# Accuracy assessment of offshore wind measurements partially compensated by MCP method considering data availability

**K Enoki[1], T Ishihara[2]**

[1] Obayashi Corporation, 4-640, Shimokiyoto, Kiyose-shi, Tokyo, 204-8558, Japan
[2] The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan

enoki.kota@obayashi.co.jp, ishihara@bridge.t.u-tokyo.ac.jp

**Abstract**. In Japan, most offshore promotion areas for bottom-fixed wind turbines are in near-shore areas, where offshore wind speeds are not uniform. The dual-scanning LiDAR system (DSL), which is a remote-sensing technology, is widely used to measure the mean wind speed, wind direction, and turbulence for resource and site assessments. However, the availability of DSL data tends to be lower than the key performance indicators proposed in the Carbon Trust in Offshore Wind Accelerator project owing to the characteristics of DSL. Missing observed data can be predicted using the measure–correlate–predict (MCP) method, and the uncertainty of the MCP method can be examined using a prediction function. However, a method to evaluate the accuracy of the final dataset comprising observed data and data predicted by the MCP method is yet to be devised. In this study, a set of formulas is proposed to evaluate the accuracy of indices, such as the coefficient of determination, slope, and offset of linear regression, for a partially complemented dataset using the MCP method. The proposed formulas are validated against on-site measurements.

## 1. Introduction

Offshore wind energy is garnering considerable attention to realize a sustainable society. In wind-energy projects, on-site measurements are conducted to assess the amount of resources and ensure the suitability of site conditions for wind turbines. Therefore, highly accurate and reliable measurements under severe offshore environments must be conducted.

Three primary technologies are used to measure offshore wind, as shown in table 1. One is to install a bottom-fixed platform in shallow water and set a met mast or vertical LiDAR (VL) on it. The accuracy and reliability of measurements achieved using the met mast and VL have been proven in the field of onshore wind energy. Anemometers attached on the mast were shown to be affected by the flow distortion around the mast, thus necessitating the appropriate corrections [1]. Meanwhile, floating LiDAR systems (FLSs) are widely used in deep water [2]. Compared with the case of met masts, the installation of FLSs is relatively inexpensive; however, the turbulence intensity observed by FLSs is contaminated by the motion of the floater due to wind and waves and must be corrected by applying motion compensation [3]. For VL measurements, the data availability tends to decrease as the measurement-height increases [4]. Furthermore, the measurements overestimated the turbulence intensity [5]. Additionally, these two methods require regular maintenance at sea, thus rendering access difficult if equipment issues occur and reducing data availability owing to long downtimes during stormy weather [6]. The third method is remote sensing from land using scanning LiDAR (SL), either

**Table 1.** Offshore wind measurement technologies.

| Platform | Bottom fixed | Floating | Onshore | |
|---|---|---|---|---|
| Type of LiDAR | VL | VL | SSL | DSL |
| Wind speed | ○ [1] | ○ [3] | ○ [7] | ○ [8] |
| Turbulence | ○ [1] | △ [3], [5] | × [9] | ○ [8] |
| Post-processed data availability | > 95 % [1] | > 85～90 % [4], [6] | > 90 % [7] | > 80 % [8] |
| References | Ishihara et al. [1] | Yamaguchi and Ishihara [3] | Mano et al. [7] | Watanabe et al. [8] |

Abbreviations VL: Vertical LiDAR. SSL: Single Scanning LiDAR. DSL: Dual Scanning LiDAR
Symbols ○: Applicable. △: Applicable with motion compensations. ×: Inapplicable.

**Table 2.** KPIs for accuracy assessment of LiDAR system [11].

| Acceptance criteria | Wind speed | | Wind direction | | |
|---|---|---|---|---|---|
| | Slope | $R^2$ | Slope | Offset | $R^2$ |
| Minimum | 0.97-1.03 | > 0.97 | 0.95-1.05 | ±10 ° | > 0.95 |
| Best practice | 0.98-1.02 | > 0.98 | 0.97-1.03 | ±5 ° | > 0.97 |

**Table 3.** KPIs for reliability assessment of LiDAR system [11].

| Stage | System data availability | | Post-processed data availability | |
|---|---|---|---|---|
| | Monthly | Overall | Monthly | Overall |
| Pre-commercial | ≧90 % | ≧95 % | ≧80 % | ≧85 % |
| Commercial | ≧95 % | ≧97 % | ≧85 % | ≧90 % |

single-scanning LiDAR (SSL) [7] or dual-scanning LiDAR (DSL) [8]. Studies showed that using a SSL system in the PPI and RHI modes does not allow one to measure the turbulence intensity owing to the low sampling rate [9]. By contrast, when the mean wind direction is parallel to the scanning beam and a fixed line-of-sight (LoS) scanning mode is used, SSL systems can be used to measure the turbulence intensity [10].

In Japan, most offshore promotion areas for bottom-fixed wind turbines are in near-shore areas, where the offshore wind speeds are not uniform owing to uneven topography and rapid changes in the land surface roughness. DSL systems are widely used to measure the mean wind speed, direction, and turbulence. The availability of SL data decreases as the measurement distance increases. In particular, DSL calculates the mean wind speed and wind direction from two LoS wind speeds. The data are loose if one wind speed is available and the other is not. The data availability of DSL may be lower than that of SSL. Additionally, the data availability of LiDAR is adversely affected by atmospheric conditions, such as rainfall, snowfall, and fog, as well as low aerosol concentrations. Thus, the data availability of DSL tends to be lower than the criteria proposed in the Carbon Trust in Offshore Wind Accelerator project to determine whether the key performance indicators (KPIs) defined in [11] are satisfied. Therefore, to achieve high data availability, the missing data must be complemented using the measure–correlate–predict (MCP) method, which predicts missing data based on the correlation with nearby wind measurements. The uncertainty of the MCP method can be examined using the constructed prediction functions. However, a method to evaluate the final dataset comprising observed data and data predicted by the MCP method is yet to be devised.

In this study, a set of formulas is proposed to evaluate the KPIs, such as the coefficient of determination ($R^2$), slope, and offset of linear regression, for the partially complemented dataset using the MCP method presented in section 2. Subsequently, it is validated based on on-site measurements presented in section 3. The conclusions are summarized in section 4.

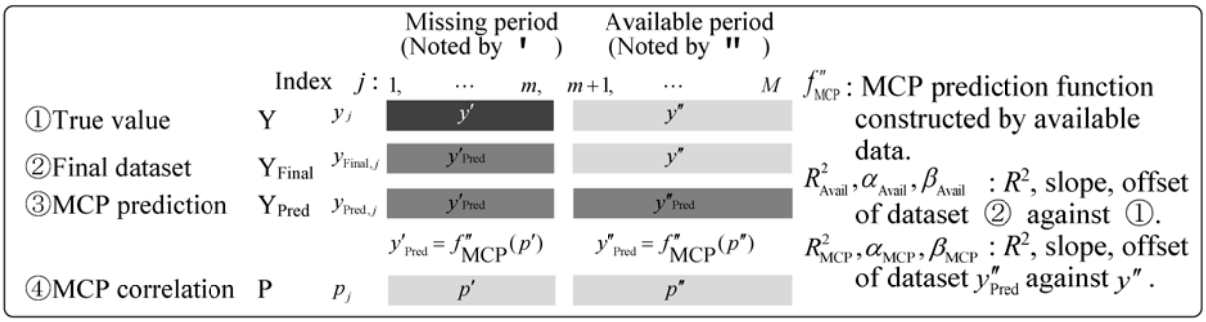


**Figure 1.** Definition of variables, symbols, and periods used in current study.

## 2. Methodology

### *2.1. KPIs for wind-speed and wind-direction measurements using remote-sensing devices*

To evaluate the accuracy and reliability of wind speed and direction measured using a remote-sensing device (RSD), such as VL, FLSs, and SL, a KPI-based assessment proposed by the Carbon Trust [11] is typically used. Measurements obtained using RSDs are compared with those obtained using pre-calibrated instruments. The KPIs and related acceptance criteria are presented in table 2. The accuracy of the RSD is evaluated based on the KPI, such as the slope, offset, and $R^2$ of the regression line. The reliability of the observation system, such as system availability before post-processing and the data availability after post-processing, is evaluated based on the KPIs, as shown in table 3. At the "pre-commercial" stage, an RSD must satisfy the "minimum" accuracy criterion at the least, whereas at the "commercial" stage, the RSD must satisfy the "best practice" accuracy criterion. However, wind-speed measurements performed using a DSL system are more likely to have missing data than those performed using a met mast or VL. Complementation with the MCP method is necessary to satisfy the KPIs of data availability. Therefore, the final dataset comprising the observed data and data predicted by the MCP method must be evaluated.

### *2.2. Mean value and standard deviation considering data availability*

Observations with missing data are generally complemented with nearby observations using the MCP method. The final dataset obtained through observations comprises data measured during the available period and data predicted using the MCP method during the missing period. This section describes the derivation of the formulas for evaluating the final dataset.

Figure 1 presents the definitions of the variables, symbols, and periods used in this study. The maximum value of index $j$ representing the observation period Y is defined as $M$. Values from 1 to $m$ represent missing data, and those from $m$ to $M$ represent available data. Where appropriate, the variables are denoted by "prime" for missing periods and "double prime" for available periods. The missing data are complemented with nearby observations. We assumed that the $R^2$, slope, offset, and other statistical indices during the missing period can be represented by the indicators of the dataset used to construct the MCP method.

If the availability of the reference data to complement missing data is 100 %, then the mean and standard deviation of the final dataset are derived as a function of the data availability of the observations. Hence, the mean value can be expressed as follows:

$$\bar{y} = \sum_{j=1}^{M} y_j / M = (M-m)M^{-1} \sum_{j=m+1}^{M} y_j / (M-m) + mM^{-1} \sum_{j=1}^{m} y_j / m \quad (1)$$

By substituting the value $y' \simeq y'_{\mathrm{Pred}} = f''_{\mathrm{MCP}}(p')$ predicted by the MCP method as an approximate value for the missing period and using the data availability $\zeta$ ($m / M = 1 - \zeta$), the following equation is obtained:

$$\bar{y} \simeq \bar{y}_{\text{Final}} = \zeta \bar{y}'' + (1-\zeta)\bar{y}'_{\text{Pred}}, \tag{2}$$

where $\bar{y}''$ is the mean value over the available period, and $\bar{y}'_{\text{Pred}}$ is the mean value predicted by the MCP method over the missing period.

In this study, the standard deviation $\sigma_y$ is defined as the square root of the variance $\sigma_y^2$ as follows:

$$\sigma_y^2 = M^{-1}\sum_{j=1}^{M}(y_j - \bar{y})^2 \tag{3}$$

By substituting the mean value $\bar{y}$ obtained using equation (2) into equation (3), $\sigma_y$ is rewritten as:

$$\begin{aligned}\sigma_y^2 &= M^{-1}\sum_{j=1}^{m}\left((y_j - \bar{y}') + \zeta(\bar{y}'_{\text{Pred}} - \bar{y}'')\right)^2 + M^{-1}\sum_{j=m+1}^{M}\left((y_j - \bar{y}'') + (1-\zeta)(\bar{y}'' - \bar{y}'_{\text{Pred}})\right)^2 \\ &= (1-\zeta)(\overline{y^2}' - 2\bar{y}'\bar{y}'_{\text{Pred}} + \bar{y}'^2_{\text{Pred}}) - (1-\zeta)\zeta(2\bar{y}''\bar{y}' - 2\bar{y}'\bar{y}'_{\text{Pred}} + \bar{y}'^2_{\text{Pred}} - \bar{y}''^2) + \zeta\sigma_y''^2.\end{aligned} \tag{4}$$

Since $y_i$ in the missing period is unknown, $\bar{y}' \simeq \bar{y}'_{\text{Pred}}$, $\overline{y^2}' \simeq \overline{y^2_{\text{Pred}}}'$, and $\sigma_y'^2 \simeq \sigma'^2_{\text{Pred}}$ are applied.

$$\sigma_y^2 = \sigma^2_{\text{Pred}} = (1-\zeta)\sigma'^2_{\text{Pred}} + \zeta(1-\zeta)(\bar{y}'' - \bar{y}'_{\text{Pred}})^2 + \zeta\sigma_y''^2 \tag{5}$$

The ratio of the predicted value to the observed value decreases as the data availability $\zeta$ increases.

### *2.3. $R^2$ considering data availability*

The $R^2$ of a regression equation for dataset $(y_{\text{Final},i}, y_i)$ $(i = 1 \ldots n)$ is expressed as

$$R^2 \equiv \text{ESS}/\text{TSS} = 1 - \text{RSS}/\text{TSS}, \tag{6}$$

where the total sum of squares TSS = $\Sigma_i^n (y_i - \bar{y})^2$ comprises the explained sum of squares ESS = $\Sigma_i^n (y_{\text{Final},i} - \bar{y})^2$, which can be estimated using the regression equation, and the residual sum of squares RSS = $\Sigma_i^n (y_i - y_{\text{Final},i})^2$. $\bar{y}$ and $y_{\text{Final}}$ denote mean and predicted values of $y$, respectively. The regression equation for the available period is $y_{\text{Final}} = y$, and the following holds:

$$\sum_{j=m+1}^{M}(y_j - y_{\text{Final},j})^2 = 0. \tag{7}$$

$R^2_{\text{MCP}}$ predicted in the available period is expressed as

$$R^2_{\text{MCP}} = 1 - \sum_{j=m+1}^{M}(y_j - y_{\text{Pred},j})^2 / \sum_{j=m+1}^{M}(y_j - \bar{y}''_{\text{Pred}})^2. \tag{8}$$

Assuming that the RSS/TSS ratio in the construction period is equal to that in the prediction period, the following equation is derived:

$$\sum_{j=m+1}^{M}(y_j - y_{\text{Pred},j})^2 / \sum_{j=m+1}^{M}(y_j - \bar{y}''_{\text{Pred}})^2 = \sum_{j=1}^{m}(y_j - y_{\text{Pred},j})^2 / \sum_{j=1}^{m}(y_j - \bar{y}'_{\text{Pred}})^2 \tag{9}$$

Using equations (7) and (9), the $R^2_{\text{Avail}}$ of the final dataset can be transformed as follows:

$$R^2_{\text{Avail}} = 1 - \frac{\sum_{j=1}^{M}(y_j - y_{\text{Final},j})^2}{\sum_{j=1}^{M}(y_j - \bar{y})^2} = 1 - \frac{m\sum_{j=1}^{m}(y_j - \bar{y}'_{\text{Pred}})^2 m^{-1}\sum_{j=m+1}^{M}(y_j - y_{\text{Pred},j})^2}{M\sum_{j=1}^{M}(y_j - \bar{y})^2 M^{-1}\sum_{j=m+1}^{M}(y_j - \bar{y}''_{\text{Pred}})^2} \tag{10}$$

Subsequently, $R^2_{\text{Avail}}$ is expressed as a function of data availability $\zeta$ using equation (8) as follows:

$$R^2_{\text{Avail}} = 1 - (1-\zeta)\sigma'^2_{\text{Pred}}\sigma_y^{-2}(1 - R^2_{\text{MCP}}). \tag{11}$$

Because $\sigma'^2_{\mathrm{Pred}}\sigma_y^{-2}$ is almost equal to 1, the $R^2$ of the final dataset is expressed as

$$R^2_{\mathrm{Avail}} = 1-(1-\zeta)(1-R^2_{\mathrm{MCP}})\,. \tag{12}$$

The $R^2_{\mathrm{Avail}}$ of the final dataset depends on the data availability $\zeta$ and $R^2_{\mathrm{MCP}}$ obtained via the MCP method.

### *2.4. Slope and offset considering data availability*

The slope and offset are derived while considering data availability. The regression equation for the KPI evaluation is $y=\alpha_{\mathrm{Avail}}.y_{\mathrm{Final}}+\beta_{\mathrm{Avail}}$. The RSS of the final dataset and the dataset of actual values is expressed as

$$\mathrm{RSS}=\sum_{j=1}^{M}(y_j-\alpha_{\mathrm{Avail}}y_{\mathrm{Final},j}-\beta_{\mathrm{Avail}})^2. \tag{13}$$

$\alpha_{\mathrm{Avail}}$ and $\beta_{\mathrm{Avail}}$ are expressed as shown in [12] by minimizing the RSS.

$$\alpha_{\mathrm{Avail}}=\sum_{j=1}^{M}(y_{\mathrm{Final},j}-\overline{y}_{\mathrm{Final}})(y_j-\overline{y})\Big/\sum_{j=1}^{M}(y_{\mathrm{Final},j}-\overline{y}_{\mathrm{Final}})^2 \tag{14}$$

$$\beta_{\mathrm{Avail}}=\overline{y}-\alpha_{\mathrm{Avail}}\overline{y}_{\mathrm{Final}} \tag{15}$$

$\alpha_{\mathrm{Avail}}$ can be simplified using $\sigma_y^2\simeq\sigma^2_{\mathrm{Final}}=\sum_{j=1}^{M}(y_{\mathrm{Final},j}-\overline{y}_{\mathrm{Final}})^2M^{-1}$, equation (2), and $\overline{y}'\simeq\overline{y}'_{\mathrm{Pred}}$.

$$\begin{aligned}\alpha_{\mathrm{Avail}} &= M^{-1}\sigma^{-2}_{\mathrm{Final}}\sum_{j=1}^{m}\left[(y_{\mathrm{Pred},j}-\overline{y}'_{\mathrm{Pred}})(y_j-\overline{y}'_{\mathrm{Pred}})+\zeta^2(\overline{y}''-\overline{y}'_{\mathrm{Pred}})^2\right]+\\ &\quad M^{-1}\sigma^{-2}_{\mathrm{Final}}\sum_{j=m+1}^{M}\left[(y_{\mathrm{Pred},j}-\overline{y}''_{\mathrm{Pred}})(y_j-\overline{y}''_{\mathrm{Pred}})+(1-\zeta)^2(\overline{y}''_{\mathrm{Pred}}-\overline{y}'_{\mathrm{Pred}})^2\right]\\ &=(1-\zeta)\alpha'_{y,\,\mathrm{Final}}\sigma'^2_y\sigma^{-2}_{\mathrm{Final}}+\zeta\alpha''_{y,\mathrm{Final}}\sigma''^2_y\sigma^{-2}_{\mathrm{Final}}+(1-\zeta)\zeta(\overline{y}''-\overline{y}'_{\mathrm{Pred}})^2\,\sigma^{-2}_{\mathrm{Final}}\end{aligned} \tag{16}$$

$\alpha'_{y,\,\mathrm{Final}}$ in the first term is the slope of the regression line during the missing period and is unavailable; the slope of the regression line $\alpha_{\mathrm{MCP}}$ is used for $\alpha'_{y,\,\mathrm{Final}}$. In the available period, the regression line is $y=y_{\mathrm{Final}}$, and $\alpha''_{y,\,\mathrm{Final}}$ in the second term is 1. Thus, $\alpha_{\mathrm{Avail}}$ is expressed as

$$\alpha_{\mathrm{Avail}}=1-(1-\zeta)(1-\alpha_{\mathrm{MCP}})\sigma'^2_{\mathrm{Pred}}\sigma^{-2}_{\mathrm{Final}}\,. \tag{17}$$

By assuming that $\sigma'^2_{\mathrm{Pred}}\sigma^{-2}_{\mathrm{Final}}$ is almost equal to 1, $\alpha_{\mathrm{Avail}}$ is simplified as

$$\alpha_{\mathrm{Avail}}=1-(1-\zeta)(1-\alpha_{\mathrm{MCP}})\,. \tag{18}$$

Equation (18) is only used for evaluating the wind speed, where linear regression without offset is used. Subsequently, $\beta_{\mathrm{Avail}}$ can be obtained as follows:

$$\beta_{\mathrm{Avail}}=\overline{y}-\alpha_{\mathrm{Avail}}\overline{y}_{\mathrm{Final}}\simeq(1-\alpha_{\mathrm{Avail}})\overline{y}=(1-\zeta)(1-\alpha_{\mathrm{MCP}})\overline{y}=y\,\overline{y}''^{-1}(1-\zeta)(\overline{y}''-\alpha_{\mathrm{MCP}}\overline{y}'')\,, \tag{19}$$

where the relationship between $\overline{y}\simeq\overline{y}_{\mathrm{Final}}$ and equation (18) is used. As $\beta_{\mathrm{MCP}}=\overline{y}''-\alpha_{\mathrm{MCP}}\overline{y}''_{\mathrm{Pred}}$ and $\overline{y}''\simeq\overline{y}''_{\mathrm{Pred}}\simeq\overline{y}$, $\beta_{\mathrm{Avail}}$ is derived as

$$\beta_{\mathrm{Avail}}=(1-\zeta)\beta_{\mathrm{MCP}}\,. \tag{20}$$

In this study, $\alpha_{\mathrm{Avail}}$ and $\beta_{\mathrm{Avail}}$ are evaluated using equations (18) and (20), respectively.

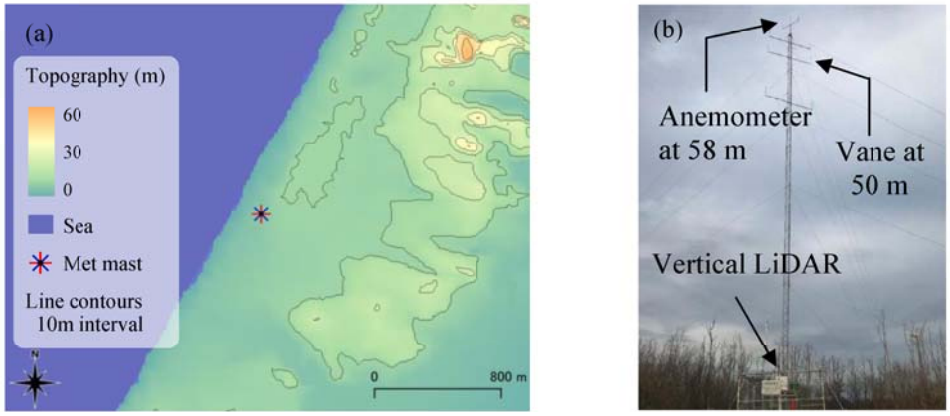


**Figure 2.** Measurement site information. (a) Topography and (b) met mast and vertical LiDAR

**Table 4.** Summary of on-site measurements.

| Location | N 40.05025 °, E 139.93232 ° | |
|---|---|---|
| Measured items | Mean and standard deviation of wind speed and mean direction in 10 minutes | |
| Observation | Met mast | Vertical LiDAR |
| Instruments | Anemometer (NRG Class1). Vane (NRG 200M) | Vertical LiDAR (Leosphere Wind Cube V2) |
| Heights | 58m (Wind speed), 50m (Direction) | 40/50/58/70/90/110/130/140/150/170/190m |

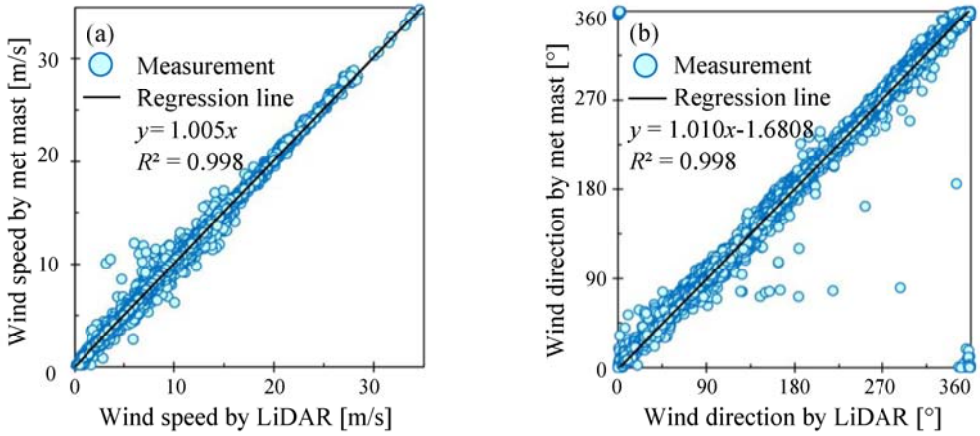


**Figure 3.** Comparison of (a) mean wind speed at 58 m and (b) wind direction at 50 m obtained from met mast and VL.

## 3. Results and discussions

### *3.1. On-site measurements for verification*

Figure 2 and table 4 show an overview of the on-site measurements using the VL and met mast. The location is approximately 150 m inland from the Sea of Japan coastline in Akita Prefecture. During one year of observation, the VL showed good agreement with the met mast at a horizontal distance of 15 m. The correlation between the VL and mast is shown in figure 3. The tower complies with IEC61400-50-1 and was equipped with two anemometers at the top, thus eliminating the shadow effect via the appropriate selection of an anemometer based on the wind direction.

**Table 5.** Statistics of wind data at different heights against at target height 58 m.

| Height | Availability | Mean speed | | Mean direction | | | Standard deviation | |
|---|---|---|---|---|---|---|---|---|
| [m] | [%] | Slope | $R^2$ | Slope | Offset | $R^2$ | Slope | $R^2$ |
| 190 | 96.44 | 1.094 | 0.866 | 0.927 | 21.65 | 0.937 | 1.013 | 0.786 |
| 150 | 98.33 | 1.077 | 0.913 | 0.953 | 14.11 | 0.965 | 1.002 | 0.849 |
| 90 | 99.16 | 1.038 | 0.981 | 0.983 | 5.30 | 0.994 | 1.003 | 0.942 |
| 58 | 99.21 | 1.000 | 1.000 | 1.000 | 0.00 | 1.000 | 1.000 | 1.000 |

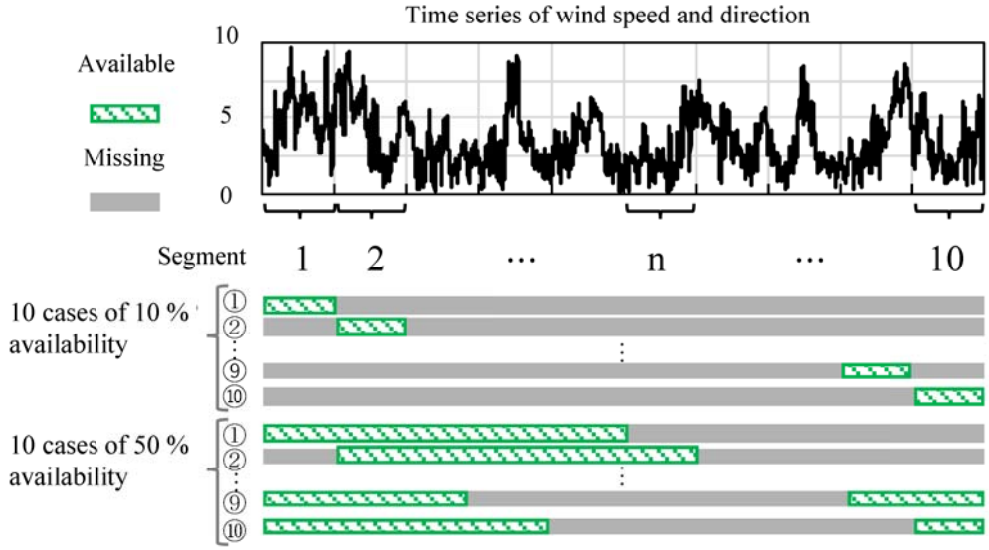


**Figure 4.** Examples of data segmentation for validation.

In this study, the data observed at 58 m by the VL is regarded as the actual values. The data obtained at 90, 150, and 190 m are used as reference data for complementation. Table 5 shows the annual data availability and the statistical indices of the reference data against the data at 58 m. The statistical indices deviate from perfect agreement as the altitude of the reference data increases.

*3.2. Evaluation of mean value and standard deviation*

To examine the applicability of the proposed formulas to various data availabilities and the statistical properties of complementation, the following data are prepared. Figure 4 shows the configuration of the available data for 10 % and 50 % availabilities. The time-series data at 58 m are partitioned into 10 segments. A unit segment accounts for 10 % of the total data and the availability of each case is set to $n$ ×10 % ($n$ = 1…9). Consecutive intervals are regarded as available intervals and are represented by shaded areas, as shown in figure 4. The start and end of the interval are treated consecutively, and the number of cases for each data availability is maintained at 10. Because the extracted intervals corresponded to the available period, the observed data at 58 m and the reference data at each altitude in these intervals are used to construct a prediction model for the MCP method. The data predicted by the MCP method are used for the missing intervals in the final dataset. The observed data are used directly for available intervals. Finally, the $R^2$, $\alpha$, and $\beta$ values of the final dataset are calculated and compared with those evaluated using the proposed formulas.

Linear regression using the MCP method, as expressed in equation (21), is performed to predict the mean wind speed $U$. The bias correction method, as expressed in equation (22), is adopted for the mean

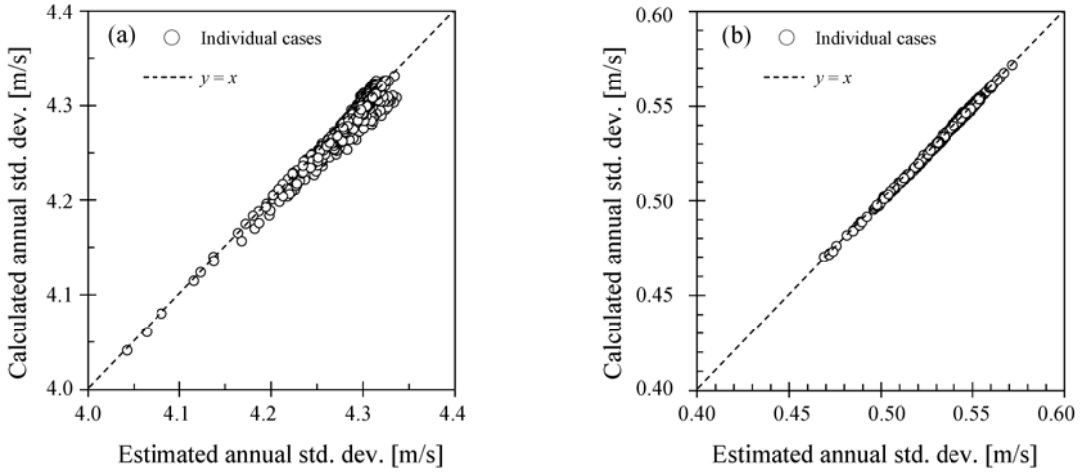


**Figure 5.** Comparison of annual standard deviation of (a) mean wind speed (b) standard deviation between calculations from time series and estimations using equation (5).

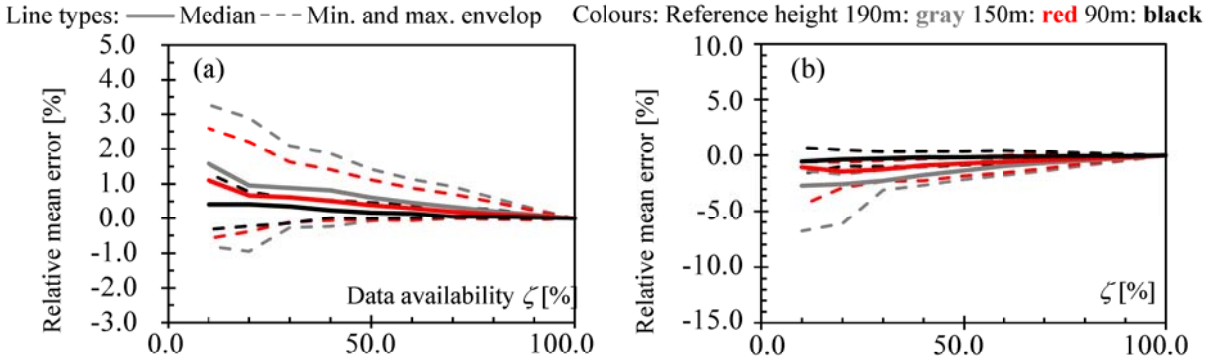


**Figure 6.** Relative mean errors of final dataset against actual value. (a) Annual mean and (b) annual standard deviation for mean wind speed.

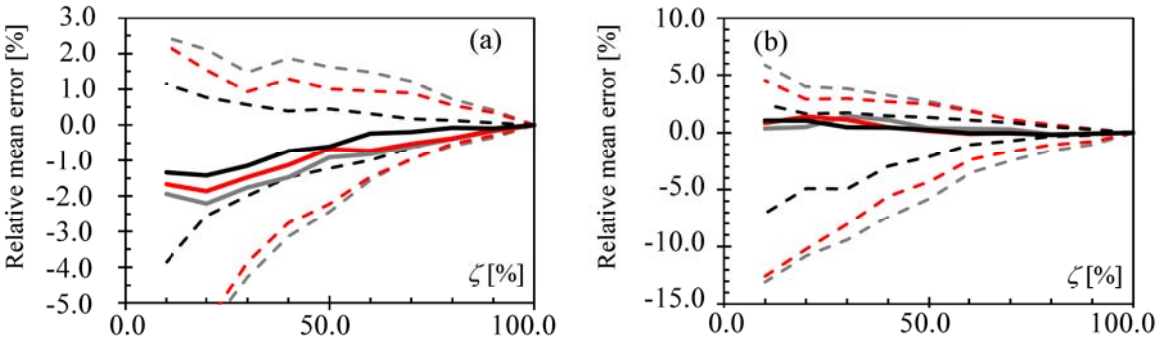


**Figure 7.** Relative mean errors of final dataset against actual value. (a) Annual mean and (b) annual standard deviation for standard deviation of 10-min wind speed. Legend shown in figure 6 is adopted.

wind direction $\theta$, whereas the double-bias correction method, as expressed in equation (23), is employed for the standard deviation $\sigma$.

$$U_{\mathrm{Pred}} = a(\theta_{\mathrm{P}})U_{\mathrm{P}} + b(\theta_{\mathrm{P}}) \tag{21}$$

$$\theta_{\mathrm{Pred}} = \theta_{\mathrm{P}} + \Delta\theta(\theta_{\mathrm{P}}) \tag{22}$$

$$\sigma_{\mathrm{Pred}} = c(\theta_{\mathrm{P}})\sigma_{\mathrm{P}} + d(\theta_{\mathrm{P}}) \tag{23}$$

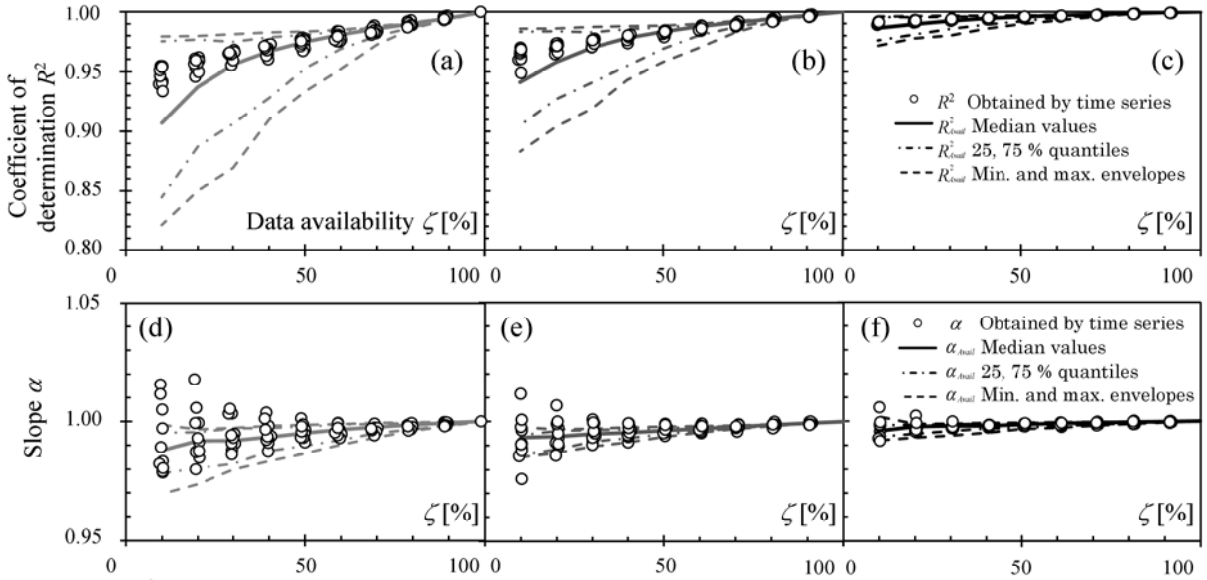


**Figure 8.** $R^2$ and slope of time series (symbols) and estimations by proposed formulas (lines) for mean wind speed. (a)-(c) evaluation of $R^2$; (d)-(f) slope. Left column shows data at 190 m obtained using MCP method, middle and right columns present data at 150 and 90 m, respectively.

where the subscripts Pred and P denote the predicted and observed values, respectively. The coefficients $a$, $b$, $c$, $d$, and $\Delta\theta$ in equations (21)–(23) are estimated for each of the 16 wind directions during the available period. The coefficients for wind direction are interpolated from the coefficients for adjacent wind directions. In practice, multiple reference observations may be available simultaneously; however, the proposed formulas can be applied using an approach similar to the MCP method for multiple wind directions.

The annual mean and standard deviation estimated using equations (2) and (5) are compared with those calculated directly from the time series. The estimated annual standard deviation, as shown in figure 5, agrees well with the observed standard deviation. The annual mean and standard deviation of wind speed of the final dataset are compared with actual values, and the relative error as a function of data availability is shown in figures 6 and 7. For 10 % availability, the median as well as the minimum and maximum ranges of the relative error against the observation at 58 m are shown by lines. The predictions with high availability agree well with the measurements. However, when the availability is low, the error in the final dataset varies significantly, depending on the configuration of the missing period. For a specific availability, the closer the reference altitude is to 58 m, the more accurate is the dataset because the reference data are more correlated with the data at 58 m.

### *3.3. Verification of proposed formulas*

The $R^2$, slope, and offset calculated from the predicted and actual values are compared with the $R_{\mathrm{Avail}}$, $\alpha_{\mathrm{Avail}}$, and $\beta_{\mathrm{Avail}}$ evaluated using equations (12), (18), and (20), respectively, for the mean wind speed (see figure 8) and mean wind direction (see figure 9). For each availability, the values of $R_{\mathrm{Avail}}$, $\alpha_{\mathrm{Avail}}$, and $\beta_{\mathrm{Avail}}$ predicted by the proposed formulas are shown as lines connecting the median, 25 % and 75 % quantiles, and the minimum and maximum values.

The actual distribution of $R^2$ with respect to the mean wind speed shows a higher degree of scatter as the reference altitude increases, as shown in figures 8 (a)–(c). Similarly, the proposed formulas exhibit a higher degree of scatter compare with the actual distribution. In case (c), the errors in the estimated mean and standard deviation are small, as shown in figure 6; therefore, the actual and predicted values match well. This is because if the correlation between the reference and observed data is high, then the

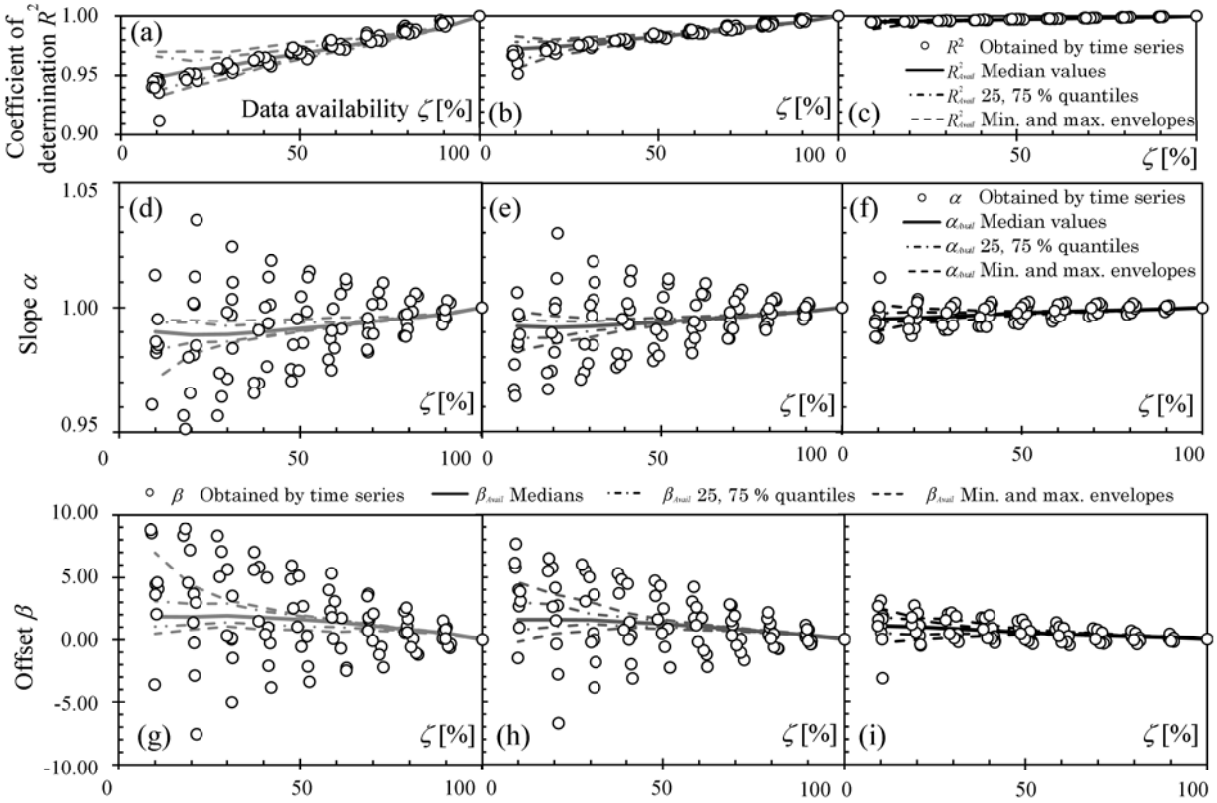


**Figure 9.** $R^2$, slope and offset of time series (symbols) and estimations by proposed formulas (lines) for mean wind direction. (a)-(c) evaluation of $R^2$; (d)-(f) slope; (g)-(i) offset. Left column shows data at 190 m obtained using MCP method, middle and right column present 150 m and 90 m, respectively.

assumptions used to derive the proposed formulas are valid. The slope for wind speed is shown in figures 8 (d)–(f). Because the prediction function shown in equation (21) is based on the least-squares method, $\alpha_{MCP}$ is approximately 1 in each wind direction. Therefore, the results obtained using equation (18) tend to show a value of approximately 1. The actual $\alpha$ is scattered, as shown in figure 8, when the data availability is low, whereas it converges to the value predicted by the proposed formula when the data availability is high.

Figures 9 (a)–(i) show $R^2$ with respect to the mean wind direction. Compared with the wind speed, the evaluated values agree well with the actual values at any reference altitude. This is because the correlation of the wind direction at each reference altitude is higher than that of the wind speed, as listed in table 5. As shown in figures 9 (d)–(i), $\alpha$ and $\beta$ exhibit a high degree of scatter when data availability is low, whereas their degree of scatter decreases as data availability increases. The slope predicted by the proposed formula becomes more accurate as the correlation between the reference and observed data at 58 m increases. Similar to the slope, the offset predicted by the proposed formula deviates less from the actual value when the data availability is low; however, as the data availability approaches 100 %, the predicted value approaches the actual value.

Finally, the bin method is used to examine the certainty of the proposed formulas by analyzing the difference between the actual and predicted values. The 99% quantile is calculated by assuming that the bin differences follow a normal distribution. Two-sided 99% quantiles are used for the slope and offset. Only the lower side is considered for $R^2$. When the data availability exceeds 40 %, the null hypothesis of normality for the frequency distribution of the differences is rejected in the $\chi^2$ test at a significance

**Table 6.** Certainity of KPIs estimatied using proposed equation.

(a) Wind speed

| Accuracy | Best Practice | | | | | | Minimum | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| KPI | $R^2 > 0.98$ | | | | | | $R^2 > 0.97$ | | | | | |
| $\zeta$[%] | 40 | 50 | 60 | 70 | 80 | 90 | 40 | 50 | 60 | 70 | 80 | 90 |
| $R^2_{MCP}$ 0.965 | | | | | | | | | | | | |
| $R^2_{MCP}$ 0.975 | | | | | | | | | | | | |
| $R^2_{MCP}$ 0.985 | | | | | | | | | | | | |
| $R^2_{MCP}$ 0.995 | | | | | | | | | | | | |
| KPI | $\alpha$ : 0.98～1.02 | | | | | | $\alpha$ : 0.97～1.03 | | | | | |
| $\zeta$[%] | 40 | 50 | 60 | 70 | 80 | 90 | 40 | 50 | 60 | 70 | 80 | 90 |
| $\alpha_{MCP}$ 0.9875 | | | | | | | | | | | | |
| $\alpha_{MCP}$ 0.9925 | | | | | | | | | | | | |
| $\alpha_{MCP}$ 0.9975 | | | | | | | | | | | | |
| $\alpha_{MCP}$ 1.0025 | | | | | | | | | | | | |

(b) Wind direction

| Accuracy | Best Practice | | | | | | Minimum | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| KPI | $R^2 > 0.97$ | | | | | | $R^2 > 0.95$ | | | | | |
| $\zeta$[%] | 40 | 50 | 60 | 70 | 80 | 90 | 40 | 50 | 60 | 70 | 80 | 90 |
| $R^2_{MCP}$ 0.965 | | | | | | | | | | | | |
| $R^2_{MCP}$ 0.975 | | | | | | | | | | | | |
| $R^2_{MCP}$ 0.985 | | | | | | | | | | | | |
| $R^2_{MCP}$ 0.995 | | | | | | | | | | | | |
| KPI | $\alpha$ : 0.97～1.03 | | | | | | $\alpha$ : 0.95～1.05 | | | | | |
| $\zeta$[%] | 40 | 50 | 60 | 70 | 80 | 90 | 40 | 50 | 60 | 70 | 80 | 90 |
| $\alpha_{MCP}$ 0.9875 | | | | | | | | | | | | |
| $\alpha_{MCP}$ 0.9925 | | | | | | | | | | | | |
| $\alpha_{MCP}$ 0.9975 | | | | | | | | | | | | |
| $\alpha_{MCP}$ 1.0025 | | | | | | | | | | | | |
| KPI | $\beta$ : -5～+5 | | | | | | $\beta$ : -10～+10 | | | | | |
| $\zeta$[%] | 40 | 50 | 60 | 70 | 80 | 90 | 40 | 50 | 60 | 70 | 80 | 90 |
| $\beta_{MCP}$ -1 | | | | | | | | | | | | |
| $\beta_{MCP}$ 1 | | | | | | | | | | | | |
| $\beta_{MCP}$ 3 | | | | | | | | | | | | |

Annotations
Predictive (> 99 % reliablility)
Less predictive (< 99 % reliablility)

level of 99.99 %. In this study, a KPI is defined as highly predictive if the 99 % quantile of predicted values within a bin satisfies the set criteria.

Table 6 shows that when the data availability exceeds 60 % for both the wind speed and direction, the confidence interval of the result is 99 % and the final dataset is accurately evaluated by the proposed formulas.

## 4. Conclusions

In this study, formulas for evaluating the $R^2$, slope, and offset of the final dataset after complementation using the MCP method are proposed by considering the data availability of observations. The proposed formulas are then validated against on-site measurements. Formulas for the mean and standard deviation of the post-completion data are derived to account for data availability. Results show that the final dataset is accurately evaluated using the proposed formulas when the availability exceeded 60 % for both the wind speed and direction. Verification against on-site measurements shows that when the data availability of the dataset before complementation is high or when the correlation between the reference data for complementation and the target dataset is high, the KPIs of the final dataset partially complemented by the MCP method is accurately evaluated by the proposed formulas.

## References

[1] Ishihara T, Kawatake T, Arakawa H and Yamaguchi A 2019 Wind climate assessment by using meteorological mast and LiDAR *Proc. of 41st Wind Energy Symp.* 41 pp. 54-7
[2] Gottschall J, Gribben B, Stein D and Würth I 2017 Floating lidar as an advanced offshore wind speed measurement technique: current technology status and gap analysis in regard to full maturity *Wiley Interdiscip. Rev. Energy Environ.* 6 e250
[3] Yamaguchi A and Ishihara T 2016 A new motion compensation algorithm of floating lidar system for the assessment of turbulence intensity *J. Phys. Conf. Ser.* 753 072034
[4] Goit J P, Yamaguchi A and Ishihara T 2020 Measurement and prediction of wind fields at an offshore site by scanning doppler LiDAR and WRF *Atmosphere* 11 442 pp. 1-20
[5] Kelberlau F, Neshaug V, Lønseth L, Bracchi T and Mann J 2020 Taking the motion out of floating lidar: turbulence intensity estimates with a continuous-wave wind lidar *Remote Sens.* 12 898
[6] Kikuchi Y and Ishihara T 2016 Assessment of weather downtime for the construction of offshore wind farm by using wind and wave simulations *J. Phys. Conf. Ser.* 753 092016
[7] Mano A, Ueno A, Itozaki S and Ishihara T 2023 A study of data availability and measurement accuracy of single scanning LiDAR for offshore wind assessment *J. JWEA* **47** pp. 44-54
[8] Watanabe K, Takakuwa S, Hemmi C and Ishihara T 2021 A study of offshore wind assessment using dual-Doppler scanning lidars *J. JWEA* **45** pp. 40-8

[9] Newsom R K, Berg L K, Shaw W J and Fischer M L 2015 Turbine-scale wind field measurements using dual-Doppler lidar *Wind Energy* **18** pp. 219-35
[10] Cheynet E, Jakobsen J B, Snæbjörnsson J, Reuder J, Kumer V, and Svardal B 2017 Assessing the potential of a commercial pulsed lidar for wind characterisation at a bridge site *J. Wind Eng. Indust. Aerodyn.* **161** pp. 17-26.
[11] Carbon Trust 2018 Offshore Wind Accelerator Roadmap for the Commercial Acceptance of Floating Lidar Technology Version 2
[12] Matthias L and Focken U 2006 *Physical Approach to Short-term Wind Power Prediction* 10.1007/3-540-31106-8 (Springer)