The Crucial Records Number to Retrieve Offshore Directional Wind Distribution

X Zhu[1, 2], Z Li[1] and X Yang[1]

[1] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, 100101 Beijing, China

[2] University of Chinese Academy of Sciences 100040 Beijing, China

zhuxh@radi.ac.cn

Abstract. The wind energy production estimates are very important to a wind power project. And, the remote sensing technique has been widely used to obtain the offshore wind speed and direction which could be used to calculate the wind energy of potential wind farm. However, the directional wind energy distributions are rarely studied, which also play important roles in analysis of wind farms' potential power. In this article, the minimum number of records to obtain offshore directional wind distribution is stated by simulation experiment on In-situ dataset. The NDBC buoy dataset is randomly and multiply sampled to build new dataset under different numbers of observation records, which vary from 21 to 800. The resample under the same number of observation is repeated for 100 times to build dataset group. The directional wind distribution of new dataset is compared with the one of original buoy dataset, and errors made by dataset with fewer records are calculated. Besides, the 10th largest error in the sampled dataset group, which have the same number of observation records, is regarded as the error bound for those dataset. The change rule of the error bound is shown by fitted curves. Based on the fitted curves, minimum number of records is calculated. By this simulation experiment, the minimum number of records to represent wind direction frequency is 350, and 800 for annual direction distributions of wind energy density. To reduce the number of records needed in retrieval, some methods are discussed and tested.

1. Introduction

1.1 *Research Background*

Wind energy brings about ecological and social benefits and contributes to sustainable energy development. Using electricity from wind turbine instead of electricity from coal power plant could save an average of about 820-910 tonnes of CO_2 for every GWh [1]. The wind is usually stronger and steadier in offshore areas, making offshore wind energy a major trend in wind power industry [2]. And, offshore directional wind distribution plays a valuable role in this trend. Using wind direction distributions information to analysis potential wind power of a specific site plays an important role when selecting the location of new wind energy facilities, and has impacts on both, price of generated energy and return of investment [3]. Some researches about how to use the multiple satellite data to assess offshore wind resources have been done [4]. The mean wind speed and wind power density are the focus of attention in those studies. The parameters of directional wind distribution, however, have not been sufficiently studied.

1.2 *Parameters*

According to the Chinese national standard GB/T 18710-2002, the direction distributions used to investigate wind farm site are annual wind direction frequency (DF) and annual direction distributions of wind energy density (DDED). Both parameters could be shown by rose diagram. (Showed in Figure 1) Those two distributions both have 16 directions.

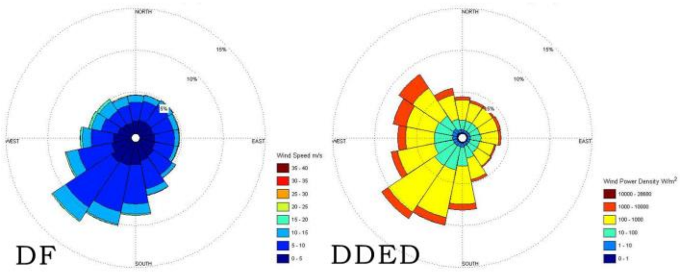

Figure 1. Rose diagram of two wind direction distributions

The equation of annual wind direction frequency:

$$DF_j = \frac{n_j}{\sum_{j=1}^{16} n_j} \tag{1}$$

where DF_j = the frequency of direction j(When j=1, j means north. When j=5, j means east, etc.)

n_j = the number of wind records whose directions are j

The equation of annual direction distributions of wind energy density:

$$DDED_j = \frac{D_j}{D} \tag{2}$$

$$D = \sum_{j=1}^{16} D_j \tag{3}$$

$$D_j = \frac{1}{2}\sum_{i=1}^{m} (\rho)(v_{ji}^3) t_{ji} \tag{4}$$

where $DDED_j$ = the wind energy density distribution of direction j

D_j = the wind energy density of direction j

D = the total wind energy density

ρ = the air density

v_{ji} = the wind speed of record NO.i in direction j

t_{ji} = the duration represented by the record

m= the records number of direction j

1.3 *Research Purposes*

In this paper, how many records are needed to retrieve reliable directional wind distributions is discussed. Moreover, some approaches to reduce the record number in retrieval are presented. And, the results of those approaches are also showed.

2. Data Description

2.1 *In-situ buoy measurements*

The In-situ wind measurements at 3~10 m above the ocean surface, with mean values at time intervals of 1 h are provided by National Data Buoy Centre (NOAA/NDBC). The detail of station information is described in Table 1. Those four buoys include diverse offshore wind condition, giving this study

enough representative records. All the in-situ data used in this study were checked for quality control to delete inaccurate and unrealistic records [5].
The criteria to discard unreliable records:

$$|\vec{u}| < 0.4ms^{-1}; |\vec{u}| > 75ms^{-1} \tag{5}$$

$$\text{When} \quad |\vec{u_i}| > 15ms^{-1}, |\vec{u_i}| > |\overrightarrow{u_{i-1}}| \times 1.5 \;\; \text{and} \;\; |\vec{u_i}| > |\overrightarrow{u_{i+1}}| \times 1.5 \tag{6}$$

$$0^{\circ} > \theta > 360^{\circ} \tag{7}$$

where, $\vec{u}$ = the wind vector
θ = the angle of wind direction

Table 1. The description of In-situ wind measurements

Station	41048	42040	46001	46086
Location	31.87°N 69.57°W Atlantic	29.21 °N 88.21 °W Gulf of Mexico	56.30 °N 147.92 °W Gulf of Alaska	32.49 °N 118.04 °W Pacific
Sea water depth	5340 m	164.6 m	4206 m	1828.8 m
Anemometer height	5 m	10 m	5 m	5 m
Buoy	3-meter discus	10-meter discus	6-meter NOMAD	3-meter discus
Payload	ARES	ARES	ARES	ARES
Data period	2008-2014	2005-2014	2005-2013	2005-2014

2.2 *ASCAT wind observations*
The ASCAT L2 12.5-km swath near real-time wind products are provided by the NASA/JPL Physical Oceanography Distributed Active Archive Centre [6].The wind products contain quality control information, according to which unreliable records will be deleted. Because of the orbit characteristics of the satellite, MetOp, which carries the ASCAT, one specific sits' wind records only include several specific hours' data. In this study, ASCAT wind products in 2010 are used to analyse the effectiveness of approaches to reduce the minimum number of records.

2.3 *Data analysis*
To get the basic understanding of those data, the directional wind distribution parameters of those buoy observation dataset groups are calculated. And, circular statistics of those distributions are also analysed, which include average, standard deviation, skewness and kurtosis. These analysis results are showed in Figure 2 and Table 2. Moreover, the time series characteristic of ASCAT data is analysed in Figure 3, which prepares for the research work of over-sample.

Table 2. The circular statistics of In-situ wind datasets

Station	41048	42040	46001	46086
		Circle statistics of DF		
Mean	212.57°	100.42°	226.00°	293.30°
Std dev	72.39°	72.88°	68.87°	46.20°
Skewness	0	0.06	0.12	0.06
Kurtosis	0.12	0.04	0.02	0.46
		Circle statistics of DDEF		
Mean	245.18°	61.35°	226.28°	292.94°
Std dev	70.55°	71.76°	69.86°	34.44°
Skewness	-0.06	0.02	0.20	0.11
Kurtosis	0.04	-0.15	-0.07	0.71

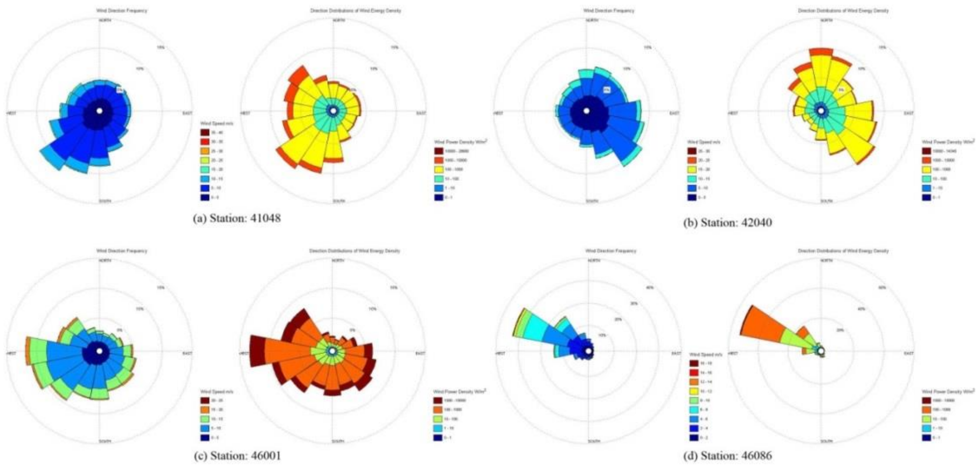

Figure 2. The DF and DDED of those stations

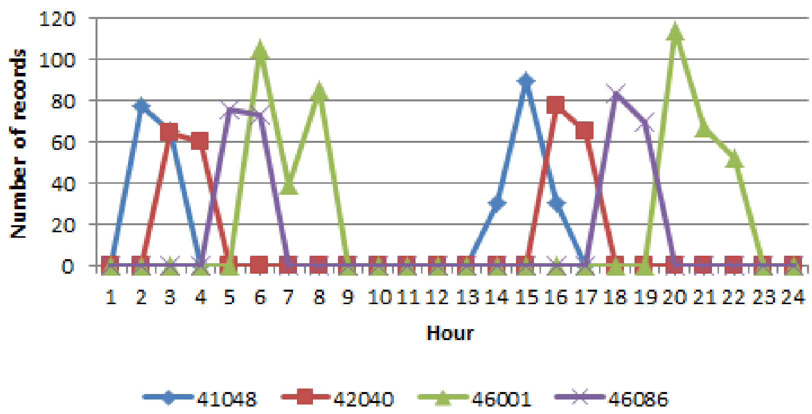

Figure 3. The time series of ASCAT data

3. Methodology

3.1 *Fitting of errors change and finding the crucial records number*

To examine the influence of the record number on retrieval accuracy, the dataset from In-situ buoy measurements was randomly and multiply sampled for a range of number of observations from n=21 (assumed to be the lower bound on the dataset likely to be obtained using remote sensing) to n=0.1 of the actual number of observations available from buoy for the annual data collection period [7]. The resampling was undertaken 100 times for each n to build 100 data sets for every n, number of records. Those new-built data sets' offshore directional wind distributions, DF and DDED, are calculated and compared with the one of the original dataset, which is regarded as real distribution. In this compare, the total difference of distribution's 16 directions is summed and normalized to demonstrate the error caused by fewer observations. Obviously, the error reduces swiftly with the number of records' increase. To avoid the extreme condition, the 10th largest error in the group of 100 new-built data sets of the same n is regarded as this records number's error bound at a confidence level of 90%. One sample is presented in Figure 4(a) and (c).

Different types of fitted curves are tested to give a good fit to the change rule of error bound. In these fitted curves, the 2-terms exponential function gives superior and robust fit. The fitted curve of one sample is presented in Figure 4(b) and (d). With this function, the minimum number of records at which the error bound decline to below 10% could be calculated. 2-terms exponential function:

$$f(x) = a \times e^{bx} + c \times e^{dx} \tag{8}$$

where, a, b, c and d are parameters in fitted function

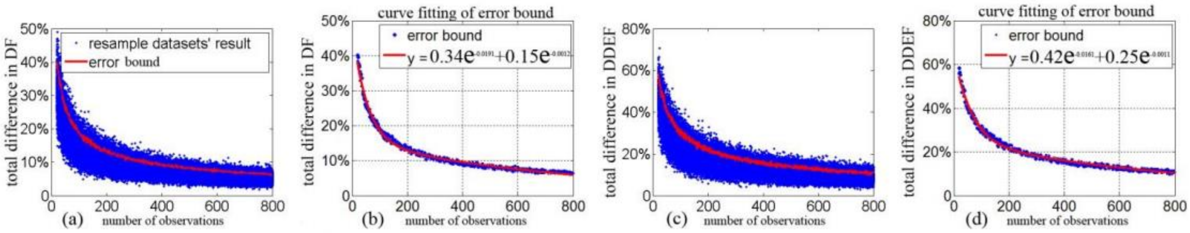

Figure 4. The error bound of sampled dataset and fitted curve

3.2 *Over-sample*

The different revisiting period of satellites makes them observe the same site different times in one year, and, the polar orbit satellite usually revisiting a specific place at fixed local time. As a result of this, when multiple satellite data is used to represent directional wind distribution, the number of records of different hours might be imbalanced. Because wind observations at some offshore area exhibit significant diurnal cycles due largely to advective effects [8], this imbalance data might distort the retrieval results. To handle with imbalanced data, the sub-dataset of hour with fewer records is over-sampled. Moreover, over-sampling also adds new records to dataset which might reduce the number of records of original dataset needed to obtain reliable distribution.

The over-sample technique used in this research work is improved SMOTE over-sample algorithm [9]. In the improved SMOTE method, new synthetic samples created by randomly interpolating pairs of selected nearest neighbours are added into minority sub-datasets. The number of new synthetic samples and the way to find the nearest record are changed in test to find the best over-sample method which could be used to reduce the minimum number of records. In the test, the imbalanced experiment dataset is built by conditionally sampling from in-situ records, which imitates the time series of ASCAT data.

The new synthetic sample $S_j (\mathrm{j} = 1, 2, 3 \ldots\ldots \mathrm{m})$:

$$S_j = x_i + \mathrm{rand}\,(0,1) \times (x_{ij} - x_i) \tag{9}$$

where, m = the oversample multiple

x_i = the record in minority sub-dataset

x_{ij} = the No. j nearest record in the same sub-dataset

$\mathrm{rand}\,(0,1)$ = a number randomly chosen from (0,1)

3.3 *Weighting*

By analysing the In-situ data, the seasonal cycle and diurnal cycle of offshore wind in those areas are discovered, which are illustrated in Figure 5. The diversity of wind in difference months and hours means the records in some sub-dataset are more representative than others. Using the more representative sub-dataset to represent directional wind distribution will get more reliable result. In this condition, putting difference weight on records according to their representativeness could be a method to get better retrieval result with fewer records.

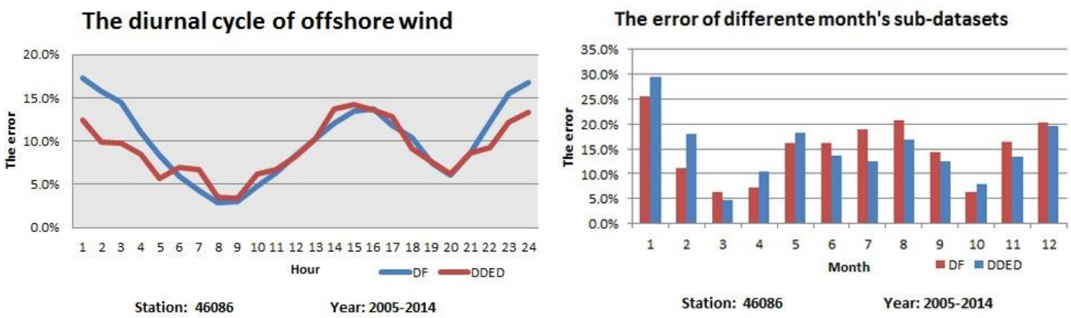

Figure 5. The seasonal cycle and diurnal cycles of offshore wind

The new equation for DF with weighting:

$$n_j = \sum_{i=1}^{k} \varphi(t) \; ; DF_j = \frac{n_j}{\sum_{j=1}^{m} n_j} \tag{10}$$

where DF_j = the frequency of direction j(When j=1, j means north. When j=5,j means east, and so on.)

n_j = the number of wind records whose direction

$\varphi(t)$ = the function to calculate the weight of one record according to its time

The new equation for DDED with weighting:

$$DDED_j = \frac{D_j}{D}; \; D = \sum_{j=1}^{16} D_j \tag{11}$$

$$D_j = \frac{1}{2}\sum_{i=1}^{m}(\rho)(v_{ji}^3)\psi(t) \tag{12}$$

where $DDED_j$ = the wind energy density distribution of direction j

D_j = the wind energy density of direction j

D = the total wind energy density

ρ = the air density

v_{ji} = the wind speed of record i in direction j

t_{ji} = the duration represented by the record

m= the records number of direction j

$\psi(t)$ = the function to calculate the weight of one record according to its time

The weighting function $\varphi(t)$ or $\psi(t)$:

$$\varphi(t) \text{ or } (t) = 1 + \frac{weight(t) - minimumweight}{maximalweight - minimumweight}\; ; weight(t) = \frac{1}{error_{month}(t)} \times \frac{1}{error_{hour}(t)} \tag{13}$$

where $error_{month}(t)$ = the DF or DDED error of the month t sub-dataset

$error_{hour}(t)$ = the DF or DDED error of the hour t sub-dataset

$weight(t)$ = the weight of the record whose time is t

$minimumweight$ = the minimum weight in every record's weight

$maximalweight$ = the maximal weight in every record's weight

4. Results

4.1 *Minimum number of records*

The results of the minimum number of records needed to represent DF and DDED realistically are showed in Table 3. The results are different for different offshore areas because of these areas' characteristic. Generally, when n >350, the error for DF declines to below 10% at an acceptable confidence level. And, to make the error under 10% for DDED, the n need to be more than 800.

Table 3. The minimum number of records to get reliable retrieval DF and DDEF

	Station 41048		Station 42040		Station 46001		Station 46086	
	DF	DDED	DF	DDED	DF	DDED	DF	DDED
2005	/	/	363.85	1102.10	344.55	637.38	263.42	569.54
2006	/	/	370.50	843.84	344.21	689.07	266.41	518.73
2007	/	/	365.03	728.29	308.93	605.94	259.12	528.17
2008	367.99	808.52	368.21	903.55	353.66	780.59	264.49	567.19
2009	366.15	785.23	330.37	586.52	367.28	777.51	255.03	541.13
2010	361.72	752.46	360.63	747.23	357.81	722.62	267.12	645.36
2011	371.71	1041.09	374.51	845.73	325.42	615.48	267.02	575.45
2012	365.69	860.04	375.46	NULL [a]	368.85	790.12	272.58	562.45
2013	372.71	785.60	366.07	NULL [a]	334.02	585.23	285.39	495.39
2014	369.68	806.60	368.09	800.10	/	/	265.62	533.95
Average	367.95	834.22	364.272	819.67	344.97	689.32	266.62	553.736
	Average for DF		335.95		Average for DDEF		724.23	

[a] The crucial number could not be calculated because of the interference of hurricane records

4.2 *The influences of over-sample*
Because of the representativeness diversity of different hour sub-dataset, the imbalanced dataset's influence is complex. When the majority sub-datasets are more representative than minority sub-datasets, the retrieval result's error will be smaller. So, the over-sample method could reduce the minimum number of records only in condition that the imbalanced data's minority sub-datasets are much more representative than majority counterparts. One example is showed in Figure 6.

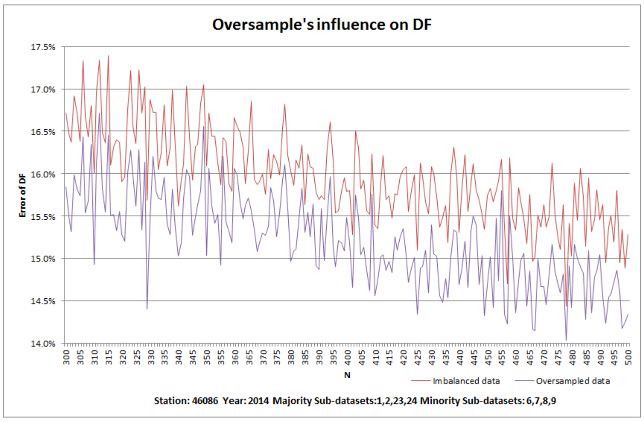

Figure 6. The over-sample's influence

4.3 *The influences of weighting*
The weighting's influence on the minimum number of records largely depends on the weighting function φ(t) or ψ(t). The weighting function use in this study could reduce the minimum number of records for retrieval of DF or DDED for the buoy data tested in research. Result of this is demonstrated in Table 4. The weighting method reduces 15 records for DF retrieval and 33 for DDED in this test.

Table 4. The weighting's influence on crucial records number

		2005	2006	2007	2008	2009	2010	2011	2012	2013
DF	Original	363.85	370.5	365.03	368.21	330.37	360.63	374.51	375.46	366.07
Station:	Weighting	346.83	358.78	350.27	349.24	329.63	347.52	355.75	353.63	345.82
	Reduced	17.02	11.72	14.76	18.97	0.74	13.11	18.76	21.83	20.25
42040	Average	15.19								
DDED	Original	569.54	518.73	528.17	567.19	541.13	645.36	575.45	562.45	495.39
Station:	Weighting	514.33	493.60	492.10	569.23	532.68	593.33	531.89	557.75	443.06
	Reduced	55.21	25.13	36.07	-2.04	8.45	52.03	43.56	4.70	52.33
46086	Average	33.28								

5.Discussion
With the changing of methods of over-sample or function of weighting, the retrieval results and the minimum number of records also changed. To get better retrieval results, much more weighting functions and over-sample methods should be tested and compared in further study.
Those methods used in this work rely on basic understanding of specific area's wind. So, when use those methods to obtain directional wind distribution from satellite data, historical In-situ observations are needed to build weighting function or decided whether over-sample is helpful in retrieval. Moreover, In-Situ observations usually only demonstrate the wind condition of a specific location.

Whether the information from this specific location could be used to build weighting function for near area is needed to answer in further research.

6. Conclusions

By this simulation experiment, the minimum number of records for annual wind direction frequency is 350, and 800 for annual direction distributions of wind energy density. And, the over-sample for minority sub-datasets of imbalanced data and using weighting function could have positive influences on retrieval of directional wind distribution and reduce the records needed in retrieval.

References

[1] Nguyen K Q 2007 Wind energy in Vietnam: Resource assessment, development status and future implications *Energy Policy* **35** pp. 1405-1413

[2] Zhao S, Jiang B, Xu H., Li S. and Ding J 2010 Exploration and Application of Ocean Wind Energy Resources in Coastal Sea of China *Ocean Technology* **29** pp. 117-121.

[3] Heckenbergerova J, Musilek P and Krömer P 2015 Optimization of Wind Direction Distribution Parameters Using Particle Swarm Optimization *Afro-European Conf. for Industrial Advancement* vol 334, ed A Abraham, P Krömer and V Snasel (Gewerbestrasse: Springer International Publishing) pp. 15-26.

[4] Chang R, Zhu R, Badger M, Hasager C B, Xing X and Jiang Y 2015 Offshore Wind Resources Assessment from Multiple Satellite Data and WRF Modeling over South China Sea *Remote Sens.* **7** pp. 467-487

[5] Doubrawa P, Barthelmie R J, Pryor S C, Hasager C B, Badger M and Karagali I 2015 Satellite winds as a tool for offshore wind resource assessment: The Great Lakes Wind Atlas *Remote Sens. Environ.* **168** pp. 349-359

[6] Yang X., Li X, Pichel W G and Li Z 2011 Comparison of Ocean Surface Winds From ENVISAT ASAR, MetOp ASCAT Scatterometer, Buoy Measurements, and NOGAPS Model *Ieee T. Geosci. Remote.* **49** pp. 4743-4750

[7] Barthelmie R J and Pryor S C 2003 Can satellite sampling of offshore wind speeds realistically represent wind speed distributions *J. Appl. Meteorol.* **42** pp. 83-94

[8] Barthelmie R J, Grisogono B and Pryor S C 1996 Observations and simulations of diurnal cycles of near-surface wind speeds over land and sea *J. Geophys. Res.* **101** pp. 21327-21337

[9] Gu Q 2009 Research of Machine Learning in Imbalanced Data Set and its Application in Geosciences Data Processing, China University of Geosciences

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 41371355.