On the Estimation of the Distribution of Power Generated at a Wind Farm using Forecast Data

Siobhan Devin, University College Cork. Conor Houghton, Trinity College Dublin.
David Ramsey, University of Limerick. Eliza Loza Reyes, University of Bath.
Conor Sweeney, University College Dublin. Eddie Wilson, University of Bristol.

Report from the 62nd European Study with Industry,
University of Limerick, Jan. 2008

1 Introduction

New regulations on the supply of electrical energy in Ireland have introduced a system of suppliers bidding proposals for the cost of supply for each 30 minute period of the following day. In order to effectively bid and hedge against adverse events, it is necessary to assess the uncertainty of supply from wind energy farms. Hence, the aim of this study group was to estimate the distribution of power supply from a given set of wind farms a day in advance given forecast data.

2 Background

The amount of power generated at a wind farm depends on several factors most importantly the turbines used and the speed of the wind. The relationship between the speed of the wind and the amount of power is not a straightforward one. At low wind speeds the turbines will not rotate and at high wind speeds the turbines will disengage, in order to protect against damage from extreme weather conditions. Turbines are designed to work most efficiently when the wind speed is around its median speed and the power produced increases most rapidly with increasing wind speed in this region.

The speed of the wind depends, among other things, on the local terrain and vegetation and the height of the turbines. Due to the large number of factors affecting the power generated at each site, we decided to first develop a simple model that would estimate the distribution of the power supplied at regional or national level directly using historical data regarding windspeed forecasts and actual power output.

Haslett and Raftery (1989) by looking at the wind speeds recorded at Irish meteorological stations noted that the coefficient of correlation between the wind speed at two such sites in Ireland falls exponentially with the distance between the two sites. They estimate that the coefficient of correlation between the average daily wind speeds observed at two sites that are 100km apart is almost 0.9.

Windfarms are specifically placed in order to ensure a high median wind speed (on plateaus and in coastal areas). Although, this is not the case for meteorological stations, in order to be representative the wind speed must also be measured in an area that is unsheltered from the wind. These facts ensure that there will be a large correlation between the wind speed at a meteorological station and the wind speed at the wind farms in that region. Hence, (forecasted) wind speed at the meteorological station will be a predictor of the power output from the wind farms in that region.

Another feature of the data Haslett and Raftery highlight is what they term long-memory dependence. This is to say that there are significant correlations between wind speeds measured at time t_0 and time t_0+t_1 for a given site, even for relatively large t_1. This means that in order to get any reasonable estimate

of the wind speed at a site one or more days into the future using observed wind speeds, a reasonably complex time series model is required.

It was felt that the reason for this lies in the fact that the evolution of the weather occurs on a much larger than national scale and any predictions should be based on such a model. Since weather forecasts are based on such models, we proposed that the estimation of the distribution of the future power output should be based on weather forecasts rather than observed data. Such forecasts may be understood as a concise way of describing the recent evolution of weather patterns and this allows us to greatly simplify the model.

There are other seasonal factors that should be taken into account. Firstly, wind speed depends strongly on the time of day. The highest average speeds are observed at midday and the lowest average speeds are observed early in the morning. Also, wind speeds are on average 25% lower in summer than in winter. Hence, our estimation of the conditional distribution of power output given our explanatory variables could be improved by taking the time of day and the time of year into account.

3 The Simplest Model Proposed

The basic model aims to estimate the distribution of the total power output of any set of wind farms in a given region based on the forecast for wind speed at the local meteological station and historical data regarding actual power output. Assume that we wish to estimate the distribution of the power output k days in the future. Suppose we have observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where x_i is the forecast of the wind speed on day i at the meteorological station (made k days in advance) and y_i is the actually observed power output from the wind farms on day i. Ideally, these forecasts should be made for the same time of day and at similar times of the year (say in the same month).

Suppose $f_{X,Y}(x, y)$ is the (real, but unknown) joint density function for the forecast wind speed at the regional station X and the actual power output Y. Let $f_X(x)$ be the density function of the forecast speed. The conditional density of the power output given the forecast speed is given by

$$f_{Y|X}(y|x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)}, & f_X(x) > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Our goal is to estimate $f_{Y|X}(y|x)$ using kernel density estimation. Our estimates of the joint density function and this conditional density function will be denoted by $\hat{f}_{X,Y}(x, y)$ and $\hat{f}_{Y|X}(y|x)$, respectively.

3.1 Kernel Density Estimation for the Distribution of a Single Random Variable

The idea of kernel density estimation can be thought of as a method of producing a smoothed histogram which estimates the density function of a random variable (or of several random variables). Firstly, we describe the intuition behind using kernel density estimation to estimate the density function of a single random variable.

Suppose we have observations $x_1, x_2, \ldots, x_n$ of a random variable X (say wind speed). X will be placed on the horizontal axis of the graph and the estimate of the density function will be placed on the vertical axis. We place a curve around each data point x_i, such that each curve attains its highest point at the corresponding value of x_i, is symmetric about x_i and the area under each curve is $\frac{1}{n}$. The estimate of the density function of X is obtained by summing these curves. This ensures that the total area under the estimate of the density function is 1. This is illustrated in Figure 1.

It should be noted that the most important factor in the choice of the kernel is what is termed the bandwidth, which may be thought of as the degree of smoothing used. The larger the bandwidth the more the density estimate is smoothed. The form of the kernel corresponds to some density function and

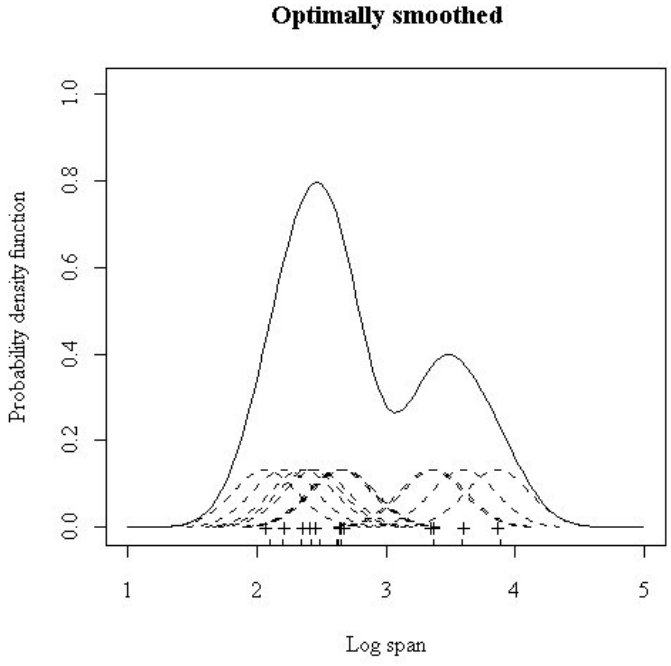

Figure 1: Use of Kernel Density Estimation to Estimate the Density Function of a Random Variable

the bandwidth may be thought of as the standard deviation of the density function used to define the kernel. In the diagram above, the kernel is based on the normal distribution (is bell-shaped). The form of the distribution used to define the kernel is not of great importance for samples of reasonable size, but the bandwith (degree of smoothing used) is of crucial importance. The use of too large a bandwidth leads to the estimate of the density function being over-smoothed, i.e. the density function will tend to be less concentrated than it should be. The use of too small a bandwidth leads to there being too many peaks in the estimate of the density function (undersmoothing). This is illustrated in Figures 2 and 3.

Statistical packages implementing kernel density estimation use methods to choose a near optimal bandwidth. The optimal bandwidth minimises the expected square error (mean integrated square error - MISE) from estimating the density function f (which is of course unknown). In order to do this, an initial bandwidth (based on a distribution whose standard deviation is a fraction of the standard deviation observed in the sample) is chosen to estimate the density function. This estimate is then used to estimate the MISE and choose a better bandwidth.

3.2 Kernel Density Estimation for the Joint Distribution of Two Random Variables

The idea of kernel estimation for joint distributions is similar. Suppose we wish to estimate the joint distribution of the variables X and Y. The graph describing the estimate of the joint distribution will be three-dimensional with the variables X and Y corresponding to the horizontal axes and the joint density corresponds to the vertical axis. In simplistic terms, a pair of observations (x, y) is more

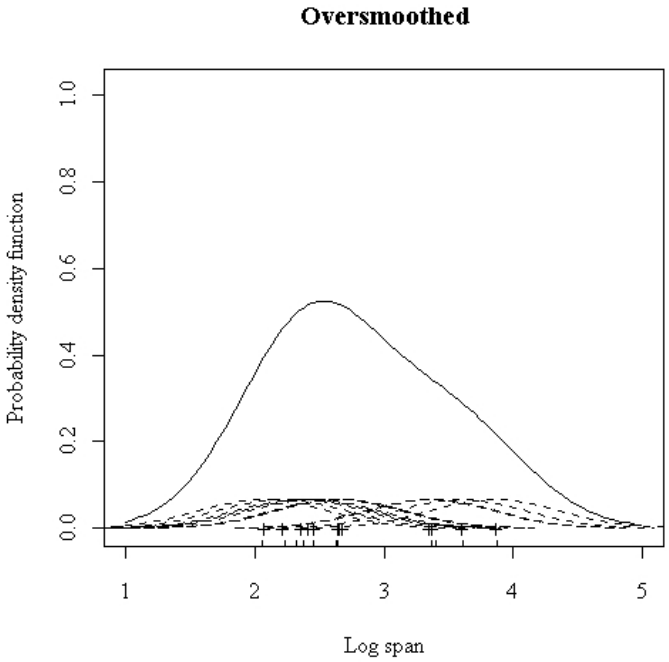

Figure 2: Oversmoothing the Estimate of the Density Function of a Random Variable

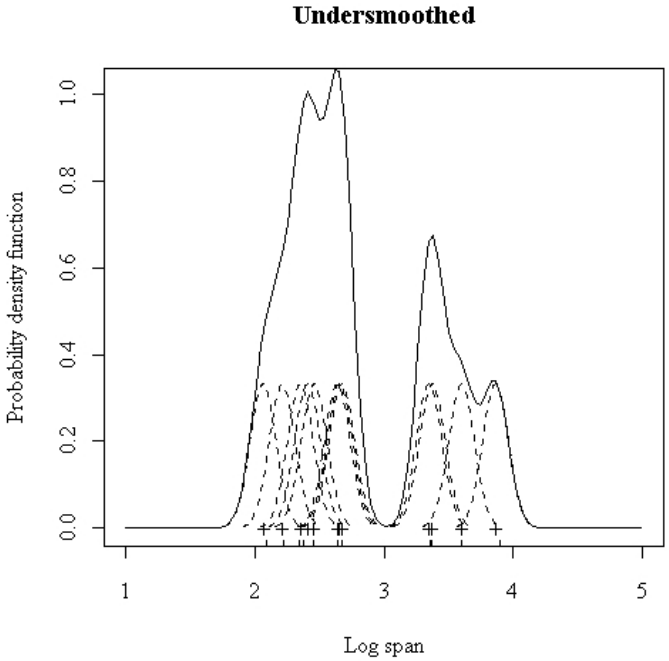

Figure 3: Undersmoothing the Estimate of the Density Function of a Random Variable

likely the greater the value of the density function at (x, y). Suppose we have n pairs of observations $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$. In the context of this report, x_i might denote the forecast for the windspeed for day i and y_i the observed power output of a wind farm at that time. A curved surface is drawn above each point (x_i, y_i), such that the volume under the surface is $\frac{1}{n}$, the maximum height of the surface is attained at (x_i, y_i) and the surface is symmetric around (x_i, y_i). The estimate of the joint density function is the sum of the heights of these surfaces. Hence, the volume under the estimator of the joint density function is 1. The method used to choose a near optimal bandwidth for the kernel used is similar to the one used in the estimation of the distribution of a single random variable. It should be noted that the optimal kernel for estimating a joint density function reflects the correlation between the two variables.

3.3 Estimation of the Conditional Density Function of Variable Y given the Value of Variable X

The conditional density function of the variable Y given the value of the variable X is given by $f_{Y|X}(y|x)$, where

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

It should be noted that if $f_X(x) = 0$, then $f_{Y|X}(y|x)$ is defined to be 0. Given observations $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$, we estimated the conditional density function using

$$\hat{f}_{Y|X}(y|x) = \frac{\hat{f}_{X,Y}(x, y)}{\hat{f}_X(x)},$$

where $\hat{f}_X(x)$ is the kernel density estimate of the density function of X. It should be noted that this conditional density may be estimated directly without having to separately estimate $f_X(x)$ by using the following estimator

$$\hat{f}_X(x) = \int_{-\infty}^{\infty} \hat{f}_{X,Y}(x, y)dy.$$

4 Initial Results

It was impossible for us to collect appropriate data regarding weather forecasts and the power output of wind farms during the few days in which the study group worked. Hence, it was decided that data on observed wind speeds would be used to illustrate the ideas behind the method. The wind speed at a given time on a given day was used to forecast the wind speed at that time on the next day. The data used were the wind times observed at 6am, midday, 6pm and midnight at Cork airport on each day over a period starting from Jan. 1st, 2004 (in total just over 3 years of data).

Initially, the analysis was carried out on the data as a whole (i.e. we did not take into account the dependency of the wind speed on the time of day or season). Secondly, we split the data into four groups according to the time of day at which observations were made (i.e. we took into account the fact that wind speed depends on the time of day, but not the dependency upon season). It was decided not to split the data up according to season (e.g. according to month), since the long-term dependency apparent in weather patterns would mean that the amount of data available in each subsample would be rather small for our purposes.

The data were analysed using code written by Duong (2007) for the R package to carry out kernel density estimation of the joint distribution of several random variables.

4.1 Results for the Data Taken as a Whole

The top graph in Figure 4 shows the contour plot for the estimated joint density of wind speeds on successive days. The orientation of the distribution is clearly from bottom left to top right, which shows

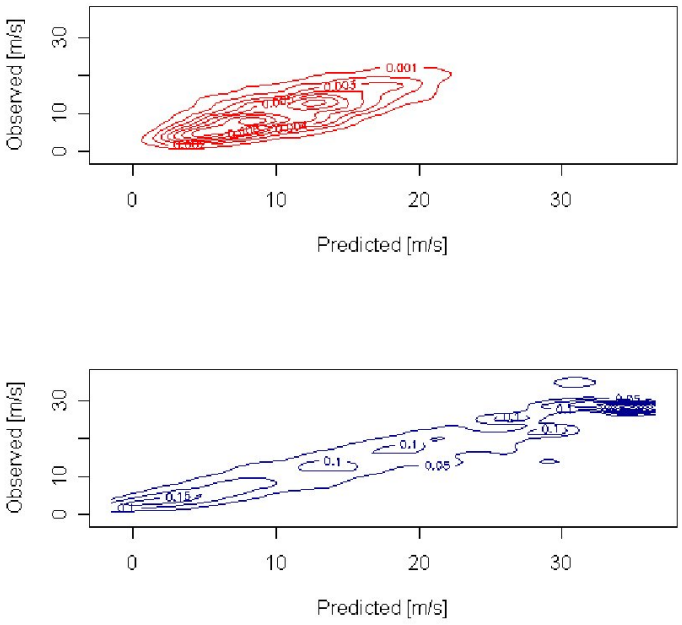

Figure 4: Estimation of the joint density of predicted and observed wind speed

that today's wind speed is a predictor of tomorrow's windspeed. The maximum point of the density function is attained at around 8m/s.

The lower graph in Figure 4 is a cross-sectional view of the estimated conditional density of today's wind speed given yesterday's wind speed. In order to interpret the graph, we inspect vertical cross-sections of this graph. For example, taking yesterday's wind speed (the prediction of today's) windspeed to be 5m/s, the conditional density of today's wind speed increases to its maximum at around 5m/s and then decreases. As yesterday's (the prediction of) windspeed increases, the point at which the maximum of the conditional density of today's wind speed also increases. This confirms the clear positive correlation between today's and yesterday's wind speed. It should also be noted that this relation becomes less clear at extreme wind speeds (above 20m/s). This is due to the fact that since there are less observations of such wind speeds, estimation of the conditional distribution of today's wind speed given yesterday's wind speed becomes much less accurate. One possible solution of this problem would be to group extreme wind speeds together. This problem will be addressed in context of the model proposed in the conclusion.

Figures 5, 6 and 7 give the estimated conditional density of today's windspeed given yesterday's wind speed was 10m/s (a commonly observed wind speed) for a) all 4 times of the day, b) 6am and c) midday, respectively. These conditional density functions are, as we would expect, slightly right-skewed and have maximum points around 9-10m/s. The time of the day does not have a big influence of this distribution, but it can be seen that there is a longer tail for the conditional distribution for the wind speed at midday than for the wind speed at 6am. This indicates that when only a small amount of data is available, it would be reasonable to group different times of the day together, in order to estimate wind speed given a prediction (given the prediction was reasonably common). However, taking account of the time of day would enable more accurate estimation of the conditional distributions when a large amount of data was available.

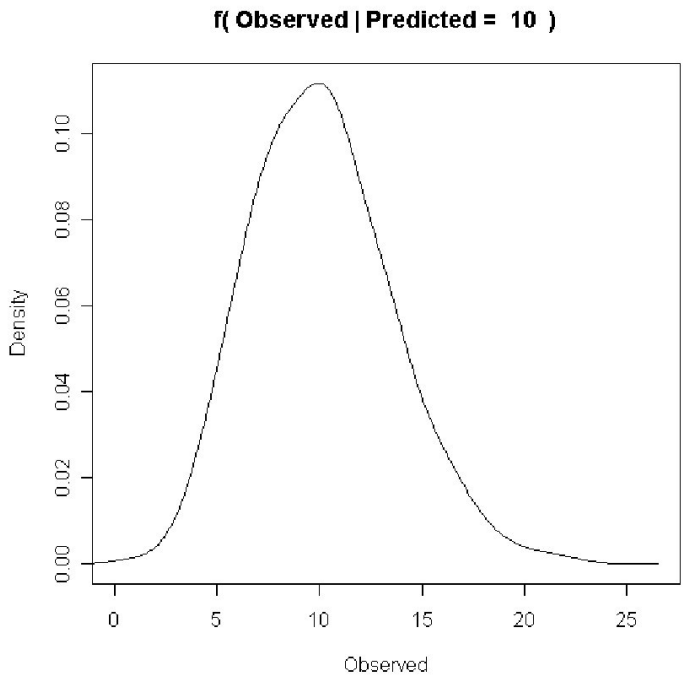

Figure 5: Conditional distribution of today's wind speed given the predicted windspeed was 10m/s (all times of day)

Figure 8 shows the estimate of the conditional distribution of today's wind speed given yesterday's wind speed was 1 m/s. This illustrates that the estimation of the conditional distribution of today's wind speed is inaccurate for extreme predictions, since there are very few such predictions. In this case when limited data is available, then extreme predictions should be appropriately grouped together. This will be considered in the conclusion.

4.2 Interpretation of Results

There is a visible positive correlation between wind speeds observed on successive days. This can be seen clearly from the graph of the contours of the conditional distributions for the observed wind speed given the prediction (yesterday's wind speed). The coefficient of correlation between the predictions and the observations is 0.7228.

The model we propose predicts power output on the basis of the actual forecast from the previous day. This forecast will be a better predictor of tomorrow's wind speed at a weather station than the previous day's wind speed. However, the wind speed at the weather station is not a perfect predictor of the power output of local wind farms. Thus, it is not clear whether the correlation between forecast wind speed and power output should be less than or greater than the correlation between wind speeds at the weather station on successive days (estimated to be 0.7228).

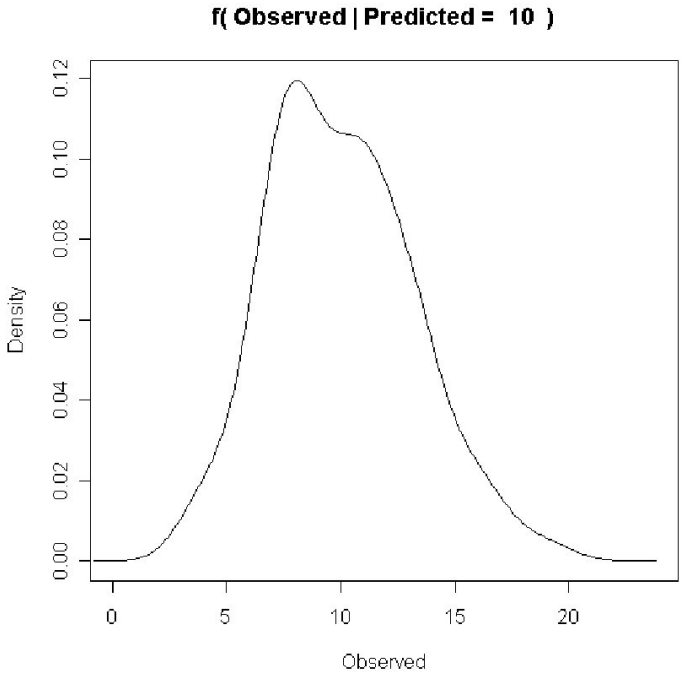

Figure 6: Conditional distribution of today's wind speed given the predicted windspeed was 10m/s (6am)

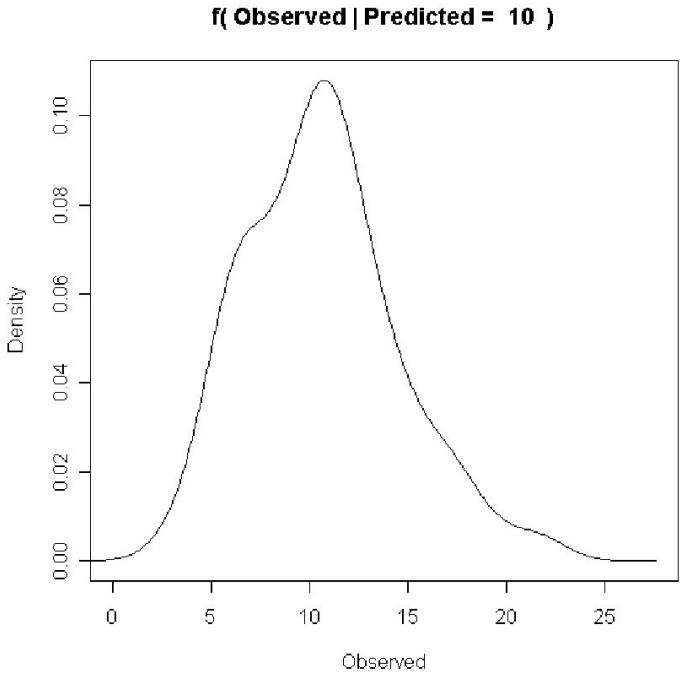

Figure 7: Conditional distribution of today's wind speed given the predicted windspeed was 10m/s (midday)

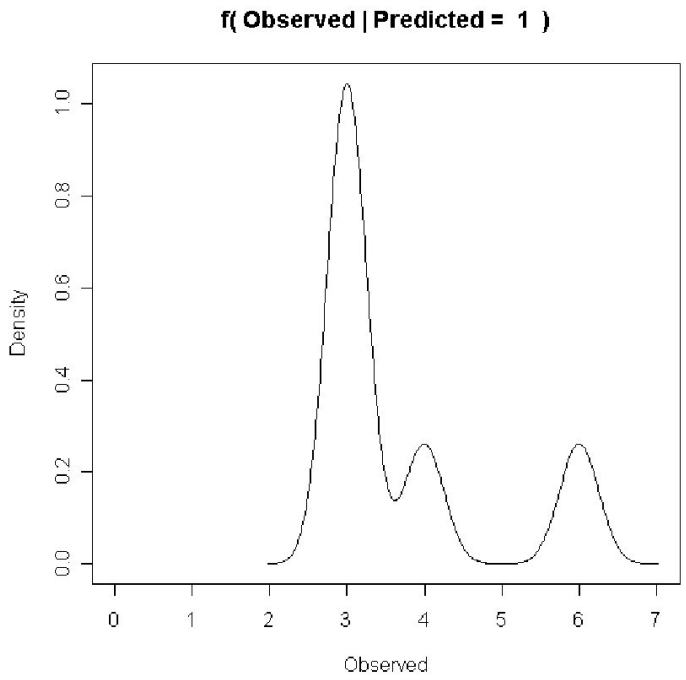

Figure 8: Conditional distribution of today's wind speed given the predicted windspeed was 1m/s (6am)

It should be noted that although the method used calculated "near optimal" bandwidths, the joint distributions had several maximum points. It is assumed that this is due to the fact that the observed windspeeds are given to the nearest metre per second. This means that there will tend to be local maxima of the estimated density function at points where both variables take integer values (i.e. rounding the observations to the nearest m/s leads to undersmoothing of the estimated density curve).

In relation to the model proposed, we wish to estimate the conditional distribution of the power output of a farm or region based on the wind speed forecast. The observed power output can be measured very accurately, while the wind speed forecast is generally to the closest metre per second. In this case it may be best to directly estimate the conditional distribution of the power output given the forecast for the wind speed is k metres per second. In order to do this, we would take all the observations of the power output given such a forecast. The estimation of this conditional distribution can then be done by kernel density estimation of the distribution of a single random variable. In the case a forecast wind speed is rare, we may group similar forecasts. The undersmoothing problem observed above should not occur.

5 Advantages and Disadvantages of the Model

It should be noted here that the model has not been tested in any way, due to the impossibility of collecting appropriate data in the time available. However, one obvious advantage of the model is its simplicity. The method can also be used at differing levels. It can not only be used to estimate the power output of one wind farm given forecast data, but also estimate the total power output of wind farms in a given region. This data driven method takes into account the correlations between the outputs of the individual wind farms in a very natural way, without any complex analysis.

Due to the geography of Ireland (it is not a large country and the vast majority of wind farms lie in the west of the country), it should be relatively simple to extend the model to one that predicts the output of wind farms on a national scale. This may be done by using historical data on the total output of

wind farms and forecast data from several weather stations and estimating the conditional distribution of power output given these combined forecasts. One problem here might be the choice of an appropriate set of weather stations. However, there are a relatively small number of weather stations spread reasonably evenly over the country. Hence, it should not be a great problem to choose a small number of weather stations such that we have a representative forecast for areas in which wind farms are located.

However, there are drawbacks of such a data driven approach. Firstly, a large amount of data on the output of wind farms and weather forecast data are needed. Secondly, this historical data only refers to the set of wind farms that are currently in existence and new wind farms will come online in the near future. Obviously no historical data is available for these wind farms. The method would have to be adapted to deal with this problem. In the short term, we may predict the power output of a wind farm on the basis of the power output of a local wind farm that uses the same or similar turbines. Such estimation will not be ideal due to the different location of these two wind farms. In the medium term, we may estimate the relationship between the power output of these two wind farms and use this to predict the power output of a relatively new wind farm. In the long term, the historical data available for this wind farm will enable its inclusion into the standard model described above.

Another problem is the short time scale required in the bidding process. A bid must be offered for each 30 minute period. Weather fronts will pass through the wind farms in a region at slightly different times, which makes prediction of power output more difficult. Such volatility, however, should be apparent in the forecast for a given period. It seems possible that taking account of this volatility will make estimation of the distribution of the power output more accurate. The power output may not only be affected by wind speed, but also by the direction of the wind. Hence, using the forecasted direction of the wind may be a predictor of power output.

6 Conclusions and Recommendations

It must be stressed that the model has not been tested and such testing is necessary before implementation. In order to test this model, it is necessary to collect data regarding the power output of wind farms at both local and national levels and forecasts for wind speeds at weather stations selected to cover the areas in which wind farms are situated. Such data should cover a period of a least a couple of years in order to estimate the conditional distribution of power output given forecast data on the basis of historical data.

Although wind speeds do show a lot of variation according to season and time of day, conditioning on the forecast wind speed will take account of a large part of this variation. It thus seems that unless a very large amount of data is available, there is no great advantage of estimating the conditional distribution of power output according to the time of day or season of the year.

One major problem of such a data driven approach is the fact that the number of wind farms is continually increasing. New wind farms cannot be fully integrated into such a model and if initial results show that the basic model can accurately estimate the conditional distribution of the power output of a well established wind farm or set of well established wind farms, work should be done on ways of integrating new wind farms into a more generalised version.

In this case, research should also be done on whether information on the volatility of the weather and other factors, such as wind direction, can be used in order to improve the model.

References

[1] Duong T. (2007), ks: Kernel Density Estimation and Kernel Discriminant Analysis for Multivariate Data in R. *Journal of Statistical Software*, **21**, issue i07.

[2] Haslett J. and Raftery A. E. (1989), Space-Time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource. *Applied Statistics*, **38**, 1-50.

[3] Pinson P. and Kariniotakis (2004), On-line Assessment of Prediction Risk for Wind Power Production Forecasts. *Wind Energy*, 7:119-132.

[4] Wand M. P. and Jones M. C. (1995), *Kernel Smoothing: Monographs on Statistics and Applied Probability 60*, Chapman & Hall, London.