for each day, for each member and for each time scale. The scores were then averaged for each time scale.

The forecasts and observations were decomposed in a two step procedure:

Step 1

The time series from both observation and simulations were decomposed into a set of intrinsic mode functions (IMFs) using the ensemble empirical mode decomposition technique. The first IMF extracts the component of the fastest oscillations from the time series, the second the second fastest oscillation and so forth. The updated EEMD methodology as described above was used in all the decompositions.

Step 2

The largest amplitude and the associated period for each IMF was extracted for each ensemble member and the observations for each of the 20 days in the study period using the fast Fourier transform.

Results and Discussion

The periods from the FFT analysis were sorted into time scale bins following Vincent et al. (2009), as given in table 2. It should be pointed out that timescale 6 is not well represented in the time series as only 1 single cycle is possible within the simulation period. The timescales 5 and 6 are included only for completeness and comparison with earlier work. In agreement with the results in section 2, it is seen that the short time scales are found as the dominant time scale in the model much less frequently that in the model. These results, however, should be treated with caution because of the shortness of the time series, and the fact that the fast Fourier transform will not necessarily identify the most important time scale in a non-stationary time series.

Time scale 1	24-180 minutes
Time scale 2	24-120 minutes
Time scale 3	60-120 minutes
Time scale 4	120-240 minutes
Time scale 5	240-320 minutes
Time Scale 6	180-600 minutes

Table 2: Time scales studied in the 24 hour forecasts

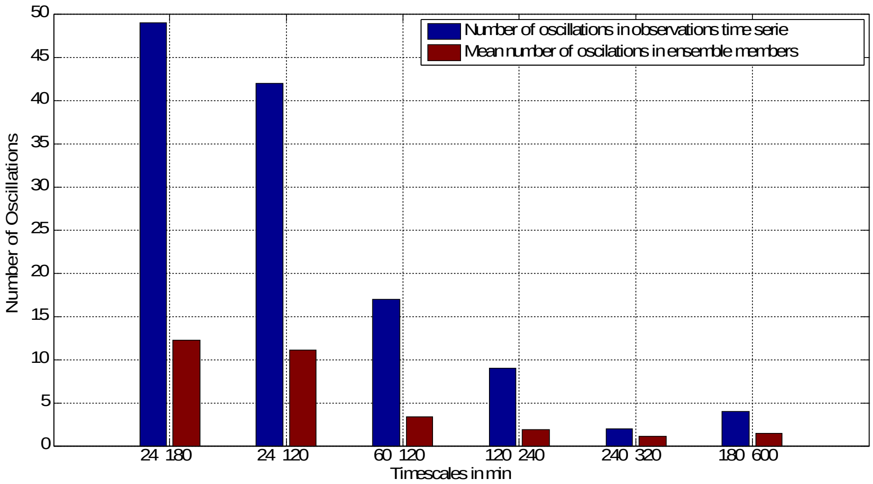

Figure 7: The number of times dominant frequencies with the time scale 1-6 occurred in the observations (blue bar) and the mean number of times the same scale showed up in the ensemble members (brown bar).

Figure 7 shows the number of times that a significant peak in the oscillations is observed in the observed and forecast IMFs. For example, on the time scale of 24-180 minutes, significant peaks in the spectra of the observed IMFs are seen in 48 times (i.e. these time scales appear 48 times in the combined IMFs of the 20 days), while significant peaks were only found, on average, in 12 of the 20 forecasts. Each forecast has 8 IMF and can therefore have more than one significant peak within a the same time interval. Figure 7 also shows that in many of the 24 hour forecasts periods, there were significant fluctuations in the observed time series that were not actually captured by the model. If the forecast IMFs did have a spectral peak in the right place, this analysis only shows that the fluctuations exist at some time within the 24 hour period; it does not assess the timing of the fluctuations. Although not shown here, some members of the MSEPs captured the shorter oscillations more frequent which supports the findings from Vincent et al (2009).

Validation scores.

Since it was shown in section 2 that the modelled and forecast IMFs show nearly identical periods, an attempt was made to quantify how well the 24 hour MSEPS forecasts performed within each time scale bin. The validation score was the different between the dominant periods in the observation and the forecast, scaled by the dominant period of the observation and is given in equation 1. This score shows how closely the period of the model matched that of the observations for the most important oscillations.

$$Error_{relative} = \frac{period_{observation} - period_{forecast}}{period_{observation}} \qquad \text{Equation (1)}$$

Each MSEPs member may have numerous hits within a time scale and the members relative errors are therefore averaged together. In the averaging procedure only relative error between -1 and 1 are consider as larger errors are consider to be artifacts of the procedure. The relative error for all the MSEPS members on time scale 1-4 are shown below in figure 8.

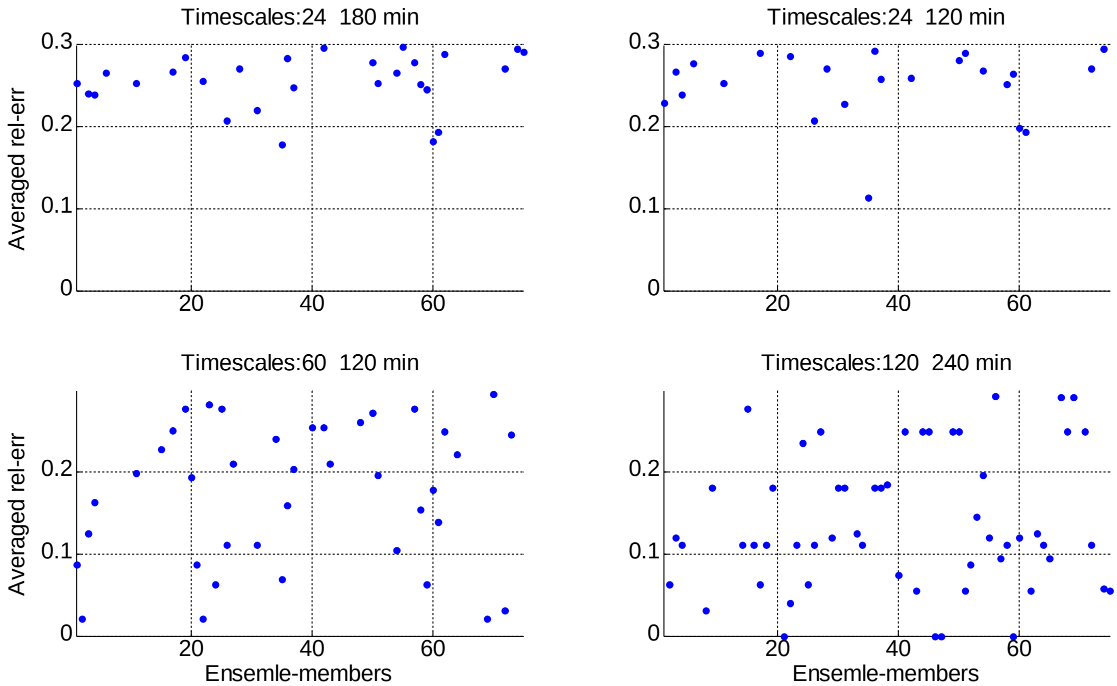

Figure 8: Average squared MSEPS errors for the 20 day period for each time scale bin

In figure 9, the errors shown in figure 8 are averaged for all ensemble members. It is seen that the relative errors decrease for time scales longer than 2 hours, which supprts the overall picture that emerges from the analysis in sections 2 and 3. That is, that below about 2-3 hours, the fluctuations in the model are mostly random fluctuations that do not have much bearing on the physical fluctuations observed in the atmosphere. For time scales longer than this, the timing and amplitude of the fluctuations both begin to show increasing skill. As before, care should be taken when interpreting the last time scale, since it is too long to be properly represented in a 24 hour forecast.

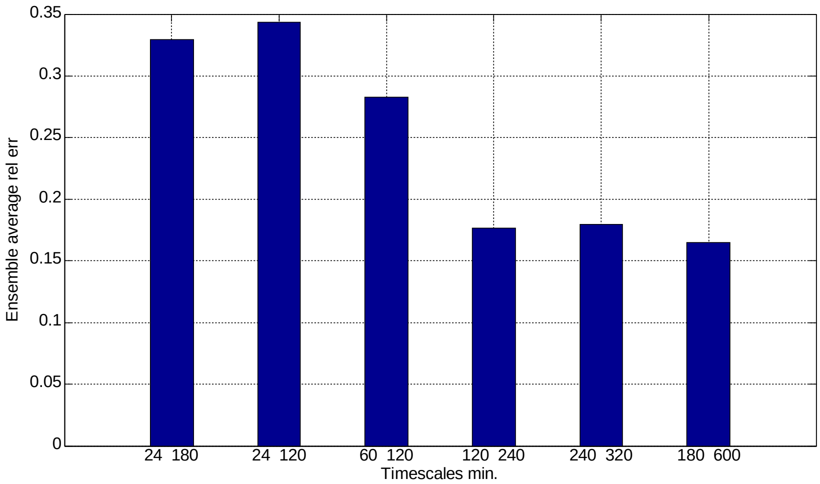

Figure 9: Average relative error (as defined in equation 1) for all MSEPS members in each time scale bin.

Section 4: Case study of a 24 hour forecast of a day with regular fluctuations in wind speed

We have already shown that the variance in the model is greatly suppressed for the time scales of less than about 3 hours, but we wish to investigate whether any realistic physical signals are captured by the model on these scales. To investigate this, a case study day was chosen when regular and significant fluctuations with a time scale of about 1 hour were observed in the meteorological mast observations from Horns Rev.

The comparison is done for time scales in the 1-2 hour range on a day that hosts interesting observed fluctuations on these time scales. The day is 11 of January 2004, where warm air is advected over a relatively cold North Sea. This may have generated a stable stratified boundary layer and the fluctuations could be associated with the stable stratification.

The upper panels in figure 10 show the observed time series for the selected day. It is seen that there are regular fluctuations in the wind speed with a period of around hour throughout this day. The first 10 members from the MSEPS ensemble are also shown. It is seen that the members capture only a very small proportion of the mesoscale variability. Therefore, it is not necessarily reasonable to be searching for mesoscale variability in a model of horizontal grid spacing 5 km. In the lower panels of figure 10, the observed time series is shown together with the third and fourth IMFs from the decomposed time series. These two IMFs contain the fluctuations with time scale close to 1 hour.

Even though the MSEPS ensemble doesn’t capture the fluctuations explicitly, there may be some dampened mesoscale variability signal that can be uncovered by decomposing the forecast time series, and this will be explored in the following sections.

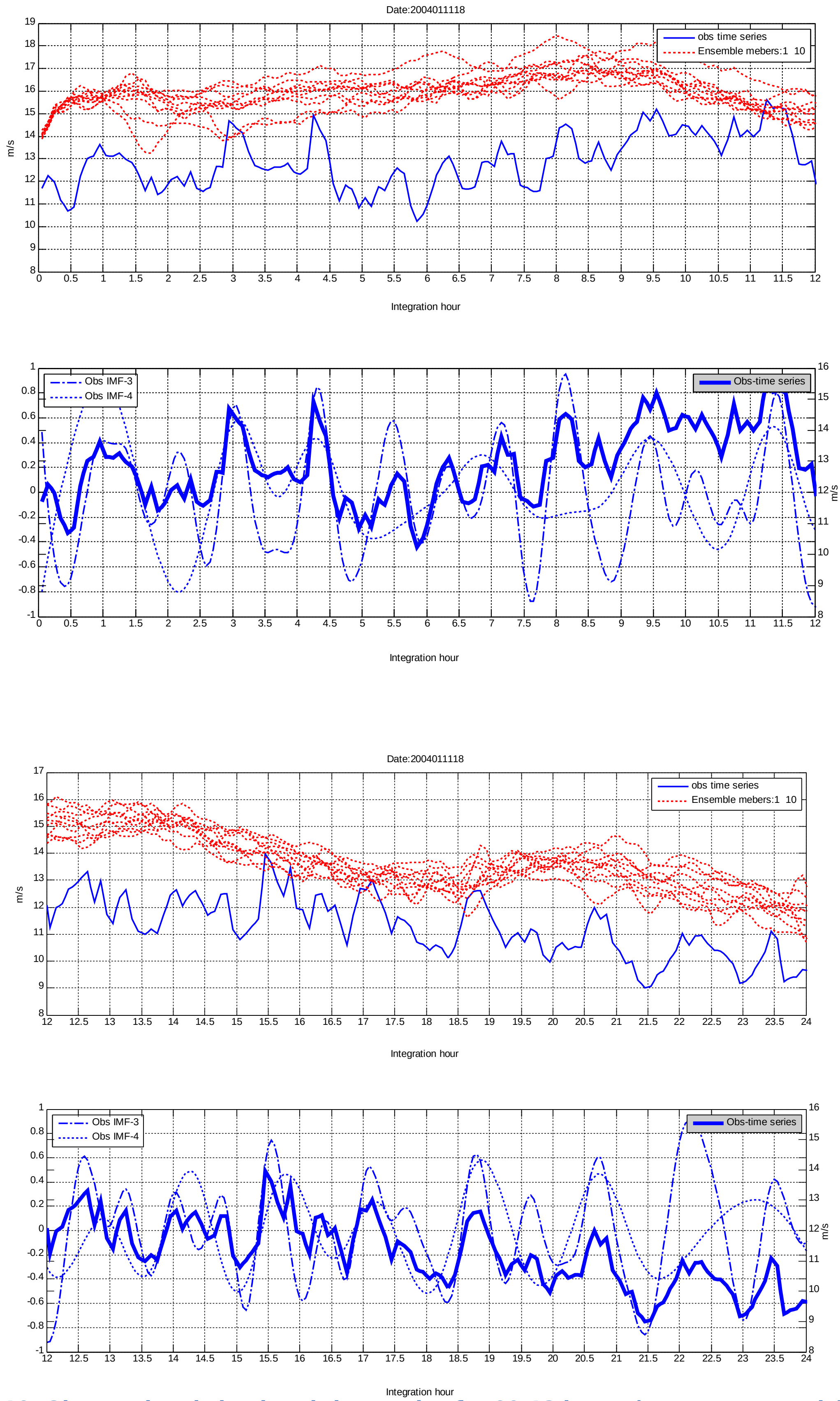

Figure 10: Observed and simulated time series for 00-12 hours (upper two panels) and 12-24 hours (lower two panels). For each 12 hour segment, the upper plot shows the observed time series and the first 10 ensemble members, while the lower plot shows the third and fourth IMFs from the observed time series, which contain most of the variability signal.

The observed IMFs were then compared with the simulated IMFs. The third and fourth IMFs from the observed time series are shown in figure 11, together with the third and fourth IMFs from two of the MSEPS ensemble members. The blue arrows indicate three distinct features in the observed IMFs – at hours 1-3 and 14-16 in IMF 4, and at hours 8-10 in IMF3. We can indeed see there are analogous features (albeit damped, note the different axis scaling for forecast and observed IMFs) in the forecast IMFs, but it may just be by chance. Overall, the variability in the MSEPS model is greatly damped compared with that in the observations.

If the features compared in this way are related to the same atmospheric behavior, then direct spectral comparison therefore is meaningful. In this case, the model performance can be considered ‘good’ when the feature is both in the observation and in the simulations occurring within a reasonable small time interval. However, the primary result from this analysis is that in this case, the model does not really capture fluctuations on time scales as short as 1-2 hours.

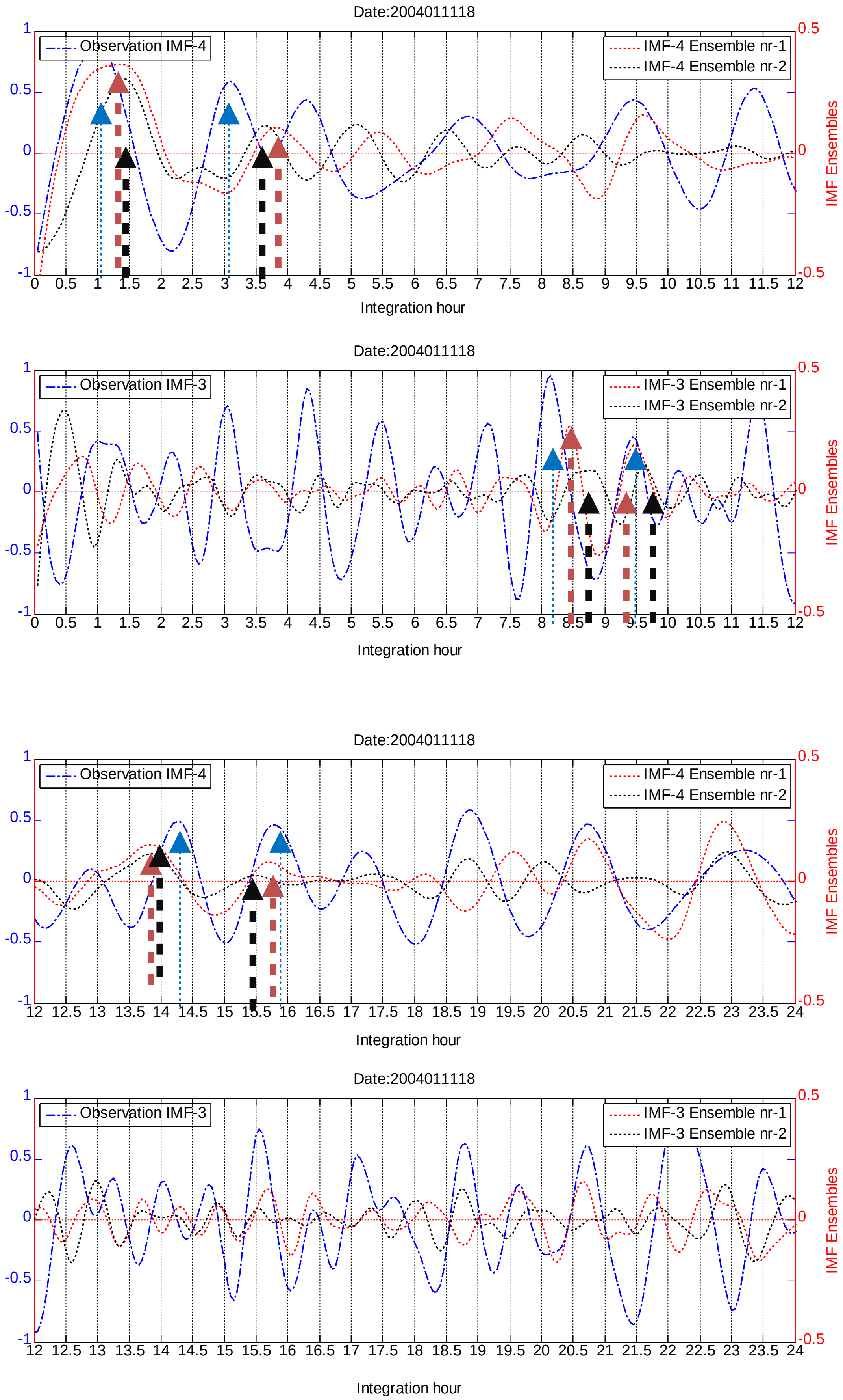

Figure 11: Forecast and observed IMFs 3 and 4

Conclusions and final remarks

The MSEPS ensemble starts to predict variance with realistic amplitudes for time scales longer than about 3 hours. The variance in the model for time scales shorter than 2-3 hours is greatly damped compared to what is seen in observations. This is not a failure or shortcoming of the model, but simply relates to the fact that a mesoscale model of resolution 5 km should not be expected to contain the same amount of variance as the real atmosphere on these scales. For longer timescales, the model appeared to capture the amplitude and phase of the significant fluctuations very well in the 61 day test period that was studied here.

The ensemble empirical mode decomposition was an interesting way to analyse the variance in the MSEPS model. In particular, analysis of the warm run experiments clearly differentiated the poorly forecast timescales from the well forecast time scales. The ensemble empirical mode decomposition, which was first reported in the literature during the past year, seems to be a much more promising model verification tool than the original empirical model decomposition and Hilbert transform.

The ensemble empirical mode decomposition was also applied to a 20 day period of individual 24 hour forecasts. We tried to quantify the periods that were well captured in these forecasts, and to provide a case study that illustrated the performance of the model. These experiments were partially successful, but it was difficult to study such short time series.

Spectral verification of mesoscale models is still a new and under-developed area of model validation. Although it was not possible for us to produce a finished or operational-style verification product, we believe that we have gained extra insight into the methodology for scale based verification. Future work on these areas involved further automation and quantification of the results presented here, such that robust verification statistics could be developed for scale based verification. Scale based verification remains an imperative aspect of mesoscale model development, because traditional verification scores cannot reward forecasts that show high skill in mesoscale variance.

References

Vincent, C., Draxl, C., Giebel, G., Pinson, P., Jørgensen, J., and Möhrlen, C.,. 2009. Spectral verification of a mesoscale ensemble. EWEC'09, European Wind Energy Conference, Marseille, France, 2009

Flandrin, P., Rilling, G., and Goncalves, P. 2004. Empirical mode decomposition as a filter bank. IEEE Signal Process. Lett., 112-114.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N-C., Tung, C.C., and Liu. H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Roy. Soc. A., 454:903-995

Wu, Z and Huang N. 2004. A study of the characteristics of white noise using the empirical mode decomposition method. Proc. Roy. Soc. A., 460, 1597-1611.

Wu, Z and Huang, N. 2009. Ensemble Empirical mode decomposition: A noise assisted data analysis method. Adv. Adaptive Data Anal. 1:1-41.