Reconstructing long-term wind data at an offshore met-mast location using cyclostationary empirical orthogonal functions

Ji-Young Kim, Kwang-Yul Kim

Korea Electric Power Research Institute, Daejeon, Republic of Korea
School of Earth and Environmental Science, Seoul National University, Seoul, Republic of Korea

Article info
Article history:
Received 22 June 2015
Received in revised form 28 July 2016
Accepted 28 July 2016

Keywords:
Offshore met-mast
Wind data
Reconstructing long-term data
CSEOF

Abstract
For the development of a wind power plant, plant design and its project feasibility analysis are implemented with wind data observed by a met-mast at a target location. Since observation period/time of a met-mast is normally about one year before the plant design, correlation with long-term data existing in the neighborhood of the target location is used for hindcasting past met-mast data to reduce uncertainty in the feasibility analysis, which is called the Measure-Correlate-Predict (MCP) method. In this study, cyclostationary empirical orthogonal function (CSEOF) analysis as a new approach is employed to extend the 1.5-year offshore met-mast HeMOSU-1 data into 34-year long-term data based on the MERRA reanalysis dataset. Both the one- and two-dimensional CSEOF results are compared with that of the widely-used MCP method. The CSEOF method shows a similar level of accuracy to the existing method for mean wind speed, while the former exhibits a slightly better accuracy for the frequency distribution of wind speed and the capacity factor as an index related to the estimation of wind power generation. In additional hypothetical test based on reanalysis datasets, the 1D-CSEOF method shows, in general, a better performance than the conventional MCP method in terms of the accuracy of statistical properties of wind.

1. Introduction
Wind power is an energy source that is in the spotlight, spurring the development of new and renewable energy all around the world. Korea is also experiencing a shortage of electricity, but it is not easy to construct thermal or nuclear power plants anymore because of its small land area. Considering the geographic characteristics of the Korean Peninsula surrounded by the seas on the three sides, there is a rising interest in the development of the offshore areas with relatively rich resources. The Korean government has recently established a development plan for a 2.5 GW offshore wind power plant in the West (Yellow) Sea as an effort to leap as a leading country in offshore wind power. The project and research now has been proceeding with the phase 1 of the test site construction on a scale of 100 MW. An optimal location that is economically feasible and available for a large-scale development has been chosen considering wind resource distribution according to a numerical wind resource map, topographic characteristics such as sea floor depth, and distance from the coastline (Kim et al., 2013). Korea's first offshore met-mast, HeMOSU-1, has been set up and operated near Buan-gun, Jeollabuk-do, and Yeonggwang-gun, Jeollanam-do, where the selected target site was placed in 2010 as shown in Fig. 1. Various meteorological data including wind direction and speed observed at HeMOSU-1 are used in assessing the project feasibility, wind farm layout design, and turbine structure design (Oh et al., 2012). When evaluating the feasibility of wind power project or estimating design load for structure, it is important to characterize the long-term trends and year-to-year variability of observational data.

Since met-mast data generally has a short observation period, around one to two years, past data are reconstructed by extrapolation through regression, known as the measure-correlate-predict (MCP) technique, using neighborhood long-term observation data. Several MCP algorithms have been suggested based on the linear regression method proposed by Derrick (1992), which estimates the relationship between the wind speeds at target and reference sites. Rogers et al. (2005) compared four MCP algorithms in terms of the accuracy of the estimated wind speed and the effect of the length of employed data on the accuracy of estimation. Thøgersen et al. (2007) implemented general-purpose software for four of the MCP methods including a linear regression model; this model is universally accepted so far. Various methods were developed to overcome limitations of the linear regression method, but they do not consistently show improved results in the context...
of the target data characteristics (García-Rojo, 2004; Sreevalsan et al., 2007; Clive, 2008; Carta and Velázquez, 2011; Saavedra-Moreno et al., 2013).

Since most existing methods consider correlation of the wind speed for each bin of wind directions, it is difficult to understand how the principal physical components, such as the diurnal cycle of wind, in the data are calibrated by the regression method. In the present study, a new approach is developed by using cyclostationary empirical orthogonal functions (CSEOFs) in order to carry out regression analysis with respect to individual processes instead of individual wind direction bins. Kim et al. (1996); Kim and North (1997); Kim and Wu (1999), and Kim and Chung (2001) introduced the concept of CSEOF analysis to identify and extract individual spatio-temporal modes from a given dataset.

By carrying out CSEOF analysis, data are decomposed into cyclostationary loading vectors (CSLVs), each of which is modulated by corresponding amplitude or principal component (PC) time series following the naming convention of Kim et al. (1996). Each CSLV reflects a distinct process in the data and is rendered as temporally-varying spatial patterns together with their amplitude fluctuations on longer time scales. This provides a unique means of matching common processes between the target and reference data in terms of their evolution patterns and amplitudes.

Hamlington et al. (2011) used the CSEOF technique to extend significantly the period of sea level reconstructions based on tide gauge data. In a similar manner, the CSEOF technique is applied in this study for the purpose of extending the short-term measurements at HeMOSU-1 in conjunction with long-term reference data. By carrying out regression analysis in CSEOF space and comparing the regressed CSLVs against those derived from the target data, the accuracy of the reference data can be addressed more clearly in the context of major physical modes in the data.

Long-term hindcasting or forecasting of wind data is very difficult due to its irregular variation and various attempts are needed to improve the accuracy of estimation. As mentioned above, a new approach is introduced in the present study as an alternative to the existing MCP methods and to improve the estimation accuracy. This new concept will be addressed in detail in the Method section.

2. Data

HeMOSU-1, which is an offshore met-mast to promote the west coast offshore wind power project, is located on the West Sea about 30 km offshore as shown in Fig. 1. It performs various meteorological observations including wind direction and speed at several altitudes as shown in Fig. 2. HeMOSU-1 has accumulated approximately 2.5 years of observational data starting from February 2011. Since there are few offshore meteorological observatories operating for an appreciable length of time, it is unavoidable to use reanalysis reference data to estimate the long-term wind characteristics at the test site.

In this study, Modern Era-Retrospective analysis for Research and Application (MERRA) data are used as a reference dataset, which is of relatively high spatial and temporal resolutions among the long-term reanalysis datasets (GMAO, 2014). MERRA is a re-processed dataset, into which various satellite-era datasets were blended by using the GEOS-5 atmospheric data assimilation system at Global Modeling and Assimilation Office (GMAO) of NASA. In this study, reference data are used in two different manners—only single-point data near HeMOSU-1 and 2-dimensional spatial array to consider the spatial distribution around the Korean Peninsula. A MERRA grid point at approximately 12.2 km west outbound from HeMOSU-1 is used for a one-dimensional (1D) analysis (see Fig. 1b and Table 1). For a 2-dimensional (2D) analysis, the domain includes offshore area around the Korean Peninsula (32°~39°N × 120°~135.33°E; see Table 1). Data at 50 m above the sea level, which is closest to the hub height of the wind turbine, are used. The details of the MERRA reference data are shown in Table 1. Meanwhile, the observational data at 46.3 m height from the sea level are taken from HeMOSU-1; this level is closest to the elevation of the MERRA data used in the present study. Since there are missing data due to data logger defect for about one month from June 5 to July 14, 2011, only the data after the missing period are used for training, leaving the earlier data before the missing period for verification. We also used the 1.5° × 1.5° European Center for Medium-Range Weather Forecasts Reanalysis (ERA) Interim product (Dee et al., 2011) in a hypothetical test.

3. Method

3.1. CSEOF analysis

It is useful to decompose complex physical system into simpler basis functions in order to understand the basic physical processes in the physical system. The CSEOF technique is introduced to separate individual physical processes from the datasets. This technique has been used in many previous studies and the details can be found in Kim et al. (1996, 2015); Kim and North (1997); Kim and Wu (1999), and Kim and Chung (2001). In CSEOF analysis, space-time data, \(T(r, t)\), are expressed as a unique linear superposition

![Fig. 1. Locations of the measurements (HeMOSU-1) and one-point reference (MERRA).](image-url)
where $LV_n(r, t)$ and $PC_n(t)$ are cyclostationary loading vectors (CSLVs) and principal component (PC) time series, respectively. Unlike EOF analysis, LVs are time dependent and periodic in CSEOF analysis, that is,

$$LV_n(r, t) = LV_n(r, t+d),$$

where $d$ is called the nested period and represents the periodicity of the space-time covariance function. Each CSLV expresses a temporally evolving physical process, and corresponding PC time series denotes the amplitude modulation of the physical process. In this study, the nested period is set to 24 h in order to understand the diurnal structures of variation of wind. The corresponding PC time series, then, show the amplitude fluctuations on time scales longer than a day.

### 3.2. Regression analysis in CSEOF space

Similar to the conventional MCP method, regression analysis is conducted between two sets of CSEOF PC time series to generate physically consistent evolution patterns between two variables or datasets. This procedure is called the regression analysis in CSEOF space. A crucial difference from the existing methods is that regression analysis is conducted on the PC time series of CSEOF modes instead of the raw data. In other words, amplitude variation of a physical process in one variable is matched with that in another variable. This is a sensible approach since physical processes are rendered differently in two different variables (datasets), and henceforth statistical relationship should be identified for each mode (physical process) instead of regression relationship between two raw datasets.

Let $T(r, t)$ be a target variable and $P(r, t)$ be a predictor variable. CSEOF analysis yields:

$$T(r, t) = \sum_n LV_n(r, t)PC_n(t),$$

and

$$P(r, t) = \sum_n C_n(r, t)B_n(t),$$

where $B_n(r, t)$ and $C_n(r, t)$ are respectively the CSLVs of the target and predictor variables, and $T_n(t)$ and $P_n(t)$ are respectively the PC time series of the target and predictor variables, $n$ is the mode number, and $r$ and $t$ denote space and time. In general, there is no one-to-one correspondence between the two sets of PC time series. That is, $T_n(t)$ and $P_n(t)$ are generally not identical for each $n$. Thus, regression analysis should be conducted in CSEOF space to derive physically consistent evolutions of the predictor variable with those of the target variable.

The procedure consists of two steps:

$$T_n(t) = \sum_{m=1}^{M} a_{nm}^m B_n(t) + \epsilon_n^m(t),$$

where $a_{nm}^m$ are the regression coefficients that are determined through regression analysis in CSEOF space.
and
\[ D_n(r, t) = \sum_{m=1}^{M} a_{nm}^o c_m(r, t), \]
(6)
where \( a_{nm}^o \) are regression coefficients and \( e \equiv (t) \) is regression error time series. As a result of (5) and (6), one can write the predictor variable as
\[ P(r, t) = \sum_n D_n(r, t) t_n(t). \]
(7)

The right-hand side is not exactly equal to the left-hand side because the regression error in (5) is not exactly zero. In fact, this deviation becomes greater as regression error variance, which is measured by \( 1 - R^2 \) value, increases. The \( R^2 \) value is defined by
\[ R^2 = 1 - \frac{\sum e_n(t) e_n(t)}{\sum t_n(t) t_n(t)}, \]
(8)
which measures the magnitude of regression error variance with respect to that of the total variance of the target time series. Thus, regression error variance approaches zero as \( R \) approaches 1. In CSEOF analysis, \( R^2 \) values are, in general, close to unity, which implies that physical relationship between two variables is well established.

Forecasting (hindcasting) wind field at a met-mast location proceeds in the following manner. Once the regression relationship in (5) is determined for the training period, met-mast data can be extended by using
\[ \hat{t}_n(r, t) = \sum_n B_n(r, t) \hat{t}_n(t), \]
(9)
where \( B_n(r, t) \) is the CSLV derived from the target variable (met-mast), and \( \hat{t}_n(t) \) is the target PC time series estimated from the predictor PC time series of the reference data (MERRA) by using (5). While the loading vectors, representing specific physical evolutions, derived from the target and the reference data as in (3) and (7) may not necessarily be identical, long-term modulation may be close to each other if the reference dataset is reasonably accurate; this, of course, is an important caveat. Another important caveat is that physical processes, \( B_n(r, t) \), remain identical in the forecasting (hindcasting) period. Thus, the amplitude time series can be derived from the reference data for the prediction interval, which in conjunction with the target loading vectors are used to generate “synthetic” wind field for the prediction interval. We will test the validity and accuracy of this procedure in the next section.

4. Results

4.1. Comparison between the target and reference data

Two datasets, HeMOSU-1 measurements and MERRA reanalysis data, are compared to evaluate their proximity before regression analysis. A strong correlation is anticipated since the observation site is located offshore with little influence from the Korean Peninsula or China, and the MERRA grid point is relatively close to the observation site. Fig. 3 shows the time series of the wind components at the reference point against the target data (HeMOSU-1). As shown, correlations for both the wind components are reasonably high. The reference data, however, tend to show lower amplitudes than the target data, which is a general characteristic of model datasets. Fig. 4 shows the spatial patterns of correlations, which show the spatial correlation scales of the zonal and meridional winds around the Korean Peninsula with respect to the target time series. As shown in Fig. 4, relatively high correlation values are found over a domain of considerable size; correlation length scale increases slightly over the marginal seas due most likely to weaker influence by topography. The zonal component is weaker than the meridional component of wind (Fig. 3), and the former is less correlated with the target data than the latter (Fig. 4). Fig. 4 shows that the reference data in the region of high correlations can be used to improve long-term hindcasting based on one-point reference data.

Since regression analysis in CSEOF space is conducted for individual modes, CSLV and PC time series of each mode are compared between the target data and reference data. Because of the relatively short length of the target data, it is not plausible to compare the long-term trend in the datasets. Instead, short-term (daily) physical evolutions including the diurnal cycle and longer-term evolutions of their amplitudes have been extracted by using the nested period of one day. The first two modes together explain more than 70% of the total variability both in the target and reference data. As shown in Fig. 5, the first CSEOF mode represents a predominant meridional wind throughout the day whereas the second CSEOF mode depicts a predominant zonal wind throughout the day both in the target and reference datasets. Diurnal variation is relatively small in the two modes.

The PC time series measuring the amplitudes of the LVs are also similar between the two datasets for the first two modes (see Table 2). Fig. 6 shows the PC time series of the first CSEOF mode. As can be seen, the first PC time series derived from the 1D MERRA data is similar in evolution and magnitude to that of the target data. For the 2D MERRA data, on the other hand, the amplitude of the PC time series often differs appreciably from the target data while the evolution pattern is similar. This may reflect that the strength of wind often varies irregularly in the analysis domain. Considering the complex terrain of the Korean Peninsula, such non-uniform variation of the wind speed should be expected. The PC time series of the third CSEOF mode derived from the 2D MERRA data exhibits negligible correlation with that of the target data; this indicates that there is, in general, no one-to-one correspondence between two sets of CSEOF modes derived from two
Regression analysis in CSEOF space is needed to establish one-to-one correspondence between the two sets of CSEOF modes.

4.2. Regression analysis

Regression analysis in CSEOF space was conducted as delineated in (5) and (6). Regression coefficients were computed for each PC time series over the training period, in which the target and the reference data coexist. It needs to be determined how many target PC time series should be used for regression in order to reconstruct optimal wind data. Correlation coefficients between the target and the reconstructed data for each wind component were calculated as a function of the number of modes used for regression (Fig. 7 and Table 3). In reconstructing the wind field, the regressed PC time series are used in conjunction with the target CSLVs as in (9). In other words, regressed PC time series are constructed based on (5) from the reference data and are multiplied with the target CSLVs in order to generate data over the validation or prediction period. As can be seen in Table 3, increase in correlation for each wind component is less than 0.01 when additional mode is added beyond the mode number 5; results are similar for both the 1D and 2D cases. The first 5 modes of the target data together explain more than 90% of the total variability. Consequently, the first 5 modes are used to reconstruct the regressed data for both the 1D and 2D cases. Specifically, correlations of \( u \) and \( v \) increased respectively by 5.6% and 4.3% for MERRA-1D analysis, and 27.6% and 14.0% for MERRA-2D analysis based on the 5-mode reconstructions (Table 3).

As should be expected, correlation increases significantly after the regression as a comparison between Figs. 6 and 8 shows (see also Tables 2 and 4). The correlation for the verification period is not less than that of the training period for the MERRA-1D analysis. This reflects that the regression relationship established over the training period holds fairly well over the verification period. Meanwhile, correlation over the verification period is somewhat lower than that over the training period for the third CSEOF mode in the MERRA-2D analysis; on the other hand, it is a significant improvement over the value before regression. It is worth mentioning that the PC time series of the MERRA-2D data can be fitted to the PC time series of HeMOSU-1 data nearly as well as those of the MERRA-1D data via regression in CSEOF space. Thus, in theory, it is possible to generate reasonable spatial patterns based on measurements at a single station, or use neighboring measurements to improve reconstruction at a target station.

The results of regression in CSEOF space can easily be understood by comparing Figs. 9 and 10, and Figs. 11 and 12 (see also Table 3). The regression relationship is fairly reasonable and the accuracy of the 5-mode reconstructions is generally improved both for the MERRA-1D and MERRA-2D experiments. Figs. 10 and 12 clearly show that the regression in CSEOF space improved the 5-mode reconstructions both in terms of the correlation and the magnitude of the wind components in both the verification and training periods. When using the MERRA-1D reference data, correlation of \( u \) and \( v \) increased respectively by 4.5% and 3.5%, and 1.6% and 1.0% for the MERRA-2D data (Table 5). Although the increase after the regression is relatively low, there is not much room for a significant improvement; correlation of the reference data is already rather high before regression. It should be noted that correlation of \( v \) is over 90% after the regression for the MERRA-1D data. When using the MERRA-2D data, correlation is somewhat lower than the MERRA-1D data, but correlation over the analysis domain increased after regression as a comparison between Figs. 6 and 8 shows (see also Table 3). It has been demonstrated that a wind field having a higher correlation than the raw reference data could be reconstructed based on regression analysis in CSEOF space. In this section, the performance of the CSEOF-based method is compared with that of the widely-used MCP method. In the MCP method, linear regression is conducted between two sets of wind speed data according to the wind directional sector. Fig. 14 shows the monthly mean values of the reconstructed wind speed after regression against those of the HeMOSU data. As shown in Fig. 14a, all three results are fairly consistent with the target data and reproduce the wind-speed variability in a reasonable manner. A close examination shows that the period of the best performance differs from one method to another. There were some missing data in June and July 2011; as a result, wind speed is underestimated in this period. During a strong typhoon, Boloven, reaching a maximum wind speed of 50 m s\(^{-1}\) in August 2012, wind speed again is underestimated. Long-term variability at the target station is estimated from the MERRA-1D data as shown in Fig. 14b. As shown in the figure, it is possible to improve reliability and assess uncertainty of the expected energy production based on a longer estimation of wind data in feasibility assessment of wind power development. Various statistics have been compared in order to evaluate the three different methods quantitatively (Table 6). Mean Bias in the MERRA raw data (reference) has been reduced in all three methods. The regressed wind also exhibits slightly higher correlations.
and lower relative RMSE except for the 2D CSEOF method. For the MERRA-2D data, CSEOF modes depict variability over a wide area and may not solely reflect variability at the met-mast location. As a result, both correlation and RMSE deteriorate slightly. Skewness measures the asymmetry of the probability distribution; positive skewness indicates that the distribution has longer or fatter tail on the right side and vice versa. All the data including the MERRA raw data (reference) exhibit the same sign of skewness. Kurtosis is a measure of “tailedness” of a probability distribution; distributions with kurtosis less than 3 are associated with fewer and less extremes and vice versa. The CSEOF-based methods produce the same sign of “excess” kurtosis (kurtosis minus 3) for all three variables (zonal wind, meridional wind, and wind speed).

Fig. 15 shows a comparison of the HeMOSU data and the regressed data from the MCP and CSEOF-1D methods. The figure shows that both the methods reproduce the histograms of wind reasonably. The

![Loading vectors for zonal (left) and meridional (right) components of wind in the target and the reference data for the first two CSEOF modes.](image)

### Table 2
Correlation of the PC time series between the target and reference data for the first five CSEOF modes: $R_1$ denotes the verification period and $R_2$ the training period.

<table>
<thead>
<tr>
<th>Mode</th>
<th>With MERRA-1D</th>
<th>With MERRA-2D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R_1$</td>
<td>$R_2$</td>
</tr>
<tr>
<td>1</td>
<td>0.9397</td>
<td>0.9063</td>
</tr>
<tr>
<td>2</td>
<td>0.8086</td>
<td>0.8905</td>
</tr>
<tr>
<td>3</td>
<td>0.6518</td>
<td>0.7633</td>
</tr>
<tr>
<td>4</td>
<td>0.6647</td>
<td>0.7195</td>
</tr>
<tr>
<td>5</td>
<td>0.5788</td>
<td>0.5972</td>
</tr>
</tbody>
</table>

Fig. 5. The PC time series of the first CSEOF mode for the target and the (a) MERRA-1D, and (b) MERRA-2D reference data.
The histogram of wind speed is skewed to the right and exhibits a long tail. The histograms of zonal and meridional components of wind are nearly Gaussian with small excess kurtosis. The directional distribution of winds from the regressed data also looks reasonably similar to that of the HeMOSU (target) data. Winds are predominantly in the

Table 3

<table>
<thead>
<tr>
<th>Mode number</th>
<th>Target vs. reconstructed reference data (before regression)</th>
<th>Target vs. reconstructed regressed data (after regression)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MERRA-1D</td>
<td>MERRA-2D</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>v</td>
</tr>
<tr>
<td>1</td>
<td>0.5256</td>
<td>0.7614</td>
</tr>
<tr>
<td>2</td>
<td>0.7111</td>
<td>0.8284</td>
</tr>
<tr>
<td>3</td>
<td>0.7589</td>
<td>0.8451</td>
</tr>
<tr>
<td>4</td>
<td>0.7745</td>
<td>0.8616</td>
</tr>
<tr>
<td>5</td>
<td>0.7798</td>
<td>0.8704</td>
</tr>
<tr>
<td>6</td>
<td>0.7871</td>
<td>0.8739</td>
</tr>
<tr>
<td>7</td>
<td>0.7887</td>
<td>0.8757</td>
</tr>
</tbody>
</table>

Table 4

<table>
<thead>
<tr>
<th>Mode number</th>
<th>Target vs. reconstructed reference data (before regression)</th>
<th>Target vs. reconstructed regressed data (after regression)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MERRA-1D</td>
<td>MERRA-2D</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>v</td>
</tr>
<tr>
<td>1</td>
<td>0.9717</td>
<td>0.9548</td>
</tr>
<tr>
<td>2</td>
<td>0.8926</td>
<td>0.9143</td>
</tr>
<tr>
<td>3</td>
<td>0.7011</td>
<td>0.7886</td>
</tr>
<tr>
<td>4</td>
<td>0.7333</td>
<td>0.7439</td>
</tr>
<tr>
<td>5</td>
<td>0.6723</td>
<td>0.6712</td>
</tr>
</tbody>
</table>

Fig. 7. Correlation coefficients between the target and the regressed data as a function of the number of modes used for reconstruction for each wind component: (a) MERRA-1D, and (b) MERRA-2D data.

Fig. 8. A comparison of the target and the regressed PC time series for the first CSEOF mode: (a) MERRA-1D, and (b) MERRA-2D data.

Fig. 9. Comparison between the HeMOSU and the MERRA-1D data based on the 5-mode reconstructions before regression: (a) zonal wind, and (b) meridional wind.

The histogram of wind speed is skewed to the right and exhibits a long tail. The histograms of zonal and meridional components of wind are nearly Gaussian with small excess kurtosis. The directional distribution of winds from the regressed data also looks reasonably similar to that of the HeMOSU (target) data. Winds are predominantly in the
east-south-east (ESE) direction with slightly less frequent easterlies (W direction). The wind speed and its standard deviation as a function of wind direction also look fairly reasonable with strong winds and their variability in the W direction.

Actual wind power generation is assessed for each method. Since the frequency distribution of wind speed is used when estimating the amount of generation, the frequency distribution of wind speed and the resulting power generation are compared for individual estimates. A wind speed frequency distribution model most frequently used in estimating wind power generation is the Weibull distribution (IEC, 2005a):

$$f(u) = \frac{k}{c} \left( \frac{u}{c} \right)^{k-1} \exp \left[ -\left( \frac{u}{c} \right)^k \right].$$

(10)

where $u$ is wind speed, $c$ is the scale factor, $k$ is the shape factor.

Amount of power generation is calculated by the Annual Energy Production (AEP) (IEC, 2005b):

$$\text{AEP} = 24 \times 365 \times \sum_{u=0}^{\text{cutout}} P(u) f(u),$$

(11)

where $P(u)$ is turbine output (kW) when wind speed is $u$, and $u_{\text{cutout}}$ is the maximum wind speed for turbine operation. The capacity factor, then, is the ratio of the calculated AEP at the target site according to the wind distribution given by (10) to the theoretical maximum AEP; it is considered a crucial indicator for feasibility assessment of wind power plants. Comparison of the mean wind speed, parameters for the Weibull distribution, and the

Fig. 10. Comparison between the HeMOSU and the MERRA-1d data based on the 5-mode reconstructions after regression: (a) zonal and (b) meridional wind components.

Fig. 11. Comparison between the HeMOSU and the MERRA-2d data based on the 5-mode reconstructions before regression: (a) zonal wind, and (b) meridional wind.

Fig. 12. Comparison between the HeMOSU and the MERRA-2d data based on the 5-mode reconstructions after regression: (a) zonal and (b) meridional wind components.

Table 5

<table>
<thead>
<tr>
<th>Variable</th>
<th>Before regression</th>
<th>After regression, MERRA-1d</th>
<th>After regression, MERRA-2d</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$</td>
<td>$R_1$</td>
<td>$R_2$</td>
<td>$R_1$</td>
</tr>
<tr>
<td>$R_1$</td>
<td>0.7289</td>
<td>0.7882</td>
<td>0.8375</td>
</tr>
<tr>
<td>$R_2$</td>
<td>0.8987</td>
<td>0.8775</td>
<td>0.9227</td>
</tr>
</tbody>
</table>
The capacity factor is calculated by applying the power curve of a 3 MW turbine of company D. As can be seen in Fig. 16, the frequency distribution of the estimated wind speed is reasonably similar to that of the HeMOSU data. As can be seen in Table 7, the CSEOF reconstruction based on the MERRA-1D data is closest to the target data. The accuracy of capacity factor derived from the existing MCP method is 94.8% whereas the 1D CSEOF method yields 96.5%; there is a slight improvement in the accuracy of the reconstructed wind speed. Overall accuracy is lowest for the wind field generated by the CSEOF method based on the MERRA-2D data. Nonetheless, the accuracy of estimated capacity factor is 93.5%, indicating that the wind field based on the CSEOF analysis of the MERRA-2D data is still useful for estimating wind power generation.

### Table 6

Mean bias, correlation, relative RMSE, skewness, and kurtosis of the hourly zonal wind, meridional wind, and wind speed for the MERRA raw data and the 5-mode reconstructions of regressed data derived from the three different methods in comparison with the HeMOSU (target) data. This comparison is for the verification period.

<table>
<thead>
<tr>
<th>Data</th>
<th>Mean Bias</th>
<th>Corr.</th>
<th>Rel. RMSE</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>HeMOSU</td>
<td></td>
<td>0.692</td>
<td></td>
<td>0.494</td>
<td></td>
</tr>
<tr>
<td>MERRA raw</td>
<td></td>
<td>0.459</td>
<td>0.729</td>
<td>0.441</td>
<td>0.390</td>
</tr>
<tr>
<td>MCP</td>
<td></td>
<td>0.206</td>
<td>0.825</td>
<td>0.357</td>
<td>-0.392</td>
</tr>
<tr>
<td>1D-CSEOF</td>
<td></td>
<td>0.387</td>
<td>0.837</td>
<td>0.395</td>
<td>0.615</td>
</tr>
<tr>
<td>2D-CSEOF</td>
<td></td>
<td>0.103</td>
<td>0.891</td>
<td>0.355</td>
<td>-0.488</td>
</tr>
</tbody>
</table>

Fig. 13. Correlation of (a) zonal and (b) meridional wind components between the HeMOSU and the 5-mode reconstruction based on the regressed MERRA-2d data.

Fig. 14. Comparison of monthly mean wind speed between the target (HeMOSU) and the regressed data for (a) the training and verification periods and (b) the prediction period.
4.4. Additional test in different settings

As shown in the test results derived from the offshore met-mast and MERRA data, the performance of the new CSEOF method applied to the MERRA-1D data is as good as the conventional MCP method. In order to confirm that the performance of the 1D CSEOF method is not merely coincidental, additional hypothetical test was conducted. Since we do not have another set of tower measurements, MERRA data near the HeMOSU location is used as the target variable and ERA-Interim data (Dee et al., 2011) as the reference variable (Fig. 17 and Table 8). This is a rather stringent test, since the two locations, separated by ~400 km, exhibit fairly distinct wind patterns. MERRA data at station 1 are generated for the period of 2000–2015 using the ERA-Interim data at station 2 as the predictor variable using both the CSEOF and conventional MCP methods. Data for a 2-year period (2014–2015) are used for training.

Fig. 18 shows the zonal and meridional components of wind reproduced by the two methods in comparison with the target and reference data for a two-month period. Correlations of $u$ and $v$ increased respectively by 12.0% and 10.6% for the results by CSEOF method, whereas correlation of $u$ decreased by 3.6% and $v$ increased by 5.9% for the results by MCP method (Table 9). Relative RMSE also decreased slightly for both the modes. Table 10 shows a comparison between the reconstructed wind speed against the actual wind speed in terms of the mean and standard deviation. The statistics comparing the sample means is defined by the two-sample $t$ statistic:

![Figure 15](image1)

![Figure 16](image2)
where $\bar{x}_1$ and $\bar{x}_2$ are sample means, $S_1^2$ and $S_2^2$ are sample variances, $\mu_1$ and $\mu_2$ are true means, and $n$ and $m$ are sample sizes. The degree of freedom is $k$, which is either the smaller of $n-1$ and $m-1$ or the calculated degree of freedom. The test statistic

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2}$$

has an $F$ distribution with $n-1$ and $m-1$ degrees of freedom. Here $\sigma_1^2$ and $\sigma_2^2$ are true variances of two samples. The degree of freedom for both distributions is assumed to be 5845 ($\frac{1}{7} \times 40,912$), since the autocorrelation time scale is close to 7 points and total data points over the valid period are 40,912. The mean of wind speed for the validation period is respectively 5.369, 5.380 and 5.100 for the MERRA data, the reconstructed data based on the CSEOF method, and the regressed data based on the MCP method. The variance of the wind speed is given by 8.476, 7.973, and 4.114, respectively. Then, the test results are summarized as in Table 10.

As can be seen in the table, the wind speed data derived from the CSEOF method agree better with the MERRA data.

Fig. 19 shows the comparison of the MERRA station 1 data (target) and the regressed data from the ERA-Interim station 2 data (reference) based on the MCP and CSEOF-1D methods. The histograms of the zonal and meridional components of wind show that the MCP method significantly underestimates the very weak wind events and overestimates the slightly stronger wind events. This is the reason for a significant overestimation in the wind speed range of $3-7 \text{ m s}^{-1}$. The directional distributions of winds from the regressed data look different from that of the MERRA data.
CSEOF method produces a higher occurrence of wind in the SE direction and to a lesser extent in the SSW direction instead of prevalent wind in the SSE direction. The MCP method produces a large fraction of wind in the NE direction and to a lesser extent in the SW direction, which is not consistent with the MERRA data (target). The angular distribution of wind speed seems reasonable although the direction of maximum wind speed differs among the three datasets. The regressed data generally underestimates the standard deviation of wind speed. This underestimation is more significant for the MCP method.

5. Summary and conclusions

In order to extend the wind measurements from an offshore met-mast, HeMOSU-1, a new method was developed to construct wind at the target position based on the MERRA data. The MERRA data, which is a long-term reanalysis dataset, were used as the reference data in the present study. Based on the regression analysis in CSEOF space, wind field was constructed at the target location using the reference dataset. The accuracy of the constructed wind field was evaluated against that of the commonly used MCP method.

For the wind field based on the regression analysis in CSEOF space, correlation of \( u \) and \( v \) increased from that of the raw reference data by 4.5% and 3.5% respectively when using the MERRA-1D data, and 1.6% and 1.0% respectively when using the MERRA-2D data. It should be pointed out that the improvement in correlation is small, since the raw MERRA data are already fairly close to the met-mast data. A comparison with the existing MCP method shows that the accuracy, measured in terms of correlation and RMSE, of the reconstructed wind field is similar between the two methods. In the accuracy of the Weibull frequency distribution of wind speed, the CSEOF-based method is quite comparable to the MCP method. The capacity factor, an important indicator for wind power generation, is slightly improved by CSEOF analysis of the MERRA-1D data.

CSEOF analysis based on the MERRA-2D data results in the lowest accuracy, even lower than the raw MERRA data. CSEOF modes derived from the MERRA-2D data reflect variability not only at the met-mast location but also in the entire data domain. It is obvious, therefore, that the long-term variability in the PC time series derived from the MERRA-2D data may not necessarily be strongly correlated with the variability at the met-mast location. Thus, regression error, in general, is larger for 2D CSEOF analysis than for 1D analysis. It should be pointed out, however, that the emphasis of the 2D CSEOF analysis is different from that of the 1D CSEOF analysis. In some applications, wind field and other physical variables in the area surrounding the target location provide beneficial information for power plant operations. Through 2D CSEOF analysis, optimal locations of future met-masts can be determined without much redundancy with existing met-masts.

As demonstrated in this study, the performance of the new CSEOF method applied to the MERRA-1D data is as good as the conventional MCP method in explaining the wind variability at the HeMOSU site. As a new method, however, more tests and comparisons in different settings are certainly necessary. In particular, the CSEOF-based method, both one- and two-dimensional versions, needs to be evaluated in more stringent settings, such as the estimation of wind field at a mountainous location, which is currently not possible due to the lack of available met-mast measurements around the Korean Peninsula. In a hypothetical and more stringent experiment, the 1D-CSEOF method reproduces wind in the MERRA data (target) reasonably based on the ERA-Interim reference data located ~400 km away from the target location. The 1D-CSEOF method generally performs better than...
the MCP method in terms of the veracity of histograms and directional properties of wind.

Acknowledgments

This work was supported by a New & Renewable Energy of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government Ministry of Knowledge Economy (Project no. 20113040020010).

References