Wind Speed Modelling and Short-term Prediction using Wavelets

Katherine Hunt and Guy P Nason
Department of Mathematics, University Walk, Bristol. BS8 1TW
e-mail: makh@bristol.ac.uk and magpn@bristol.ac.uk

Abstract

The mathematical method of wavelets is explained and used to predict wind conditions using short-term data collected at a site and referred to long term data from, say, a meteorological station. We model the response time-series in terms of a multiscale wavelet decomposition of the explanatory time-series. Preliminary results of this method, using hourly 10 minute averaged data from six locations in the British Isles, allow comparison with a linear regression method in terms of prediction errors over 21 days.

1 Introduction

This paper notifies our initial work using the mathematical method of wavelets for the accurate prediction of the wind regime at a proposed wind farm site. Suppose a small amount of wind speed data has been collected at a site and that a nearby reference location, for example a meteorological station, provides long term data. We wish to model the data at the target site, known as the *response* time-series (y_t), in terms of the data at the reference location, known as the *explanatory* times series (x_t).

Predicting future wind speeds at a target site using data from a reference location is known as *measure-correlate-predict*. Linear regression is a popular industry method for constructing the statistical model which will predict future values of a response time-series, see for example Derrick (1992). As an alternative, we use *wavelet* methods which take a multiresolution approach. We model the response time-series in terms of a multiscale wavelet decomposition of the explanatory time-series.

We provide some preliminary results of this method using hourly 10 minute averaged data from six locations in the British Isles and show how our model compares with a linear regression method in terms of prediction errors over 21 days. We aim to provide a full analysis on a longer time-series at a later date. For a more detailed discussion of the methodology refer to Nason and Sapatinas (2001).

2 A brief introduction to wavelets

We provide only a brief introduction to wavelets here: a more detailed exposition can be found in Burrus and Guo (1998) for example.

Multiresolution analysis provides the framework for examining functions at different scales. In a multiresolution analysis a *father wavelet*, $\phi(x)$, is a function constructed to approximate general functions to a certain scale by using shifted copies of itself. The *mother wavelets*, $\psi(x)$, derived from the father wavelets, represent the difference between father wavelet approximations at two different scales.

A mother wavelet is a localised oscillating function from which a family of wavelets, $\psi_{j,k}(x)$, can be constructed by dilation and translation, i.e.

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \text{ for integers } j,\ k. \qquad (1)$$

The *dilation* parameter j controls the scale (or size) of the wavelet and the *translation* parameter k controls the location of the wavelet.

For suitable mother wavelets, $\psi(x)$, the set $\{\psi_{j,k}(x)\}_{j,k}$ provides a basis that can be used to approximate functions, i.e.

$$f(x) = \sum_j \sum_k d_{j,k}\psi_{j,k}(x), \text{ for integers } j,\ k, \qquad (2)$$

where $d_{j,k}$ are the *wavelet coefficients*.

One of the reasons wavelets have been successful in fields such as image compression is their ability to efficiently represent all manner of complicated signals. Wavelets are particularly effective at representing signals with discontinuities due to their excellent localisation properties.

For some time-series *wavelet packets* may be of more use as they provide a wider choice of decompositions of the frequency domain. A wavelet packet is a particular linear combination of wavelets that retains many of the orthogonality, smoothness and localisation properties of wavelets (Wickerhauser, (1994)).

Using wavelets (or wavelet packets) in our model allows us to attach a physical interpretation to the model. For example, a wavelet packet is given in Figure 1 which could be used to represent daily variation in wind speeds over the past 2.75 days. The wavelet packet provides valuable information about which components in the explanatory time series drive the response time series, i.e. which types of oscillatory behaviour in x_t influence y_t.

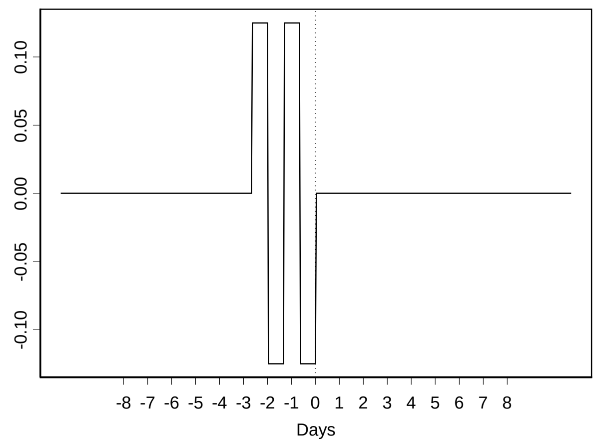

Figure 1: Haar wavelet packet at dilation scale level 3 and translation index 2, which shows approximate daily oscillation over the last 2.75 days.

3 A wavelet measure-correlate-predict model

To represent the explanatory time-series in terms of wavelet packets we compute a time ordered non-decimated wavelet packet transform in *S-Plus* (Mathsoft, Seattle, U.S.) using the freeware *WaveThresh* package (Nason, 1993-9). This calculates wavelet packet coefficients for the j scales and k locations in the analysis. These coefficients are stacked together to form a multivariate time-series matrix. Each variable in the multivariate time-series matrix corresponds to a particular wavelet packet and quantifies how similar the time-series is to the wavelet packet at each time point.

Principal Components Analysis (PCA) is performed on the multivariate time-series matrix because of some high correlations between the wavelet packets. The resulting linear combinations of wavelet packets, the *principal components* (pc), are uncorrelated and are such that a few usually will explain most of the variation in the explanatory time-series.

We model the response time-series in terms of the principal components (pc). We assume that the residuals of the fitted model follow the normal distribution and this is reasonable in each example we have seen. We use the routines available in *S-Plus* for fitting a standard multivariate linear regression.

4 Case studies

We illustrate our model with the following three examples. See Section 5 for the prediction results. For a further example refer to Nason and Sapatinas (2001).

4.1 St Bees Head and Valley

For our first example we use data from St Bees Head meteorological station as the response time-series and data from Valley meteorological station as the explanatory time-series. St Bees Head is a coastal location in Cumbria (Northwest England) and Valley is a coastal location in Gwynedd (North Wales). The sites are approximately 136 km (85 miles) apart on a Northeast-Southwest alignment and the correlation between the time-series is 0.658. A time-series plot of the data is given in Figure 2.

We perform the modelling procedure described in Section 3 to fit the model;

$$\text{St Bees Head} \sim \text{pc } 1 + \text{pc } 5 + \text{direction factor.} \qquad (3)$$

We have chosen to include the first principal component (pc 1) and the fifth principal component (pc 5) in the model. Remember the principal components represent linear combinations of non-decimated wavelet packet coefficients - so (3) expresses a statistical relationship between the response time series, St Bees Head, to the coefficients through the principal components. The direction factor is added to permit the statistical model to take account of the variability in wind direction. The factor itself simply records the direction bin of the wind: there are 12 bins corresponding to twelve 30° direction sectors. To help us choose which components to include in the model we used a stepwise analysis to determine which components are most useful for the prediction of wind speeds at the target site. We adjust the estimate of the wind speed at the target site by a given

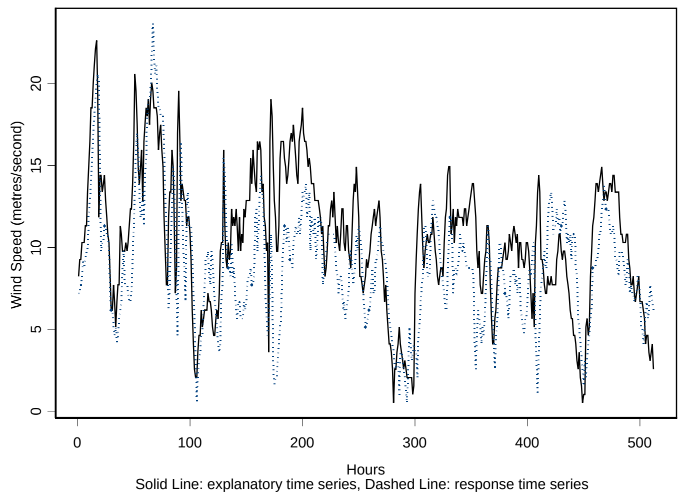

Figure 2: Time-series plot of the St Bees Head and Valley data.

direction factor. This is simply a constant that is determined by the direction of the wind at the reference location.

The first principal component is a mixture of father wavelets and wavelet packets. Father wavelets average wind speeds over a given time determined by the dilation level and translation index of the father wavelet. Wavelet packets capture a wide variety of high and low frequency oscillations in the data, the exact frequency determined by the dilation and translation index of the packet. The second principal component is made up of mother wavelets. Mother wavelets capture up to one oscillation over a given time determined by the dilation and translation index of the mother wavelet. The direction factor is significant when the wind is from a South or Northeast direction. The Southern direction factor corresponds to prevailing winds and the Northeast direction factor corresponds to the alignment of the sites.

In Figure 3 we have plotted a graphical representation of the principal components. As an example, we suggest an approximate physical interpretation for principal component 5. In the first two hours there is a negative relationship between the two sites. As the sites are 136 km apart, unless the winds were extremely strong, we would not initially expect to see the same behaviour at the two sites. After two hours there is a gradual build up of the positive relationship between the two sites as wind speeds from the reference location reach the target site. This continues up until 16 hours, by which time the weather system from Valley will have passed over St Bees Head.

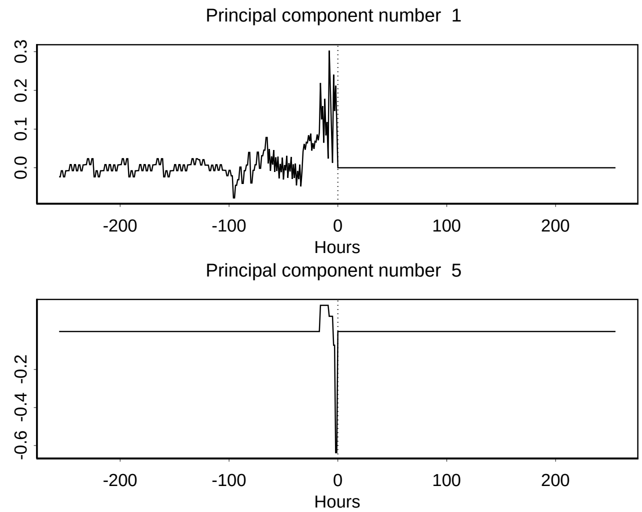

Figure 3: Principal components from the St Bees Head and Valley model.

4.2 Milford Haven and Aberdaron

For our second example we use data from Milford Haven meteorological station as the response time-series and data from Aberdaron meteorological station as the explanatory time-series. Milford Haven is on the Southwest coast of Wales and Aberdaron is on the Northwest coast of Wales. Aberdaron is approximately 136 km (85 miles) to the North of Milford Haven and the correlation between the time-series is 0.762. A time-series plot is given in Figure 4.

We explored various model relationships and the following model was found to fit well;

$$\text{Milford Haven} \sim \text{pc } 10 + \text{pc } 41 + \text{direction factor}.$$

The forty-first component is similar to the first component picked out in the St Bees Head and Valley example which represents a mixture of father wavelets and wavelet packets. The tenth principal component is a mixture of wavelet packets. These components are picking up more high frequency characteristics in the data than the components in the previous model. The direction factor is significant when the wind direction is from the Northwest, which we can again relate to site alignment, or from the Southwest (the direction of prevailing winds).

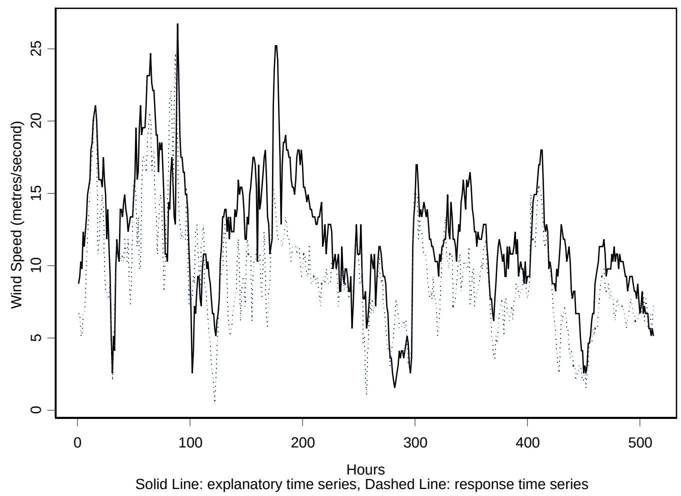

Figure 4: Time-series plot of the Milford Haven and Aberdaron data.

4.3 Walney Island and Ronaldsway

For our third example we take Walney Island on the Northwest coast of England as the response time-series and Ronaldsway on the Southwest coast of the Isle of Man as the explanatory time-series. The two meteorological stations are approximately 64 km (40 miles) apart on an East-West alignment. The correlation between the two time-series is 0.697 and a time-series plot is given in Figure 5.

For this data we fit the model;

$$\text{Walney Island} \sim \text{pc } 1 + \text{pc } 4 + \text{direction factor}.$$

The principal components are similar to those illustrated in the first example. They pick out similar wavelets and wavelet packets that represent the major characteristics of the wind speed data but with some variation that may be specific to local conditions. The direction factor is significant for winds from a Southern to a Northwest direction encompassing both prevailing wind and site alignment direction.

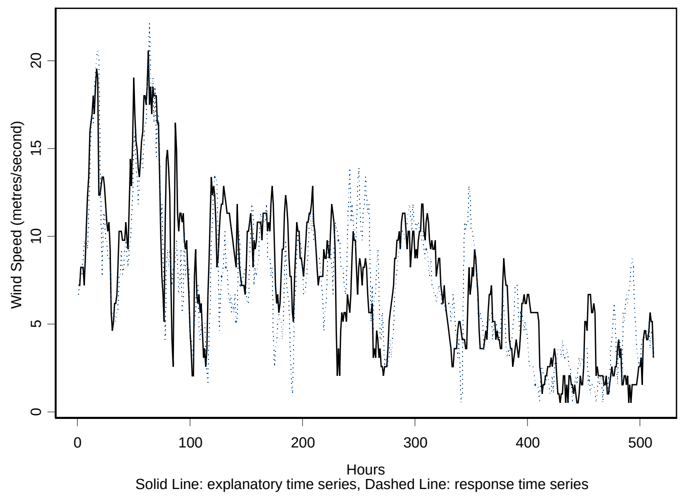

Figure 5: Time-series plot of the Walney Island and Ronaldsway data.

5 Predicting future wind speeds

We measure the accuracy of our method by looking at the mean residual sum of squares (MRSS) error of the predictions generated by our model. The models were fitted using a training dataset of 512 observations and a consecutive data run on 512 observations was used for the prediction. To evaluate our model we compare our predictions with those generated by the simple linear regression model.

It can be seen from Table 1 that the wavelets perform significantly better over the early prediction period. The results are closer for the longer prediction period as only a small stretch of data is used to formulate the models. We believe these preliminary results suggest our wavelet method is worth investigating as an alternative method for the short-term prediction of wind speeds.

6 Conclusions and future work

We have proposed an alternative method for the prediction of wind speeds at a target site using wind speeds from a reference location. The preliminary results on relatively short time-series have been encouraging. Wavelet methods have been shown to provide more reliable estimates than a prediction method using

Location	Forecast Time Period	Wavelets	Linear Regression
St Bees Head	5 days	6.20	8.08
and	10 days	5.40	6.89
Valley	21 days	5.46	6.48
Milford Haven	5 days	4.80	16.69
and	10 days	4.36	12.74
Aberdaron	21 days	6.95	9.09
Walney Island	5 days	3.29	7.00
and	10 days	3.59	6.57
Ronaldsway	21 days	4.39	5.51

Table 1: Mean Residual Sum of Squares (MRSS) of the predictions.

simple linear regression over 21 days. Our wavelet models also have the added bonus of often being physically interpretable.

We appreciate that we have to continue the work for much longer time periods than 21 days. However the computational complexity of fitting the model with, for example, one year's data is high, so we wish to make our early results known at this stage. We intend to provide a full analysis on a longer time-series in the future.

We are grateful to the British Atmospheric Data Centre (BADC) which provided us with access to the Met Office Land Surface Observation Stations Data. Hunt gratefully acknowledges the support of an EPSRC studentship and Nason is supported by EPSRC grant GR/M10229 and Advanced Research Fellowship AF1664.

The software used in the note for the wavelet analysis may be obtained from *http://www.stats.bris.ac.uk/∼wavethresh/*

References

Burrus, C. S., R. A. Gopinath, & Guo, H. 1998. *Introduction to wavelets and wavelet transforms: A primer.* Upper Saddle River, NJ: Prentice Hall.

Derrick, A. 1992. The development of the measure-correlate-predict strategy for site assessment. *Proc. 14th British wind energy association conference*, 259–265.

Nason, G. P., & Sapatinas, T. 2001. Wavelet packet transfer function modelling of nonstationary time series. *Statistics and Computing*, (to appear).

Wickerhauser, M. V. 1994. *Adapted wavelet analysis from theory to software.* Wellesley, Massachusetts: A K Peters.