A case based reasoning data fusion scheme: application to offshore wind energy resource mapping

M. B. Ben Ticha and T. Ranchin
Centre Energétique et Procédés
Ecole des Mines de Paris
Sophia Antipolis, France.

mohamed-bassam.ben_ticha@ensmp.fr

Abstract - A data fusion scheme is proposed for wind energy resource mapping at high spatial resolution. The resource assessment is based on wind speed and direction measurements. Remotely sensed data is a solution to get wind observations offshore. However, high spatial resolution data do not have a sufficient repetitiveness to establish reliable wind energy resource maps. The scheme proposed in this paper uses these measurements as typical situations which have to be merged with low spatial resolution data having a sufficient temporal repetitiveness. The fusion process builds a library of typical cases. To these typical cases are associated typical fields representing the information to be merged with the corresponding low spatial resolution data. In this paper, we give, firstly, the general fusion scheme. Then, we present the different tools needed by this process. We focus particularly on the definition of the typical situations. The retrieval of these situations is achieved by a classification process. Finally, some prospects are given.

Keywords: case based reasoning, classification, wind energy, resource mapping.

1 Introduction

In Earth Observation, high spatial resolution data is generally associated to a low temporal repetitiveness. Much larger low spatial resolution data sets could be available. Data fusion could be a means to benefit from the high repetitiveness of the latter and the high spatial resolution of the first to build a high spatial resolution data set with the best temporal repetitiveness. On the basis of the belief that for two similar low spatial resolutions the high spatial resolution structures are similar, one can design a fusion scheme with few high spatial resolution data needed. In this paper we present such a fusion scheme. This scheme is applied for offshore wind energy mapping.

In the framework of offshore wind energy development, wind resource mapping is essential. In fact, this will allow a better sitting of windmills and thus improve their outcome. The wind variation is, traditionally, described using the two-parameter Weibull probability density function [1]. This function is given in equation 1, \( v \) being wind speed, \( A \) Weibull scale parameter and \( k \) Weibull shape parameter.

\[
p(v) = \frac{k}{A} \left( \frac{v}{A} \right)^{k-1} e^{-\left( \frac{v}{A} \right)^k}
\]

The wind resource at a given location is fully described by \( A \) and \( k \). To assess wind resource one should compute \( A \) and \( k \) for each direction sector. The requested spatial resolution for wind energy resource mapping is on the range of the kilometre.

Onshore, Weibull parameters are assessed by interpolating long-term data available from meteorological stations nearby the prospected site with data acquired by a mast installed on this site for one year [2]. Offshore, in-situ measurements are very costly (750,000 euros vs. 15,000 euros onshore [3]) and do not represent the spatial variability of wind. Remotely sensed data could be an accurate and economic way to access wind data offshore. Previous studies showed the capabilities of some instruments to measure the wind, especially, radar scatterometers and synthetic aperture radars (SAR). SAR have a high spatial resolution (few hundred of meters) but they have an insufficient repetitiveness for establishing wind statistics. Scatterometer have a sufficient repetitiveness (two measurements a day) but a low spatial resolution. Thus, no source could be used alone to assess wind resource.

A solution is to merge data sets. In this paper, we present a scheme for fusion of these data. This scheme could be used in any analogue problem where high spatial resolution data has a low temporal repetitiveness and where more large low spatial data sets are available.

In this scheme, SAR measurements serve as typical high resolution wind variations whereas scatterometers will give the temporal repetitiveness needed to assess the Weibull parameters of the zone.

In section 1, we give an overview of the available data sources. Then, in section 2, we present the data fusion scheme. In section 3, we give a classification method for wind typical situations extraction.

2 Data sources

Two main instruments were retained by previous studies for wind energy resource assessment: scatterometer
and synthetic aperture radar [4]. These two instruments are active sensors. The radar response from sea depends on waves instantaneously generated by wind [5]. Therefore, by an appropriate processing of satellite raw data, we can extract wind measurements. Scatterometer and SAR data are similar. However, these instruments have different spatial resolution and temporal repetitiveness. Hereafter, we present the main spaceborne scatterometers and SAR launched in the last decade and their characteristics.

2.1 Scatterometers
A summary of scatterometers launched since 1991 is given in table 1. The spatial resolution of scatterometer wind measurements is between 25 and 50 km. An important characteristic of these instruments is their revisit period, i.e. the period between two successive measurements at the same point. This period depends on the repeat cycle of the satellite, i.e. the period between two successive identical orbits, and the swath width, i.e. the total width of the area covered by the sensor on the ground. Thanks to a large swath width, revisit period is lower than the repeat cycle of the satellite. Since June 1999, two scatterometer measurements are available per day. Due to noise introduced by land, scatterometer measurements are located at least 25 km off the coast. For retrieving wind speed and direction from raw measurements, empirical models are used: CMOD4 [6], CMOD5 [7], CMOD-ifr2 [8].

Table 1: Wind scatterometers summary [9]

<table>
<thead>
<tr>
<th>Satellite</th>
<th>Swath width (km)</th>
<th>Spatial resolution (m)</th>
<th>Repeat cycle (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERS-1</td>
<td>100</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>ERS-2</td>
<td>100</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>ADEOS</td>
<td>2*600</td>
<td>25/50</td>
<td>41</td>
</tr>
<tr>
<td>QuikSCAT</td>
<td>1800</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>ADEOS-2</td>
<td>1800</td>
<td>50</td>
<td>4</td>
</tr>
</tbody>
</table>

2.2 Synthetic aperture radars
A summary of spaceborne synthetic aperture radars is given in table 2. Synthetic aperture radars images have a typical resolution of a few tens of meters. However, there is not enough synthetic aperture radar images archived to establish accurate wind statistics based only on them [10]. This is due to the low swath width of these instruments. Previous studies have shown the possibility of adapting the empirical algorithm CMOD used for scatterometers to wind speed retrieval from SAR images [11].

Table 2: Synthetic aperture radars summary [9]

<table>
<thead>
<tr>
<th>Satellite</th>
<th>Swath width (km)</th>
<th>Spatial resolution (m)</th>
<th>Repeat cycle (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERS–1</td>
<td>100</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>JERS–2</td>
<td>75</td>
<td>18</td>
<td>44</td>
</tr>
<tr>
<td>ERS–2</td>
<td>100</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>Radarsat</td>
<td>100-170</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>Envisat</td>
<td>100</td>
<td>30</td>
<td>35</td>
</tr>
</tbody>
</table>

2.3 Conclusions on available data
For the assessment of wind statistics, with an uncertainty of ±10% at a confidence level of 90%, on the order of 250 independent observations are required to fit the Weibull parameters [12]. Scatterometer temporal coverage fulfills this requirement. The spatial resolution of SAR measurements is sufficient for wind energy resource mapping. Concerning the homogeneity of the two data sets, the major difference is the spatial coverage of scatterometers. In fact SAR covers coastal areas which is not the case of scatterometers.

3 Data fusion scheme
For the schematisation of a fusion process we use the fusion cell scheme defined by [13]. The fusion process is presented by a box having as input the different data to be merged. The output is the expected result and depends on the fusion level. Two other inputs are represented by this scheme: external knowledge and auxiliary information which would be useful to the fusion process.

As showed in section 1, scatterometer data, which has a sufficient temporal repetitiveness, is located far away from the coast. It does not cover the most interesting area for windmills siting. Fluid dynamics laws could be used as an external knowledge to propagate measurements to coastal areas. The propagation of the measurements near the coast is a part of the data alignment process.

The error associated to wind resource assessed by the fusion process is dependent on the number of available measurements and the quality of data. Thus it is dependent on the data given to the fusion process. Therefore, confidence or uncertainty intervals have to be associated to wind energy resource maps produced by this process.

Figure 1 illustrates the proposed fusion process. The data fusion process gets the two data sets as inputs and produces a wind resource map as principal output. Case based reasoning rules are given to this process as auxiliary information. In fact, the different data given to the fusion process have different dates. These rules define how the data should be fused.

In case based reasoning, a solution is generated by comparing the problem to a set of resolved problems. Data fusion is based, in this context, on the assumption that for two similar situations at low spatial resolution this similarity is valid at high spatial resolution. This assumption is used in the downscaling of
Wind fields [14]. Downscaling uses relationships between large scale (low spatial resolution) features and local features to get a local climatology. The assessment of the high spatial resolution could be achieved in such a scheme by the nesting of numerical models with a finer resolution into a coarser one. The global model, with low spatial resolution, serving as a constraint to the high spatial resolution one. However, running a numerical model is costly and the lack of high spatial resolution initial and boundary conditions could have a great impact on the quality of the results [15]. In the case of data fusion, there is no need for initial conditions or boundary conditions. Data fusion could benefit from the available high spatial resolution observations to synthesize a large high spatial resolution data set. As for downscaling, this data fusion process is based on the fact that wind climatology over a given region can be described by a few typical wind situations [16].

In a case based reasoning scheme, an important task is to build a library of typical cases. Each typical case is characterized by a specific wind flow behaviour over the interest area. These typical cases are representative of several similar wind field observations. This task is achieved by a classification of the set of observations. The result of this procedure is a partition of the initial low spatial resolution data set into several classes. Each class is characterised by its corresponding high spatial resolution situations. The next step is the association of the high spatial resolution data and the low spatial typical situation. Thus, we can retrieve the typical high resolution behaviour of wind flow corresponding to each low spatial resolution typical situation. In this step we need the fluid dynamics laws for the alignment task. In fact, high spatial resolution data covers coastal areas whereas low spatial resolution data is available far away from the coast. After this step we get high spatial resolution typical fields. These typical fields represent the information to be "added", after being adjusted, to low spatial resolution measurements to get the high spatial resolution wind estimations permitting the computation of high spatial resolution wind statistics. This leads to the detailed data fusion process scheme presented in figure 2.

An important component of this data fusion scheme is the classification process. In the next section, we focus on this process and give an adaptation of existing wind classes generation schemes for the specific purpose of data fusion.

4 Data classification

Wind classification is conducted to represent typical wind situations, i.e. typical wind flow behaviour over interest area. In the graphic of the case based data fusion scheme presented above, the clustering algorithm should satisfy three conditions:

- stability of classes;
- minimization of the maximum distance between two elements of the same class;
- minimization of the number of classes.
Several authors have already build up classifications of wind fields at different scales [17, 18]. The number of wind classes is unknown a priori. As [17] we use a two-stage procedure for wind classes generation. In the first stage we focus on the determination of the most convenient number of classes. Two criteria have been taken into account for the choice of this number:

- the representativity of the class: situations having a few members have to be classed as outliers,
- each wind direction sector have to be represented in the final typical situations.

The output of this first stage is \( k \) typical situations and a set of wind fields classed as outliers. The aim of the second stage of this scheme is to assign each wind field classed as an outlier to the most similar typical situation. This assignement is done because it is not likely that a high spatial resolution measurement is available for such a rare situation.

An important output of the clustering algorithm is the frequency of occurrence of each class. This representativity index would be important for the evaluation of the fusion process and for the computation of wind statistics. All these steps lead to the classification scheme given in figure 3.

\[
d_{C_{t_1},C_{t_2}} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \left[ (\tilde{u}_{t_1,ij} - \tilde{u}_{t_2,ij})^2 + (\tilde{v}_{t_1,ij} - \tilde{v}_{t_2,ij})^2 \right]^{\frac{1}{2}}
\]

\[
(\tilde{u}_{t_1,ij}, \tilde{v}_{t_1,ij}) = \frac{1}{s_{t_1}} (u_{t_1,ij}, v_{t_1,ij})
\]

\[
s_{t_1} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} (u_{t_1,ij}^2 + v_{t_1,ij}^2)^{\frac{1}{2}}
\]

The result of this classification procedure is a set of typical situations. These situations are differentiated by the behaviour of the wind flow over the interest area. We applied this classification scheme to practical data. The result is that 28% of the data was classed as outliers. The number of typical situations was of 36. These situations are presented in figure 4. The error associated to these typical situation have to be assessed. This error would reflect the representativity of wind climatology over a specific area by the typical cases resulting from the classification of observations data set. The next step is to associate these typical cases to high spatial resolution observations. Thus a library of typical fields, permitting to retrieve high spatial resolution data, could be constructed.

5 Conclusions and prospects

In this paper, we presented a case based reasoning data fusion scheme. In this scheme, data is partitioned in several classes. For each class a typical field represents the information to be merged with the low spatial resolution data. This scheme is applied to offshore wind energy resource mapping. We presented the different data sources useful for offshore wind energy mapping. Based on the assumption that wind observations could be classified into few classes, the case based data fusion scheme could permit a better evaluation of high spatial resolution wind statistics. The classification scheme
proposed in this paper does not need an a priori knowledge of wind climatology over the interest area. It generates typical situations representing typical wind flow behaviour over interest region. The next step is the extraction of the typical fields permitting, after an appropriate adjustment, to synthesize high spatial resolution data from the low resolution one. Another important task is the error assessment. Wind resource maps generated by this process will be of a great interest for offshore wind energy sector. The fusion scheme developed in this paper could be useful in many other fields where high spatial resolution data is associated to low temporal repetitiveness.

References


