Correlation analysis between UK onshore and offshore wind speeds

This item was submitted to Loughborough University's Institutional Repository by the/an author.


Additional Information:

- This is a conference paper.

Metadata Record: https://dspace.lboro.ac.uk/2134/20577

Version: Published

Publisher: European Wind Energy Association

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
CORRELATION ANALYSIS BETWEEN UK ONSHORE AND OFFSHORE WIND SPEEDS

Simon Watson and Rolando Soler-Bientz
Centre for Renewable Energy Systems Technology (CREST)
Wolfson School of Engineering, Holywell Park, Loughborough University, LE11 3TU, UK

Abstract
The work presented in this paper is an analysis of the correlation between wind speeds at offshore and coastal/inland sites to determine the expected accuracy of long term correction techniques for offshore resource assessment such as measure-correlate-predict (MCP). In particular, data from offshore mast masts at three different heights and five different surface stations have been studied. Different coastal-offshore, inland-offshore and coastal-onshore combinations have been analysed. Correlations based on one half of the available data and prediction accuracy based on the second half of the data are presented. The correlation coefficient, the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) were used as metrics in this case. Although RMSE of predictions is greatest for offshore sites, the MAPE is lowest for two of the three regions studied. An MAPE of 25-30% may be expected in terms of wind speed prediction accuracy based on statistical (turbulent) variation.

Keywords: MCP, Correlation, RMSE, MAPE, Offshore Wind

1. Introduction
Improving the accuracy of offshore wind resource estimation is important in reducing the overall levelised cost of offshore wind power. Yearly wind power density estimates can range in uncertainty from 10% for an onshore wind farm to 30% for an offshore site [1]. This is particularly pertinent in the UK where offshore wind development is moving ever further from the coast in deeper water giving rise to greater costs as the foundations become larger and hub heights increase.

In order to reduce the potential cost of an offshore wind farm, it is important to increase confidence in the prediction of the long term offshore wind speed. A common way of making a long term prediction is to use a technique known as Measure-Correlate-Predict (MCP), e.g. [2] whereby a series of short-term measurements (12-18 months) is adjusted to the expected long term value using historic data from a nearby site, traditionally a meteorological station, though increasing reanalysis data are being used for this purpose. There are a number of uncertainties when calculating an MCP long term average [3] and the correlation between predictor and candidate is one of these. This will be influenced by the prevailing climate at the two sites which ideally should have the same characteristics.

In this paper, we consider three UK offshore regions containing an offshore mast and five meteorological stations: three coastal sites and two inland. We look at the correlation in wind speeds between combinations of the sites and use a correlation (neglecting the effect of direction) determined from half of each data set for the pair which is then applied to predict the second half of the data. The correlation coefficient, the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) were computed in order to evaluate the correlations and predictions for each pair of sites. Conclusions are drawn as to the likely accuracy of MCP prediction when using either an inland or coastal site to predict the long term climate at an offshore site compared with predicting site combinations on land.

2. Study sites
Figure 1 below shows the locations of the three offshore masts which have been used to define three study regions: South-East (around the London Array Mast), East (around the Race Bank Mast) and North-West (around one of the Shell Flats Masts).
Along with the offshore mast shown in Figure 1, a total of five surface stations were included for each study region. Figure 2 below shows the geographical area of the three study regions. The locations of the offshore masts and surface stations are shown. The surface stations are identified by their unique source identifier as used by the UK Met. Office.

3. Data

Table 1 shows the time periods of data available for the offshore masts that were used in this work. In addition, the distance to the coast and the direction of closest coastal location is given. Table 2 shows the heights of the data used in this work and the combinations of sites studied offshore and onshore (coastal and inland).

![Figure 2: Geographical areas of the study regions. Locations of the offshore masts and surface stations have been included for each region. The surface stations are identified by their unique source identifier.](image)

### Table 1

<table>
<thead>
<tr>
<th>Region</th>
<th>Offshore Met Mast</th>
<th>Met surface stations (SS) ID</th>
<th>Concurrent time period</th>
<th>Closest coastal distance [km]</th>
<th>Closest coastal direction [Degrees]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SE</td>
<td>London Array (LA)</td>
<td>504 498 775</td>
<td>440 19188</td>
<td>2004-12 / 2011-12</td>
<td>26 305</td>
</tr>
</tbody>
</table>

The concurrent periods were divided into two equal parts to create working datasets of 42 months for the South-West region, 24 months for the East region and 30 months for the North-West region. Then first half of the dataset was used as a reference dataset to: 1) calculate the correlation between the wind speed at two heights; 2) calculate a linear fit to predict the wind speed at a target site from a predictor site (neglecting the effect of direction). The accuracy of prediction of the mean wind speed at the target site as compared with the second half of each dataset was evaluated in terms of the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE). The prediction method was then reversed using the second half of each dataset to determine the correlation and the first half of the dataset for prediction validation.

4. Results

4.1 Assessment of the correlations

The correlations were assessed using the Pearson correlation coefficient. This coefficient is a measure of the degree of linear dependence between two variables (values close to +1/-1 indicate total correlation while values close to 0 means no correlation). Considering two datasets of n elements measured concurrently: \{x_1, \ldots, x_n\} and \{y_1, \ldots, y_n\}, the correlation coefficient \(R_{xy}\) can be defined then as:

\[
R_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}
\]
Figure 3 shows the correlation between the three offshore masts and the onshore stations using data from the lowest offshore mast height. In each case, the highest correlation is seen for the coastal sites. In general, there is up to a 15% increase in correlation coefficient for the coastal sites compared with the inland sites which would suggest the potential influence of geographical and climatic conditions which may differ between inland and coastal locations.

Figure 3 also includes the correlation coefficients for both halves of the concurrent period. It can be seen that the correlations were better for LA and SF when the second half of the concurrent dataset was used as the reference to produce the correlation coefficient. This suggests that seasonal and annual variation in site correlations can be significant. This effect is largest in the case of the SF mast. It is worth noting that the different trends in correlation by time period are not consistent between the three offshore masts.

### 4.2 Assessment of the predictions

Assessment was made of the wind speed predictions using the linear correlation from one half of the data to predict the second half of the data. No account was taken of direction dependence. The accuracy of prediction was evaluated using two parameters: the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE).

The Root Mean Square Error (RMSE) is a measure of the difference between the predicted values $v_{p,i}$ and the measured values $v_{m,i}$. Considering a concurrent period with $N_{cp}$ values then the RMSE can be defined as:

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N_{cp}} (v_{p,i} - v_{m,i})^2}{N_{cp}}} \quad (2)$$

The Mean Absolute Percentage Error (MAPE) is a similar metric which is normalised for the magnitude of the measured value given as below:

$$MAPE = \frac{100}{N_{cp}} \sum_{i=0}^{N_{cp}} \left| \frac{v_{m,i} - v_{p,i}}{v_{m,i}} \right| \quad (3)$$

Both RMSE and MAPE were computed for the predictions of one half of the concurrent period by means of the linear fit generated with the other half of the concurrent period. Figure 4 shows the results for each study region and period predicted.

The results show broadly the same trends as seen for the correlation coefficients in that the prediction error is smaller when using coastal sites as the predictor. There is some geographical variation in the level of MAPE where the masts off the East Coast at Race Bank and the North-West Coast showed MAPE values of 23-33%,
depending on site and time period) with London Array off the South-East Coast showing an MAPE of 28-37%.

![Root Mean Square Error (RMSE)](image)

![Mean Absolute Percentage Error (MAPE)](image)

Figure 4. RMSE and MAPE for the wind speed predictions at the lowest height of each offshore mast (LA_20, RB_15 and SF_20) using the coastal and inland surface stations as reference sites.

4.3 Correlation coefficient for combinations of all sites

In this section, we consider the correlation between all sites in each geographical region. In addition, correlations are calculated between the wind speed measured at different heights on the offshore masts (top, bottom and an intermediate height) as a ‘base case’. The results are shown in Figure 5.

As might be expected, correlations between wind speeds at different heights on the offshore masts are high. It is difficult to determine other definitive trends. As was seen before, the highest correlations with the offshore mast data are with coastal sites. There is evidence for the SE and S areas, that onshore correlations are highest between inland sites though this trend is not seen for the NW region with no clear difference observed.

![Correlation coefficient computed with the second half of the concurrent period](image)
4.4 RMSE of predictions for combinations of all sites

Figure 6 shows the RMSE metric derived by calculating a linear fit based on one half the data and using this to predict the wind speeds during the second half. There is a marked increase in the RMSE for the offshore sites predicted using onshore data compared with combinations of the onshore sites, although this may be due to higher mean wind speeds at the offshore sites. RMSE errors for predictions at the coastal sites are higher than the inland sites for the SE and E regions, though again wind speeds are higher at these sites. This trend is less obvious for the NW region. It is interesting to note that for the NW region second period, that the highest height wind speeds (82m) on the Shell Flats mast are better predicted (in absolute values) from some of the coastal sites than from measurements lower down the mast.
4.4 MAPE of predictions for combinations of all sites

The MAPE metric normalizes for the mean wind speeds observed at each site and thus might be expected to better show any trends between inland-costal-offshore prediction accuracy. Figure 7 shows the MAPE for combinations of sites in each of the regions. The MAPE is lowest when predicting the offshore mast wind speeds in most cases for the E and NW regions. This is not so clearly the case for the SE region. There is some evidence to suggest that for the SE and NW regions at least, the MAPE is greater for the coastal sites. For the E region, the MAPE is greatest for the inland sites. For the NW region, in this case, there is no evidence to suggest a better prediction of the 82m wind speeds from coastal data when using the MAPE as a metric.
5. Conclusions

From the preliminary analysis undertaken in this research it can be concluded:

• Correlations between offshore and onshore sites would seem to be greatest when the onshore site is coastal;
• RMSE of predictions are significantly higher for offshore sites, though lower when using a coastal site as a reference site compared to an inland site;
• MAPE of predictions for offshore sites is lower for two of the three regions compared with the inland sites, indicating that the higher speeds at offshore and coastal sites is a large contributor to the higher RMSE values observed;
• MAPE values of 25-30% can be expected for predictions of offshore wind speeds based on statistical (turbulent) variability.

References
